



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DE RUSSAS
CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

NATÁLIA LÍDIA LIMA COELHO

**USO DE PROGRAMAÇÃO LINEAR INTEIRA PARA ANÁLISE DE AGRUPAMENTOS
NA CÂMARA DOS DEPUTADOS**

RUSSAS

2023

NATÁLIA LÍDIA LIMA COELHO

USO DE PROGRAMAÇÃO LINEAR INTEIRA PARA ANÁLISE DE AGRUPAMENTOS NA
CÂMARA DOS DEPUTADOS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação do Campus de Russas da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Ciência da Computação.

Orientadora: Prof. Dra. Tatiane Fernandes Figueiredo

RUSSAS

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

C618u Coelho, Natália Lídia Lima.
Uso de programação linear inteira para análise de agrupamentos na Câmara dos Deputados / Natália Lídia Lima Coelho. – 2023.
28 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Russas, Curso de Ciência da Computação, Russas, 2023.
Orientação: Profa. Dra. Tatiane Fernandes Figueiredo.

1. parlamentos políticos. 2. programação linear inteira. 3. problemas de agrupamento. I. Título.
CDD 005

NATÁLIA LÍDIA LIMA COELHO

USO DE PROGRAMAÇÃO LINEAR INTEIRA PARA ANÁLISE DE AGRUPAMENTOS NA
CÂMARA DOS DEPUTADOS

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Ciência da Computação
do Campus de Russas da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em Ciência da Computação.

Aprovada em:

BANCA EXAMINADORA

Prof. Dra. Tatiane Fernandes
Figueiredo (Orientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. Pablo Luiz Braga Soares
Universidade Federal do Ceará (UFC)

Prof. Dr. Marcio Costa Santos
Universidade Federal de Minas Gerais (UFMG)

AGRADECIMENTOS

Agradeço primeiramente à minha família, em especial meus pais, Fernanda e Válder, que sabendo da importância da educação, mesmo diante de tantas dificuldades, sempre fizeram tudo o que foi possível para que eu pudesse ter os estudos como prioridade. Obrigada por serem minha base e por todos os esforços que fizeram para que eu pudesse chegar a este ponto, espero ainda trazer muitas realizações para vocês.

Aos amigos que me acompanharam durante a graduação, João Victor, Wilkinis, Dhioleno, Milene, Thiago, Adryan, Anne, Hanna, Lucas e João Pedro. Obrigada pela companhia nos bons e maus momentos, por todo o amor, suporte e incentivo. Por muitas vezes vocês foram muito mais do que amigos e por isso sou muito feliz e grata por tê-los comigo neste período.

Aos amigos de longa data, Tiago e Lukas, que mesmo com a rotina e a distância, de alguma forma se mantiveram presentes. Obrigada por todo o amor e torcida durante esta fase.

À minha orientadora Tatiane Fernandes, por toda sua dedicação na orientação deste trabalho. Obrigada por acreditar em mim, por toda ajuda, todos os ensinamentos e conselhos. Te vejo como um grande exemplo e fonte de inspiração, espero realizar ainda mais trabalhos com você daqui para frente.

Agradeço também aos demais professores do campus, por toda a excelência e dedicação no repasse do conhecimento nas disciplinas do curso. Agradeço em especial à professora Anna Beatriz, por me acolher em meu primeiro projeto no campus e me introduzir às pesquisas acadêmicas, você também é um grande exemplo para mim.

Agradeço à professora Rosa Figueiredo e ao pesquisador Mário Levorato pelo fornecimento dos dados da Câmara dos Deputados utilizados nesta pesquisa.

Por fim, agradeço aos demais que contribuíram direta ou indiretamente na minha graduação, obrigada por todo o suporte e incentivo que me fizeram seguir esta jornada.

RESUMO

Compreender o comportamento em parlamentos políticos tem sido foco de diversos estudos na literatura. Trabalhos recentes têm aplicado técnicas de agrupamento usando dados obtidos por meio da votação de propostas ocorridas em um determinado período de tempo em parlamentos políticos. Neste trabalho, propõe-se um estudo sobre os agrupamentos na Câmara dos Deputados do Brasil. Para isso, os grupos foram gerados por meio da aplicação de um modelo de Programação Linear Inteira. Utilizando o conceito de grafos de sinal para representar as relações entre políticos, o problema em estudo foi modelado como um Problema de Formação de Equipe Competitiva. A análise realizada a partir dos resultados obtidos permitiu a identificação de fatores que possivelmente influenciaram, ou não, os votos dos agrupamentos políticos criados.

Palavras-chave: parlamentos políticos; programação linear inteira; problemas de agrupamento.

ABSTRACT

Understanding the behavior in political parliaments has been the focus of several studies in the literature. Recent works have applied clustering techniques using data obtained through the voting of proposals occurred within a given period of time in political parliaments. In this work, a study on groupings in the Brazilian Chamber of Deputies is proposed. For this, the groups were generated through the application of an Integer Linear Programming model. Using the concept of signed graphs to represent the relationships between politicians, the problem under study was modeled as a Competitive Team Formation Problem. The analysis carried out from the results obtained allowed the identification of factors that possibly influenced, or not, the votes of the created political groupings.

Palavras-chave: political parliaments; integer linear programming; grouping problems.

LISTA DE FIGURAS

Figura 1 – Exemplo de solução para o CTFP.	16
Figura 2 – Etapas da metodologia.	21
Figura 3 – Amostra da matriz deputado \times estado para instância do ano de 2010.	22
Figura 4 – Relação entre a demanda de deputados e o total para cada estado no ano de 2010.	23
Figura 5 – Número de deputados de cada estado por equipe.	24
Figura 6 – Número de deputados de cada partido por equipe.	24

LISTA DE ABREVIATURAS E SIGLAS

CTFP	<i>Competitive Team Formation Problem</i>
PL	Programação Linear
PLI	Programação Linear Inteira
PLIB	Programação Linear Inteira Binária
PLIM	Programação Linear Inteira Mista
TFP	<i>Team Formation Problem</i>
TFSN	<i>Team Formation in Signed Networks</i>

