



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS**  
**DEPARTAMENTO DE COMPUTAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA COMPUTAÇÃO**

**JOSÉ ROBERTTY DE FREITAS COSTA**

**O PROBLEMA DA K-FLORESTA GERADORA BALANCEADA EM ARESTAS**

**FORTALEZA**

**2024**

JOSÉ ROBERTTY DE FREITAS COSTA

O PROBLEMA DA K-FLORESTA GERADORA BALANCEADA EM ARESTAS

Dissertação apresentada ao Programa de Pós-Graduação em Ciências da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Ciência da Computação. Área de concentração: Teoria da Computação.

Orientador: Prof. Dr. Manoel Bezerra Campêlo Neto.

FORTALEZA

2024

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)

---

C873p Costa, José Robertty de Freitas.

O Problema da k-Floresta Geradora Balanceada em Arestas / José Robertty de Freitas Costa. – 2024.  
57 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Ciência da Computação, Fortaleza, 2024.

Orientação: Prof. Dr. Manoel Bezerra Campêlo Neto.

1. k-floresta geradora balanceada em arestas. 2. Programação matemática. 3. Algoritmos heurísticos. I. Título.

CDD 005

---

JOSÉ ROBERTTY DE FREITAS COSTA

O PROBLEMA DA K-FLORESTA GERADORA BALANCEADA EM ARESTAS

Dissertação apresentada ao Programa de Pós-Graduação em Ciências da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre em Ciência da Computação. Área de concentração: Teoria da Computação.

Aprovada em: 29/04/2021

BANCA EXAMINADORA

---

Prof. Dr. Manoel Bezerra Campêlo Neto (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Fábio Carlos Sousa Dias  
Universidade Federal do Ceará (UFC)

---

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

---

Prof. Dr. Rafael Castro de Andrade  
Universidade Federal do Ceará (UFC)

A Deus.

Aos meus pais, Francisco e Neide.

A minha esposa, Thyara.

## **AGRADECIMENTOS**

Agradeço a Deus, por me dar forças e saúde durante esta caminhada.

Agradeço aos meus pais, Francisco e Neide, por todo sacrifício feito por eles, para que eu pudesse ter acesso a uma educação de qualidade.

Agradeço a minha esposa, Thyara Araújo, por estar sempre ao meu lado me ajudando e me dando forças em todos os momentos.

Agradeço ao professor Manoel Campêlo, pela sua dedicação e paciência ao me orientar neste trabalho. Além de ser um ótimo professor e pesquisador, é uma pessoa excelente.

Agradeço aos professores Dr. Rafael Castro de Andrade e Dr. Victor Almeida Campos, por estarem sempre dispostos a contribuir com o trabalho desenvolvido.

Agradeço aos membros da banca, pela disponibilidade em participar da banca desse trabalho, e pelas suas colaborações e sugestões.

Agradeço aos meus amigos Gabriel Sousa, Emanuel Elias, Claro Henrique, Fábio Dias, Atílio Gomes, Paulo Miranda, Marcelo Martins e Lucas Cruz, que me ajudaram de forma significativa.

Agradeço aos professores Dr. Antônio Clecio Fontelles Thomaz e Dr. Carlos Arthur Sobreira Rocha por toda a ajuda fornecida.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

“An equation means nothing to me unless it expresses a thought of God.”

(Srinivasa Ramanujan)

## RESUMO

O problema da  $k$ -Floresta Geradora Balanceada em Arestas ( $k$ -FGBA) consiste em, dados um grafo  $G$  não direcionado com pesos nas arestas e um inteiro  $k$ , encontrar uma floresta geradora de  $G$  com  $k$  árvores, de forma a minimizar o peso da árvore mais pesada. No caso particular  $k = 1$ , o problema se resume a determinar uma Árvore Geradora Mínima. O problema foi introduzido em 2017, motivado por aplicação em topologia computacional. O  $k$ -FGBA já foi demonstrado ser NP-Difícil mesmo para  $k = 2$  ou para o caso de pesos unitários. Neste trabalho apresentamos e provamos limitantes para o problema e um algoritmo de Programação Dinâmica para o caso onde o grafo de entrada é um caminho. Desenvolvemos duas novas heurísticas para o problema. Além disso, propomos três formulações de Programação Linear Inteira Mista (PLIM) para o  $k$ -FGBA, junto com desigualdades válidas e um procedimento de fixação de variáveis que objetivam fortalecer os modelos. Geramos um conjunto de instâncias de forma aleatória com as quais avaliamos a qualidade dos procedimentos e formulações propostas. Os testes computacionais sugerem que as heurísticas propostas podem ser adotadas como bons limitantes superiores para o problema. Ademais, as formulações  $F\_REP^+$  e  $F\_ROT^+$  se destacaram, conseguindo encontrar uma solução ótima dentro do limite de tempo estabelecido em 69,05% e 73,87%, respectivamente, das instâncias avaliadas.

**Palavras-chave:**  $k$ -floresta geradora balanceada em arestas; programação matemática; algoritmos heurísticos.



## ABSTRACT

The Edge Balanced Spanning  $k$ -Forest ( $k$ -EBSF) problem consists in, given an undirected edge-weighted graph  $G$  and an integer  $k$ , finding a spanning forest of  $G$  with  $k$  trees, in order to minimize the weight of the heaviest tree. In the particular case  $k = 1$ , the problem comes down to determining a Minimum Spanning Tree. The problem was introduced in 2017 motivated by applications in computational topology. The  $k$ -EBSF has already been shown to be NP-Hard even for  $k = 2$  or for the unit weight case. In this work, we present and prove bounds for the problem and a Dynamic Programming algorithm for the case where the input graph is a path. We introduce two new heuristics for the problem. Furthermore, we propose three Mixed Integer Linear Programming (MILP) formulations for the  $k$ -EBSF together with valid inequalities and a variable fixing procedure that aim to strengthen the models. We generated a set of instances to evaluate the quality of the proposed procedures and formulations. Computational tests suggest that the proposed heuristics can be adopted as good upper bounds for the problem. In addition, the formulations  $F\_REP^+$  and  $F\_ROT^+$  stood out, managing to find an optimal solution within the time limit in 69.05% and 73.87%, respectively, of the evaluated instances.

**Keywords:** edge-balanced spanning  $k$ -Forest; mathematical programming; heuristic algorithms.

## LISTA DE FIGURAS

Figura 1 – Exemplo de Instância e Solução para k-FGBA . . . . .	15
Figura 2 – Exemplo de Grafo Simples . . . . .	19
Figura 3 – Exemplos de Grafos . . . . .	20
Figura 4 – Exemplo de Árvore Geradora Mínima . . . . .	22
Figura 5 – Exemplo de Aplicação do Algoritmo Kruskal . . . . .	23
Figura 6 – Digrafo Simétrico Associado . . . . .	25
Figura 7 – Exemplo de particionamento da solução em vértices e em arestas para $k = 3$	35
Figura 8 – Particionamento dos vértices via rotulação de árvores . . . . .	36
Figura 9 – Particionamento dos vértices via vértices representantes . . . . .	36
Figura 10 – Exemplo de grafo $G'$ com $k = 3$ . . . . .	37
Figura 11 – Exemplo de solução na modelagem de particionamento de arestas . . . . .	37
Figura 12 – Solução associada à Tabela 2 . . . . .	39
Figura 13 – Comparação entre os modelos propostos em relação à densidade e ao número de vértices. . . . .	48

## LISTA DE TABELAS

Tabela 1 – Resumo dos resultados em relação ao parâmetro $k$ . . . . .	33
Tabela 2 – Exemplo de matriz $[x_{ut}]$ . . . . .	39
Tabela 3 – Comparação dos modelos em relação ao número de variáveis e restrições . .	43
Tabela 4 – Relaxação linear dos modelos em relação ao parâmetro $k$ . . . . .	44
Tabela 5 – Relaxação linear comparada ao Limite Inferior em relação ao parâmetro $k$ .	45
Tabela 6 – Resumo dos resultados em relação ao parâmetro $k$ . . . . .	47

## LISTA DE ABREVIATURAS E SIGLAS

$BCP_k$	k-Partição Conexa Balanceada
k-FGBA	k-Floresta Geradora Balanceada em Arestas
PI	Programação Inteira
PL	Programação Linear
PLIM	Programação Linear Inteira Mista
TLE	<i>Time Limit Exceeded</i>

## LISTA DE SÍMBOLOS

$V$	Conjunto de vértices
$E$	Conjunto de arestas
$A$	Conjunto de arcos
$A(G)$	Conjunto de arcos da orientação simétrica do grafo $G$
$d(u, v)$	Distância entre os vértices $u$ e $v$
$N(u)$	Vizinhança do vértice $u$ em um grafo não direcionado
$N^+(u)$	Vizinhança positiva do vértice $u$ em um grafo direcionado
$N^-(u)$	Vizinhança negativa do vértices $u$ em um grafo direcionado
$\mathbb{R}$	Conjunto dos números reais
$\mathbb{B}$	Conjunto $\{0, 1\}$
$n$	Cardinalidade de $V$
$m$	Cardinalidade de $E$

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
<b>2</b>	<b>CONCEITOS PRELIMINARES</b>	<b>17</b>
<b>2.1</b>	<b>Programação Matemática</b>	<b>17</b>
<b>2.2</b>	<b>Teoria dos Grafos</b>	<b>19</b>
<b>3</b>	<b>K-FLORESTA GERADORA BALANCEADA EM ARESTAS</b>	<b>26</b>
<b>3.1</b>	<b>Definição</b>	<b>26</b>
<b>3.2</b>	<b>Complexidade Computacional</b>	<b>26</b>
<b>3.3</b>	<b>Limitantes</b>	<b>28</b>
<b>4</b>	<b>ALGORITMOS HEURÍSTICOS</b>	<b>30</b>
<b>4.1</b>	<b>Algoritmos Propostos</b>	<b>30</b>
<b>4.2</b>	<b>Comparação dos algoritmos</b>	<b>33</b>
<b>5</b>	<b>FORMULAÇÕES MATEMÁTICAS</b>	<b>35</b>
<b>5.1</b>	<b>Estratégia de Modelagem do problema</b>	<b>35</b>
<b>5.2</b>	<b>Modelo de Árvore Rotulada</b>	<b>38</b>
<b>5.3</b>	<b>Modelo de Representantes</b>	<b>40</b>
<b>5.4</b>	<b>Modelo de Fluxo</b>	<b>41</b>
<b>5.5</b>	<b>Desigualdades Válidas</b>	<b>42</b>
<b>5.6</b>	<b>Procedimento de Fixação de Variáveis</b>	<b>42</b>
<b>5.7</b>	<b>Comparação entre as formulações</b>	<b>43</b>
<b>6</b>	<b>EXPERIMENTOS COMPUTACIONAIS</b>	<b>46</b>
<b>6.1</b>	<b>Instâncias de Teste</b>	<b>46</b>
<b>6.2</b>	<b>Ambiente Computacional</b>	<b>46</b>
<b>6.3</b>	<b>Avaliação dos Modelos</b>	<b>47</b>
<b>7</b>	<b>CONCLUSÃO</b>	<b>50</b>
	<b>REFERÊNCIAS</b>	<b>51</b>
	<b>APÊNDICE A- RESULTADOS DAS FORMULAÇÕES</b>	<b>53</b>

## 1 INTRODUÇÃO

Grafos são estruturas matemáticas abstratas que nos permitem modelar diferentes problemas. Essa estrutura consiste em um par ordenado  $(V, E)$ , onde  $V$  é o conjunto de vértices e  $E$ , o conjunto de arestas. Um vértice  $u \in V$  representa um objeto e uma aresta  $uv \in E$  representa uma relação entre dois objetos  $u, v \in V$ . Um grafo  $H(V', E')$  é chamado subgrafo gerador de um grafo  $G(V, E)$  quando  $V' = V$  e  $E' \subseteq E$ , ou seja, possui os mesmos vértices de  $G$  e um subconjunto de suas arestas.

Uma questão fundamental em grafos é como a relação definida pelas arestas se estende entre os vértices, conectando-os. Nesse sentido, define-se um caminho em um grafo  $G(V, E)$  como uma sequência de vértices distintos  $\langle v_0, v_1, \dots, v_p \rangle$  onde  $v_i v_{i+1} \in E$ , para todo  $i \in \{0, \dots, p-1\}$ . Esse caminho conecta os vértices  $v_0$  e  $v_p$ . Se há pelo menos (resp. no máximo) um caminho conectando cada par de vértices, o grafo é dito conexo (resp. acíclico). Um grafo conexo e acíclico é uma árvore. Uma floresta geradora  $F$  de  $G(V, E)$  consiste em um subgrafo gerador acíclico. Se  $F$  é também conexo, dizemos que  $F$  é uma árvore geradora.

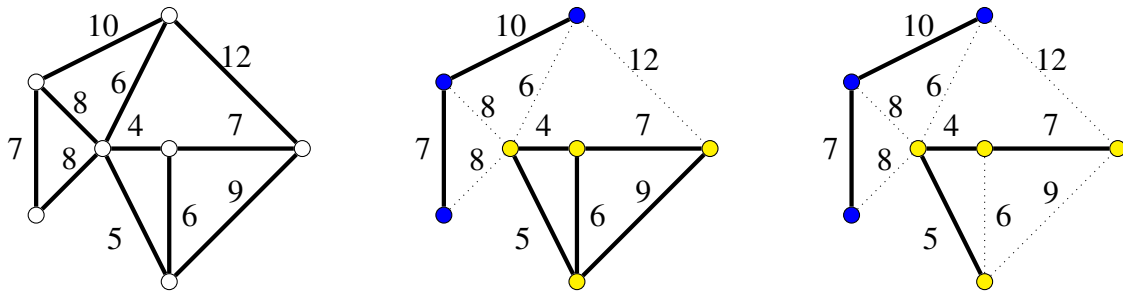
Árvores e florestas geradoras são estruturas bastante úteis para descrever e estudar conectividade mínima (sem redundância) em um universo ou subconjuntos desse universo. Diversas aplicações são encontradas na literatura para problemas que podem ser modelados por meio dessas estruturas: rede de sensores (MIN *et al.*, 2006), geometria computacional (PREPARATA; SHAMOS, 2012), análise genética (EISEN *et al.*, 1998), processamento de imagens (LOVÁSZ, 1977), e entre outras. Essa gama de aplicações leva a uma grande quantidade de variações de problemas de otimização sobre florestas e árvores.

O problema da  $k$ -Floresta Geradora Balanceada em Arestas ( $k$ -FGBA) consiste em, dados um grafo  $G$  não direcionado com pesos/custos nas arestas e um inteiro  $k \geq 1$ , encontrar uma floresta geradora de  $G$  com exatamente  $k$  árvores, de forma a minimizar o peso<sup>1</sup> da árvore mais pesada. Para  $k = 1$ , resume-se a encontrar árvore geradora mínima do grafo  $G$ .

