



**FUNDAÇÃO UNIVERSIDADE
FEDERAL DE MATO GROSSO DO SUL**

Faculdade de Computação - FACOM

ATIVIDADE ORIENTADA DE ENSINO

**Porcentagens de particionamento do relax and fix e nova proposta de
PLI para o problema da filogenia viva com politomia**

ALUNA: Maria Elisa Rodrigues Rabello

ORIENTADORA: Edna Ayako Hoshino

Campo Grande - MS

2025

1 Introdução

A filogenia foi definida pela primeira vez pelo entomólogo alemão Willi Hennig em 1950. No seu livro *Filogenia Sistemática*, Hennig descreve a filogenia como um método de análise da evolução das espécies que permite a construção de relações de parentesco entre elas [Hennig, 1999]. A partir de análises filogenéticas é possível construir um grafo que relaciona as espécies de acordo com seu parentesco. Esse grafo é chamado de árvore filogenética. As folhas dessa árvore são as espécies existentes, os nós internos representam espécies hipotéticas que deram origem a outras até a espécie representada nas folhas, e cada espécie possui apenas dois descendentes [Lenzini and Marianelli, 1997].

1.1 O problema da filogenia viva

Descrita por Telles et al. [2013], a árvore filogenética do problema da filogenia viva pode possuir em seus nós internos objetos do conjunto S de entrada, ou seja, objetos vivos podem ser ancestrais de outros objetos vivos. Agora, uma aresta (x, y) que liga dois vértices x e y pode conectar dois vértices x e y internos, x ou y pertencendo ao conjunto S e os dois vértices sendo objetos do conjunto S . Telles et al. [2013] descreve formalmente a árvore filogenética viva baseada em distância conforme segue:

Seja M^n uma matriz quadrada de ordem n , e S o conjunto de n objetos representados na matriz M , em que $M^n_{ij} \in \mathbb{R}$ é a distância entre os objetos i e j . E seja T^n uma árvore ponderada, sem raiz. T^n é uma árvore viva para M^n se T^n :

- Cada folha em T^n representa um objeto de S ,
- cada objeto em S aparece apenas uma vez em T^n
- $d^n_{ij} = M^n_{ij}$, $1 \leq i, j \leq n$, onde d^n_{xy} é a distância entre x e y em T^n .

Um vértice interno será vivo se representa um objeto em S .

E seguindo o comportamento da filogenia, quando a matriz de distância é aditiva para o problema da filogenia viva existe algoritmo que resolve o problema em tempo polinomial, mas quando a matriz não é aditiva então o problema da filogenia viva é NP-difícil provado por Araújo et al. [2017].

1.2 O problema da filogenia viva com politomia baseada em distância

Robinson and Foulds [1981] demonstra que as árvores filogenéticas com politomia são isomorfas às árvores filogenéticas sem politomias se o conjunto de objetos presentes nas duas árvores são os mesmos e preservarem a posição dos vértices que representam os objetos do conjunto S . Os autores inclusive consideram a presença de nós vivos nas árvores. É demonstrado que é possível chegar a árvore sem politomia a partir da árvore com politomia e vice-versa com o mesmo número α de passos preservando a posição dos vértices do conjunto S garantido o isomorfismo. E sendo árvores isomórficas construir as duas é igualmente difícil.

Uma árvore filogenética com politomia é uma árvore multifurcada. Isso significa dizer que seus nós internos podem possuir mais de três vizinhos, diferente da árvore bifurcada tradicional em problemas de filogenia e filogenia viva.

Combinada com a filogenia viva, agora o problema pode gerar uma árvore em que os nós internos não só podem ser objetos vivos, como também podem ser relacionados com mais de 3 objetos.

1.3 Relax & fix

A heurística *relax & fix* consiste em particionar o conjunto de variáveis inteiras de um problema em conjuntos menores. O problema então é resolvido iterativamente em que, a cada iteração, somente as variáveis

associadas à parte atual são consideradas inteiras. A cada iteração, as variáveis correspondentes às partes de iterações anteriores tem seus valores fixados, ou seja, não serão mais alterados. A ideia do método é que, ao final das iterações, todas as variáveis terão seus valores fixados e, portanto, uma solução viável será encontrada, como demonstrado na Figura 1. No entanto, pode ocorrer da sequência de variáveis fixadas levar a uma situação em que não há uma solução viável [Pochet and Wolsey, 2006].