SUMÁRIO

1	INTRODUÇÃO	9
2	OBJETIVOS	11
2.1	Objetivo geral	11
2.2	Objetivos específicos	11
3	FUNDAMENTAÇÃO TEÓRICA	12
3.1	Teoria dos Grafos	12
3.2	Programação Linear	13
3.2.1	<i>Programação Linear Inteira</i>	14
3.2.2	<i>O Problema de Formação de Equipes Competitivas</i>	14
3.2.2.1	<i>Modelo de Programação Linear Inteira para o CTFP</i>	15
4	TRABALHOS RELACIONADOS	17
4.1	Formando Equipes Compatíveis em Redes de Sinais	17
4.2	Análise de Grafos de Sinais para a Interpretação do Comportamento Eleitoral	18
4.3	Análise do Balanço Estrutural do Congresso Brasileiro	19
4.4	Relacionamento entre trabalhos	20
5	METODOLOGIA	21
5.1	Coleta de dados	21
5.2	Geração e padronização das instâncias	22
5.3	Implementação do modelo de PLI para agrupamentos em parlamentos políticos	23
5.4	Análise das soluções obtidas	23
6	RESULTADOS	24
6.1	Solução obtida	24
6.2	Análise da solução	25
7	CONCLUSÕES E TRABALHOS FUTUROS	27
	REFERÊNCIAS	28

1 INTRODUÇÃO

Atualmente, estudos científicos para compreender as interações e comportamentos de indivíduos em agrupamentos políticos têm sido foco de pesquisadores da área de humanas e exatas. Levorato e Frota (2017) apresentaram um trabalho onde é realizada uma análise de grupos do Congresso Nacional Brasileiro. Para tal, os autores utilizaram algoritmos heurísticos para a construção de agrupamentos. De forma similar, Arinik *et al.* (2017) realizaram uma análise quantitativa e qualitativa aprofundada a partir dos resultados obtidos de algoritmos heurísticos e exatos para a criação de agrupamentos, utilizando dados de votação do Parlamento Europeu.

Embora ambos autores tenham apresentados importantes *insights* sobre as bases de dados estudadas, é importante mencionar que as análises dos agrupamentos criados não consideraram informações adicionais de cada um dos políticos envolvidos. Informações como formação científica, idade, cidade ou estado residente, podem ajudar na compreensão dos votos e auto-organização dos políticos em um parlamento. Buscando contribuir com essas lacunas científicas ainda existentes na literatura, este trabalho propõe a aplicação de um modelo de Programação Linear Inteira para resolução do Problema de Formação de Equipes Competitivas com objetivo de realizar e estudar agrupamentos em parlamentos políticos. O estudo tem como objetivo central a análise do Congresso Nacional Brasileiro, mais especificamente da Câmara dos Deputados, utilizando dados de votações em temas gerais ocorridas no ano de 2010.

Assim, para criação dos agrupamentos estudados nesta pesquisa foram utilizados dados da Câmara dos Deputados obtidos a partir do trabalho de Levorato e Frota (2017), sendo criada a partir dos grafos disponibilizados pelos autores, uma entrada padronizada para o Problema de Formação de Equipes Competitivas, considerando também características pessoais de cada deputado, neste caso o partido e o estado de residência. Os resultados obtidos a partir da aplicação do modelo foram analisados para identificação de fatores que possivelmente influenciaram a criação dos agrupamentos resultantes, e consequentemente, os votos dos políticos presentes nestes agrupamentos.

Para uma melhor organização, esta monografia está segmentada conforme a seguir: no Capítulo 2 serão apresentados os objetivos gerais e específicos deste trabalho; o Capítulo 3 irá tratar das teorias que formam a base necessária para realização do estudo desenvolvido; no Capítulo 4 serão discutidos trabalhos que tratam de questões relacionadas ao problema trabalhado; o Capítulo 5 irá detalhar as etapas metodológicas seguidas para realização desta análise; o Capítulo 6 irá apresentar a solução obtida e as análises realizadas a partir dela; por fim,

o Capítulo 7 discorrerá sobre a conclusões e os trabalhos futuros.

2 OBJETIVOS

2.1 Objetivo geral

Aplicar um modelo de Programação Linear Inteira para gerar e analisar agrupamentos na Câmara dos Deputados.

2.2 Objetivos específicos

- Estudar e atualizar instâncias geradas a partir dos dados da Câmara existentes na literatura;
- Aplicar o modelo para o Problema de Formação de Equipes Competitivas para gerar agrupamentos de políticos;
- Analisar os resultados obtidos a fim de obter *insights* sobre os grupos criados.

3 FUNDAMENTAÇÃO TEÓRICA

Neste Capítulo serão tratados os conceitos necessários para o entendimento das técnicas utilizadas para realizar a análise em questão. Nas Seções 3.1 e 3.2 serão tratadas as definições referentes à Teoria dos Grafos e Programação Linear, respectivamente. Na Subseção 3.2.1 será abordado o conceito de Programação Linear Inteira, e na Subseção 3.2.2 será apresentado o Problema de Formação de Equipes Competitivas.

3.1 Teoria dos Grafos

Um *grafo* G é composto por um conjunto finito e não vazio, denotado por V , de elementos chamados vértices, e um conjunto de pares distintos subconjuntos de V , denotado por E , chamado de arestas. Escreve-se $G = (V, E)$ para indicar que o grafo G possui o conjunto V de vértices e o conjunto E de arestas, podendo também escrever V como $V(G)$ e E como $E(G)$ para enfatizar que os conjuntos V e E fazem parte do grafo G . Pode se denotar como uv ou vu uma aresta $\{u, v\}$ do grafo G . Sendo $a = uv$, diz-se que a *incide* em u e v , que serão, portanto, vértices *adjacentes* (CHARTRAND *et al.*, 2010).

Define-se H como um *subgrafo* de um grafo G , caso os conjuntos de vértices e arestas de H estejam contidos nos conjuntos de vértices e arestas de G , respectivamente. Ou seja, H é subgrafo de G , se $V(H) \subseteq V(G)$ e $E(H) \subseteq E(G)$. Sendo S um subconjunto não-vazio de V de um grafo G , e $G[S]$ um subgrafo de G induzido por S , ou seja, em que S é o conjunto de vértices de $G[S]$ e cada par de vértices adjacentes em $G[S]$ são também adjacentes em G , chama-se de *subgrafo induzido* um subgrafo H de G tal que $H = G[S]$ (CHARTRAND *et al.*, 2010).