A Figura 1 apresenta um exemplo de instância e solução viável para o problema considerando  $k = 2$ . O grafo exemplo é particionado em dois subgrafos conexos, identificados pelas cores azul e amarelo na figura. Destacamos em seguida uma árvore geradora mínima de cada subgrafo, de peso 17 e 16, respectivamente. Portanto, a solução apresentada tem custo 17.

<sup>1</sup> O peso de uma árvore é a soma dos pesos de suas arestas.

Figura 1 – Exemplo de Instância e Solução para k-FGBA



Fonte: elaborado pelo autor.

Este problema foi apresentado em Madkour *et al.* (2017), sob a denominação de k-Floresta Geradora Mínima (do inglês, *Minimum Spanning k-Forest*), motivado por uma aplicação de topologia computacional do software RIVET. Nessa aplicação se necessita realizar cálculos (pesados) em vértices de um dado grafo e o cálculo em um vértice pode ser reutilizado para aquele a ser efetuado em um vértice adjacente. Deseja-se paralelizar o processo em diversas máquinas, para isso devemos dividir o grafo em  $k$  partições conexas de forma a minimizar o tempo final de execução.

Apesar da denominação inicial do problema em Madkour *et al.* (2017) ser diferente, utilizamos aqui a designação k-Floresta Geradora Balanceada em Arestas por considerá-la mais representativa da descrição do problema e pela denominação similar vir sendo empregada mais recentemente para nomear problemas semelhantes (ver por exemplo (ANDERSSON *et al.*, 2003; CHEN *et al.*, 2020; MIYAZAWA *et al.*, 2020; WANG *et al.*, 2013)).

O problema k-FGBA foi demonstrado ser NP-Difícil já para  $k = 2$  (MADKOUR *et al.*, 2017) ou mesmo que os pesos sejam unitários (VAISHALI *et al.*, 2018). No trabalho de Madkour *et al.* (2017) são apresentadas duas heurísticas para o problema. Vaishali *et al.* (2018) apresentam um algoritmo polinomial para quando o grafo é uma árvore. Em (ANDERSSON *et al.*, 2003), o k-FBGA é estudado quando os vértices são pontos em um plano e o custo das arestas refere-se à distância euclidiana entre seus vértices extremos. O trabalho apresenta algoritmos aproximativos para os casos  $k = 2$  e  $k \geq 3$ .

Uma variação do problema, apresentada em Guttmann-Beck e Hassin (1997), visa, dado um grafo completo com pesos nas arestas e um inteiro  $k$ , determinar  $k$  subconjuntos disjuntos de vértices de mesmo tamanho, minimizando a árvore geradora mínima de maior custo. Este problema pode ser visto como o k-FGBA com a adição da restrição de que as partições de vértices possuam mesmo tamanho. Este caso também é NP-Difícil, e na literatura podemos



encontrar algoritmos aproximativos (GUTTMANN-BECK; HASSIN, 1997; GUTTMANN-BECK; HASSIN, 1998).

Outras variações também são mencionadas na literatura, porém ainda muito pouco ou nada exploradas. Estas variações referem-se a modificações na função objetivo, tais como, maximizar o custo da árvore de menor custo ou minimizar a diferença de custo das árvores de maior e menor custo.

O k-FGBA está bastante relacionado com o problema da k-Partição Conexa Balanceada ( $BCP_k$ ). Assim como o k-FGBA, o  $BCP_k$  busca determinar uma floresta com  $k$  árvores de forma a balancear os custos das árvores, ou seja, procurando equilibrar os custos das árvores. Porém, neste problema os custos estão associados aos vértices (e não às arestas). O  $BCP_k$  é bastante explorado na literatura, possuindo algumas variações, dadas pela modificação da função objetivo. As versões mais estudadas consistem em minimizar a maior componente (min-max  $BCP_k$ ) ou maximizar a menor componente (max-min  $BCP_k$ ). Os dois problemas já foram demonstrados ser NP-Difíceis mesmo para grafos bipartidos com custos unitários (DYER; FRIEZE, 1985). Ademais, já foram propostos algoritmos aproximativos para estes problemas para os casos gerais e para casos com  $k$  fixo (CHEN *et al.*, 2020).

O objetivo deste trabalho é propor métodos heurísticos e exatos para o problema k-FGBA e comparar os mesmos com as abordagens já existentes na literatura.

O restante do texto está organizado como segue. O Capítulo 2 apresenta alguns conceitos preliminares considerados relevantes para o entendimento do trabalho. No Capítulo 3 é apresentada a definição formal do problema e resultados teóricos relevantes. No Capítulo 4 são apresentadas e analisadas as heurísticas propostas. Já no Capítulo 5 são discutidas formulações e desigualdades válidas. No Capítulo 6 são apresentados os experimentos computacionais realizados e seus respectivos resultados. Por fim, no Capítulo 7 são apresentadas as conclusões e apontados trabalhos futuros.

## 2 CONCEITOS PRELIMINARES

Neste capítulo são apresentados os principais conceitos utilizados no decorrer do texto. Na Seção 2.1 são abordados os conceitos relacionados à Programação Linear/Inteira e Desigualdades Válidas. Na Seção 2.2 são apresentados conceitos e notações de Teoria dos Grafos. Detalhes mais aprofundados nestes temas podem ser obtidos em Cormen *et al.* (2002), Hillier e Lieberman (2013), West *et al.* (2001), Wolsey (1998) e Wu e Chao (2004).

### 2.1 Programação Matemática

Problemas de otimização são aqueles que envolvem encontrar a melhor solução possível para uma determinada situação, considerando certas restrições. Usualmente, esses problemas podem ser descritos por um modelo matemático, consistindo de expressões matemáticas definidas a partir de variáveis (de decisão). Tais expressões descrevem tanto as restrições do problema quanto o ranqueamento de suas soluções viáveis. Nesse contexto, denominamos solução qualquer atribuição de valores às variáveis do modelo matemático. Uma solução é dita viável quando a mesma atende a todas as restrições, e inviável, caso contrário. Uma solução é dita ótima se ela possui o maior (no caso de maximização) ou menor (no caso de minimização) valor possível de função objetivo (que ranqueia as soluções) dentre as soluções viáveis.

Programação Linear (PL) é uma subárea da otimização matemática. O termo linear refere-se ao fato de que todas as expressões matemáticas usadas no modelo são lineares nas variáveis. Além disso, neste tipo de modelo, todas as variáveis utilizadas são contínuas. Há um número finito de variáveis e restrições. Mantendo essas propriedades principais, podemos obter diferentes modelos matemáticos, porém todos eles podem ser convertidos para formas específicas com características adicionais.

Em particular, na forma canônica, todas as restrições do problema são expressas como desigualdades. Especificamente, em um problema de maximização, essas restrições são do tipo "menor ou igual a" ( $\leq$ ). Além disso, as variáveis são não negativas. Sendo assim, considerando  $x \in \mathbb{R}^n$  o vetor das variáveis,  $c \in \mathbb{R}^n$  e  $b \in \mathbb{R}^m$  vetores constantes com coeficientes da função objetivo e termos independentes das restrições, e  $A \in \mathbb{R}^{m \times n}$  a matriz de coeficientes das restrições, o modelo geral de PL na forma canônica é dado por:

$$\begin{aligned}
 & \max \quad c^T x \\
 & \text{sujeito a} \\
 & \quad Ax \leq b \\
 & \quad x \geq 0
 \end{aligned} \tag{2.1}$$

Mostra-se que as variáveis não nulas de uma solução ótima (caso exista alguma) são dadas pela resolução de um sistema linear  $A'x' = b'$ , onde  $A'$  é uma submatriz quadrada de  $A$  e  $b'$  o subvetor correspondente em  $b$ . Soluções assim obtidas, são chamadas básicas. A partir dessa propriedade, em 1947, George Dantzig criou um algoritmo eficiente, conhecido como Simplex, para a resolução de problemas de PL. O Simplex é um método iterativo exato que busca resolver de forma ótima problemas de PL. O algoritmo possui um procedimento que permite identificar, a partir de qualquer solução viável básica, uma nova solução básica viável melhor, ou permite terminar a busca quando não for mais possível encontrar tal solução. Há outros métodos para resolver problemas de PL, como os métodos de pontos interiores.

Modelos de Programação Linear Inteira, ou somente Programação Inteira (PI), são modelos de PL acrescidos de restrições de integralidade sobre as variáveis, ou seja, estas precisam assumir um valor inteiro. Outro caso particular é quando as variáveis são inteiras e restritas aos valores zero ou um, esse caso é conhecido como Programação Linear Inteira Binária. Quando no modelo há algumas variáveis contínuas e outras inteiras, temos o que denominamos de Programação Linear Inteira Mista (PLIM).

Um conceito muito importante em PI é o de relaxação linear. Dado um modelo de PI, a relaxação linear é obtida removendo as restrições de integralidade das variáveis. Note que uma solução ótima da relaxação linear pode ser obtida utilizando algoritmos como o Simplex. Além disso, a relaxação linear nos fornece limitantes para o problema original, que são bastante úteis nos métodos de resolução em PI. Tipicamente, esses métodos realizam buscas na região das soluções viáveis, e algumas sub-regiões podem ser descartadas a partir desses limitantes.

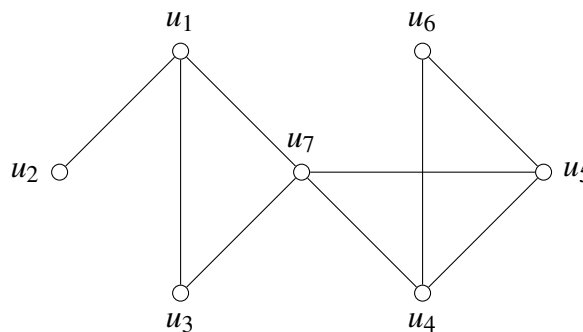
Apesar dos modelos matemáticos representarem muitos problemas de forma precisa, as soluções destes modelos podem ser extremamente custosas, principalmente em relação a tempo computacional. Com a finalidade de contornar esta situação, podemos modificar o conjunto de restrições do modelo. Neste contexto, uma das estratégias utilizadas é a adição de desigualdades válidas.

Considere  $P$  um problema de Programação Inteira e  $P'$  o modelo de Programação Inteira obtido adicionando uma restrição qualquer. Esta restrição é dita desigualdade válida se, e somente se, toda solução viável em  $P$  também é viável em  $P'$ . Note que, embora essa restrição adicional não elimine soluções viáveis do problema de programação inteira, ela pode descartar soluções viáveis de sua relaxação linear, ou seja, pode haver solução que seja viável na relaxação linear de  $P$ , mas que não seja viável na relaxação linear de  $P'$ . Esse fato pode levar à melhoria dos limitantes e, assim, contribuir para a resolução mais eficiente do modelo.

## 2.2 Teoria dos Grafos

Um grafo simples  $G$  consiste em um par ordenado  $(V, E)$ , onde  $V$  é um conjunto não vazio de vértices e  $E$  conjunto de arestas. Um vértice  $u \in V$  representa um objeto e uma aresta  $\{u, v\} \in E$  (denotada simplesmente por  $uv$ ) representa uma relação entre dois vértices distintos  $u, v \in V$ . A Figura 2 exibe uma representação gráfica de um grafo simples, onde  $V = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$  e  $E = \{u_1u_2, u_1u_3, u_1u_7, u_3u_7, u_4u_5, u_4u_6, u_4u_7, u_5u_6, u_5u_7\}$ . Dois vértices  $u, v \in V$  são ditos adjacentes quando existe  $uv \in E$ . Denotamos  $N_G(u)$  o subconjunto de vértices adjacentes a  $u$  no grafo  $G$ , ou seja,  $N_G(u) = \{v \in V : uv \in E\}$ . Comumente utiliza-se  $n = |V|$  e  $m = |E|$ .

Figura 2 – Exemplo de Grafo Simples



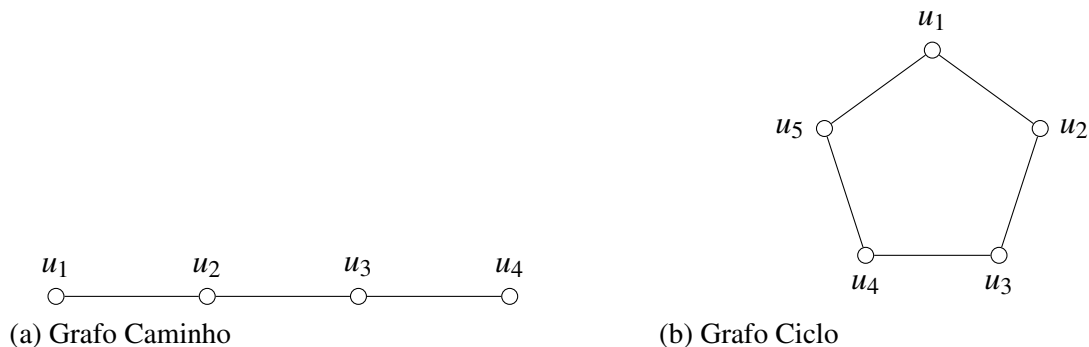
Fonte: elaborado pelo autor.

Considere um grafo  $G(V, E)$ . Um grafo  $H(V', E')$  é dito subgrafo de  $G$  quando  $V' \subseteq V$  e  $E' \subseteq E$ . Além disso, dizemos que  $H$  é subgrafo gerador de  $G$  quando  $V' = V$ .  $H$  é dito subgrafo induzido por  $V'$  se, e somente se,  $H$  é subgrafo e para todos  $u, v \in V'$  temos que  $uv \in E'$  se  $uv \in E$ . Denotamos por  $G[V']$  o subgrafo de  $G$  induzido por  $V' \subseteq V$ . Um subgrafo  $H$  é dito clique quando, para todo,  $u, v \in V' : u \neq v$ , temos que  $uv \in E'$ .

Um caminho é um grafo cujos vértices podem ser dispostos em uma sequência linear

de modo que dois vértices são adjacentes se, e somente se, eles são consecutivos na sequência. O primeiro e o último vértices da sequência são os extremos do caminho. Já um ciclo consiste em um grafo cujos vértices podem ser dispostos em uma sequência cíclica de modo que dois vértices são adjacentes se, e somente se, eles são consecutivos na sequência. A Figura 3 apresenta exemplos de grafos caminho e ciclo.

Figura 3 – Exemplos de Grafos



Fonte: elaborado pelo autor.

Um  $(u, v)$ -caminho em um grafo consiste em um subgrafo caminho cujos extremos são os vértices  $u$  e  $v$ . Um grafo é dito conexo quando, para quaisquer  $u, v \in V$ , existe pelo menos um  $(u, v)$ -caminho. Um grafo é dito acíclico se não contém um subgrafo ciclo. Todo grafo acíclico é denominado floresta. Todo grafo acíclico e conexo é denominado árvore. O Teorema 1 apresenta um conjunto de proposições que caracterizam uma árvore.