Figura 1: Modelo de execução do *relax & fix*.

1.4 Modelo de PLI utilizado com o relax & fix

O modelo de programação linear inteiro (PLI) a seguir foi utilizado implementado para testar as matheurísticas propostas descritas na próxima seção.

Dados um conjunto S de objetos de estudo, um grafo orientado $G = (V \cup H, A)$, no qual V é o conjunto de objetos em S e H o conjunto de objetos hipotéticos e A o conjunto de arcos (i, j) e (j, i) , uma função de distância $M : V \times V \mapsto \mathbb{R}^+$ que atribui uma distância a cada par de vértice em V de G , o problema irá buscar uma árvore T de G e uma função de peso $c : A \mapsto \mathbb{R}^+$ tal que:

- $V \subseteq V[T]$,
- $d(i) \geq 3$ para todo vértice interno i em T e

- $\sum_{i,j \in T} |M_{ij} - p_{ij}|$ seja mínima, onde p_{ij} é o custo somado dos pesos das arestas do único caminho em T que liga os vértices i e j

Os vértices em H podem ou não aparecer em T .

As seguintes notações são adotadas no modelo:

- M_{ij} é a distância entre os objetos i e j da matriz M ;
- V representa o conjunto dos objetos vivos;
- H representa os objetos hipotéticos;
- $V' = V \cup H$;
- $\delta^-(i)$ arestas que entram em i ;
- $\delta^+(i)$ arestas que saem de i .

As variáveis de decisão do modelo com solução (T, c) , onde T é a árvore filogenética e c a função de distância, são:

- x_a é uma variável binária em que x_a é igual a 1 se, e somente se, a aresta a for pertencente à árvore;
- y_a^i é o fluxo no arco a para o vértice i , representando que o arco participa do caminho da raiz até o vértice i ;
- z_i é uma variável binária tal que z_i é igual a 1 se e somente se i é um vértice interno em T ;
- g_i^{kl} indica que i é o último ancestral comum entre k e l ;
- p_{ij} é a distância entre um vértice i e um vértice j na árvore.
- d_{ij} é a diferença entre M_{ij} e p_{ij} ;

Considere que a raiz r da árvore é um vértice arbitrário de T (e que não tem relação com ancestralidade, sendo selecionado aleatoriamente para realizar a construção da árvore, que será uma árvore filogenética viva com politomia sem raiz), considere também que fluxo é a existência de um caminho de r até um vértice i e que o último ancestral comum é o último vértice pelo qual passa um fluxo comum entre i e j . O modelo de programação linear inteira para o problema da filogenia viva com politomia é definido como se segue:

$$\begin{aligned}
\min \quad & \sum_{i \in V} \sum_{j \in V} d_{ij} \\
s.a. \quad & \sum_{a \in \delta^-(k)} y_a^k - \sum_{a \in \delta^+(k)} y_a^k = 1, & \forall k \in V, k \neq r \quad (01) \\
& \sum_{a \in \delta^-(i)} y_a^k - \sum_{a \in \delta^+(i)} y_a^k = 0, & \forall k \in V, i \in V', i \neq k \quad (02) \\
& y_a^k \leq x_a, & \forall a \in V' \times V', \forall k \in V, k \neq r \quad (03) \\
& \sum_{a \in \delta^-(r)} x_a = 0, & (04) \\
& \sum_{a \in \delta^-(i)} x_a = z_i, & \forall i \in H \quad (05) \\
& \sum_{a \in \delta^+(r)} x_a \geq 1 + 2z_r & (06) \\
& \sum_{a \in \delta^+(r)} x_a \leq |V'|z_r + 1, & (07) \\
& \sum_{a \in \delta^+(i)} x_a \geq 2z_i, & \forall i \in V' \quad (08) \\
& \sum_{a \in \delta^+(i)} x_a \leq |V'|z_i, & \forall i \in V' \quad (09) \\
& d_{ij} \leq M_{ij} - p_{ij}, & \forall i, \forall j \in V \quad (10) \\
& d_{ij} \leq p_{ij} - M_{ij}, & \forall i, \forall j \in V \quad (11) \\
& p_{rk} \geq p_{ri} + p_{ij} + p_{jk} - M(1 - y_{ij}^k), & \forall k \in V \setminus \{r\}, \forall i, j \in V', i \neq j, i \neq k, j \neq k \quad (12) \\
& p_{rk} \leq p_{ri} + p_{ij} + p_{jk} + M(1 - y_{ij}^k), & \forall k \in V \setminus \{r\}, \forall i, j \in V', i \neq j, i \neq k, j \neq k \quad (13) \\
& p_{kl} \geq p_{ik} + p_{il} - M(1 - g_i^{kl}), & \forall k, l \in V \setminus \{r\}, k \neq l, \forall i \in V', i \neq k, i \neq l \quad (14) \\
& p_{kl} \leq p_{ik} + p_{il} + M(1 - g_i^{kl}), & \forall k, l \in V \setminus \{r\}, k \neq l, \forall i \in V', i \neq k, i \neq l \quad (15) \\
& g_i^{kl} \geq (y_{ij}^k - y_{ij}^l) + (\sum_{x \in V'} y_{xi}^l - 1), & \forall k, l \in V \setminus \{r\}, k \neq l, \forall i \in V', i \neq k, i \neq l \quad (16) \\
& g_i^{kl} \geq (y_{ij}^l - y_{ij}^k) + (\sum_{x \in V \cup H} y_{xi}^k - 1), & \forall k, l \in V \setminus \{r\}, k \neq l, \forall i \in V', i \neq k, i \neq l \quad (17) \\
& \sum_{a \in \delta^-(i)} x_a = 1 & \forall i \in V \quad (18) \\
& x_{ij} + x_{ji} \leq 1, & \forall i, \forall j \in V \quad (19) \\
& x_a, g_k^{ij}, z_i \in \{0, 1\} & (20) \\
& y_a^k \in \mathbb{Z} & (21) \\
& p_{ij}, d_{ij} \in \mathbb{R}, & \forall i \in V, \forall j \in V \quad (22)
\end{aligned}$$

As restrições (01) e (02) aplicam restrição de fluxo, garantindo uma unidade de fluxo ao vértice k , saindo de r até k . A restrição (03) garante a existência do fluxo para o vértice (caso $y_a^k = 1$), ou seja, caso o arco a de G exista na árvore T . As restrições (04), (06) e (07) dizem respeito à raiz da árvore. A restrição (04) garante que nenhum fluxo chegue a raiz, a (06) garante que da raiz saiam pelo menos 3 arcos com fluxo e a (07) garante que, caso a raiz seja um vértice folha, que saia apenas um arco com fluxo da raiz. A restrição (05), obriga que todo vértice hipotético tenha um arco de fluxo chegando nele e a restrição (18) aplica o mesmo aos vértices de objetos reais. As restrições (08), (09) referem-se a todos os vértices menos a raiz, com a (08) garantindo que

pelo menos dois arcos de fluxos saiam dos vértices internos e a (09) que caso sejam folhas, não tenham fluxos saindo dos mesmos. As restrições (10) e (11) forçam a distância entre dois vértices reais a ser o módulo da diferença entre a distância na árvore e a distância na matriz M . As restrições (12) e (13) forçam que a distância da raiz a um vértice qualquer seja igual à soma das distâncias de todas as arestas no caminho entre raiz e dito vértice. Já as restrições (14) e (15) forçam que a distância entre dois vértices, em que nenhum deles é a raiz, seja a soma da distância das arestas no caminho entre o último ancestral comum e eles. As restrições (16) e (17) definem que o último ancestral i comum a dois vértices k e l é aquele que recebe fluxo para os dois e passa fluxo para um através de um arco $[i, k]$ e para o outro através de $[i, l]$. A restrição (19) força a chegada de um fluxo ao vértice j a partir de i . Por fim, as restrições (20), (21) e (22) definem o domínio das variáveis.

Esta atividade orientada de ensino teve como objetivo analisar o comportamento e resultados do particionamento do conjunto de variáveis com diferentes valores percentuais e desenvolver um novo PLI para o problema da filogenia viva com politomia.

2 Trabalho desenvolvido

2.1 Particionamento do relax & fix

Sendo $N = 2 \times$ número de vértices, $m = N(N - 1)$, t o tamanho de cada parte do conjunto de partições dado por $t = per \times m$, onde per é um valor percentual e K a quantidade de partes do conjunto de partições dado por $K = 1/per$, as matheurísticas a seguir foram implementadas utilizando o relax & fix.

