Capítulo 2

Grafos, Cortes, Subgrupos e Árvores Filogenéticas

Uma árvore filogenética é um grafo não orientado, conexo e acíclico com no máximo um vértice de grau 2. Os vértices de grau um de uma árvore filogenética chamados de *folhas* e representam seres vivos. Todos os demais vértices de uma árvore filogenética são denominados *nós internos* e representam espécies ancestrais das espécies representadas pelas folhas. Um dos nós internos pode ser identificado como o ancestral mais antigo de todo o conjunto das folhas da árvore. Neste caso, este nó é chamado de *raiz* da árvore e a árvore é uma árvore filogenética com raiz. Uma árvore filogenética com um vértice de graus dois é sempre uma árvore filogenética com raiz e a raiz é o vértice de grau dois. As arestas de uma árvore filogenética, por sua vez, podem ser diferenciadas entre *arestas folha*, que são as arestas adjacentes a uma folha, e *arestas internas*, que são todas as demais.

Um vértice de grau maior ou igual a quatro é chamado de uma *politomia* [9] e uma árvore sem politomias é chamada de uma *árvore filogenética completamente resolvida*. Uma árvore filogenética com politomias é denominada *parcialmente resolvida*. A Figura 2.1 mostra um exemplo de árvore filogenética parcialmente resolvida com raiz.

As folhas de uma árvore filogenética são rotuladas por elementos de um conjunto L de espécies. A relação entre folhas e elementos de L é dada por uma função bijetora, ou seja, toda folha está associada a um rótulo e cada rótulo está associado a uma única folha. Denotamos por $\mathcal{T}(L)$ o conjunto de todas as árvores filogenéticas cujas folhas são rotuladas por elementos de L. Devido a esta relação tão estreita dos elementos de L com as folhas de uma árvore filogenética, para evitar complicações desnecessárias neste texto, não faremos distinção entre os elementos de L e as folhas que eles rotulam. Ao longo de todo o texto, os elementos de L serão considerados as folhas das árvores estudadas, portanto, para qualquer árvore T em $\mathcal{T}(L)$, para fins de simplificação, consideraremos

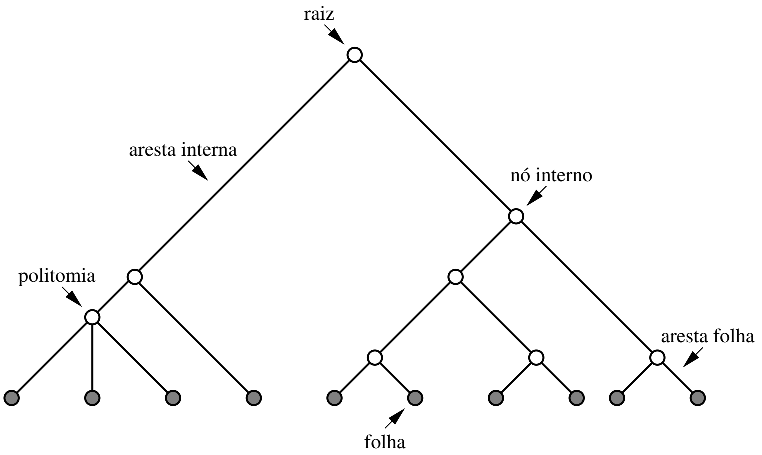

Figura 2.1: *Um exemplo de árvore filogenética parcialmente resolvida com raiz, indicando alguns dos vértices e arestas especiais.*

válida a relação:

$$L = \{v \in V(T) \mid d(v) = 1\}$$

Muitas vezes é interessante poder ordenar os elementos de L, porém os elementos de L não têm uma ordem natural bem definida. Todavia é sempre possível definir uma função $\varphi : L \mapsto \mathbb{N}_L$, onde $\mathbb{N}_L \subset \mathbb{N}$ é o conjunto dos $|L|$ primeiros números naturais. Graças à função φ é possível ordenar qualquer subconjunto de L, pois dados quaiquer dois elementos a e b de L, se $a \neq b$, então $\varphi(a) < \varphi(b)$ ou $\varphi(a) < \varphi(b)$.