É denominado de *grafo de sinais* um grafo $G = (V, E, sign)$ que possui V como o conjunto de vértices, E como o conjunto de arestas e a função $sign : E \rightarrow \{+, -\}$ que vincula a cada aresta um sinal positivo ou negativo. Uma aresta $uv \in E$ que possua $sign(uv) = +$ será chamada de positiva e, analogamente, será chamada de negativa caso $sign(uv) = -$. Utiliza-se a notação E^+ e E^- para denotar os subconjuntos de arestas positivas e negativas de G , respectivamente, fazendo com que $E = E^+ \cup E^-$ e $E^+ \cap E^- = \emptyset$ (FIGUEIREDO, 2021).

3.2 Programação Linear

A Programação Linear (PL) é um método que utiliza de um modelo matemático para solucionar problemas que competem à Pesquisa Operacional, tais problemas que envolvem comumente a questão de alocar da melhor forma recursos limitados entre tarefas que competem entre si. Desta forma, a Programação Linear pode ser definida como um planejamento de atividades realizado para obter o melhor resultado (ótimo) dentre as alternativas possíveis (HILLIER; LIEBERMAN, 2013). O modelo matemático utilizado na Programação Linear pode ser escrito na forma de um conjunto de equações e inequações lineares e é composto por elementos chamados: *função objetivo*, *variáveis de decisão* e *restrições*. Um modelo pode ser representado da forma generalizada abaixo, com o objetivo de escolher valores para $x_1, x_2, x_3, \dots, x_n$ de forma a:

$$\text{Maximizar ou Minimizar } Z = c_1x_1 + c_2x_2 + \dots + c_nx_n$$

Sujeito às restrições:

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n < b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n < b_2$$

...

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n < b_m$$

$$x_i \geq 0 \text{ e } b_j \geq 0, \forall i = \{1, 2, \dots, n\} \text{ e } j = \{1, 2, \dots, m\}$$

Tal que:

- Z é a função objetivo que deve ser maximizada ou minimizada;
- x_i são as variáveis que representam quantidades de recursos do elemento i ;
- c_i são coeficientes de ganho ou custo de i ;
- b_j é a quantidade disponível correspondente ao recurso j ;
- a_{ij} é a quantidade de recursos de i que uma unidade j irá consumir.

Dada esta forma padrão, qualquer problema, não somente os de alocações de recursos entre atividades, em que o modelo matemático que o descreva se ajuste neste formato (e suas variações válidas como: restrições na forma de equações, ou do tipo “maior ou igual a que”, e a não restrição de sinal para algumas variáveis de decisão) é considerado um problema de Programação Linear (HILLIER; LIEBERMAN, 2013).

3.2.1 Programação Linear Inteira

A Programação Linear Inteira (PLI) se trata de uma variante mais específica da Programação Linear, em que o modelo matemático contém restrições exigindo a integralidade de todas as suas variáveis. Pode se denotar como problema de Programação Linear Inteira Mista (PLIM) os casos que somente algumas variáveis possuem a restrição de integralidade. Por fim, é denotado como problema de Programação Linear Inteira Binária (PLIB) os casos em que as variáveis são restritas aos valores inteiros 0 ou 1, somente (WOLSEY, 2020).

O modelo geral apresentado anteriormente pode ser convertido em um modelo PLI adicionando a restrição:

$$x_i \in \mathbb{Z}, \forall_i = \{1, 2, \dots, n\}.$$

3.2.2 O Problema de Formação de Equipes Competitivas

O Problema de Formação de Equipes, do inglês *Team Formation Problem* (TFP), é definido por Lappas *et al.* (2009) como o problema de encontrar um subconjunto de indivíduos em uma rede social, onde cada indivíduo possui uma determinada habilidade, que possam se comunicar efetivamente em uma equipe e possuam as habilidades necessárias para realizar uma determinada tarefa. Neste caso, a efetividade da comunicação é medida através da métrica de *custo de comunicação* baseada na distância entre os indivíduos na rede. Desta definição surgem várias situações que resultam na criação de diferentes variações do TFP, dentre elas o Problema de Formação de Equipes Competitivas, do inglês *Competitive Team Formation Problem* (CTFP).

Figueiredo (2021) define o CTFP como uma variante do TFP que busca maximizar a afinidade dentro das equipes e a competitividade entre elas. A rede social pode ser representada como um grafo de sinais, onde os indivíduos são representados pelos vértices e as relações de afinidade e competitividade entre eles por meio de arestas com peso positivo (+) e negativo (-), respectivamente. O problema utiliza da *Teoria do Balanço Estrutural* para definir o custo de comunicação dentro das equipes e entre elas. A Teoria do Balanço Estrutural, proposta por Heider (1946), descreve relações antagônicas, como gostar/desgostar ou amor/ódio, entre indivíduos de um mesmo grupo social e propõe que estes tendem a se organizar em redes sociais onde situações de conflito são evitadas, a fim de chegar a um estado de balanço. Com base nesta teoria, o CTFP impõe a restrição de que os indivíduos em uma mesma equipe possuam apenas relacionamentos positivos entre si e que os indivíduos em equipes diferentes possuam apenas

relações negativas, com o objetivo de incentivar a afinidade intra-equipes e, ao mesmo tempo, a competitividade entre elas.

O CTFP é formalmente definido abaixo:

Problema 3.2.2.1 *O Problema de Formação de Equipes Competitivas (CTFP)*

Entrada: Uma tupla $\vartheta = (G, S, s, T, t)$, onde:

- O grafo de sinais $G = (V, E, \text{sign})$, em que o conjunto de vértices representa os indivíduos na rede social e as arestas seus relacionamentos;
- O conjunto de habilidades S ;
- A função $s : V \rightarrow 2^S$ que retorna o conjunto $s(u) \subseteq S$ de habilidades do indivíduo $u \in V$;
- O conjunto de equipes T ;
- A função de demanda $t : T \times S \rightarrow \mathbb{N}$ que retorna o número $t(j, s)$ de indivíduos necessários para a habilidade $s \in S$ na equipe $t \in T$.

Questão: Determinar $|T|$ subgrafos induzidos disjuntos, onde:

1. Cada vértice $u \in V$ pertença à no máximo um subgrafo, exercendo somente uma habilidade $s \in s(u)$;
2. Cada subgrafo j possui no mínimo $t(j, s)$ vértices exercendo a habilidade s , para todo $j \in T$ e todo $s \in S$;
3. Cada aresta uv em que seus vértices u e v pertençam a um mesmo subgrafo seja positiva, enquanto cada aresta uv de u e v pertencentes a subgrafos distintos seja negativa, para todo $uv \in E$;
4. O somatório de vértices alocados aos subgrafos seja maximizado ou a verificação de que as equipes não podem ser formadas.