2.1.1 Matheurística sequencial

A heurística sequencial particiona o conjunto de variáveis seguindo a ordem em que é construído. Dado um conjunto ordenado $Vars$ que contém as v variáveis, um conjunto ordenado $Part$ vazio que armazenará as partes com cada subconjunto de variáveis, essa matheurística irá armazenar as t primeiras variáveis de $Vars$ na primeira parte de $Parts$, as seguintes t variáveis na segunda parte até que todas as K partes tenham sido preenchidas.

2.1.2 Matheurística aleatória

A heurística aleatória organiza o conjunto de variáveis de forma aleatória. Dados os mesmo conjuntos e valores anteriormente mencionados, a heurística aleatória seleciona uma variável aleatoriamente de $Vars$ e adiciona na primeira parte de $Part$, seleciona aleatoriamente outra variável de $Vars$ e novamente adiciona na primeira parte de $Part$ até que a primeira parte e as demais estejam preenchidas com todas as variáveis.

2.2 Novo modelo de PLI

O seguinte modelo de PLI foi proposto, mas não implementado.

Seguindo as definições e as notações do modelo descrito na introdução, as variáveis de decisão são definidas como se segue:

- x_a é uma variável binária em que x_a é igual a 1 se, e somente se, a aresta a for pertencente à árvore;
- y_a^i é o fluxo no arco a para o vértice i ;
- $\hat{y}_a^{ij}$ é o fluxo no arco a comum para i e j ;
- z_i é uma variável binária tal que z_i é igual a 1 se e somente se i é um vértice interno em T ;
- g_k^{il} indica que i é o último ancestral comum entre k e l ;

- c_{ij} representa o custo do caminho i, j na árvore T ;
- d_{ij} é a diferença em valor absoluto entre M_{ij} e c_{ij} ;

Considere que a raiz r da árvore é um vértice arbitrário de T (e que não tem relação com ancestralidade, sendo selecionado aleatoriamente para realizar a construção da árvore, que será uma árvore filogenética viva com politomia sem raiz), considere também que fluxo é a existência de um caminho de r até um vértice i e que o último ancestral comum é o último vértice pelo qual passa um fluxo comum entre i e j . O novo modelo de programação linear inteira para o problema da filogenia viva com politomia é definido como se segue:

$$\begin{aligned}
\min \quad & \sum_{i \in V} \sum_{j \in V} d_{ij} \\
s.a. \quad & \sum_{a \in \delta^-(k)} y_a^k - \sum_{a \in \delta^+(k)} y_a^k = 1, & \forall k \in V, k \neq r \quad (01) \\
& \sum_{a \in \delta^-(i)} y_a^k - \sum_{a \in \delta^+(i)} y_a^k = 0, & \forall k \in V, i \in V', i \neq k \quad (02) \\
& y_a^k \leq x_a, & \forall a \in V' \times V', \forall k \in V, k \neq r \quad (03) \\
& \sum_{a \in \delta^-(r)} x_a = 0, & (04) \\
& \sum_{a \in \delta^-(i)} x_a = z_i, & \forall i \in H \quad (05) \\
& \sum_{a \in \delta^+(r)} x_a \geq 1 + 2z_r & (06) \\
& \sum_{a \in \delta^+(r)} x_a \leq |V'|z_r + 1, & (07) \\
& \sum_{a \in \delta^+(i)} x_a \geq 2z_i, & \forall i \in V' \quad (08) \\
& \sum_{a \in \delta^+(i)} x_a \leq |V'|z_i, & \forall i \in V' \quad (09) \\
& d_{ij} \leq M_{ij} - p_{ij}, & \forall i, \forall j \in V \quad (10) \\
& d_{ij} \leq p_{ij} - M_{ij}, & \forall i, \forall j \in V \quad (11) \\
& c_{ij} = c_{ri} + c_{rj} - 2 \sum_{k \in V'} c_{rk} \cdot g_k^{ij}, & \forall i, \forall j \in V \quad (12) \\
& \sum_{k \in V'} g_k^{ij} = 1, & \forall i, \forall j \in V \quad (13) \\
& \hat{y}_a^{ij} \leq y_a^i, & \forall i, \forall j \in V, \quad \forall a \in V' \times V' \quad (14) \\
& \hat{y}_a^{ij} \leq y_a^j, & \forall i, \forall j \in V, \quad \forall a \in V' \times V' \quad (15) \\
& \hat{y}_a^{ij} \geq y_a^i + y_a^j - 1, & \forall i, \forall j \in V, \quad \forall a \in V' \times V' \quad (16) \\
& g_k^{ij} = \sum_{e \in \delta^-(k)} \hat{y}_e^{ij} - \sum_{e \in \delta^+(k)} \hat{y}_e^{ij}, & \forall i, \forall j \in V, \quad \forall k \in V', \quad k \neq r \quad (17) \\
& \sum_{a \in \delta^-(i)} x_a = 1 & \forall i \in V \quad (18) \\
& x_a, z_i, g_k^{ij} \in \{0, 1\} & (19) \\
& y_a^k, \hat{y}_a^{ij} \in \mathbb{Z} & (20) \\
& c_{ij}, d_{ij} \in \mathbb{R} & (21)
\end{aligned}$$