Em algumas situações neste trabalho será necessário trabalhar com subconjuntos específicos de $\mathcal{T}(L)$. Denotaremos por $\mathcal{T}_U(L)$ o maior subconjunto de $\mathcal{T}(L)$ que contém apenas árvores filogenéticas sem raiz; e por $\mathcal{T}_R(L)$ o maior subconjunto de $\mathcal{T}(L)$ que contém apenas árvores filogenéticas com raiz. A notação $\mathcal{T}^*(L)$ será usada para o maior subconjunto de $\mathcal{T}(L)$ composto apenas por árvores filogenéticas completamente resolvidas. As composições $\mathcal{T}_U^*(L)$ e $\mathcal{T}_R^*(L)$ também são possíveis e denotam respectivamente os maiores subconjuntos de $\mathcal{T}^*(L)$ formados unicamente por árvores sem raiz e com raiz. Claramente $\mathcal{T}_U^*(L)$ e $\mathcal{T}_R^*(L)$ formam uma bipartição do conjunto $\mathcal{T}^*(L)$.

A definição de árvores filogenéticas como grafos é provavelmente a mais utilizada. Quando trabalhamos com consensos entre árvores filogenéticas, no entanto, como veremos nos próximos capítulos, geralmente temos que "desmontar" um grupo de árvores filogenéticas e decidir quais partes de quais árvores escolheremos para formar a árvore consenso. Este tipo de operação pode começar a ser um pouco complicada usando grafos. Felizmente há definições diferentes de árvores filogenéticas usando conceitos de conjuntos que são mais simples de ser usados e que não só permitem simplificar a definição de consenso entre árvores, como também permite definir medidas de distâncias entre árvores filogenética. Uma destas medidas é apresentada na Seção 2.1, logo a seguir.

2.1 Cortes e Sistemas de Cortes

As definições de corte, sistema de cortes e compatibilidade entre cortes apresentadas neste trabalho foram extraídas de uma monografia de Andreas Dress [3, Sessão 2.5].

Um *corte*, ou *split* em inglês, $S = \{A, B\}$ de um conjunto qualquer X é uma bipartição de X em dois subconjuntos não vazios A e B. O conjunto de todos os cortes possíveis do conjunto X é denotado por $\mathcal{S}(X)$ e qualquer subconjunto $\mathcal{S}$ de $\mathcal{S}(X)$ é denominado um *sistema de cortes* definido sobre X. Dois cortes $S, S' \in \mathcal{S}(X)$ são chamados *compatíveis* se e somente se existem subconjuntos $A \in S$ e $A' \in S'$ tais que $A \cap A' = \emptyset$; caso contrário, S e S' são chamados *incompatíveis*.

Dada uma árvore filogenética $T \in \mathcal{T}(L)$, para qualquer aresta $e \in E(T)$, escolhendo-se um vértice qualquer $v \in V(T)$ tal que v é uma folha de T, é possível dividir o conjunto L em dois subconjuntos:

$$\begin{aligned} A &= \{x \in L \mid \text{o caminho de } x \text{ a } v \text{ em } T \text{ não passa por } e\} \\ B &= \{x \in L \mid \text{o caminho de } x \text{ a } v \text{ em } T \text{ passa por } e\} \end{aligned}$$

É simples perceber que $A \cup B$ é o conjunto de folhas de T e que $A \cap B = \emptyset$. Desta forma, $S = \{A, B\}$ é um corte do conjunto de folhas de T, logo, do conjunto L. Note que a escolha arbitrária do vértice não influi na composição do corte. Assim, podemos definir o sistema de cortes $\mathcal{S}_T(L)$, composto por todos os cortes de L relacionados a arestas de T.

Um resultado importante relacionando cortes e árvores filogenéticas é que, se $S, S' \in \mathcal{S}_T(L)$, então S é compatível com S'. Outro resultado igualmente importante é que, qualquer que seja $T \in \mathcal{T}(L)$, a árvore T pode ser determinado a partir de $\mathcal{S}_T(L)$. Além disso, também é sabido que tanto o número de arestas de uma árvore filogenética completamente resolvida sem raiz com n folhas quanto a maior cardinalidade possível para um conjunto de cortes compatíveis dois a dois definidos sobre um mesmo conjunto de n elementos é $2n - 3$.

Dado $l \in L$, um corte do tipo $\{\{l\}, L \setminus \{l\}\}$ é chamado *corte trivial* de L. O sistema de cortes formado por todos cortes triviais de um conjunto L é denotado por $\mathcal{S}^*(L)$. Denotamos por $\Theta(L)$ o conjunto de todos os sistemas de cortes definidos sobre L compostos apenas por cortes compatíveis e que obrigatoriamente contém $\mathcal{S}^*(L)$. Definimos a relação $\alpha \subseteq \Theta(L) \times \Theta(L)$ da seguinte maneira:

$$\alpha = \{(\mathcal{S}, \mathcal{S}') \in \Theta(L) \times \Theta(L) \mid \mathcal{S}' \subset \mathcal{S} \text{ e } |\mathcal{S}'| = |\mathcal{S}| - 1\}$$