**Teorema 1** (WEST, 2001). *Temos que as seguintes proposições são equivalentes:*

1.  $G$  é acíclico e conexo.
2.  $G$  é conexo e  $|E| = |V| - 1$
3.  $G$  é acíclico e  $|E| = |V| - 1$
4. Para todo  $u, v \in V$ ,  $G$  possui exatamente um  $(u, v)$ -caminho.

Um grafo é dito ponderado (em arestas) quando existe uma função  $c: E \rightarrow \mathbb{R}$  que atribui um peso para cada aresta. O peso (ou custo) de um subgrafo ponderado é a soma dos pesos de suas arestas. A distância entre dois vértices  $u, v$  em um grafo ponderado, denotada por  $d(u, v)$ , é o menor peso (custo) de um  $(u, v)$ -caminho no grafo.

Existem diversos algoritmos na literatura para determinar um  $(u, v)$ -caminho (de peso) mínimo em grafos onde  $c(uv) \geq 0$ , para todo  $uv \in E$ , ou em grafos que não possuem ciclo (de peso) negativo. Em particular, o algoritmo Dijkstra permite calcular o menor caminho de um

vértice específico a cada um dos demais vértices. Já o algoritmo Floyd-Warshall permite calcular o menor caminho entre todos os pares de vértices do grafo.

O Algoritmo 1 apresenta o pseudocódigo do algoritmo Floyd-Warshall, que possui como entrada um grafo ponderado  $G$  e retorna a matriz de distância entre os pares de vértices.

---

**Algoritmo 1:** Floyd-Warshall

---

```

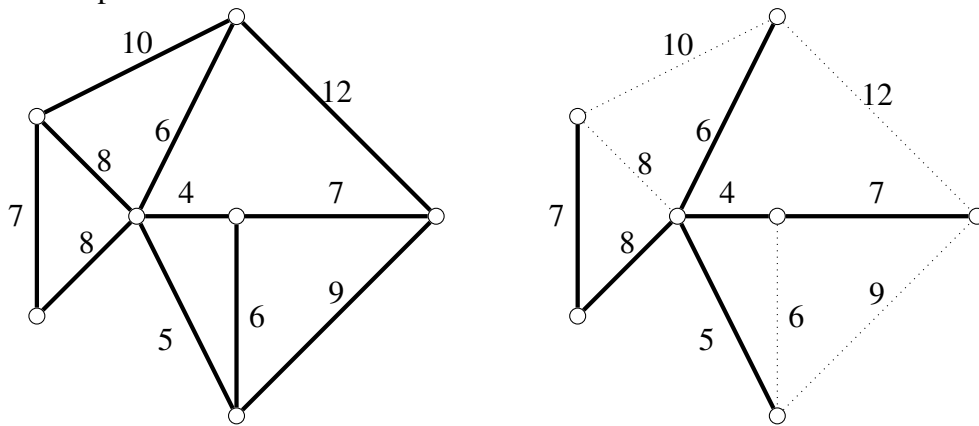
1 início
2   Para  $u \in V$  faça
3      $D_{uu} \leftarrow 0$ 
4     Para  $v \in V \setminus \{u\}$  faça
5       Se  $uv \in E$  então
6          $D_{uv} \leftarrow c(uv)$ 
7       Senão
8          $D_{uv} \leftarrow \infty$ 
9     Para  $u \in V$  faça
10      Para  $v \in V$  faça
11        Para  $w \in V$  faça
12          Se  $D_{uv} > D_{uw} + D_{wv}$  então
13             $D_{uv} \leftarrow D_{uw} + D_{wv}$ 
14  Retorna  $D$ 

```

---

Denominamos árvore geradora de um grafo, um subgrafo gerador que é uma árvore. Uma árvore geradora mínima em um grafo ponderado  $G$ , consiste na árvore geradora de menor peso (custo) possível. A Figura 4 apresenta um exemplo de um grafo e de uma a árvore geradora mínima, que possui custo total igual a 37.

Figura 4 – Exemplo de Árvore Geradora Mínima



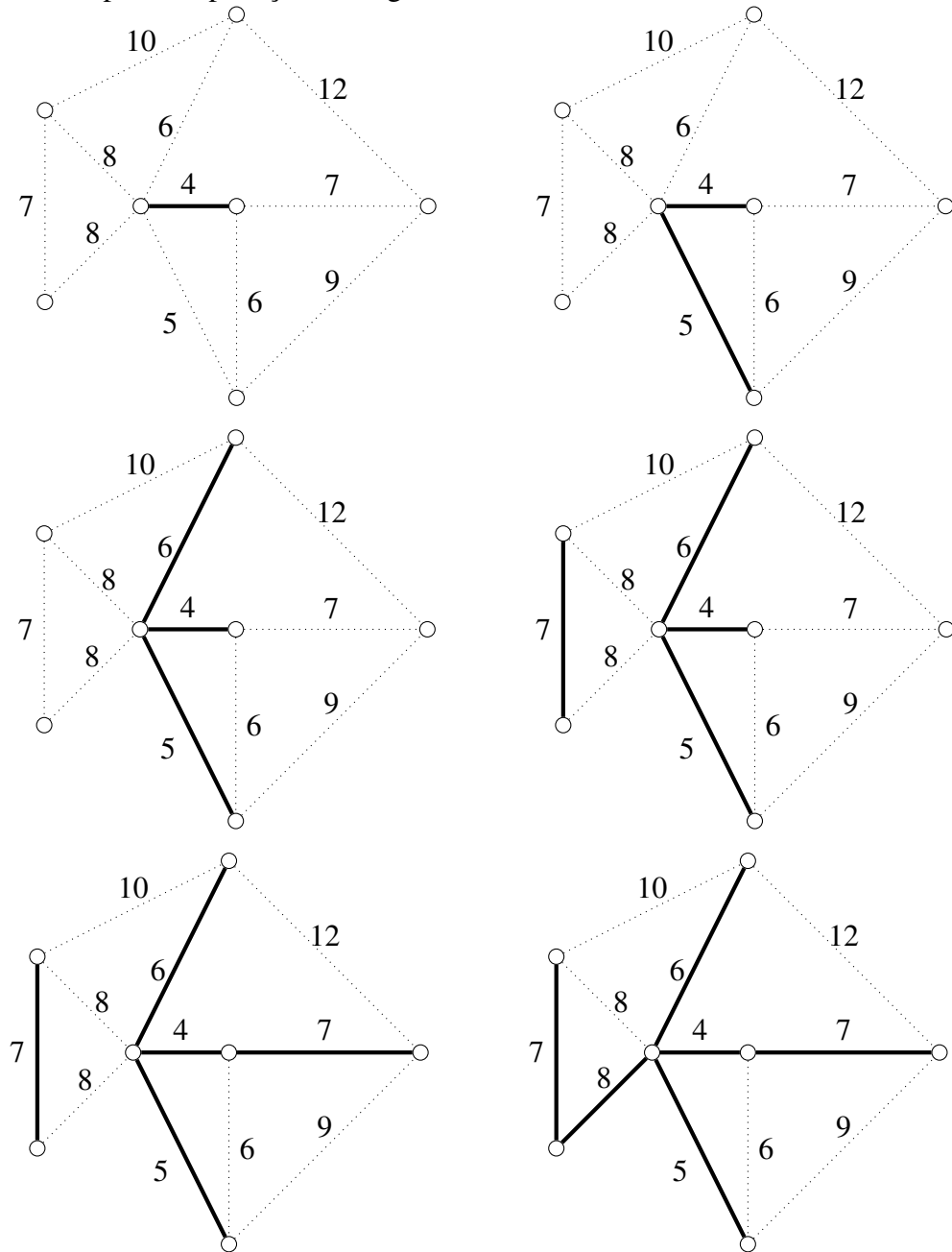
Fonte: elaborado pelo autor.

Os algoritmos Kruskal e Prim são duas estratégias iterativas para resolver o problema da árvore geradora mínima em um grafo. Começando de um grafo gerador sem arestas, a ideia do algoritmo Kruskal é inserir, a cada passo, a aresta de menor peso que não gere um ciclo com as já inseridas na solução. Enquanto isso, o algoritmo de Prim começa de um vértice inicial qualquer e, em cada iteração, seleciona o vértice mais próximo do conjunto de vértices já escolhidos, expandindo a árvore geradora formada através da aresta de menor custo que conecta esse vértice a tal conjunto.

A Figura 5 ilustra a aplicação do Algoritmo Kruskal ao grafo da Figura 4. Note que, a cada iteração de Kruskal, tem-se uma floresta geradora de peso mínimo, que vai, ao final, se transformar numa árvore. A implementação deste algoritmo pode ser feita usando uma estrutura de conjuntos disjuntos, cada um deles representando uma árvore da floresta.

O Algoritmo 2 apresenta o pseudocódigo do algoritmo Kruskal. Para a representação e operações de conjuntos disjuntos, o algoritmo faz uso das operações UNION e FIND, onde temos duas operações FIND-SET e UNION. O FIND-SET recebe como parâmetro um vértice  $u$  e retorna o vértice que representa o conjunto em que  $u$  está contido; inicialmente cada vértice representa o seu próprio conjunto. A operação UNION recebe como parâmetro dois vértices e realiza a união dos dois conjuntos aos quais os vértices pertencem.

Figura 5 – Exemplo de Aplicação do Algoritmo Kruskal



Fonte: elaborado pelo autor.



---

**Algoritmo 2: Kruskal**


---

```

1 início
2   Crie um vetor  $T$  com  $n$  posições
3   Para  $u \in V$  faça
4      $T[u] \leftarrow u$ 
5   Ordene as arestas  $e_1, e_2, \dots, e_m$ , onde  $c(e_i) \leq c(e_j) \forall i, j \in [m]: i \leq j$ .
6   Para  $i \leftarrow 1, 2, \dots, m$  faça
7     Seja  $uv = e_i$ 
8     Se  $FIND-SET(u) \neq FIND-SET(v)$  então
9        $UNION(u, v, T)$ 
10  Retorna  $T$ 

```

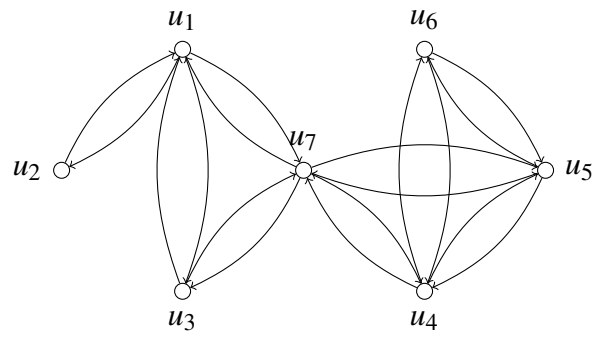
---

Um grafo direcionado  $D$ , também denominado digrafo, consiste em um par ordenado  $(V, A)$ , onde  $V$  é o conjunto de vértices e  $A$  o conjunto de arcos. Um arco consiste em um par ordenado de vértices  $(u, v)$ , representando um relacionamento direcionado de  $u$  para  $v$ . Diferentemente de um grafo simples, um digrafo descreve relacionamentos possivelmente não simétricos entre objetos (vértices). Dizemos que o arco  $(u, v)$  sai do vértice  $u$  e entra no vértice  $v$ . Definimos o conjunto  $N^+(u)$ , como os vértices  $v \in V$ , onde  $(u, v) \in A$ . De forma similar, definimos  $N^-(u)$ , como os vértices  $v \in V$ , onde  $(v, u) \in A$ .

Um caminho direcionado é um digrafo cujos vértices podem ser dispostos em uma sequência linear, de modo que existe arco de  $u$  para  $v$  se, e somente se,  $v$  é consecutivo a  $u$  na sequência. Um  $(u, v)$ -caminho em um digrafo consiste em um subgrafo caminho direcionado de  $u$  para  $v$ . Uma arborescência é um grafo direcionado, onde existe um vértice especial  $r$ , chamado raiz, tal que há um único  $(r, v)$ -caminho para cada vértice  $v$ . Note que, transformando cada arco de uma arborescência numa aresta, obtemos uma árvore. Similarmente, escolhido um nó arbitrário  $r$  de uma árvore  $T$ , podemos atribuir uma orientação a cada aresta de  $T$  para obtermos uma arborescência  $\vec{T}$ , de modo que todo  $(r, v)$ -caminho em  $T$  corresponde a um  $(r, v)$ -caminho direcionado em  $\vec{T}$ .

Um digrafo  $D$  é dito simétrico quando  $(u, v) \in A$  se, e somente se,  $(v, u) \in A$ . Todo grafo simples  $G(V, E)$  possui um digrafo simétrico  $D(V, A)$  associado, onde cada aresta  $uv \in E$  origina dois arcos  $(u, v)$  e  $(v, u)$  em  $A$ , ou seja,  $A = \{(u, v), (v, u) : uv \in E\}$ . Ao longo do texto, usamos o conjunto  $A(G) = \{(u, v), (v, u) : uv \in E\}$  para representar a orientação simétrica de  $G$ . A Figura 6 apresenta o digrafo simétrico associado ao grafo da Figura 2.

Figura 6 – Digrafo Simétrico Associado



Fonte: elaborado pelo autor.

### 3 K-FLORESTA GERADORA BALANCEADA EM ARESTAS

Este presente capítulo apresenta conceitos e resultados teóricos do problema k-FGBA. Na Seção 3.1 é apresentada a definição formal do problema. Na Seção 3.2, discutimos a relação entre a k-FGBA e o Problema da k-Partição Conexa Balanceada ( $BCP_k$ ), destacando sua equivalência em determinadas condições. Além disso, abordamos a complexidade computacional do problema, apresentando um algoritmo exato para caminhos. Por fim, na Seção 3.3, exploramos limitantes teóricos associados ao problema k-FGBA.

#### 3.1 Definição

Uma k-Floresta pode ser definida como um grafo acíclico que contém exatamente  $k$  árvores. O problema da k-Floresta Geradora Balanceada em Arestas (k-FGBA) pode ser definido da seguinte forma:

**Definição 2.** *Dados um grafo  $G$  não direcionado com pesos/custos nas arestas e um inteiro  $k$ , o problema k-FGBA busca determinar uma k-Floresta geradora de forma a minimizar o peso da árvore mais pesada.*

Podemos também pensar o problema como um particionamento do conjunto de vértices em  $k$  conjuntos  $V_1, V_2, \dots, V_k$ , disjuntos dois a dois, onde  $V_i \neq \emptyset$  e  $G[V_i]$  é conexo para todo  $i \in [k]$ , de modo a minimizar a função  $\max\{W(V_i) : i \in [k]\}$ , sendo  $W(V_i)$  o custo de uma árvore geradora mínima de  $G[V_i]$ .