Por fim, adota-se também a notação $s(j) = \{s \in S : t(j, s) > 0\}, \forall j \in T$ para determinar o conjunto de habilidades requeridas pela equipe j .

3.2.2.1 *Modelo de Programação Linear Inteira para o CTFP*

Levando em consideração variável binária:

$$x_{us}^j = \begin{cases} 1, & \text{se o indivíduo } u \text{ está na equipe } j \text{ com a habilidade } s, \\ 0, & \text{caso contrário.} \end{cases} \quad \forall u \in V, \forall j \in T, \forall s \in u(s).$$

O modelo PLI para o CTFP é definido como:

$$(CTFP) \max \sum_{j \in T} \sum_{u \in V} \sum_{s \in s(u)} x_{us}^j \quad (3.1)$$

$$\text{S.A: } \sum_{s \in s(u)} x_{us}^j + \sum_{s \in s(v)} x_{vs}^j \leq 1, \quad \forall uv \in E^-, \forall j \in T \quad (3.2)$$

$$\sum_{s \in s(u)} x_{us}^i + \sum_{s \in s(v)} x_{vs}^j \leq 1, \quad \forall uv \in E^+, \forall i, j \in T, i \neq j \quad (3.3)$$

$$\sum_{j \in T} \sum_{s \in s(u)} x_{us}^j \leq 1, \quad \forall u \in V \quad (3.4)$$

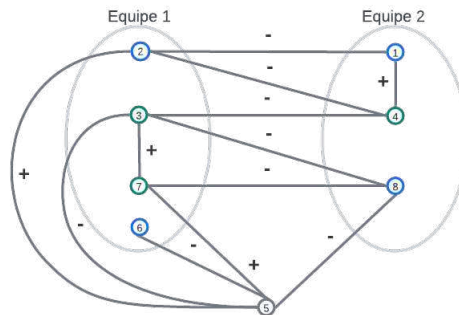
$$\sum_{u \in V: s \in s(u)} x_{us}^j \geq t(j, s), \quad \forall j \in T, \forall s \in s(j) \quad (3.5)$$

$$x_{us}^j \in \{0, 1\}, \forall j \in T, \quad \forall u \in V, \forall s \in s(u) \quad (3.6)$$

A função objetivo (3.1) maximiza o número de indivíduos com as habilidades necessárias em uma equipe. As restrições (3.2) e (3.3) asseguram que dentro de cada equipe só haja indivíduos com relacionamentos positivos e que entre equipes os indivíduos só possuam ligações negativas, respectivamente. A restrição (3.4) garante que um indivíduo só faça parte de uma equipe com somente uma habilidade. A penúltima restrição (3.5) faz com que cada equipe possua o mínimo de indivíduos necessários para cada habilidade. Por fim, a restrição (3.6) garante que a binaridade das variáveis.

A Figura 1 apresenta um exemplo de solução para o CTFP em uma instância de oito indivíduos em que foram requisitadas duas equipes e duas habilidades em cada equipe. Nesta instância era demandado pelo menos um indivíduo para cada habilidade em cada equipe e todos eles possuíssem ambas as habilidades. Como resultado, foram alocados sete dos oito indivíduos, sendo que os indivíduos 1, 2, 6 e 8 foram alocados na habilidade 1 (representada pela cor azul), e os indivíduos 3, 4 e 7 na habilidade 2 (representada pela cor verde), em suas respectivas equipes.

Figura 1 – Exemplo de solução para o CTFP.



Fonte: Elaborado pela autora.

4 TRABALHOS RELACIONADOS

Neste Capítulo serão apresentados trabalhos relacionados ao problema proposto, com o objetivo de estudar problemas similares tratados em outras obras. Na Seção 4.1 será apresentada a definição para o Problema de Formação de Equipes em redes de sinais, assim como diversas definições de compatibilidade entre pares de indivíduos em uma rede e a apresentação de algoritmos heurísticos para solucionar o problema. A Seção 4.2 apresentará como grafos de sinais podem ser utilizados para entender o comportamento eleitoral, utilizando dados do Parlamento Europeu. Por fim, a Seção 4.3 apresentará uma análise do comportamento de políticos brasileiros realizada através de algoritmos de *clusterização* para redes de sinais.

4.1 Formando Equipes Compatíveis em Redes de Sinais

Analisando que os trabalhos anteriores relacionados ao TFP consideravam que todos os relacionamentos em uma rede eram positivos, Kouvatis *et al.* (2020) introduz o problema de Formação de Equipes em Redes de Sinais, do inglês *Team Formation in Signed Networks* (TFSN), onde é considerado que em uma rede social possam existir relacionamentos negativos, assim como positivos. Desta forma, tendo uma rede social representada por um grafo de sinais, um conjunto de habilidades necessários para realizar uma tarefa e uma relação de compatibilidade entre pares de indivíduos, o TFSN pode ser definido como o problema de encontrar subgrafos que contenham pares de indivíduos que sejam compatíveis e possuam as habilidades necessárias, minimizando um determinado custo de comunicação entre os indivíduos.

Kouvatis *et al.* (2020) definiram também um conjunto de relações de compatibilidade para medir o custo de comunicação entre indivíduos no TFSN. Utilizando da Teoria do Balanço Estrutural de Heider (1946), as relações são definidas, em suma, pela paridade de arestas negativas existentes em caminhos (caminhos mínimos e caminhos estruturalmente balanceados) entre dois indivíduos. Foi apresentada uma comparação entre as relações de compatibilidade, com o objetivo de mostrar o aumento da porcentagem de indivíduos compatíveis encontrados à medida que a noção de compatibilidade era relaxada.

Por fim, os autores apresentaram algoritmos heurísticos para solucionar o problema definido, utilizando dados de redes sociais online para realizar os experimentos. Foram apresentados dois algoritmos, o LCMD, que seleciona o indivíduo com a distância mínima, e o LCMC, que seleciona o usuário mais compatível com o time existente. Após a análise dos resultados

obtidos pelos algoritmos, foi constatado que o LCMD obteve melhores resultados.