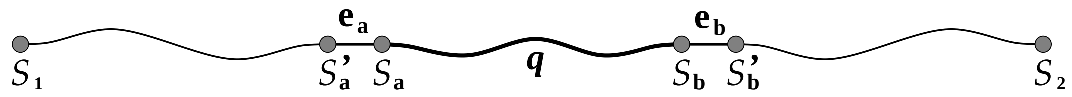

Figura 2.2: *Caminho encontrado em $G_{\Theta(L)}$ durante a prova do Lema 2.1.3*

A relação α relaciona um sistema de cortes qualquer $\mathcal{S}$ a todos os sistemas $\mathcal{S}'$ que podem ser obtidos de $\mathcal{S}$ pela remoção de um de seus cortes. Uma vez definida a relação α, é possível definir um grafo $G_{\Theta(L)}$ da seguinte maneira:

$$\begin{aligned} V(G_{\Theta(L)}) &= \Theta(L) \\ E(G_{\Theta(L)}) &= \{\mathcal{S}\mathcal{S}' \mid (\mathcal{S},\mathcal{S}') \in \alpha \text{ ou } (\mathcal{S}',\mathcal{S}) \in \alpha\} \end{aligned}$$

Um vértice em $G_{\Theta(L)}$ tem a propriedade de diferir de seus vizinhos por exatamente um corte. Assim, qualquer caminho entre dois vértices neste grafo define uma seqüência de inserções e remoções de cortes que transforma o sistema de cortes de uma das extremidades do caminho no sistema de cortes da outra extremindade.

Teorema 2.1.1 *Qualquer que seja $\mathcal{S} \in \Theta(L)$, existe um caminho $\mathcal{S}^* \rightsquigarrow \mathcal{S}$ em $G_{\Theta(L)}$.*

Prova Provaremos o teorema por indução no tamanho n do conjunto $|\mathcal{S} \setminus \mathcal{S}^*|$.

Base: Para $n = 0$ a hipótese se verifica, pois o único conjunto em $\Theta(L)$ para o qual $|\mathcal{S} \setminus \mathcal{S}^*| = 0$ é o próprio conjunto $\mathcal{S}^*$.

Hipótese de Indução: Se $|\mathcal{S} \setminus \mathcal{S}^*| < n$, então existe um caminho $\mathcal{S}^* \rightsquigarrow \mathcal{S}$ em $G_{\Theta(L)}$.

Passo: Seja $\mathcal{S}$ um sistema de cortes tal que $|\mathcal{S} \setminus \mathcal{S}^*| = n \geq 1$. Seja $\mathcal{S}'$ um sistema de cortes tal que $|\mathcal{S} \setminus \mathcal{S}'| = 1$. Note que $|\mathcal{S}' \setminus \mathcal{S}^*| = n-1$, portanto, por hipótese de indução, existe um caminho $\mathcal{S}^* \rightsquigarrow \mathcal{S}'$ em $G_{\Theta(L)}$. Além disso, $|\mathcal{S} \setminus \mathcal{S}'| = 1$, portanto, $(\mathcal{S},\mathcal{S}') \in \alpha$ e $\mathcal{S}\mathcal{S}' \in E(G_{\Theta(L)})$. Assim, existe um caminho $\mathcal{S}^* \rightsquigarrow \mathcal{S}$ em $G_{\Theta(L)}$.

□

Teorema 2.1.2 *$G_{\Theta(L)}$ é conexo.*

Prova Por conseqüência do Teorema 2.1.1, para quaisquer $\mathcal{S}_1, \mathcal{S}_2 \in \Theta(L)$, existem os caminhos $\mathcal{S}^* \rightsquigarrow \mathcal{S}_1$ e $\mathcal{S}^* \rightsquigarrow \mathcal{S}_2$ em $G_{\Theta(L)}$. Como $G_{\Theta(L)}$ é um grafo não-orientado, existe também o caminho $\mathcal{S}_1 \rightsquigarrow \mathcal{S}^* \rightsquigarrow \mathcal{S}_2$.

□

Sejam T e U duas árvores pertencentes a $\mathcal{T}(L)$. Seja $\rho(T,U)$ o comprimento do menor caminho que liga $\mathcal{S}_U(L)$ a $\mathcal{S}_T(L)$ em $G_{\Theta(L)}$. Provarei a seguir que $\rho(T,U) = |\mathcal{S}_T(L)| + |\mathcal{S}_U(L)| - 2\,|\mathcal{S}_T(L) \cap \mathcal{S}_U(L)|$.