#### 3.2 Complexidade Computacional

Em Madkour *et al.* (2017), o problema k-FGBA é demonstrado ser NP-Difícil já para o caso  $k = 2$ . Vaishali *et al.* (2018) demonstram que esta complexidade se mantém a mesma para o caso onde as arestas possuem peso unitário. Em Andersson *et al.* (2003) são apresentados algoritmos aproximativos para o caso onde os vértices são pontos no plano e o custo das arestas refere-se à distância euclidiana entre eles.

O Problema da k-Partição Conexa Balanceada ( $BCP_k$ ) consiste em uma variação do problema k-FGBA, onde também desejamos uma k-floresta geradora cuja árvore mais pesada tenha o menor peso, porém agora considerando que os pesos/custos são associados aos vértices. Os problemas k-FGBA e  $BCP_k$  são equivalentes quando os pesos (em cada problema) são

unitários, pois neste caso o peso em vértices e o peso em arestas de cada árvore difere sempre em uma unidade.

Em Vaishali *et al.* (2018) é apresentado um algoritmo exato de complexidade  $O(kn^3)$  para o problema k-FGBA no caso onde o grafo de entrada é uma árvore (ponderada) com  $n$  vértices. Em Frederickson e Zhou (2017) é apresentado um algoritmo  $O(n)$  para o problema  $BCP_k$  quando o grafo de entrada é uma árvore com pesos unitários nos vértices. Pela equivalência citada acima, concluímos que existe um algoritmo  $O(n)$  para k-FGBA quando o grafo de entrada é uma árvore com pesos unitários (nas arestas).

Neste contexto propomos um algoritmo de Programação Dinâmica para o caso onde o grafo de entrada  $G$  é um caminho com pesos arbitrários. O Algoritmo 3 apresenta um pseudocódigo do procedimento proposto. Ele se baseia no fato de que resolver k-FGBA em um grafo caminho consiste em determinar  $k - 1$  arestas a serem removidas de forma a minimizar o subcaminho de maior custo. Isso nos leva a seguinte equação de recorrência associada ao problema.

Sejam  $e_1, e_2, \dots, e_m$  as arestas do caminho, de acordo com a sequência de vértices. Para  $j \geq i$ , sejam  $P_{ij}$  o subcaminho contendo as arestas  $e_i, \dots, e_j$  e  $c(P_{ij})$  o peso deste subcaminho. Por conveniência, se  $j < i$ ,  $P_{ij}$  representa um caminho sem arestas, com peso zero. Considere ainda  $T(i, j)$  o maior peso de um subcaminho de  $P_{im}$  retirando  $j$  arestas, para  $i \in \{1, \dots, m\}$  e  $j \in \{0, \dots, m - i + 1\}$ ,  $j < k$ . Para todo  $i \in \{1, \dots, m\}$ , temos então que:

$$T(i, 0) = c(P_{im}) \quad (3.1)$$

$$T(i, j) = \min_{\ell=i, \dots, m-j+1} \max\{c(P_{i, \ell-1}), T(\ell+1, j-1)\} \quad \forall j \in \{1, \dots, m - i + 1\}, j < k \quad (3.2)$$

A expressão (3.1) é trivial, pois, quando não se retiram arestas ( $j = 0$ ),  $T(i, j)$  é o próprio peso do caminho de  $e_i$  a  $e_m$ . Se  $e_\ell$  é a primeira aresta removida de  $P_{im}$ , então  $T_{ij}$  é dado pelo maior valor entre o peso do subcaminho  $e_i, \dots, e_{\ell-1}$  e o subcaminho mais pesado de  $P_{\ell+1, m}$  com  $j - 1$  arestas retiradas. A expressão em (3.2) considera todos os possíveis valores para  $\ell$  e escolhe a melhor possibilidade.

O algoritmo implementa essa recorrência. Inicialmente, ele irá calcular, no vetor  $sum$ , a soma acumulada dos custos das arestas. Assim, obtemos  $c(P_{ij}) = sum[j] - sum[i - 1]$ . As linhas 4–6 correspondem à base da expressão recursiva. A entrada  $T[m + 1, 0]$  é inicializada por conveniência. As outras entradas da tabela, para  $j \in \{1, \dots, k - 1\}$  e  $i \in \{1, \dots, m - j + 1\}$ , são determinadas no laço das linhas 7–16, em ordem crescente de  $j$  e decrescente de  $i$ .  $S[i][j]$  guarda

o índice  $\ell$  da aresta  $e_\ell$  que fornece o valor de  $T[i, j]$  na expressão (3.2). O algoritmo apresentado possui complexidade  $O(kn^2)$ .

---

**Algoritmo 3:** Algoritmo de Programação Dinâmica para Caminhos

---

```

1 início
2   Construa o vetor sum, onde  $sum[0] = 0$  e  $sum[i] = \sum_{1 \leq \ell \leq i} c(e_\ell) \forall i \in \{1, \dots, |E|\}$ 
3   Construa a tabela  $T$  de dimensão  $(m+1) \times k$ 
4   Para  $i \leftarrow 1, 2, \dots, m$  faça
5     |  $T[i][0] \leftarrow sum[m] - sum[i-1]$ ;
6    $T[m+1][0] \leftarrow \infty$ 
7   Para  $j \leftarrow 1, 2, \dots, k-1$  faça
8     | Para  $i \leftarrow m-j+1, \dots, 1$  faça
9       |  $T[i][j] \leftarrow \infty, S[i][j] \leftarrow -1$ 
10      | Para  $\ell \leftarrow i, i+1, \dots, m-j+1$  faça
11        |  $aux \leftarrow \max\{sum[\ell-1] - sum[i-1], T[\ell+1][j-1]\}$ 
12        | Se  $T[i][j] > aux$  então
13          |  $T[i][j] \leftarrow aux$ 
14          |  $S[i][j] \leftarrow \ell$ 
15   Retorna  $S, T[1][k-1]$ 

```

---

### 3.3 Limitantes

Considere  $F^{G,k}$  a  $k$ -floresta geradora de  $G$  encontrada pelo algoritmo Kruskal quando a solução (parcial) possui exatamente  $n - k$  arestas. Seja  $w(F^{G,k})$  o peso da árvore mais pesada em  $F^{G,k}$  e  $c(F^{G,k}) = \sum_{e \in E(F^{G,k})} c(e)$ . O Teorema 3 apresenta limitantes para o problema  $k$ -FGBA. Em particular, o limitante inferior baseia-se no fato de que o algoritmo de Kruskal, a cada  $x$  arestas escolhidas, obtém a  $(n - x)$ -floresta geradora de menor peso. Por completude, apresentamos uma prova para esse resultado.

**Teorema 3.** *Seja  $OPT(G, k)$  o valor da solução ótima do problema  $k$ -FGBA para a instância  $(G, k)$ . Então  $\frac{c(F^{G,k})}{k} \leq OPT(G, k) \leq w(F^{G,k})$ .*

*Demonstração.* Inicialmente observe que  $F^{G,k}$  é um grafo acíclico, pois o algoritmo Kruskal não insere nenhuma aresta que gere ciclo, logo é uma floresta geradora. Como possui exatamente  $n - k$  arestas, temos que  $F^{G,k}$  é uma  $k$ -floresta, logo é uma solução viável para a instância  $(G, k)$ . Como o  $k$ -FGBA é um problema de minimização e  $F^{G,k}$  é uma solução viável, temos que  $OPT(G, k) \leq w(F^{G,k})$ .

Queremos provar que  $\frac{c(F^{G,k})}{k} \leq OPT(G,k)$ . Seja  $F^* \neq F^{G,k}$  uma solução ótima do problema k-FGBA para a instância  $(G,k)$ , ou seja,  $OPT(G,k) = w(F^*)$ .

Considere  $e_1, e_2, \dots, e_{n-k} \in E(F^{G,k})$  tal que  $c(e_1) \leq c(e_2) \leq \dots \leq c(e_{n-k})$ . Além disso,  $e'_1, e'_2, \dots, e'_{n-k} \in E(F^*)$ , tal que  $c(e'_1) \leq c(e'_2) \leq \dots \leq c(e'_{n-k})$ . Construa uma k-Floresta  $\bar{F}^*$  a partir da solução  $F^*$ , utilizando o seguinte procedimento:

1. Inicialmente, sejam  $e'_1, e'_2, \dots, e'_{n-k}$  as arestas de  $\bar{F}^*$ .
2. Selecione o menor índice  $i$ , tal que  $e_i \neq e'_i$ . Pela construção do algoritmo Kruskal, temos que  $c(e_i) \leq c(e'_i)$ . Caso não exista tal índice, finalize o procedimento.
3. Caso  $E(\bar{F}^*) \cup \{e_i, e'_i\}$  não gere ciclo, faça  $e'_i \leftarrow e_i$ , ou seja,  $E(\bar{F}^*) = E(\bar{F}^*) \cup \{e_i\} \setminus \{e'_i\}$ .
4. Caso contrário, por Kruskal,  $E(\bar{F}^*) \cup \{e_i, e'_i\}$  contém único ciclo, que envolve aresta  $e'_j$  com  $j \geq i$ . Faça  $(e'_i, \dots, e'_{j-1}, e'_j) \leftarrow (e_i, e'_i, \dots, e'_{j-1})$ , ou seja,  $E(\bar{F}^*) = E(\bar{F}^*) \cup \{e_i, e'_i\} \setminus \{e'_j\}$ .  
Volte ao passo 2.

Ao final do procedimento temos que  $\bar{F}^* = F^{G,k}$ . Note que sempre trocamos as arestas de forma a manter  $\bar{F}^*$  uma floresta e a nunca aumentar o seu custo total. Logo, no final,  $c(\bar{F}^*) = c(F^{G,k}) \leq c(F^*)$ . Observe que  $c(F^*) \leq k \cdot w(F^*) = k \cdot OPT(G,k)$ , pois cada uma das  $k$  subárvores em  $F^*$  tem custo menor ou igual a  $w(F^*)$ . Portanto,  $\frac{c(F^{G,k})}{k} \leq OPT(G,k)$ .  $\square$

## 4 ALGORITMOS HEURÍSTICOS

Neste capítulo apresentamos algoritmos heurísticos para o problema k-FGBA. Na Seção 4.1 são citados os algoritmos já existentes na literatura, bem como são apresentados os algoritmos propostos neste trabalho. Já na Seção 4.2 é apresentada uma avaliação experimental visando comparar os algoritmos propostos com aqueles da literatura (MADKOUR *et al.*, 2017).

### 4.1 Algoritmos Propostos

No trabalho de Madkour *et al.* (2017) são apresentadas duas heurísticas para o problema explorado. A primeira heurística, que será denominada de H\_NCCuts, consiste em um algoritmo espectral cuja ideia é determinar  $k$  clusters, onde as arestas internas são de baixo custo e as externas de alto custo. Como medida de similaridade entre dois vértices  $u, v \in V$ , o algoritmo adota  $L - d(u, v)$ , onde  $L = \max\{d(w, z) : w, z \in V\}$ . É construído uma matriz de similaridade  $n \times n$ . O algoritmo irá calcular o autovetor desta matriz e irá dividir o conjunto de vértices em dois de acordo com a positividade dos componentes do autovetor. A heurística utiliza convergência de matrizes cujos limitantes de tempo são pouco precisos, mas a mesma faz uso do algoritmo Floyd-Warshall para determinar a menor distância entre todos os pares de vértices. Logo o algoritmo possui complexidade de tempo de  $\Omega(n^3)$ .

A segunda heurística, que denominaremos por H\_DPCuts, trata-se de um algoritmo de Programação Dinâmica. Inicialmente é obtida uma árvore geradora mínima do grafo  $G$ , através de um algoritmo Kruskal ligeiramente modificado que visa reduzir o grau dos vértices da árvore. Após isso, o algoritmo determina o caminho de maior peso na árvore. O procedimento irá empregar a técnica de Programação Dinâmica, utilizando um conjunto de penalidades e descontos, para que a remoção de arestas deste caminho resulte em subárvores de custos próximos a  $\frac{c(T)}{k}$ , onde  $c(T)$  é o custo da árvore geradora mínima  $T$  do grafo  $G$ . Caso o algoritmo retorne menos que  $k$  subárvores, então o mesmo procedimento é aplicado na subárvore de maior custo. Caso o algoritmo encontre um número maior que  $k$  de subárvores, deve-se adicionar arestas que conectam subárvores de menor custo. O algoritmo possui complexidade de tempo de  $O(m^2)$ .

Neste trabalho propomos duas novas heurísticas para o problema. O primeiro procedimento, denominado H\_INS, possui como ideia central inserir, a cada iteração, uma aresta de forma gulosa, ou seja, a aresta que minimiza a função objetivo de uma iteração para outra. O algoritmo finaliza quando inserimos as  $n - k$  arestas. O pseudocódigo desta heurística é

apresentado no Algoritmo 4. O vetor  $T$  irá armazenar os representantes dos conjuntos disjuntos de vértices, e o vetor  $Z$  guardará o custo da árvore que contém cada vértice representante. O vetor  $T$  é atualizado a cada operação UNION. As operações de UNION-FIND foram implementadas utilizando a estratégia de *path compression* (CORMEN *et al.*, 2002). H\_INS possui complexidade de tempo  $O((n - k)m \log n)$ .

---

**Algoritmo 4:** Heurística H\_INS

---

```

1 início
2   Crie os vetores  $T$  e  $Z$  com  $n$  posições cada
3   Para  $u \in V$  faça
4      $T[u] \leftarrow u$ 
5      $Z[u] \leftarrow 0$ 
6   Para  $i \leftarrow 1, 2, \dots, n - k$  faça
7     Selecione  $uv = e_j$ , com  $j \in [m]$ , tal que  $\text{FIND-SET}(u) \neq \text{FIND-SET}(v)$ , de forma
      a minimizar  $Z[\text{FIND-SET}(u)] + Z[\text{FIND-SET}(v)] + c(uv)$ 
8      $Z[\text{FIND-SET}(u)] \leftarrow Z[\text{FIND-SET}(u)] + Z[\text{FIND-SET}(v)] + c(uv)$ 
9     UNION( $u, v, T$ ) // Acrescenta ao conjunto de  $u$  os elementos do
      conjunto de  $v$  e atualiza os representantes
10  Retorna  $T$ 

```

---

A segunda heurística se baseia na intuição de que vértices distantes no grafo não devem pertencer à mesma árvore em uma solução ótima do problema k-FGBA. Desta forma, primeiro propomos um procedimento que visa determinar vértices que possivelmente não estão em uma mesma árvore. Dado o grafo de entrada  $G(V, E)$  e um valor positivo  $\lambda$ , considere  $G_\lambda = (V, E_\lambda)$ , onde  $E_\lambda = \{uv \in E : d(u, v) > \lambda\}$ . Assim, uma clique de tamanho pelo menos  $k$  no grafo  $G_\lambda$  pode significar  $k$  vértices que possivelmente estão em árvores disjuntas. Note que podemos tomar  $\lambda \in [l, L]$ , onde  $l = \min\{d(u, v) : u, v \in G\}$  e  $L = \max\{d(u, v) : u, v \in G\}$ .