As restrições (01) e (02) aplicam restrição de fluxo, garantindo uma unidade de fluxo ao vértice k , saindo de r até k . A restrição (03) impede a existência de fluxo no arco a (ou seja $y_a^k = 1$), caso o arco a não faça parte da árvore T . As restrições (04), (06) e (07) dizem respeito à raiz da árvore. A restrição (04) garante que nenhum fluxo chegue a raiz, a (06) garante que da raiz saiam pelo menos 3 arcos com fluxo e a (07) garante que, caso a raiz seja um vértice folha, que saia apenas um arco com fluxo da raiz. A restrição (05), obriga que todo vértice

hipotético que seja interno à árvore tenha um arco chegando nele e a restrição (18) aplica o mesmo aos vértices de objetos reais. As restrições (08), (09) referem-se a todos os vértices menos a raiz, com a (08) garantindo que pelo menos dois arcos de fluxos saiam dos vértices internos e a (09) que caso sejam folhas, não tenham fluxos saindo dos mesmos. As restrições (10) e (11) forçam a distância entre dois vértices reais a ser o módulo da diferença entre a distância na árvore e a distância na matriz M . A restrição (12) informa o resultado do custo c_{ij} para as três possibilidades de ancestral comum. O somatório (13) estabelece que k é o último ancestral comum entre os vértices i e j . O sistema constituído pelas restrições (14), (15) e (16) refere-se ao controle de fluxo comum (entre i e j) para a árvore orientada. A restrição (17) apresenta a definição do último ancestral comum. A restrição (18) força a chegada de um fluxo da raiz a cada objeto i . As restrições (19) a (21) representam o domínio das variáveis.

3 Resultados

A seguir são apresentados os resultados da implementação da divisão das partições do relax & fix com valores percentuais. O tamanho das partes foi estabelecido com 3%, 5% e 10% do valor de $m = N \times (N - 1)$, onde N é duas vezes o número de vértices, sendo cada porcentagem aplicada as duas matheurísticas.

A tabela a seguir apresenta os valores obtidos para os limitantes para cada matheurística em cada teste com uma porcentagem de partição diferente. Como o limitante inferior foi igual para todos os teste, a coluna Limitante inferior diz respeito a todos os teste para as duas matheurísticas. As colunas relacionadas ao limitante superior seguem o padrão: perx_h, onde x é o valor percentual (3, 5 e 10 por cento) e h a sigla da matheurística (seq para sequencial e ale para aleatória).

Tabela 1: Limitantes inferiores e superiores para os três teste com os três valores de porcentagens diferentes

Instância	Limitante inferior	Limitante superior					
	Todos	per3_seq	per3_ale	per5_seq	per5_ale	per10_seq	per10_ale
d_10_2_1_124_466_1	637,00	*	*	*	*	*	*
d_10_4_1_209_928_1	0,00	*	*	*	*	*	*
d_4_1_1	0,00	0	0	0	3	0	0
d_6_2_1_1	0,00	571,80	*	*	*	*	*
d_6_2_1_10	2,80	508,38	*	555,78	*	*	*
d_6_2_1_2	1,70	43,17	43,17	43,17	43,17	43,17	43,17
d_6_2_1_3	0,00	309,60	*	537,60	*	*	*
d_6_2_1_4	0,00	481,60	807,40	417,50	*	*	*
d_6_2_1_5	0,00	554,80	*	*	*	*	*
d_6_2_1_6	0,00	434,60	*	*	*	*	*
d_6_2_1_7	0,00	201,58	*	*	*	495,28	*
d_6_2_1_8	0,00	712,58	*	*	*	*	*
d_6_2_1_9	2,80	691,08	*	*	*	*	*
d_6_4_1_1	0,00	*	*	*	*	*	*
d_6_4_1_10	3,24	*	*	*	*	*	*
d_6_4_1_2	2,70	*	*	*	*	*	*
d_6_4_1_3	0,00	*	*	*	*	*	*
d_6_4_1_4	0,00	*	*	*	*	*	*
d_6_4_1_5	0,00	*	*	*	*	*	*
d_6_4_1_6	0,00	*	*	*	*	*	*
d_6_4_1_7	0,00	*	*	*	*	*	*
d_6_4_1_8	2,30	*	*	*	*	*	*
d_6_4_1_9	3,24	*	*	*	*	*	*
Média	49,98	450,92	425,28	388,51	43,17	269,22	43,17