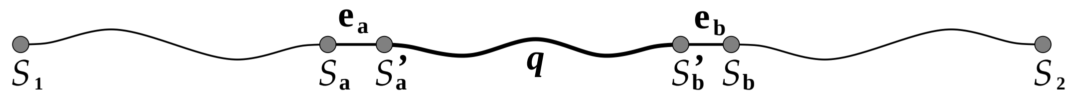

Figura 2.3: *Caminho encontrado em $G_{\Theta(L)}$ durante a prova do Lema 2.1.4*

Lema 2.1.3 *Seja p um caminho de comprimento mínimo em $G_{\Theta(L)}$ que une os vértices $\mathcal{S}_1$ e $\mathcal{S}_2$. Então p não contém nenhuma aresta $\mathcal{S}'\mathcal{S}$ tal que $\mathcal{S} \setminus \mathcal{S}' = \{x\}$ e $x \notin \mathcal{S}_1 \cup \mathcal{S}_2$.*

Prova Seja E_x o subconjunto de $E(G_{\Theta(L)})$ definido da seguinte forma:

$$E_x = \{\mathcal{S}'\mathcal{S} \in E(G_{\Theta(L)}) \mid \mathcal{S} \setminus \mathcal{S}' = \{x\} \text{ e } x \notin \mathcal{S}_1 \cup \mathcal{S}_2\}.$$

Suponha que haja um caminho p de comprimento mínimo entre $\mathcal{S}_1$ e $\mathcal{S}_2$ que contenha ao menos uma aresta $e \in E_x$. Seja $e_a = \mathcal{S}'_a\mathcal{S}_a \in E_x$ a aresta em E_x mais próxima de $\mathcal{S}_1$ em p. Note que como nenhum dos sistemas de cortes das extremidades do caminho contém o corte x, então deve haver outra aresta pertencente a E_x entre $\mathcal{S}_a$ e $\mathcal{S}_2$. Seja $e_b = \mathcal{S}'_b\mathcal{S}_b \in E_x$ a aresta em E_x mais próxima de $\mathcal{S}_a$. Seja q o sub-caminho de p que une $\mathcal{S}_a$ e $\mathcal{S}_b$. Temos então a situação esquematizada pela Figura 2.2.

Note que, se p é um caminho de comprimento mínimo entre $\mathcal{S}_1$ e $\mathcal{S}_2$, então $q \cup \{\mathcal{S}'_a, \mathcal{S}'_b\}$ é o único caminho de comprimento mínimo entre $\mathcal{S}'_a$ e $\mathcal{S}'_b$ contendo tanto e_a quanto e_b. Note também que todo vértice em q contém o corte x. Assim, é possível definir o caminho $\dot{q}$ da seguinte maneira:

$$\dot{q} = \{\dot{\mathcal{S}} \mid \mathcal{S} \in q \text{ e } \dot{\mathcal{S}} = \mathcal{S} \setminus \{x\}\}$$

É fácil ver que os comprimentos de q e q' são iguais, mas $\dot{\mathcal{S}}_a = \mathcal{S}'_a$ e $\dot{\mathcal{S}}_b = \mathcal{S}'_b$. Assim, existe um caminho, $\dot{q}$, entre $\mathcal{S}'_a$ e $\mathcal{S}'_b$ mais curto que o sub-caminho de p. Portanto, nem o sub-caminho de p e, é claro, nem o próprio p são caminhos de comprimento mínimo.

□

Lema 2.1.4 *Seja p um caminho de comprimento mínimo em $G_{\Theta(L)}$ que une os vértices $\mathcal{S}_1$ e $\mathcal{S}_2$. Então p não contém nenhuma aresta $\mathcal{S}'\mathcal{S}$ tal que $\mathcal{S} \setminus \mathcal{S}' = \{x\}$ e $x \in \mathcal{S}_1 \cap \mathcal{S}_2$.*

Prova Seja E_x o subconjunto de $E(G_{\Theta(L)})$ definido da seguinte forma:

$$E_x = \{\mathcal{S}'\mathcal{S} \in E(G_{\Theta(L)}) \mid \mathcal{S} \setminus \mathcal{S}' = \{x\} \text{ e } x \in \mathcal{S}_1 \cap \mathcal{S}_2\}.$$

Suponha que haja um caminho p de comprimento mínimo entre $\mathcal{S}_1$ e $\mathcal{S}_2$ que contenha ao menos uma aresta $e \in E_x$. Seja $e_a = \mathcal{S}'_a\mathcal{S}_a \in E_x$ a aresta em E_x mais próxima de $\mathcal{S}_1$ em p. Note que como ambos os sistemas de cortes das extremidades do caminho contém o corte x, então deve haver outra aresta pertencente a E_x entre $\mathcal{S}_a$ e $\mathcal{S}_2$. Seja $e_b = \mathcal{S}'_b\mathcal{S}_b \in E_x$ a