A ideia inicial do segundo algoritmo proposto, denominado H\_Clique, é encontrar, através de busca binária, o maior valor de  $\lambda$  de forma que exista uma clique de tamanho pelo menos  $k$ . O problema desta abordagem é que determinar de forma exata a maior clique em um grafo é NP-Difícil. Desta forma, adotamos um procedimento que, a cada iteração, iniciando de um vértice  $u \in V$  como solução parcial e, utilizando o algoritmo DSATUR (BRÉLAZ, 1979), busca encontrar a maior clique no grafo  $G_\lambda[N(u)]$ . Após determinar os  $k$  vértices base, as arestas são inseridas, a partir destes vértices, seguindo uma abordagem gulosa: a cada vez insere-se a



aresta que gera o menor valor de função objetivo. O pseudocódigo de H\_Clique é apresentado no Algoritmo 5. O algoritmo proposto possui complexidade  $O(n(m + n \log n) \log L + nm \log n)$ .

---

**Algoritmo 5:** Heurística H\_Clique

---

```

1 início
2    $d \leftarrow \text{Floyd-Warshall}(G)$ 
3   Seja  $l = \min\{d(u, v) : u, v \in G\}$  e  $L = \max\{d(u, v) : u, v \in G\}$ 
4    $\text{Clique}_{max} \leftarrow \emptyset$ 
5   Enquanto  $l \leq L$  faça
6      $\lambda \leftarrow \lfloor \frac{L+l}{2} \rfloor$ 
7     Construa o grafo  $G_\lambda$ 
8      $\text{Clique} \leftarrow \emptyset$ 
9     Para  $u \in V$  faça
10       $C \leftarrow \text{DSATUR}(G_\lambda[N(u)]) \cup \{u\}$ 
11      Se  $|C| > |\text{Clique}|$  então
12         $\text{Clique} \leftarrow C$ 
13      Se  $|\text{Clique}| \geq k$  então
14         $\text{Clique}_{max} \leftarrow \text{Clique}$ 
15         $l \leftarrow \lambda + 1$ 
16      Senão
17         $L \leftarrow \lambda - 1$ 
18    Para  $u \in V$  faça
19      Se  $u \in \text{Clique}$  então
20         $T[u] \leftarrow u$ 
21      Senão
22         $T[u] \leftarrow -1$ 
23       $Z[u] \leftarrow 0$ 
24    Para  $i \leftarrow 1, 2, \dots, n - k$  faça
25      Selecione  $uv = e_j$ , com  $j \in [m]$ , tal que  $T[u] \neq -1$  ou  $T[u] = -1$ , de forma a
        minimizar  $Z[\text{FIND-SET}(u)] + c(uv)$ 
26       $T[v] \leftarrow \text{FIND-SET}(u)$ 
27       $Z[\text{FIND-SET}(u)] \leftarrow Z[\text{FIND-SET}(u)] + c(uv)$ 
28  Retorna  $T$ 

```

---

## 4.2 Comparação dos algoritmos

Com o intuito de avaliar a eficiência no que diz respeito à qualidade das soluções geradas pelas heurísticas, foram realizados experimentos computacionais. Consideramos todas as instâncias<sup>1</sup> também utilizadas nos testes finais (Capítulo 6), porém, apenas para  $k \geq 2$  e  $G$  conexo. Esta filtragem se deu pelo fato das heurísticas apresentadas em Vaishali *et al.* (2018) não serem inicialmente projetadas para o caso  $k = 1$  e para quando o grafo de entrada é desconexo. Além disso, para  $k = 1$ , o problema resume-se a encontrar uma árvore geradora mínima. Assim, nestes experimentos foram utilizadas 516 instâncias, cujo processo de geração é descrito na Seção 6.1. O ambiente computacional adotado é descrito na Seção 6.2. Vale destacar que foram utilizadas as implementações das heurísticas da literatura disponibilizadas em Vaishali *et al.* (2018).

A Tabela 1 apresenta um resumo dos testes realizados. A coluna % representa a porcentagem de instâncias, onde o algoritmo encontrou a melhor solução entre todos os procedimentos, e  $gap(\%)$  é a distância média, em percentual, que a solução da heurística está da melhor solução. A coluna  $N$  refere-se ao número de instâncias em cada caso.

Tabela 1 – Resumo dos resultados em relação ao parâmetro  $k$

Classe	$k$	$N$	H_NCuts		H_DPCuts		H_Clique		H_INS	
			$gap(\%)$	%	$gap(\%)$	%	$gap(\%)$	%	$gap(\%)$	%
$C_1$	2	40	8.33	30.00	8.83	40.00	<b>0.62</b>	<b>97.50</b>	17.45	5.00
	4	40	19.79	15.00	16.57	27.50	<b>0.00</b>	<b>100.00</b>	31.93	22.50
	8	40	43.40	22.50	48.64	25.00	<b>3.75</b>	<b>95.00</b>	27.65	47.50
	16	30	74.56	10.00	121.06	6.67	<b>7.78</b>	<b>90.00</b>	22.72	63.33
	32	20	175.00	0.00	265.00	10.00	<b>0.00</b>	<b>100.00</b>	12.50	75.00
	64	5	200.00	0.00	300.00	0.00	<b>0.00</b>	<b>100.00</b>	<b>0.00</b>	<b>100.00</b>
$C_2$	2	39	11.20	25.64	11.15	38.46	<b>1.84</b>	<b>76.92</b>	14.70	10.26
	4	39	21.95	20.51	15.94	33.33	<b>2.50</b>	<b>82.05</b>	21.52	12.82
	8	39	78.92	10.26	27.14	25.64	<b>13.06</b>	<b>79.49</b>	13.26	46.15
	16	30	104.90	6.67	58.00	23.33	33.92	53.33	<b>3.89</b>	<b>86.67</b>
	32	20	244.17	0.00	167.50	5.00	38.33	55.00	<b>2.50</b>	<b>95.00</b>
	64	5	580.00	0.00	420.00	0.00	60.00	40.00	<b>0.00</b>	<b>100.00</b>
$C_3$	2	39	13.87	20.51	12.34	41.03	<b>2.08</b>	<b>66.67</b>	18.06	7.69
	4	39	38.64	10.26	13.24	33.33	<b>3.29</b>	<b>61.54</b>	15.88	23.08
	8	39	151.28	0.00	29.24	28.21	<b>2.93</b>	<b>64.10</b>	7.47	46.15
	16	30	391.46	0.00	95.68	20.00	5.86	46.67	<b>3.31</b>	<b>56.67</b>
	32	20	1550.80	0.00	300.24	0.00	14.80	55.00	<b>2.18</b>	<b>70.00</b>
	64	5	5467.11	0.00	695.61	0.00	1.43	<b>80.00</b>	<b>0.87</b>	<b>80.00</b>

Fonte: elaborado pelo autor.

<sup>1</sup> Ver Seção 6.1 para a descrição das instâncias.

Podemos verificar que as heurísticas propostas por este trabalho se mostraram superiores em relação aos procedimentos já existentes no conjunto de instâncias analisadas. Além disso, nas classes  $C_2$  e  $C_3$ , a heurística H\_Clique apresentou melhores resultados para  $k \leq 8$ , enquanto que o procedimento H\_INS se mostrou superior ou igual para casos  $k \geq 16$ . Já na Classe  $C_1$ , onde os custos são unitários, a H\_Clique se mostrou mais eficiente, independentemente do parâmetro  $k$ . Em números, a H\_Clique encontrou a melhor solução em 76,55% das instâncias, enquanto que H\_INS em apenas 40,89%. Em relação aos tempos de execução, as heurísticas H\_NCuts, H\_DPCuts, H\_Clique e H\_INS tiveram tempos médios de 49,76 ms, 3,21 ms, 131,82 ms e 4,33 ms, respectivamente. Diante dos resultados obtidos, a heurística H\_Clique foi adotada como limite superior e utilizada em procedimentos exatos descritos a seguir.

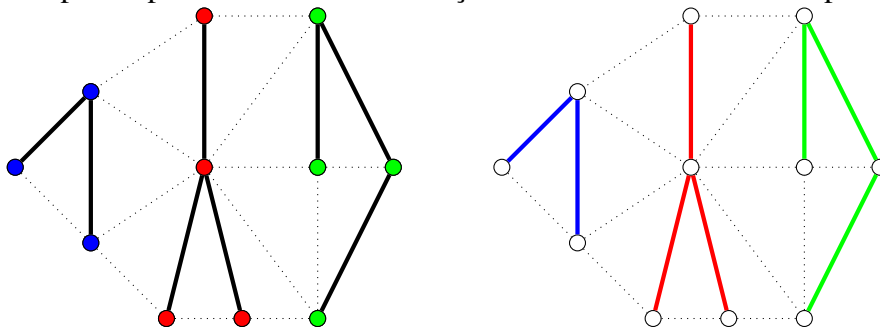
## 5 FORMULAÇÕES MATEMÁTICAS

Neste capítulo são apresentadas as formulações matemáticas propostas e algumas estratégias utilizadas para fortalecer as mesmas. Na Seção 5.1 apresentamos as estratégias de modelagem adotadas. As Seções 5.3, 5.2 e 5.4 apresentam os modelos propostos. Na Seção 5.5 são descritas desigualdades válidas para os modelos. Já na Seção 5.6 descrevemos um procedimento de fixação de variáveis.

### 5.1 Estratégia de Modelagem do problema

Um modelo matemático para o problema  $k$ -FGBA deve garantir a condição de  $k$ -floresta, e ser capaz de capturar o peso de cada árvore, afim de minimizar a de maior custo/peso. O particionamento do grafo em  $k$  árvores disjuntas pode se dar tanto em relação aos vértices como em relação às arestas. A Figura 7 apresenta como pode se dar o particionamento da solução por vértices ou por arestas, onde as partes estão diferenciadas por cores.

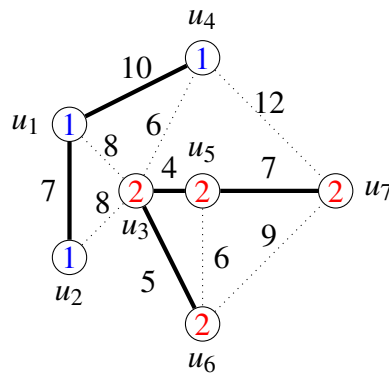
Figura 7 – Exemplo de particionamento da solução em vértices e em arestas para  $k = 3$



Fonte: elaborado pelo autor.

No que se refere ao particionamento por vértices, consideramos duas abordagens de modelagem. Em uma delas, o particionamento é feito por rotulação, ou seja, cada vértice do grafo irá receber um rótulo identificador da árvore a qual pertence. A Figura 8 apresenta um exemplo de particionamento dos vértices por rotulação em uma 2-floresta: vértices com rótulo 1 pertencem a uma árvore e vértices com rótulo 2, a outra.

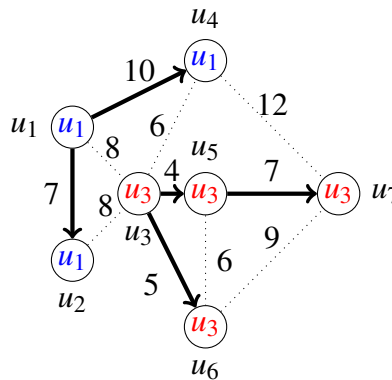
Figura 8 – Particionamento dos vértices via rotulação de árvores



Fonte: elaborado pelo autor.

Uma outra forma de particionar o conjunto de vértices é transformar cada árvore numa arborescência e usar a raiz da arborescência como identificador. Assim, cada vértice é identificado por um vértice que é o representante da arborescência a que ele pertence, no caso o vértice representante será sempre a raiz. A ideia de representantes é apresentada em Campêlo *et al.* (2008). A Figura 9 apresenta um exemplo desta modelagem.

Figura 9 – Particionamento dos vértices via vértices representantes



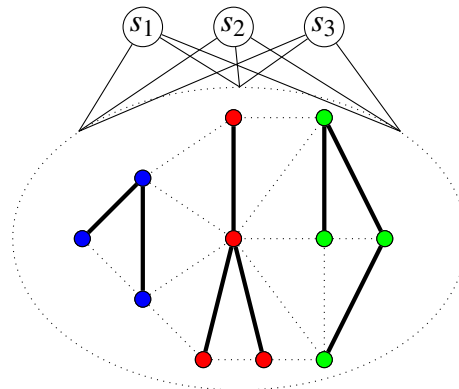
Fonte: elaborado pelo autor.

Note que, além de particionar os vértices, precisamos garantir que as arestas selecionadas (dentro de cada parte) não geram ciclo. Algumas das principais abordagens existentes na literatura para garantir soluções acíclicas são SECs (DANTZIG *et al.*, 1954), MTZ (MILLER *et al.*, 1960) e o modelo de Martin (MARTIN, 1991). Em testes computacionais prévios, o modelo MTZ se mostrou bem mais eficiente que os demais nas instâncias analisadas e, desta forma, foi utilizada neste trabalho. As duas formulações baseadas no particionamento dos vértices serão apresentadas nas seções 5.2 e 5.3.

Para a modelagem de particionamento das arestas, a estratégia parte da introdução

de  $k$  vértices universais em  $G(V, E)$ , criando um grafo  $G'(V', E')$ , onde  $V' = V \cup \{s_1, \dots, s_k\}$  e  $E' = E \cup \{s_i u : i \in [k], u \in V\}$ . Além disso, atribuem-se os custos  $c_{s_i u} = 0 \forall i \in [k] \forall u \in V$ . Esta abordagem foi apresentada em Miyazawa *et al.* (2020) para o problema  $BCP_k$ . Estratégia similar, mas usando apenas um vértice universal, foi usada em (FILHO, 2019; FIGUEREDO; CAMPÊLO, 2020) para modelar outro problema sobre floresta. Um exemplo de grafo  $G'$  é apresentado na Figura 10. Para remoção de subciclos na solução, o modelo adota restrições de fluxo (ver Seção 5.4).

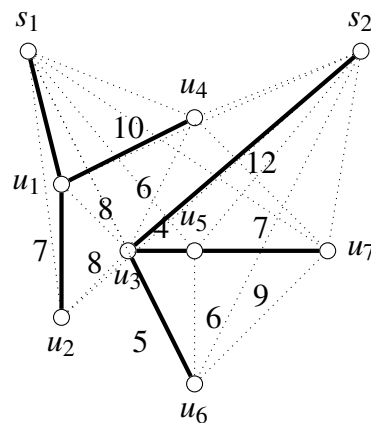
Figura 10 – Exemplo de grafo  $G'$  com  $k = 3$



Fonte: elaborado pelo autor.

A solução nesta modelagem se dá quando estabelecemos que cada vértice  $s_i$  deve ser adjacente a exatamente um outro vértice na solução, para que envia fluxo a ser distribuído entre os vértices a quem se conecta. Desta forma, a partir de um vértice  $s_i$  conseguimos determinar cada uma das  $k$  árvores. Um exemplo de solução do modelo de particionamento de arestas é apresentado na Figura 11.