4.2 Análise de Grafos de Sinais para a Interpretação do Comportamento Eleitoral

Visando realizar uma análise do comportamento de membros do Parlamento Europeu em votações, Arinik *et al.* (2017) utilizaram de técnicas de particionamento para problemas de *Correlation Clustering* e *Relaxed Correlation Clustering* e a partir dos resultados realizaram uma interpretação aprofundada de casos específicos.

Para isto, os autores realizaram uma extração dos dados dividida em quatro etapas. Para realização da extração foram utilizados dados brutos da atividade dos membros do Parlamento Europeu em sua 7^a legislatura. A primeira etapa da extração consistiu em selecionar um subconjunto dos dados, neste caso foram focadas alguns domínios políticos e países dos membros, considerando os anos da legislatura de forma separada. Na segunda etapa, cada membro foi comparado com relação à similaridade de votos, que pode variar entre +1, quando um par de membros possuem o mesmo voto, a -1, quando um membro do par votou “a favor” e outro “contra”, no caso de pelo menos um voto de “abstenção”, o valor 0 é atribuído. Esta comparação é realizada para cada proposta, com exceção daquelas que possuem até um membro ausente, e o cálculo da média entre elas resulta na similaridade média entre dois membros. A terceira etapa consiste na filtragem de valores de similaridade para remoção de valores muito próximos de 0, com objetivo de remover limitações observadas em trabalhos anteriores. Para esta filtragem foi utilizada uma abordagem automática por meio da técnica *k-means*. Por fim, a última etapa consiste na criação da rede de sinais, onde cada vértice representa um membro ativo do parlamento, e as arestas representam a similaridade existente entre um par de membros, que possuem similaridade diferente de 0. O resultado foi um total de 4150 instâncias de redes de sinais.

De posse das instâncias, os autores utilizaram de dois métodos de particionamento de grafos. Ambos os métodos utilizados para resolver os problemas de agrupamento *Correlation Clustering*, que neste caso consiste em encontrar partições no grafo em que o número de ligações positivas dentro dos grupos e negativas entre eles é maximizado (BANSAL *et al.*, 2002), e *Relaxed Correlation Clustering*, que pode permitir certos padrões de relacionamento que violam o conceito de balanço estrutural (DOREIAN; MRVAR, 2009). Para solução destes problemas foi utilizado o algoritmo heurístico *Iterated Local Search* e um algoritmo exato, onde foi verificado que heurístico acelerou o processo de particionamento e forneceu solução com

a mesma qualidade. A partir da análise qualitativa dos resultados obtidos foram identificados diversos padrões de grupos, alguns que seguiam divisões ideológicas tradicionais e outros que expuseram comportamentos mais específicos dentro do parlamento.

4.3 Análise do Balanço Estrutural do Congresso Brasileiro

Tendo como objetivo estudar o comportamento de políticos e partidos brasileiros, Levorato e Frota (2017) realizaram a extração e análise de diversas redes de sinais obtidas por meio de seções de votações realizadas no Congresso Nacional Brasileiro, utilizando as similaridades de votos entre os deputados para definir o peso das ligações no grafo que representa a rede formada entre eles.

Para a extração dos dados de votação em redes de sinais, os autores primeiramente realizaram uma comparação da similaridade de votos entre cada par de deputados, obtendo como resultado uma matriz de concordância M , onde cada elemento m_{uv} representa o nível de concordância entre dois deputados u e v considerando as propostas votadas em um período de tempo. Ao comparar os votos de dois deputados em uma proposta, a concordância entre eles pode ser determinada em um valor que varia de -1 (quando um vota “a favor” e outro “contra”) a +1 (quando ambos compartilham do mesmo voto). Neste caso, a opção de voto “obstrução” foi considerada como um voto “contra”. Para tratar a opção de “abstenção” foram consideradas duas abordagens. A primeira considera a “abstenção” como uma meia concordância, resultando no valor +0,5 quando comparada com votos “a favor”, “contra” ou outra “abstenção”. A segunda atribui valor +1 quando ambos deputados votam em “abstenção”, e valor 0 quando a “abstenção” é comparada à qualquer outro voto. Para obter concordância média entre os deputados em um período de tempo é realizado o cálculo da média dos valores de concordância para cada proposta dentro do período definido. Como resultado do uso das duas abordagens para o voto de “abstenção” foram gerados duas redes de votação para cada ano no período de 2010 a 2016.

Com os dados obtidos, os autores utilizaram do algoritmo *Iterated Local Search* para o problema de agrupamento *Correlation Clustering* com o objetivo de responder perguntas relacionadas à dinâmica da política. Os resultados obtidos serviram como base para realizar uma avaliação de aspectos como lealdade e liderança de partidos, coligações, crises e polarização, identificando aspectos como a infidelidade de partidos às suas coligações e a polarização existente entre dois grandes grupos políticos.

4.4 Relacionamento entre trabalhos

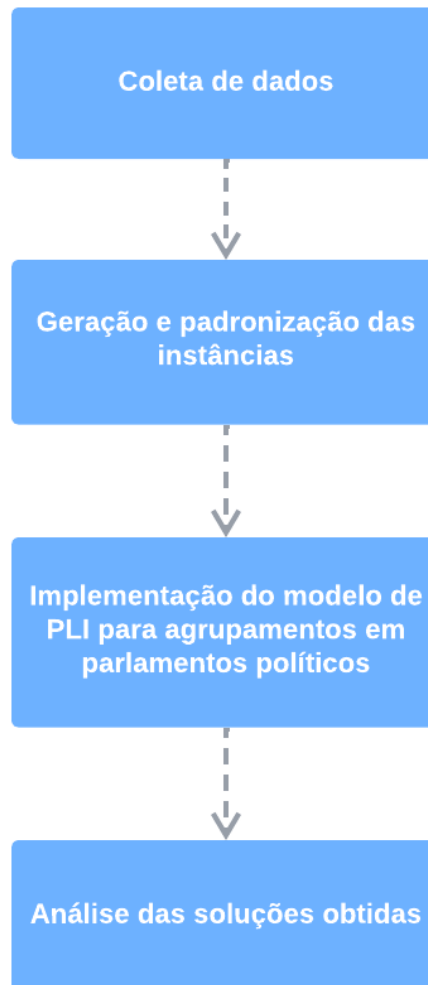
Assim como as obras de Levorato e Frota (2017) e Arinik *et al.* (2017), o presente trabalho se propõe a gerar e analisar agrupamentos criados a partir de dados de votações ocorridas em parlamentos políticos, mais especificamente, na Câmara dos Deputados do Brasil. Este trabalho busca gerar tais agrupamentos por meio da utilização de um modelo de PLI para resolução do CTFP, similar ao *Correlation Clustering* utilizado nas demais obras, que é um problema que utiliza dos conceitos de similaridade/dissimilaridade visando agrupar indivíduos com relações positivas em um mesmo conjunto, enquanto agrupa indivíduos com relações negativas em conjuntos distintos (BANSAL *et al.*, 2002). O diferencial da aplicação do CTFP é a possibilidade de requisitar que cada agrupamento (ou equipe) gerado possua um número mínimo de indivíduos com determinadas características específicas.