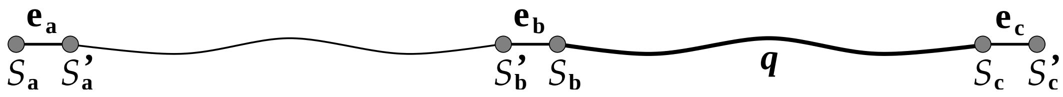

Figura 2.4: *Caminho encontrado em $G_{\Theta(L)}$ durante a prova do Lema 2.1.5*

aresta em E_x mais próxima de $\mathcal{S}'_a$. Seja q o sub-caminho de p que une $\mathcal{S}'_a$ e $\mathcal{S}'_b$. Temos então a situação esquematizada pela Figura 2.3.

Note que, se p é um caminho de comprimento mínimo entre $\mathcal{S}_1$ e $\mathcal{S}_2$, então $q \cup \{\mathcal{S}_a, \mathcal{S}_b\}$ é o único caminho de comprimento mínimo entre $\mathcal{S}_a$ e $\mathcal{S}_b$ contendo tanto e_a quanto e_b. Note também nenhum dos vértices em q contém o corte x, mas todos os cortes de ambas as extremidades são compatíveis com x, pois x está em ambos os sistemas de corte, sendo então compatível com todos os demais cortes nestes sistemas, e, segundo o Lema 2.1.3 nenhum sistema de cortes em p possui um corte que não está nem em $\mathcal{S}_1$ e nem em $\mathcal{S}_2$. Assim, x é compatvel com qualquer sistema de cortes em q, sendo possível definir o caminho $\dot{q}$ da seguinte maneira:

$$\dot{q} = \{\dot{\mathcal{S}} \mid \mathcal{S} \in q \text{ e } \dot{\mathcal{S}} = \mathcal{S} \cup \{x\}\}$$

É fácil ver que os comprimentos de q e $\dot{q}$ são iguais, mas $\dot{\mathcal{S}}'_a = \mathcal{S}_a$ e $\dot{\mathcal{S}}'_b = \mathcal{S}_b$. Assim, existe um caminho, q', entre $\mathcal{S}_a$ e $\mathcal{S}_b$ mais curto que o sub-caminho de p. Portanto, nem o sub-caminho q de p e nem o próprio p são caminhos de comprimento mínimo.

□

Lema 2.1.5 *Seja p um caminho de comprimento mínimo entre $\mathcal{S}_1$ e $\mathcal{S}_2$ em $G_{\Theta(L)}$. Para cada elemento x em $(\mathcal{S}_1 \cup \mathcal{S}_2) \setminus (\mathcal{S}_1 \cap \mathcal{S}_2)$ existe exatamente uma aresta $\mathcal{S}'\mathcal{S} \in p$ tal que $\mathcal{S} \setminus \mathcal{S}' = \{x\}$.*

Prova Seja p um caminho de comprimento mínimo entre $\mathcal{S}_1$ e $\mathcal{S}_2$ em $G_{\Theta(L)}$. Como cada elemento de $(\mathcal{S}_1 \cup \mathcal{S}_2) \setminus (\mathcal{S}_1 \cap \mathcal{S}_2)$ está em apenas um dos sistemas de cortes nas extremidades do caminho p, existe para cada um deles ao menos uma aresta cuja diferença entre as extremidades seja o próprio elemento. Resta provar que esta aresta é única.

Suponha então que exista um elemento $x \in (\mathcal{S}_1 \cup \mathcal{S}_2) \setminus (\mathcal{S}_1 \cap \mathcal{S}_2)$ e que as arestas $e_1 = \mathcal{S}'_1\mathcal{S}_1$ e $e_2 = \mathcal{S}'_2\mathcal{S}_2$ são tais que a diferença entre suas extremidades seja $\{x\}$. Além disso, suponha que não haja nenhuma aresta cuja diferença entre as extremidades seja $\{x\}$ entre e_1 a e_2 no caminho p. Lembremos que apenas uma das extremidades de p possui o corte x. Portanto, se há duas arestas cuja diferença das extremidades seja $\{x\}$, então deve haver ao menos uma terceira, caso contrário, ou ambas as extremidades de p conteriam x ou x não pertenceria a nenhuma delas. Seja $e_3 = \mathcal{S}'_3\mathcal{S}_3$ a aresta deste tipo mais próxima do caminho entre e_1 e e_2. Nomeemos estas três arestas, em relação ao menor sub-caminho de p que contenha as três, da seguinte forma:

$$\begin{array}{rcl} e_a & : & \text{A aresta mais externa cujo vértice mais externo contenha } x. \\ e_b & : & \text{A aresta central.} \\ e_c & : & \text{A aresta mais externa cujo vértice mais externo não contenha } x. \end{array}$$