Figura 11 – Exemplo de solução na modelagem de particionamento de arestas



Fonte: elaborado pelo autor.

Todos os modelos são descritos a partir da orientação simétrica do grafo de entrada, ou seja, do conjunto  $A(G) = \{(u, v), (v, u) : uv \in E\}$ .

## 5.2 Modelo de Árvore Rotulada

Para cada  $t \in [k]$ , definimos as variáveis binárias  $x_u^t$  e  $y_{uv}^t$ , indicando, respectivamente, se o vértice  $u \in V$  e o arco  $uv \in A(G)$  pertencem à árvore  $t$ . A variável  $\pi_u$  registra a distância entre  $u \in V$  e a raiz de sua árvore, e  $z$  é o maior peso de uma árvore da floresta.

$$(F\_ROT) \min z \tag{5.1}$$

$$\text{s.a: } z \geq \sum_{uv \in E} c_{uv}(y_{uv}^t + y_{vu}^t) \quad \forall t \in [k] \tag{5.2}$$

$$\sum_{u \in V} x_u^t \geq 1 \quad \forall t \in [k] \tag{5.3}$$

$$\sum_{t=1}^k x_u^t = 1 \quad \forall u \in V \tag{5.4}$$

$$y_{uv}^t + y_{vu}^t \leq x_u^t, y_{uv}^t + y_{vu}^t \leq x_v^t \quad \forall uv \in E, \forall t \in [k] \tag{5.5}$$

$$\sum_{uv \in E} (y_{uv}^t + y_{vu}^t) = \sum_{u \in V} x_u^t - 1 \quad \forall t \in [k] \tag{5.6}$$

$$\sum_{u:uv \in E} y_{uv}^t \leq x_v^t \quad \forall v \in V, \forall t \in [k] \tag{5.7}$$

$$\begin{aligned} (M-1)y_{uv}^t - (M+1)(1-y_{vu}^t) + 1 &\leq \pi_u - \pi_v \\ &\leq 1 + (M-1)(1-y_{vu}^t) - (M+1)y_{uv}^t \end{aligned} \quad \forall uv \in E, t \in [k] \tag{5.8}$$

$$\pi_u \geq 0, 0 \leq x_u^t \leq 1 \quad \forall u \in V, \forall t \in [k] \tag{5.9}$$

$$y_{uv}^t \in \mathbb{B} \quad \forall uv \in A, \forall t \in [k] \tag{5.10}$$

A restrição (5.2) junto com a função objetivo (5.1) asseguram que  $z$  seja o maior peso de árvore. Pelas restrições (5.3)-(5.4), há pelo menos um vértice em cada árvore construída e todo vértice está em exatamente uma delas. A restrição (5.5) garante que os arcos  $uv$  e  $vu$  não podem ser selecionados simultaneamente, e que só podem pertencer a uma árvore se ambos  $u$  e  $v$  pertencerem a mesma. A restrição (5.6) assegura que o número de arcos em cada árvore é uma unidade menor que sua quantidade de vértices. A clássica restrição MTZ (5.8) estabelece  $\pi_v = \pi_u + 1$  se o arco  $uv$  está numa árvore. Junto com (5.7), ela garante que os arcos escolhidos não formam ciclo (DESROCHERS; LAPORTE, 1991). Dado que o grafo induzido por  $y$  é uma floresta e que a restrição (5.5) faz  $x_u^t = x_v^t = 1$  se o arco  $uv$  pertence a árvore  $t$ , então a restrição (5.6) garante a integralidade das variáveis  $x$ . Em (F\_ROT), consideramos  $M = \lceil \frac{|V|-k}{2} \rceil$ ,

visto que a raiz de cada árvore não é escolhida a priori.

Note que uma mesma solução viável pode ser representada por meio  $k!$  soluções simétricas no modelo, pois podemos permutar os rótulos identificadores das árvores. Uma forma simples de realizar a quebra de algumas soluções simétricas consiste em ordenar as árvores pelo seu peso, ou seja, fazendo com que a árvore  $i$  tenha peso superior ou igual a árvore  $i + 1$ . Esta estratégia é explicitada na restrição (5.11).

$$\sum_{uv \in E} c_{uv}(y_{uv}^t + y_{vu}^t) \geq \sum_{uv \in E} c_{uv}(y_{uv}^{t+1} + y_{vu}^{t+1}) \quad \forall t \in [k - 1] \tag{5.11}$$

Seguindo esta estratégia, a árvore de maior custo é sempre a  $t = 1$ , logo podemos minimizar diretamente o peso desta árvore na função objetivo. Apesar desta estratégia quebrar diversas soluções simétricas em muitos casos, ela se mostra ineficiente quando temos duas ou mais árvores de mesmo peso.

Uma outra forma de eliminar soluções simétricas consiste em considerar uma ordenação do conjunto  $V$  e ordenar as árvores de acordo com seu menor vértice. A ideia é fazer com que o vértice de menor índice de uma árvore receba o menor rótulo possível. Para isso, adicionamos as seguintes restrições:

$$x_u^t = 0, \forall u \in V, \forall t \in [k] : t > u \quad \text{e} \quad \sum_{j=1}^{t-1} x_w^j \geq x_u^t - \sum_{v < u} x_v^t, \forall u, w \in V : w < u, \forall t \in [u]. \tag{5.12}$$

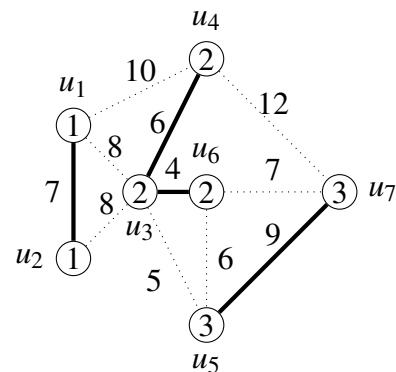
Tais restrições garantem que a matriz  $[x_{ut}]$  tem forma escada, de modo que  $u_{t+1} > u_t$ , onde  $u_t = \min\{u : x_u^t = 1\}$  é o menor vértice na árvore  $t$ . A forma escada é exemplificada na Tabela 2. Entradas iguais a um em cada coluna identificam os vértices de cada árvore. A Figura 12 apresenta a solução descrita na tabela.

Tabela 2 – Exemplo de matriz  $[x_{ut}]$

	$t_1$	$t_2$	$t_3$
$u_1$	1	0	0
$u_2$	1	0	0
$u_3$	0	1	0
$u_4$	0	1	0
$u_5$	0	0	1
$u_6$	0	1	0
$u_7$	0	0	1

Fonte: elaborado pelo autor.

Figura 12 – Solução associada à Tabela 2



Fonte: elaborado pelo autor.

Esta segunda estratégia de quebra de simetria se mostrou mais eficiente em testes computacionais preliminares, sendo assim, adotada no modelo F\_ROT.



### 5.3 Modelo de Representantes

A partir de uma ordenação do conjunto  $V$ , definimos  $x_v^u$  como a variável que determina se o vértice  $u$  representa o vértice  $v \geq u$ , enquanto  $y_{vw}^u$  indica se o arco  $vw \in A(G)$  é utilizado na árvore representada por  $u \leq v, w$ . Com isso, o representante de cada árvore é seu vértice de menor índice. A variável  $\pi_u$  registra a distância entre  $u \in V$  e a raiz de sua árvore, e  $z$  é o maior peso de uma árvore da floresta.

$$(F\_REP) \min z \tag{5.13}$$

$$\text{s.a: } z \geq \sum_{vw \in E: v, w \geq u} c_{vw} (y_{vw}^u + y_{wv}^u) \quad \forall u \in V \tag{5.14}$$

$$\sum_{u \in V} x_u^u = k \tag{5.15}$$

$$\sum_{u \in V: u \leq v} x_v^u = 1 \quad \forall v \in V \tag{5.16}$$

$$x_u^u \leq x_v^u \quad \forall u, v \in V: u \leq v \tag{5.17}$$

$$y_{vw}^u + y_{wv}^u \leq x_v^u, y_{vw}^u + y_{wv}^u \leq x_w^u \quad \forall u \in V, vw \in E: v, w \geq u \tag{5.18}$$

$$\sum_{u \in V} \sum_{v > u} y_{vu}^u = 0 \tag{5.19}$$

$$\sum_{vw \in E: v \geq u} y_{vw}^u = x_w^u \quad \forall u, w \in V: w > u \tag{5.20}$$

$$\begin{aligned} (M-1)y_{wv}^u - (M+1)(1-y_{vw}^u) + 1 &\leq \pi_w - \pi_v \\ &\leq 1 + (M-1)(1-y_{wv}^u) - (M+1)y_{wv}^u \quad \forall u \in V, vw \in E: u \leq v < w \end{aligned} \tag{5.21}$$

$$0 \leq \pi_u \leq M(1-x_u^u) \quad \forall u \in V \tag{5.22}$$

$$0 \leq x_v^u \leq 1 \quad \forall u, v \in V: u \leq v \tag{5.23}$$

$$y_{vw}^u, y_{wv}^u \in \mathbb{B} \quad \forall u \in V, \forall vw \in E: v, w \geq u \tag{5.24}$$

Essa formulação identifica cada árvore por seu vértice de menor índice, não permitindo assim soluções simétricas. As restrições (5.14), (5.16), (5.18) e (5.21) tem significado similar a (5.2), (5.4), (5.5) e (5.8), respectivamente. Pela restrição (5.15), há exatamente  $k$  vértices representantes, ou melhor,  $k$  árvores. A restrição (5.19) define  $u$  como raiz da árvore que ele representa. Já a restrição (5.20), que é a contraparte de (5.7), define exatamente um pai para todo vértice de cada árvore, exceto sua raiz. A restrição (5.17) define que se certo vértice representa um segundo vértice então ele deve ser representante de sua árvore.

A integralidade de  $x$  deve-se às restrições (5.15), (5.20) e (5.24). Em (F\_REP), usamos  $M = |V| - k$ , já que, escolhida uma árvore, sua raiz está previamente definida.

## 5.4 Modelo de Fluxo

Dada uma instância  $(G, k)$  para o problema  $k$ -FGBA, construímos o grafo orientado  $G'(V', A')$  de forma que  $V' = V \cup \{s_1, \dots, s_k\}$  e  $A' = A(G) \cup \{s_i u : i \in [k], u \in V\}$ . Além disso, qualquer aresta  $s_i u$  tem seu custo associado igual a zero. Desta forma, definimos  $y_{uv}$  como a variável que determina se o arco  $(u, v) \in A'$  foi selecionado para a solução e  $f_{uv}$  como o fluxo que passa pelo arco  $(u, v) \in A'$ . A ideia do modelo é determinar  $k$  arborescências disjuntas, onde cada uma é enraizada por um vértice  $s_i$  distinto, de forma a minimizar a de maior peso. Abaixo é apresentado o modelo F\_FLUXO.

$$(F\_FLUXO) \min \sum_{u \in N^+(s_1)} f_{s_1 u} \quad (5.25)$$

$$\text{s.a : } \sum_{u \in N^+(s_i)} f_{s_i u} \geq \sum_{u \in N^+(s_{i+1})} f_{s_{i+1} u} \quad \forall i \in [k-1] \quad (5.26)$$

$$\sum_{v \in N^-(u)} f_{vu} - \sum_{v \in N^+(u)} f_{uv} = \sum_{v \in N^-(u)} c_{vu} \cdot y_{vu} \quad \forall u \in V' \quad (5.27)$$

$$\sum_{v \in N^-(u)} y_{vu} = 1 \quad \forall u \in V \quad (5.28)$$

$$\sum_{v \in N^+(s_i)} y_{s_i v} = 1 \quad \forall i \in [k] \quad (5.29)$$

$$f_{uv} \geq c_{uv} \cdot y_{uv} \quad \forall (u, v) \in A' \quad (5.30)$$

$$f_{uv} \leq M \cdot y_{uv} \quad \forall (u, v) \in A' \quad (5.31)$$

$$y_{uv} + y_{vu} \leq 1 \quad \forall uv \in E \quad (5.32)$$

$$f_{uv} \in \mathbb{R}^+ \quad \forall (u, v) \in A' \quad (5.33)$$

$$y_{uv} \in \mathbb{B} \quad \forall (u, v) \in A' \quad (5.34)$$

A variável de fluxo  $f_{uv}$ , para cada  $(u, v) \in A'$ , vai armazenar o peso da subárvore enraizada em  $v$  mais o custo da aresta  $(u, v)$ , isto é garantido a partir das restrições (5.27) e (5.30). A restrição (5.26) assegura que a árvore que tem como raiz o vértice  $s_i$  possui peso superior ou igual a que possui como raiz o vértice  $s_{i+1}$ , isso faz com que as árvores estejam ordenadas, removendo assim soluções simétricas. Por conta desta restrição, a árvore enraizada por  $s_1$  possui o maior peso, que é minimizado em (5.25). As restrições (5.28) e (5.29) garantem que exatamente um arco chega em cada vértice  $u \in V$  e que sai exatamente um arco de cada vértice  $s_i$ ,  $i \in [k]$ , assegurando que não exista subciclo.

As restrições (5.30)-(5.31) asseguram que passa fluxo pelo arco  $(u, v)$  se, e somente se, o mesmo é selecionado para uma solução. A constante  $M$  pode ser dada por um limite

superior de solução para aquela instância. Adotamos  $M$  como o valor encontrada pela heurística H\_Clique, descrita na Seção 4.1. As restrições (5.33) e (5.34) definem os domínios das variáveis  $f$  e  $y$ , respectivamente. Apesar de não ser necessária para o modelo na integralidade, a restrição (5.32) auxilia na relaxação linear, garantindo que não podemos selecionar os arcos  $(u, v)$  e  $(v, u)$  de forma simultânea.

## 5.5 Desigualdades Válidas

Nesta seção apresentamos desigualdades válidas para os modelos apresentados nas Seções 5.2 e 5.3.

Uma primeira desigualdade, específica para o modelo F\_REP, estabelece que, se um vértice  $v$  é representado por um vértice  $u < v$ , então existe pelo menos um arco que sai do vértice  $u$ , que é raiz desta arborescência. Esta desigualdade válida é apresentada na Equação (5.35).

$$x_v^u \leq \sum_{uw \in A: u < w} y_{uw}^u \quad \forall u, v \in V: u < v \quad (5.35)$$

Considere o grafo  $G_\lambda$ , descrito na Seção 4.1 e seja  $\lambda$  o valor de uma solução viável para o k-FGBA. Se  $uv \in E(G_\lambda)$ , então  $u$  e  $v$  não podem pertencer a uma mesma árvore na solução ótima. As restrições (5.36) e (5.37) descrevem esta condição para os modelos F\_REP e F\_ROT, respectivamente. Neste trabalho foi utilizado  $\lambda$  como valor da solução retornado pela heurística H\_Clique.

$$x_u^u + x_w^u \leq 1 \quad \forall u \in V \quad \forall vw \in E(G_\lambda) \mid u \leq v, w \quad (5.36)$$

$$x_u^t + x_v^t \leq 1 \quad \forall t \in [k] \quad \forall uv \in E(G_\lambda) \quad (5.37)$$

## 5.6 Procedimento de Fixação de Variáveis

Nesta seção apresentamos um procedimento de fixação de variáveis  $x$  nos modelos F\_REP e F\_ROT. Este procedimento surge da observação que alguns pares de vértices garantidamente não estão em uma mesma árvore, quando a menor distância entre os vértices do par é maior que o valor de uma solução viável.