O trabalho de Kouvatis *et al.* (2020) se relaciona por introduzir o problema de formação de equipes utilizando grafos de sinais para representar a rede social, considerando as relações positivas e negativas entre indivíduos. O diferencial entre o TFSN e o CTFP se dá pelas definições de compatibilidade apresentadas pelo primeiro, que não consideram o problema proposto por Figueiredo (2021) no CTFP.

5 METODOLOGIA

As etapas metodológicas seguidas para a realização desta pesquisa são descritas mais detalhadamente a seguir. Abaixo é apresentado um diagrama que exemplifica o fluxo das atividades:

Figura 2 – Etapas da metodologia.



Fonte: Elaborado pela autora.

5.1 Coleta de dados

Nesta etapa foi realizada a coleta dos dados necessários para realização do agrupamento de políticos brasileiros. Os dados da Câmara dos Deputados foram obtidos através do trabalho de Levorato e Frota (2017) e contêm as seguintes informações: nome, código de cadastro na plataforma da Câmara, partido e estado, de cada deputado para o ano de 2010. Também

foram obtidos os grafos desenvolvidos na pesquisa de Levorato e Frota (2017), onde as arestas entre os deputados possuem como peso a concordância média entre eles. Como citado na Seção 4.3, para cada ano dentre 2010 e 2016 foram gerados dois grafos utilizando duas abordagens diferentes, desta forma, este trabalho utilizou apenas o grafo do ano de 2010 que considera o voto de “abstenção” como uma meia concordância (+0,5) quando comparado a qualquer outro voto. Dado que os valores dos pesos do grafo utilizado, por se tratarem de uma média, são em sua maioria valores decimais, um simples arredondamento para inteiro foi aplicado à cada um destes para transformá-lo em um grafo de sinais, conforme a definição na Seção 3.1.

5.2 Geração e padronização das instâncias

Como definido na Seção 3.2.2, uma instância é constituída por um grafo de sinais, um conjunto de habilidades e uma demanda de indivíduos para cada habilidade necessárias às equipes. Para a criação das instâncias foram utilizados os grafos de sinais descritos na Seção 5.1 e para a definição do conjunto de habilidades de cada indivíduo foi utilizada a abordagem que considera o estado do deputado como habilidade. Desta forma, utilizando dos dados de deputados obtidos, foi criado para o ano de 2010 uma matriz deputado \times habilidade, em que cada elemento x_{us} é um valor binário, indicando se o deputado u possui a habilidade s , sendo esta habilidade o estado de residência do deputado. A matriz resultante possui um tamanho de 545×27 , o que equivale ao número total de deputados federais deste ano pelo total de estados. A Figura 3 apresenta uma amostra da matriz deputado \times habilidade obtida.

Figura 3 – Amostra da matriz deputado \times estado para instância do ano de 2010.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fonte: Elaborado pela autora.

Por fim, a função de demanda de habilidades foi gerada com o objetivo de encontrar valores que não inviabilizassem a geração de uma solução e que, ao mesmo tempo, minimizassem o tempo de execução do modelo. A decisão pela minimização do tempo de execução do modelo se deu pelo fato do grafo de sinal estudado ser considerado de tamanho suficientemente grande para obter soluções ótimas em um tempo razoável (menos de 6 horas de execução). Desta forma, foi definido a criação de duas equipes. Considerando que a representação de certos estados na

Câmara é maior comparado a outros, também foi escolhido um valor proporcional ao total de deputados de cada estado. Assim, para este ano, a demanda determinada foi de aproximadamente $\frac{1}{5}$ do total de deputados de cada estado para cada uma das duas equipes, como apresentado na Figura 4.

Figura 4 – Relação entre a demanda de deputados e o total para cada estado no ano de 2010.

	Total de deputados	Demanda por equipe		Total de deputados	Demanda por equipe		Total de deputados	Demanda por equipe		Total de deputados	Demanda por equipe		Total de deputados	Demanda por equipe	
AC	8	2	CE	23	5	MG	53	11	PE	25	5	RO	10	2	
AL	9	2	DF	9	2	MS	8	2	PI	11	2	RR	9	2	
AM	8	2	ES	10	2	MT	10	2	PR	32	6	RS	32	6	
AP	9	2	GO	17	3	PA	18	4	RJ	51	10	SC	18	4	
BA	44	9	MA	20	4	PB	12	2	RN	8	2	SE	9	2	
													SP	72	14
													TO	10	2

Fonte: (COELHO *et al.*, 2023)

5.3 Implementação do modelo de PLI para agrupamentos em parlamentos políticos

Para implementação do modelo apresentado na Subseção 3.2.2.1 foi utilizada a biblioteca Python MIP¹, em sua versão 1.15.0. A execução das instâncias no modelo implementado foi realizada em um *notebook Python* no ambiente de execução em nuvem *Google Colaboratory*². O tempo total de execução desta instância foi de 1 hora e retornou como solução um total de 331 deputados alocados entre as duas equipes, de um total de 545.

5.4 Análise das soluções obtidas

Após a obtenção das soluções pelo modelo matemático para a instância em estudo, uma análise dos agrupamentos obtidos foi realizada. O foco central desta etapa foi a investigação das equipes formadas, buscando encontrar *insights* sobre a quantidade de deputados alocados por habilidade e quais questões influenciam os deputados a se agruparem ou não em um determinada equipe. Os resultados obtidos serão descritos mais detalhadamente no Capítulo 6.