Temos a situação esquematizada pela Figura 2.4. Seja q o sub-caminho de p que une $\mathcal{S}_b$ a $\mathcal{S}_c$. Como p é um caminho de comprimento mínimo, q também é. Mas, como na prova do Lema 2.1.3, aqui também é possível definir um caminho $\dot{q} = \{\dot{\mathcal{S}} \mid \mathcal{S} \in q \text{ e } \dot{\mathcal{S}} = \mathcal{S} \setminus \{x\}\}$. Como $\dot{\mathcal{S}}_b = \mathcal{S}'_b$ e $\dot{\mathcal{S}}_c = \mathcal{S}'_c$, tanto o sub-caminho de p que une $\mathcal{S}'_b$ a $\mathcal{S}'_c$ quanto o próprio caminho p não são mínimos como supunhamos.

□

Teorema 2.1.6 *Um caminho de comprimento mínimo entre dois vértices $\mathcal{S}_1$ e $\mathcal{S}_2$ em $G_{\Theta(L)}$ tem exatas $|\mathcal{S}_1| + |\mathcal{S}_2| - 2\,|\mathcal{S}_1 \cap \mathcal{S}_2|$ arestas.*

Prova Seja p um caminho de comprimento mínimo entre $\mathcal{S}_1$ e $\mathcal{S}_2$ em $G_{\Theta(L)}$. Pelo Lema 2.1.3, não há nenhuma aresta $\mathcal{S}'\mathcal{S}$ em p tal que $\mathcal{S} \setminus \mathcal{S}' = \{x\}$ e $x \notin \mathcal{S} \cup \mathcal{S}'$. Pelo Lema 2.1.4, não há nenhuma aresta $\mathcal{S}'\mathcal{S}$ em p tal que $\mathcal{S} \setminus \mathcal{S}' = \{x\}$ e $x \in \mathcal{S} \cap \mathcal{S}'$. Pelo Lema 2.1.5, para cada elemento $x \in (\mathcal{S}_1 \cup \mathcal{S}_2) \setminus (\mathcal{S}_1 \cap \mathcal{S}_2)$ há exatamente uma aresta $\mathcal{S}'\mathcal{S}$ em p tal que $\mathcal{S} \setminus \mathcal{S}' = \{x\}$. Assim, seja $\rho(\mathcal{S}_1, \mathcal{S}_2)$ o comprimento de p:

$$\begin{aligned} \rho(\mathcal{S}_1, \mathcal{S}_2) &= |(\mathcal{S}_1 \cup \mathcal{S}_2) \setminus (\mathcal{S}_1 \cap \mathcal{S}_2)| \\ &= |(\mathcal{S}_1 \setminus (\mathcal{S}_1 \cap \mathcal{S}_2)) \cup (\mathcal{S}_2 \setminus (\mathcal{S}_1 \cap \mathcal{S}_2))| \\ &= |\mathcal{S}_1 \setminus (\mathcal{S}_1 \cap \mathcal{S}_2)| + |\mathcal{S}_2 \setminus (\mathcal{S}_1 \cap \mathcal{S}_2)| \\ &= |\mathcal{S}_1| - |\mathcal{S}_1 \cap \mathcal{S}_2| + |\mathcal{S}_2| - |\mathcal{S}_1 \cap \mathcal{S}_2| \\ &= |\mathcal{S}_1| + |\mathcal{S}_2| - 2\,|\mathcal{S}_1 \cap \mathcal{S}_2| \end{aligned}$$

□

Waterman [13, Capítulo 14] apresenta outra prova para o valor da distância de cortes entre duas árvores. Robinson e Foulds [12] chegam a um valor semelhante definindo a distância entre grafos de árvores filogenéticas com base no número de operações chamadas contrações e extensões de Bourque necessárias para transformar uma árvore em outra.

2.2 Subgrupos e n-Árvores

Cortes e sistemas de cortes são elementos bem definidos para árvores filogenéticas sem raiz. No entanto, a maioria das definições de consensos entre árvores filogenéticas é definida primeiramente para um grupo de árvores filogenéticas com raiz. Esta seção se dedica à apresentação de conceitos semelhantes a cortes e sistemas de cortes, mas dedicados a árvores filogenéticas com raiz.