Para implementar essa ideia, iniciamos construindo o grafo  $G_\lambda$ , tomando  $\lambda$  como o valor da solução retornada pela heurística H\_Clique. Após isso, buscamos a maior clique possível no grafo  $G_\lambda$ . Para isso, utilizamos o mesmo método heurístico para Clique Máxima

utilizado na heurística H\_Clique. Os vértices da clique encontrada não podem pertencer a uma mesma árvore em uma solução ótima, logo podemos fixar os mesmos como raízes de árvores.

Devido à forma como o modelo de representantes é construído e ao uso de quebra de simetrias no modelo de árvore rotulada, precisamos reordenar os vértices de forma que os vértices da clique sejam os iniciais na ordem. Dessa forma, podemos fixar com valor 1 variáveis  $x$  indexadas por esses vértices, o que pode contribuir para a eficiência dos modelos F\_REP e F\_ROT.

### 5.7 Comparação entre as formulações

Uma primeira análise teórica dos modelos realizada refere-se à ordem da quantidade de restrições, número de variáveis inteiras e contínuas. De forma intuitiva, espera-se que os modelos com menos variáveis (especialmente inteiras) e restrições apresentem melhor performance. Porém, nem sempre isto se evidencia na prática. A Tabela 3 apresenta uma comparação entre os modelos propostos, onde  $v_I$  e  $v_C$  representam a quantidade de variáveis inteiras e contínuas, respectivamente, e  $r$  é o número de restrições. Podemos notar que na tabela que o modelo F\_FLUXO apresenta uma menor quantidade de variáveis inteiras e restrições, podendo ter impacto significativo em seu desempenho.

Tabela 3 – Comparação dos modelos em relação ao número de variáveis e restrições

<b>Modelo</b>	$v_I$	$v_C$	$r$
<b>F_ROT</b>	$O(m \cdot k)$	$O(n \cdot k)$	$O((m+n) \cdot k)$
<b>F_REP</b>	$O(m \cdot n)$	$O(n^2)$	$O(n \cdot m)$
<b>F_FLUXO</b>	$O(m)$	$O(m)$	$O(m+n)$

Fonte: elaborado pelo autor.

Denominamos por F\_ROT<sup>+</sup> e F\_REP<sup>+</sup> os modelos F\_ROT e F\_REP, respectivamente, com adição das desigualdades válidas (apresentadas na Seção 5.5) e aplicado o procedimento de fixação de variáveis (descrito na Seção 5.6). Como estes procedimentos visam fortalecer os modelos na sua relaxação linear, foram realizados experimentos computacionais a fim de avaliar e comparar as relaxações lineares dos modelos apresentados. Para isso, foi utilizado o conjunto de instâncias descritas na Seção 6.1.

A Tabela 4 apresenta o resumo dos resultados obtidos agrupados pelo parâmetro  $k$ , onde foram capturados: (i) o *gap* médio das relaxações, ou seja, a média da distância, em porcentagem, que o valor da relaxação do modelo está para o maior valor de relaxação obtido

entre os cinco modelos e (ii) o tempo médio (em segundos) utilizado pelo modelo para encontrar a solução ótima. O parâmetro  $N$  refere-se ao número de instâncias em cada caso.

Podemos observar que houve uma significativa melhora na relaxação e no tempo de execução dos modelos  $F\_ROT^+$  e  $F\_REP^+$  em relação aos modelos originais  $F\_ROT$  e  $F\_REP$ , indicando a efetividade dos procedimentos propostos. Podemos notar que, quanto maior o  $k$ , mais significativo foi o fortalecimento destes dois modelos. O modelo  $F\_FLUXO$  apresentou as melhores relaxações para  $k \leq 4$ , enquanto que o modelo  $F\_ROT^+$  obteve melhores resultados para os demais casos. Em relação ao tempo, o modelo  $F\_FLUXO$  foi o que obteve os menores tempos. Os modelos  $F\_ROT$  e  $F\_ROT^+$  se mostraram mais eficientes em tempo que os modelos  $F\_REP$  e  $F\_REP^+$ , para  $k \leq 8$ , e inferiores nos demais casos. Acreditamos que esta piora à medida que  $k$  cresce deve-se ao aumento significativo do número de variáveis em  $F\_ROT$  (e  $F\_ROT^+$ ), chegando a ultrapassar o número de variáveis de  $F\_REP$  (e  $F\_REP^+$ ).

Tabela 4 – Relaxação linear dos modelos em relação ao parâmetro  $k$

$k$	$N$	<b>F_ROT</b>		<b>F_ROT<sup>+</sup></b>		<b>F_REP</b>		<b>F_REP<sup>+</sup></b>		<b>F_FLUXO</b>	
		<i>gap</i>	<b>tempo</b>	<i>gap</i>	<b>tempo</b>	<i>gap</i>	<b>tempo</b>	<i>gap</i>	<b>tempo</b>	<i>gap</i>	<b>tempo</b>
1	118	1.20	<b>0.03</b>	1.20	<b>0.03</b>	0.33	0.25	0.33	0.24	<b>0.13</b>	0.05
2	120	3.71	0.33	3.54	0.33	67.38	22.13	66.44	23.56	<b>1.09</b>	<b>0.07</b>
4	120	7.87	2.91	4.22	2.95	71.03	28.47	66.12	28.85	<b>3.76</b>	<b>0.08</b>
8	120	13.31	17.09	<b>2.80</b>	16.95	62.76	27.54	52.74	27.29	13.33	<b>0.09</b>
16	90	12.11	50.69	<b>0.79</b>	40.42	53.51	37.31	45.32	33.91	13.39	<b>0.17</b>
32	60	19.75	123.37	<b>0.04</b>	54.83	43.32	46.24	29.81	20.93	21.57	<b>0.42</b>
64	15	28.29	489.57	<b>0.00</b>	115.29	35.57	55.41	11.20	6.43	29.84	<b>4.36</b>

Fonte: elaborado pelo autor.

Já a Tabela 5 apresenta os *gaps* das relaxações comparados ao limite inferior  $LIM\_INF$ , gerado pelo procedimento apresentado na Seção 3.3. Podemos observar que houve uma mudança significativa no *gap* dos modelos apenas quando  $k = 1$ , pois neste caso  $LIM\_INF$  é sempre o valor de uma solução ótima (árvore geradora mínima). Apesar de não apresentar os melhores *gaps* nos demais casos, o procedimento de limite inferior apresentou *gaps* semelhantes aos da relaxação do modelo  $F\_ROT$  e superiores aos do modelo  $F\_REP$ .

Tabela 5 – Relaxação linear comparada ao Limite Inferior em relação ao parâmetro  $k$ 

$k$	$N$	<b>F_ROT</b>	<b>F_ROT<sup>+</sup></b>	<b>F_REP</b>	<b>F_REP<sup>+</sup></b>	<b>F_FLUXO</b>	<b>LIM_INF</b>
1	118	1.42	1.42	0.55	0.55	0.35	<b>0.00</b>
2	120	3.72	3.54	67.38	66.45	<b>1.10</b>	3.36
4	120	7.87	4.22	71.03	66.12	<b>3.76</b>	7.39
8	120	13.31	<b>2.80</b>	62.76	52.74	13.33	16.05
16	90	12.11	<b>0.79</b>	53.51	45.32	13.39	15.33
32	60	19.75	<b>0.04</b>	43.32	29.81	21.57	22.41
64	15	28.29	<b>0.00</b>	35.57	11.20	29.84	29.95

Fonte: elaborado pelo autor.

## 6 EXPERIMENTOS COMPUTACIONAIS

Neste capítulo apresentamos os testes computacionais realizados. Na Seção 6.1 descrevemos o conjunto de instâncias utilizadas, bem como o seu processo de geração. Na Seção 6.2 é apresentado o ambiente computacional no qual os testes foram realizados. Por fim, na Seção 6.3 é exibida uma comparação das formulações propostas.

### 6.1 Instâncias de Teste

Neste trabalho foram geradas instâncias para o k-FGBA, cada uma consistindo de um grafo, uma ponderação de suas arestas, e um inteiro positivo. Inicialmente foi gerado um conjunto de grafos base, de maneira aleatória, usando o modelo descrito por Erdős e Rényi (1959). Estes grafos foram criados a partir de dois parâmetros: número de vértices ( $n$ ) e densidade  $d$  de arestas. Consideramos todas as combinações de  $n \in \{10, 15, 20, 30, 40, 50, 60, 80\}$  e  $d \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$ .

A partir de um mesmo grafo ponderado, foram geradas diferentes instâncias, apenas variando o valor do parâmetro  $k$ . Mais especificamente, para cada grafo, consideramos  $k = 2^r$ , com  $r = 0, 1, 2, \dots$ , tal que  $k \leq n$ . Estes valores de  $k$  foram adotados devido à aplicação prática do problema descrita na literatura. Os custos das arestas também foram gerados de forma aleatória utilizando uma distribuição uniforme. Dividimos as instâncias em três classes, de acordo com o intervalo que os custos das arestas  $c_{uv}$  poderiam assumir. A classe  $C_1$  refere-se a instâncias onde  $c_{uv} \in [1, 1]$  (ou seja, arestas de custo unitário), e as classes  $C_2$  e  $C_3$  com custos nos intervalos  $[1, 10]$  e  $[1, 1000]$ , respectivamente. Foram geradas 645 instâncias ao total. Apenas as instâncias ( $classe = C_2, d = 0.4, |V| = 10, k = 1$ ) e ( $classe = C_3, d = 0.3, |V| = 10, k = 1$ ) foram consideradas inviáveis para os testes, pois o grafo original é não conexo, resultando assim em um total de 643 instâncias avaliadas.

### 6.2 Ambiente Computacional

Os modelos e algoritmos foram implementados na linguagem C++ e executados em um computador Intel(R) Core(TM) i7-12700, com 3.60 GHz, 32 GBytes de memória RAM e usando o Ubuntu 22.04. Para a execução dos modelos foi utilizado o solver CPLEX em sua versão 22.1. Estabelecemos um tempo limite de 600 segundos para a execução de cada formulação em cada instância.

### 6.3 Avaliação dos Modelos

Nesta seção apresentamos os resultados obtidos ao avaliarmos os modelos propostos, tendo como entrada o conjunto de instâncias descritas na Seção 6.1. Buscamos avaliar os resultados observando os parâmetros importantes para a geração da instância: número de vértices, densidade do grafo e o valor  $k$ .

A Tabela 6 apresenta o resumo dos resultados agrupados pelo parâmetro  $k$ . A coluna % OPT informa a porcentagem de instâncias onde a formulação encontrou uma solução ótima. Já a coluna  $gap_I$  representa o  $gap$  de integralidade médio, ou seja, a distância média, em percentual, que a relaxação linear do modelo está da solução ótima. O  $gap_I$  é calculado apenas no subconjunto de instância, onde algum dos modelos encontrou uma solução ótima dentro do tempo limite. Cada linha da tabela refere-se às instâncias com o mesmo valor do parâmetro  $k$ . A coluna  $N$  refere-se ao número de instâncias em cada caso.

Tabela 6 – Resumo dos resultados em relação ao parâmetro  $k$

Classe	$k$	$N$	F_ROT		F_ROT <sup>+</sup>		F_REP		F_REP <sup>+</sup>		F_FLUXO	
			% OPT	$gap_I$	% OPT	$gap_I$	% OPT	$gap_I$	% OPT	$gap_I$	% OPT	$gap_I$
$C_1$	1	40	<b>100.00</b>	<b>0.00</b>	<b>100.00</b>	<b>0.00</b>	<b>100.00</b>	<b>0.00</b>	<b>100.00</b>	<b>0.00</b>	<b>100.00</b>	<b>0.00</b>
	2	40	<b>100.00</b>	<b>0.89</b>	<b>100.00</b>	<b>0.89</b>	<b>100.00</b>	74.29	100.00	74.29	87.50	<b>0.89</b>
	4	40	<b>97.50</b>	<b>5.58</b>	<b>97.50</b>	<b>5.58</b>	37.50	79.83	37.50	79.68	62.50	<b>5.58</b>
	8	40	<b>92.50</b>	<b>17.56</b>	<b>92.50</b>	<b>17.56</b>	37.50	74.69	37.50	74.24	50.00	<b>17.56</b>
	16	30	<b>80.00</b>	<b>25.57</b>	<b>80.00</b>	<b>25.57</b>	40.00	70.79	40.00	70.66	16.67	<b>25.57</b>
	32	20	65.00	<b>43.75</b>	70.00	<b>43.75</b>	<b>75.00</b>	64.09	<b>75.00</b>	64.00	0.00	<b>43.75</b>
	64	5	20.00	<b>75.00</b>	20.00	<b>75.00</b>	<b>100.00</b>	79.38	<b>100.00</b>	79.32	0.00	<b>75.00</b>
$C_2$	1	39	87.18	1.48	87.18	1.48	<b>100.00</b>	0.60	<b>100.00</b>	0.60	<b>100.00</b>	<b>0.24</b>
	2	40	<b>90.00</b>	7.28	<b>90.00</b>	6.94	82.50	66.38	82.50	65.67	77.50	<b>5.41</b>
	4	40	52.50	21.90	52.50	17.34	42.50	<b>68.26</b>	45.00	60.40	42.50	<b>20.58</b>
	8	40	45.00	44.66	<b>50.00</b>	<b>29.72</b>	<b>50.00</b>	68.72	<b>50.00</b>	54.35	32.50	47.24
	16	30	50.00	43.64	53.33	<b>40.02</b>	<b>70.00</b>	69.53	66.67	67.97	20.00	44.51
	32	20	50.00	45.83	70.00	<b>42.62</b>	80.00	64.09	<b>95.00</b>	63.22	0.00	45.89
	64	5	0.00	75.00	<b>100.00</b>	<b>74.07</b>	<b>100.00</b>	78.08	<b>100.00</b>	77.92	0.00	75.00
$C_3$	1	39	<b>100.00</b>	2.81	<b>100.00</b>	2.81	<b>100.00</b>	1.05	<b>100.00</b>	1.05	<b>100.00</b>	<b>0.83</b>
	2	40	67.50	12.84	67.50	12.60	77.50	63.10	72.50	60.66	<b>80.00</b>	<b>8.38</b>
	4	40	40.00	34.08	40.00	<b>19.44</b>	<b>50.00</b>	62.18	47.50	44.54	35.00	31.55
	8	40	37.50	55.70	47.50	<b>18.27</b>	<b>50.00</b>	66.49	<b>50.00</b>	30.50	32.50	61.81
	16	30	33.33	63.22	43.33	<b>26.44</b>	63.33	71.23	<b>66.67</b>	38.84	20.00	67.50
	32	20	35.00	71.50	75.00	<b>32.57</b>	85.00	74.25	<b>100.00</b>	43.76	10.00	74.57
	64	5	0.00	83.24	<b>100.00</b>	<b>7.34</b>	<b>100.00</b>	76.22	<b>100.00</b>	8.25	20.00	87.45

Fonte: elaborado pelo autor.