¹ <https://www.python-mip.com/>

² <https://colab.research.google.com/>

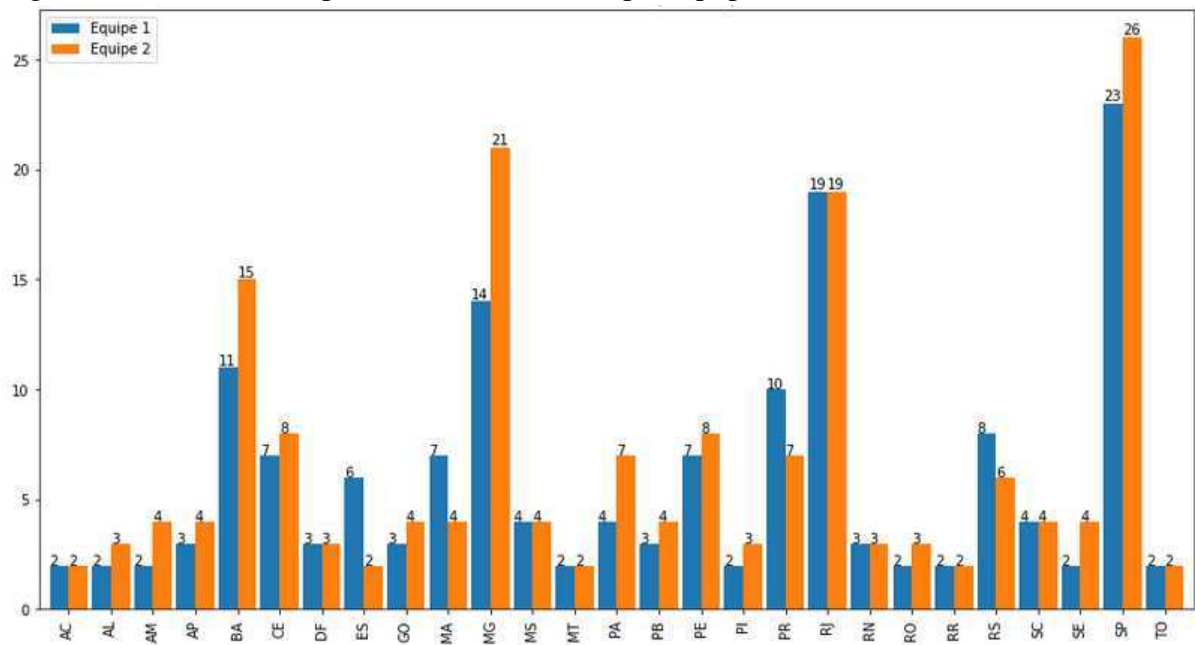
6 RESULTADOS

Neste Capítulo é apresentada a solução resultante da execução do modelo CTFP, considerando como entrada a instância descrita na Seção 5.2 e posteriormente a análise dos agrupamentos obtidos. A Seção 6.1 exibe as representações gráficas da solução enquanto a Seção 6.2 descreve os *insights* adquiridos a partir desta.

6.1 Solução obtida

Como citado na Seção 5.3, a execução do modelo retornou como resultado um total de 331 deputados alocados em uma das duas equipes, do total de 545 disponíveis. A primeira equipe alocou 157 deputados e a segunda 174. O número de deputados de cada estado em cada uma das equipes é mostrado na Figura 5.

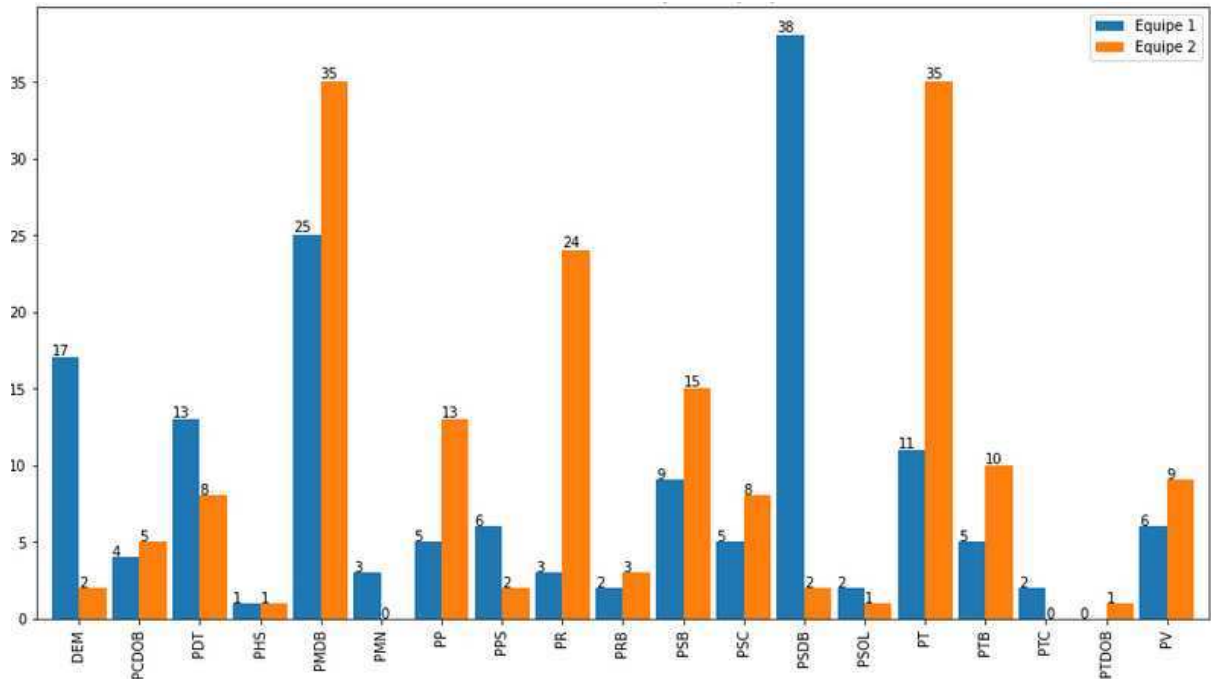
Figura 5 – Número de deputados de cada estado por equipe.



Fonte: (COELHO *et al.*, 2023)

Também foi gerado um gráfico que apresenta o número de deputados de cada partido em cada uma das equipes. O gráfico é apresentado na Figura 6.

Figura 6 – Número de deputados de cada partido por equipe.