As definições feitas na Capítulo 2 são válidas também para esta seção, porém, ao contrário da Seção 2.1, aqui não nos preocupamos com árvores filogenéticas sem raiz. As definições de n-árvore e função de altura para subgrupos foram extraídas de trabalhos de T. Margush e F. R. McMorris [7] e R. C. Powers [10].

A idéia aqui é agrupar os elementos de um conjunto qualquer X em "classes" e "subclasses" de maneira que, se a interseção de duas classes não for vazia, então uma classe é obrigatoriamente subclasse da outra, o que nos dá uma classificação hierárquica dos elementos do conjunto X, assim como, por exemplo, o sistema de classificação dos seres vivos sugerido por Linnaeus no século XVIII.

Um subconjunto de L é também denominado um *subgrupo* de L. Na literatura em inglês, podemos encontrar tanto o termo *subgroup* quanto o termo *cluster* para designar subgrupos. Um subgrupo S de L tal que $|S| \in \{1, |L|\}$ é chamado de *subgrupo trivial* de L. Todos os demais subgrupos de L são chamados de *não-triviais*. Em outras palavras, S é um subgrupo trivial de L se e somente se $S = L$ ou se $S = \{x\}$ para algum $x \in L$.

Uma família Ψ de subcojuntos de um mesmo conjunto qualquer X que satisfazem a propriedade

$$\text{Para todo } R, S \in \Psi, R \cap S \in \{R, S, \emptyset\}$$

é tradicionalmente conhecida como uma família laminar [4] sobre X e é possível descrever uma árvore filogenética com raiz usando um tipo especial de família laminar denominado n-árvore.

Conjuntos de subgrupos de um mesmo conjunto X, que obedecem a quatro critérios que apresentaremos mais adiante no texto, o que estabelece que o conjunto é uma família laminar e mais três, são chamados de *n-árvores* sobre X, como chamaremos neste trabalho, ou *n-trees*, tal como foram chamados pela primeira vez em 1972 por Bobisud e Bobisud [7].

A letra n presente do termo "n-árvore" refere-se ao número de elementos do conjunto sobre o qual a árvore está definida. Como neste trabalho as n-árvores serão definidas não só sobre L, mas também sobre outros conjuntos, como subconjuntos de L, por exemplo, não usaremos n como uma variável sem deixar claro o seu significado. Assim, a não ser que deixemos claro no texto, não usaremos a variável n como a cardinalidade de conjuntos e, na grande maioria dos casos, denotaremos a cardinalidade de um conjunto X por $|X|$, como fizemos até este ponto.

Um conjunto Ψ de subgrupos de L é denominado uma n-árvore sobre L se e somente se as quatro condições abaixo forem verificadas em Ψ:

1. $\emptyset \notin \Psi$.

2. $L \in \Psi$.
3. $\{x\} \in \Psi$ para todo $x \in L$.
4. $A \cap B \in \{A, B, \emptyset\}$ para todos os subgrupos $A, B \in \Psi$.

Seja Ψ uma n-árvore. Uma *função de altura* definida em Ψ é uma função $\eta : \Psi \to \mathbb{N}$ satisfazendo:

- $\eta(L) = 0$.
- Se $A, B \in \Psi$ e $A \subset B$, então $\eta(A) > \eta(B)$.

É possível definir várias funções de altura para os elementos de uma n-árvore. Duas funções, em especial, são bastante utilizadas. São elas:

Função de Altura Canônica: Seja Ψ uma n-árvore sobre o conjunto L. Considere uma seqüência $S_0, S_1, \ldots, S_k$ de subgrupos de Ψ onde $S_i \subset S_j$ para $0 \leq i < j \leq n$, $S_0 = S$ e $S_k = L$. Seja $\bar{k}$ o maior valor de k para o qual exista tal seqüência. A função de altura canônica é definida como $\eta_0(S) = \bar{k}$.

Função de Altura de Cardinalidade: Seja Ψ uma n-árvore sobre o conjunto L. A função de altura de cardinalidade, $\eta_\sharp$, como o próprio nome sugere, é definida como a diferença entre as cardinalidades de L e de um subgrupo qualquer $S \in \Psi$. De maneira resumida, $\eta_\sharp(S) = |L| - |S|$.