Nenhum teste achou solução em instâncias com mais de 8 vértices. A matheurística aleatória achou solução em apenas uma instância com 5% e 10% de valor percentual sobre m . A matheurística aleatória achou soluções em 5 instâncias com 8 objetos com o teste de 3%. Os valores dos limitantes superiores foram altos para todas as instâncias (com média de 450.92), menos uma (43.17).

Das soluções encontradas destaca-se a árvore gerada pelas duas matheurísticas em todos os três testes (Figura 2). Todos os testes geraram o mesmo valor de solução e a mesma árvore resultante.

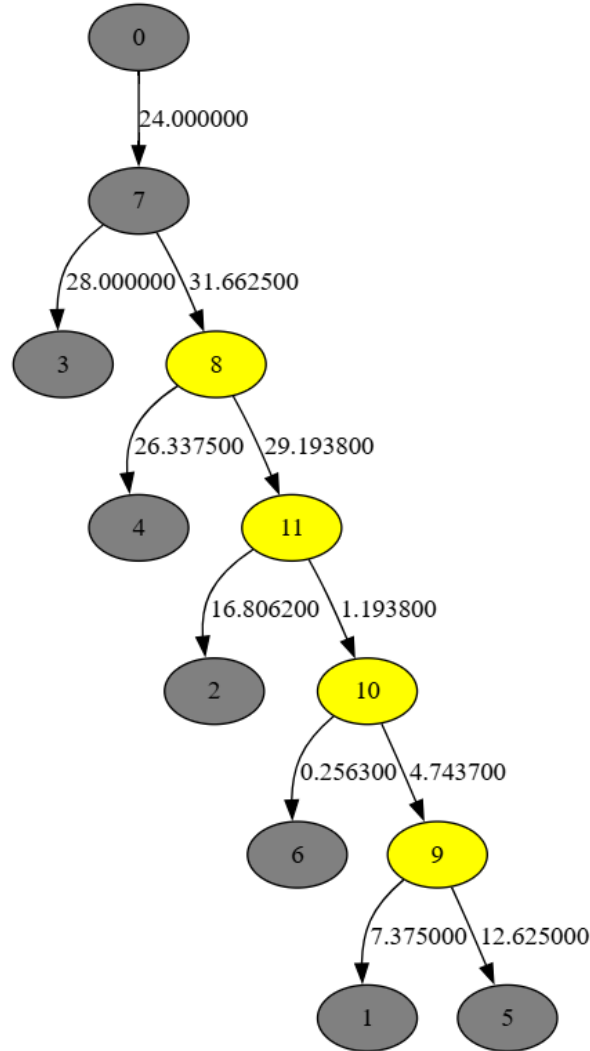


Figura 2: Árvore gerada pelas duas matheurísticas nos três testes para a instância d_6_2_1_2. Note que há apenas um objeto interno vivo (nós amarelos são nós hipotéticos) mas não há politomia.

Já as árvores a seguir foram geradas por matheurísticas diferentes. A árvore com nós internos vivos foi gerada pela matheurística aleatória com particionamento de 3% (Figura 3(A)), e as outras duas pela matheurística sequencial com 3% e 5% no valor percentual (Figuras 3 (B) e (C) respectivamente). As três árvores divergem bastante, o que do ponto de vista prático trás uma complicação para as soluções geradas, já que os valores dos limitantes para as três são ruins e não é possível dizer qual delas estaria mais próximo da árvore real.