Fonte: (COELHO *et al.*, 2023)

6.2 Análise da solução

Conforme observado na Figura 5, foi possível realizar uma distribuição equilibrada de deputados de um mesmo estado entre as equipes, como visto no estado do Rio de Janeiro, que alocou igualmente 19 deputados em ambas as equipes, indicando que há divergência de opiniões entre os deputados do estado. Este fato prediz que a região de residência pode não ter forte influência no posicionamento dos políticos nas votações parlamentares do ano de 2010.

Ao observar as equipes com relação ao partido político dos deputados alocados, pode-se notar uma forte influência de determinados partidos. Como exibido na Figura 6, pode-se citar como exemplo os partidos DEM, PMN, PSDB e PTC, os quais possuem a maioria dos deputados alocados na equipe 1, sendo um total de 60 políticos, representando 38,2% do total nessa equipe, enquanto na equipe 2, o total alocado é de apenas 4 parlamentares, representando somente 2,5% do total da equipe.

Ao realizar a análise da equipe 2, nota-se que os partidos PR e PT possuem a maioria de seus deputados alocados nela, com um total de 59 parlamentares, representando 34% do total, enquanto há apenas 14 políticos desses partidos alocados na equipe 1, 8% do total desta equipe. Além disso, pode-se notar que, apesar da influência dos partidos citados, determinados partidos ainda apresentam divergência de opiniões entre seus políticos. Podem ser citados como exemplo os partidos PCDOB, PHS, PRB e PSOL, que apresentaram uma divisão próxima de 50% de

seus deputados em cada uma das duas equipes e, além destes, há ainda outros partidos que apresentaram uma divisão quase igualitária de seus deputados entre as duas equipes. Destaca-se nesse caso o partido PMDB, que possuía uma quantidade considerável de políticos ativos na Câmara no ano de 2010, dos quais 41% dos deputados alocados foram agrupados na equipe 1, enquanto 59% foram agrupados na equipe 2, evidenciando uma divergência de opiniões existente neste partido nas votações realizadas no ano de 2010.

7 CONCLUSÕES E TRABALHOS FUTUROS

Esta monografia apresentou os resultados obtidos de um projeto de pesquisa que busca gerar, por meio da aplicação de um modelo de PLI para formação de equipes competitivas, e analisar agrupamentos de políticos da Câmara dos Deputados. Para a criação dos agrupamentos foram consideradas características pessoais dos deputados, neste caso o estado de residência e o partido político de cada um destes, além do nível de concordância/discordância dos votos de cada parlamentar nas votações parlamentares ocorridas no ano de 2010.

As análises realizadas a partir dos resultados obtidos indicaram que o estado de residência não teve grande influência nos votos dos deputados, dado que para maioria dos estados houve uma divisão balanceada de parlamentares entre as equipes. As análises também apontaram a influência de determinados partidos nos votos dos políticos, visto que certos partidos possuíam predominância de deputados alocados em somente uma das equipes, o que pode indicar a similaridade de opiniões entre os integrantes do partido. Por fim, também notaram-se casos que nem o estado, nem o partido foram influência para os votos dos parlamentares.

Os resultados obtidos por esta pesquisa foram publicados no formato de artigo científico, versão resumida, sendo apresentado no *XII Brazilian Workshop on Social Network Analysis and Mining* (BraSNAM) em João Pessoa (COELHO *et al.*, 2023). Em trabalhos futuros, pretende-se realizar a geração destes agrupamentos levando em consideração outras características como: faixa-etária, formação acadêmica, área de atuação e gênero, a fim de identificar possíveis novos padrões de voto, assim como utilizar o contexto e temas das votações analisadas. Além disso, pretende-se estender a análise para anos mais recentes e outros parlamentos, como o Parlamento Europeu.

REFERÊNCIAS

- ARINIK, N.; FIGUEIREDO, R.; LABATUT, V. Signed graph analysis for the interpretation of voting behavior. In: **International Workshop on Recommender Systems and Social Network Analysis**. [S.l.: s.n.], 2017. v. 2025. ISSN 16130073.
- BANSAL, N.; BLUM, A.; CHAWLA, S. Correlation clustering. In: **The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings**. [S.l.: s.n.], 2002. p. 238–247.
- CHARTRAND, G.; LESNIAK, L.; ZHANG, P. **Graphs & digraphs, fifth edition**. 5. ed. Caithness, UK: Whittles Publishing, 2010. (Textbooks in Mathematics).
- COELHO, N.; FIGUEIREDO, T.; FIGUEIREDO, R. Uso de programação linear inteira para geração e análise de agrupamentos de políticos da câmara dos deputados. In: **Anais do XII Brazilian Workshop on Social Network Analysis and Mining**. Porto Alegre, RS, Brasil: SBC, 2023. p. 246–251. ISSN 2595-6094. Disponível em: <<https://sol.sbc.org.br/index.php/brasnam/article/view/24807>>.
- DOREIAN, P.; MRVAR, A. Partitioning signed social networks. **Social Networks**, v. 31, n. 1, p. 1–11, 2009. ISSN 0378-8733. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0378873308000397>>.
- FIGUEIREDO, T. F. **Team formation problems: an integer linear optimization approach**. Tese (Doutorado) — Universidade Federal do Ceará, 2021.
- HEIDER, F. Attitudes and cognitive organization. **The Journal of psychology**, Taylor & Francis, v. 21, n. 1, p. 107–112, 1946.
- HILLIER, F. S.; LIEBERMAN, G. J. **Introdução à pesquisa operacional**. Porto Alegre, RS: AMGH Editora LTDA, 2013.
- KOUVATIS, I.; SEMERTZIDIS, K.; ZERVA, M.; PITOURA, E.; PANAYIOTIS, T. Forming compatible teams in signed networks. In: **Proceedings of the 23rd International Conference on Extending Database Technology (EDBT)**. [S.l.]: OpenProceedings.org, 2020. p. 363–366.
- LAPPAS, T.; LIU, K.; TERZI, E. Finding a team of experts in social networks. In: **Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2009. (KDD '09), p. 467–476. ISBN 9781605584959. Disponível em: <<https://doi.org/10.1145/1557019.1557074>>.
- LEVORATO, M.; FROTA, Y. Brazilian Congress structural balance analysis. **Journal of Interdisciplinary Methodologies and Issues in Sciences**, Vol. 2 - Graphes et systèmes sociaux, 2017. Disponível em: <<https://jimis.episciences.org/3185>>.
- WOLSEY, L. A. **Integer Programming**. 2. ed. Nashville, TN: John Wiley & Sons, 2020.