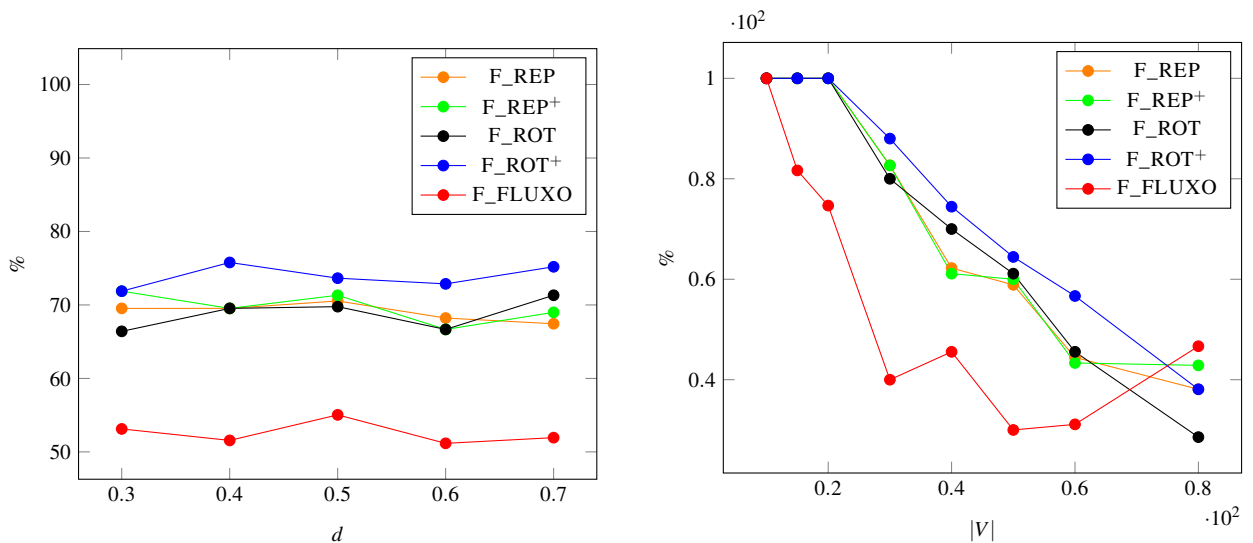
Observando os resultados da tabela, podemos verificar que, como esperado, as formulações determinam uma solução ótima para praticamente todas instâncias com  $k = 1$ . Além



disso, de modo geral, os modelos  $F\_ROT^+$  e  $F\_REP^+$  obtiveram uma solução ótima em um maior número de instâncias, enquanto que os modelos  $F\_ROT$  e  $F\_FLUXO$  obtiveram os melhores *gaps* de integralidade. Também como esperado, notamos que as instâncias são mais difíceis para valores intermediários de  $k$ . Podemos notar que o modelo  $F\_ROT$  e  $F\_ROT^+$  se destacaram positivamente na classe de instâncias  $C_1$ , onde temos custos unitários.

Uma outra análise realizada avaliou a porcentagem de soluções ótimas encontradas dentro do tempo limite, variando os parâmetros de densidade ( $d$ ) e número de vértices ( $n$ ). A Figura 13 apresenta um resumo desta avaliação. Podemos observar nos dois gráficos que as formulações  $F\_ROT^+$  e  $F\_REP^+$  encontraram um maior percentual de soluções ótimas em relação aos demais modelos. Além disso, podemos observar que a densidade do grafo parece não ter um impacto significativo em relação a determinar uma solução ótima. Por outro lado, como esperado, quanto maior o número de vértices menos instâncias foram resolvidas dentro do limite de tempo fixado.

Figura 13 – Comparação entre os modelos propostos em relação à densidade e ao número de vértices.



(a) Porcentagem de soluções ótimas em relação a  $d$

(b) Porcentagem de soluções ótimas em relação a  $|V|$

Fonte: elaborado pelo autor.

De forma geral, houve uma melhora significativa do modelo  $F\_ROT$  para o modelo  $F\_ROT^+$ . Este último conseguiu encontrar solução ótima (e comprovar otimalidade) em cerca de 7,46% de instâncias a mais que o primeiro, evidenciando que as desigualdades válidas e o procedimento de fixação de variáveis fortaleceram o modelo original. Nos modelos  $F\_REP$  e

F\_REP<sup>+</sup> essa melhora não foi tão significativa. Os modelos F\_REP e F\_REP<sup>+</sup> se mostram um pouco mais eficientes que os demais modelos nas classes  $C_2$  e  $C_3$ , onde os custos das arestas são esparsos. Já na classe  $C_1$ , onde os custos são unitários, F\_ROT e F\_ROT<sup>+</sup> apresentaram uma melhora significativa em relação aos demais modelos, conseguindo definir uma solução ótima dentro do tempo limite em mais de 90 % das instâncias avaliadas. Já o modelo F\_FLUXO se mostrou eficiente apenas para quando  $k$  assume valores pequenos ( $k < 4$ ). O modelo apresenta piores resultados à medida que o parâmetro  $k$  aumenta, isto pode estar relacionado com a quantidade de árvores de mesmo custo nessas instâncias, gerando assim soluções simétricas a serem analisadas pelo modelo.

Mais precisamente, os modelos F\_REP<sup>+</sup> e F\_ROT<sup>+</sup> definiram solução ótima em 69,05% e 73,87% das instâncias, respectivamente. Acredita-se que F\_ROT e F\_ROT<sup>+</sup> conseguiram resolver menos instâncias (com exceção da classe  $C_1$ ) para os casos onde  $k \geq 8$  devido ao número de variáveis inteiras aumentar de forma significativa à medida que  $k$  aumenta.

## 7 CONCLUSÃO

Neste trabalho, abordamos o problema da  $k$ -Floresta Geradora Balanceada em Arestas ( $k$ -FGBA), que consiste em encontrar uma floresta geradora de um grafo não direcionado com pesos nas arestas, minimizando o peso da árvore mais pesada, com um número fixo de árvores. Apresentamos novos limitantes para o problema e um algoritmo polinomial exato para o caso onde  $G$  é um caminho. Propomos duas novas heurísticas para o problema e as avaliamos computacionalmente em instâncias geradas aleatoriamente. Propomos três formulações de Programação Linear Inteira Mista (PLIM) para o problema, bem como algoritmos heurísticos e técnicas de otimização para fortalecer os modelos.

Os resultados computacionais obtidos demonstraram que as formulações propostas são eficazes na resolução do problema, encontrando soluções ótimas em uma grande parte das instâncias avaliadas. Além disso, as heurísticas desenvolvidas mostraram-se úteis como limitantes superiores para o problema.

A inclusão de desigualdades válidas e o uso do procedimento de fixação de variáveis na formulação  $F\_ROT$  levaram a uma formulação,  $F\_ROT^+$ , com expressivo melhor desempenho. De todo modo, dentre as formulações propostas, os modelos  $F\_ROT^+$  e  $F\_REP^+$  se destacaram por sua capacidade de encontrar soluções ótimas em um maior número de instâncias.

Em resumo, este trabalho contribui para o avanço no estudo e na compreensão do problema  $k$ -FGBA, apresentando novas formulações, algoritmos e técnicas que podem ser úteis em aplicações práticas em diversas áreas, como otimização combinatória e topologia computacional.

Vale destacar que resultados parciais deste trabalho já foram publicados em Costa *et al.* (2022b) e em Costa *et al.* (2022a).

Como trabalhos futuros, propomos a investigação de formulações mais elaboradas, a busca por novas desigualdades válidas e métodos de solução mais sofisticados, que melhor explorem propriedades do problema. Além disso, a realização de um estudo teórico para os casos  $k \in \{2, 3, 4\}$  e para o caso onde as arestas possuem peso unitário, com o objetivo de propor métodos mais eficientes de solução do problema para estas situações específicas.

## REFERÊNCIAS

- ANDERSSON, M.; GUDMUNDSSON, J.; LEVCOPOULOS, C.; NARASIMHAN, G. Balanced partition of minimum spanning trees. **International Journal of Computational Geometry & Applications**, World Scientific, v. 13, n. 4-5, p. 303–316, 2003.
- BRÉLAZ, D. New methods to color the vertices of a graph. **Communications of the ACM**, ACM New York, NY, USA, v. 22, n. 4, p. 251–256, 1979.
- CAMPÊLO, M.; CAMPOS, V.; CORRÊA, R. On the asymmetric representatives formulation for the vertex coloring problem. **Discrete Applied Mathematics**, v. 156, n. 7, p. 1097–1111, 2008.
- CHEN, G.; CHEN, Y.; CHEN, Z.-Z.; LIN, G.; LIU, T.; ZHANG, A. Approximation algorithms for the maximally balanced connected graph tripartition problem. **Journal of Combinatorial Optimization**, Springer, p. 1–21, 2020.
- CORMEN, T. H.; LEISERSON, C. E.; RIVEST, R. L.; STEIN, C. Algoritmos: teoria e prática. **Editora Campus**, v. 2, p. 296, 2002.
- COSTA, J. R. d. F.; SOUSA, G. H. d.; CAMPÊLO, M. Exact methods for the minimum spanning k-forest problem. In: **Proceedings of CLAIO 2022**. Buenos Aires: [S. n.], 2022. p. 52–53.
- COSTA, J. R. d. F.; SOUSA, G. H. d.; CAMPÊLO, M. Formulações para o problema da k-floresta geradora mínima. In: **Anais do VII Encontro de Teoria da Computação (ETC)**. Niterói: Sociedade Brasileira de Computação, 2022. p. 101–104. ISSN 2595-6116.
- DANTZIG, G. B.; FULKERSON, D. R.; JOHNSON, S. M. Solution of a large scale traveling salesman problem. **Operations Research**, INFORMS, v. 2, n. 4, p. 393–410, 1954.
- DESROCHERS, M.; LAPORTE, G. Improvements and extensions to the Miller-Tucker-Zemlin subtour elimination constraints. **Operations Res. Letters**, v. 10, p. 27 – 36, 1991.
- DYER, M. E.; FRIEZE, A. M. On the complexity of partitioning graphs into connected subgraphs. **Discrete Applied Mathematics**, Elsevier, v. 10, n. 2, p. 139–153, 1985.
- EISEN, M. B.; SPELLMAN, P. T.; BROWN, P. O.; BOTSTEIN, D. Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 95, n. 25, p. 14863–14868, 1998.
- ERDŐS, P.; RÉNYI, A. On random graphs. **Publicationes Mathematicae**, v. 6, p. 290–297, 1959.
- FIGUEREDO, P. J. A.; CAMPÊLO, M. O problema da floresta geradora k-rotulada. **Anais do LII SBPO**, Sobrapo, João Pessoa, v. 52, p. 127515–127526, 2020.
- FILHO, F. S. d. F. Dissertação (Mestrado em Ciência da Computação), **O problema da k-floresta com máximo número de folhas**. 2019.
- FREDERICKSON, G. N.; ZHOU, S. Optimal parametric search for path and tree partitioning. **arXiv preprint arXiv:1711.00599**, 2017.
- GUTTMANN-BECK, N.; HASSIN, R. Approximation algorithms for min–max tree partition. **Journal of Algorithms**, Elsevier, v. 24, n. 2, p. 266–286, 1997.

GUTTMANN-BECK, N.; HASSIN, R. Approximation algorithms for minimum tree partition. **Discrete Applied Mathematics**, Elsevier, v. 87, n. 2, p. 117–137, 1998.

HILLIER, F. S.; LIEBERMAN, G. J. **Introdução à pesquisa operacional**. Porto Alegre, RS: AMGH Editora Ltda, 2013.

LOVÁSZ, L. A homology theory for spanning trees of a graph. **Acta Mathematica Academiae Scientiarum Hungarica**, v. 30, p. 241–251, 1977.

MADKOUR, A.-R.; NADOLNY, P.; WRIGHT, M. Finding minimum spanning forests in a graph. **arXiv preprint arXiv:1705.00774**, 2017.

MARTIN, R. K. Using separation algorithms to generate mixed integer model reformulations. **Operations Research Letters**, v. 10, p. 119–128, 1991.

MILLER, C. E.; TUCKER, A. W.; ZEMLIN, R. A. Integer programming formulation of traveling salesman problems. **Journal of the ACM (JACM)**, v. 7, n. 4, p. 326–329, 1960.

MIN, M.; DU, H.; JIA, X.; HUANG, C. X.; HUANG, S. C.-H.; WU, W. Improving construction for connected dominating set with steiner tree in wireless sensor networks. **Journal of Global Optimization**, Springer, v. 35, n. 1, p. 111–119, 2006.

MIYAZAWA, F. K. *et al.* Cut and flow formulations for the balanced connected k-partition problem. In: **International Symposium on Combinatorial Optimization**. Cham: Springer International Publishing, 2020. p. 128–139.

PREPARATA, F. P.; SHAMOS, M. I. **Computational geometry: an introduction**. [S. l.]: Springer Science & Business Media, 2012.

VAISHALI, S.; ATULYA, M.; PUROHIT, N. Efficient algorithms for a graph partitioning problem. In: **International Workshop on Frontiers in Algorithmics**. [S. l.: s. n.], 2018. p. 29–42.

WANG, L.; ZHANG, Z.; WU, D.; WU, W.; FAN, L. Max-min weight balanced connected partition. **Journal of Global Optimization**, Springer, v. 57, n. 4, p. 1263–1275, 2013.

WEST, D. B. *et al.* **Introduction to graph theory**. [S. l.]: Prentice hall Upper Saddle River, 2001. v. 2.

WOLSEY, L. **Integer Programming**. [S. l.]: John Wiley & Sons, 1998. ISBN 978-0-471-28366-9.

WU, B. Y.; CHAO, K.-M. **Spanning trees and optimization problems**. [S. l.]: CRC Press, 2004.











Table with columns: Classe, n, d, k, F TREE (Tempo, Sol, RL), F TREE+ (Tempo, Sol, RL), F REP (Tempo, Sol, RL), F REP+ (Tempo, Sol, RL), and F FLUXO (Tempo, Sol, RL). The table contains numerous rows of data for various classes and parameters.

Continua na próxima página