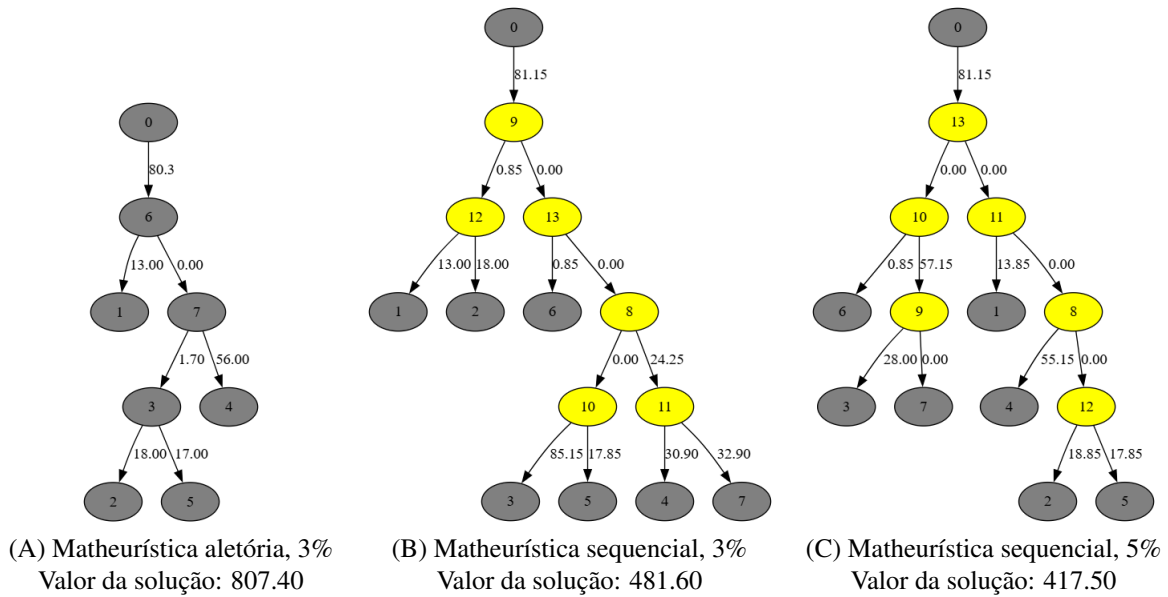


Figura 3: Árvores geradas para a instância d.6_2_1_4. Nenhuma árvore possui politomia.

4 Conclusão

De forma geral, os resultados encontrados não foram satisfatórios: solução encontrada em poucas instâncias, valores dos limitantes ruins, não se obteve ganho na implementação da politomia e do relax & fix com as matheurísticas propostas e com as porcentagens de particionamento quando comparado com os resultados apresentados por Hoshino et al. [2018]. Comparativamente, no entanto, o critério de particionamento sequencial com o percentual de 3% apresentou os melhores resultados.

5 Referências

Referências

- G. S. Araújo, G. P. Telles, M. E. M. Walter, and N. F. Almeida. Distance-based live phylogeny. In *International Conference on Bioinformatics Models, Methods and Algorithms*, volume 4, pages 196–201. SCITEPRESS, 2017.
- W. Hennig. *Phylogenetic systematics*. University of Illinois Press, 1999.
- E. A. Hoshino, B. A. Araujo, and V. O. Freitas. Algoritmos exatos para o problema da filogenia viva. In *XIX Latin-Iberoamerican Conference on Operations Research*, page 276–283, 2018.
- G. Lenzini and S. Marianelli. Algorithms for phylogeny reconstruction in a new mathematical model. *Calcolo-A Quarterly on Numerical Analysis and Theory of Computation*, 1(34):1–24, 1997.
- Y. Pochet and L. A. Wolsey. *Production planning by mixed integer programming*, volume 149. Springer, 2006.
- D. Robinson and L. Foulds. Comparison of phylogenetic trees. *Mathematical Biosciences*, 53(1):131–147, 1981. ISSN 0025-5564. doi: [https://doi.org/10.1016/0025-5564\(81\)90043-2](https://doi.org/10.1016/0025-5564(81)90043-2). URL <https://www.sciencedirect.com/science/article/pii/0025556481900432>. [Acessado em 06-12-2025].

G. P. Telles, N. F. Almeida, R. Minghim, and M. E. M. Walter. Live phylogeny. *Journal of Computational Biology*, 20(1):30–37, 2013.