Sejam Ψ e Υ n-árvores. Se todo subgrupo de Υ for subgrupo de Ψ, então dizemos que Υ é uma *n-subárvore* de Ψ. É interessante notar que não necessariamente as duas n-árvores estão definidas sobre um mesmo conjunto, mas certamente o conjunto no qual a n-árvore Υ está definida é um subgrupo de Ψ. Se S é um subgrupo pertencente a Ψ, denominamos n-subárvore *induzida por* S em Ψ, denotada por $\Psi[S]$, o conjunto de subgrupos definido da seguinte forma:

$$\Psi[S] = \{A \in \Psi \mid A \subseteq S\}.$$

Em geral, usaremos a notação $\Psi[S]$ para indicar $\{A \in \Psi \mid A \subseteq S\}$ mesmo quando Ψ é um conjunto de subgrupos que não seja uma n-árvore.

Teorema 2.2.1 *O conjunto de subgrupos $\Psi[S]$ é uma n-árvore sobre S.*

Prova Para provar que $\Psi[S]$ é uma n-árvore, temos que verificar que as quatro condições necessárias para que um conjunto de subgrupos seja uma n-árvore são válidas para o conjunto $\Psi[S]$.

(1) $\emptyset \notin \Psi[S]$**:** Esta condição também é satisfeita, pois $\Psi[S]$ é constituída apenas por elementos de Ψ e $\emptyset \notin \Psi$.

(2) $S \in \Psi[S]$**:** Esta condição é claramente satisfeita, pois $S \in \Psi$ e $S \subseteq S$.

(3) para todo $x \in S, \{x\} \in \Psi[S]$**:** Qualquer que seja $x \in S$, $\{x\} \in \Psi$, pois Ψ é uma n-árvore. Além disso, claramente $\{x\} \subseteq S$, logo, $\{x\} \in \Psi[S]$.

(4) para todo $A, B \in \Psi[S], A \cap B \in \{A, B, \emptyset\}$**:** Devido ao fato de Ψ ser uma n-árvore, esta propriedade vale para qualquer par de subgrupos $A, B \in \Psi$. Como $\Psi[S]$ é formada somente por subgrupos de Ψ, se $A, B \in \Psi[S]$, então $A, B \in \Psi$, logo, para qualquer par de subgrupos em $\Psi[S]$ a propriedade é válida.

□

Dada uma árvore $T \in \mathcal{T}^r(L)$ e um vértice $v \in V(T)$, definimos o subgrupo S_v da seguinte forma:

$$S_v = \{x \in L \mid \text{o caminho de } x \text{ à raiz de } T \text{ contém } v\}$$

Podemos ainda definir o conjunto de subgrupos relativos aos vértices de uma árvore $T \in \mathcal{T}^r(L)$, denotado por $\mathcal{N}_T$, da seguinte maneira:

$$\mathcal{N}_T = \{S_v \mid v \in T\}$$

Teorema 2.2.2 *O conjunto de subgrupos $\mathcal{N}_T$ é uma n-árvore sobre L.*

Prova Para provar que $\mathcal{N}_T$ é uma n-árvore sobre L, temos que verificar que as quatro condições necessárias para que um conjunto de subgrupos seja uma n-árvore são válidas para o conjunto $\mathcal{N}_T$.

(1) $\emptyset \notin \mathcal{N}_T$**:** Se existisse um vértice v para o qual S_v fosse vazio em T, então v seria um vértice que não está conectado a nenhuma das folhas de T. Mas T é um grafo conexo, logo, para quaisquer dois vértices u e v em T existe ao menos um caminho de u a v em T, logo, não existe nenhum vértice v em $V(T)$ para o qual S_v é vazio e $\emptyset \notin \mathcal{N}_T$.

(2) $L \in \mathcal{N}_T$**:** Seja $r \in V(T)$ a raiz da árvore T. Como r está no caminho de qualquer folha até a raiz, então $S_r = L$. Assim, $L \in \mathcal{N}_T$.

(3) para todo $x \in L, \{x\} \in \mathcal{N}_T$**:** Se $v \in L$, então $d(v) = 1$ pela própria definição de L. Se o grau de v é 1, então o único caminho que contém uma folha, v e a raiz de T é o caminho da raiz a v. Assim, se $v \in L$, então $S_v = \{v\}$. Portanto, para todo elemento v de L existe um subgrupo $\{v\}$ em $\mathcal{N}_T$.

(4) para todo $S_u, S_v \in \mathcal{N}_T, S_u \cap S_v \in \{S_u, S_v, \emptyset\}$: Tome dois vértices $u, v \in V(T)$ quaisquer. Se $S_u \cap S_v \neq \emptyset$, existe $w \in S_u \cap S_v$. Mas, como T é um grafo acíclico, só há um caminho da raiz a w e ambos u e v pertencem a este caminho. No entanto, se na ida de w à raiz u aparece antes de v, então $S_u \subset S_v$, caso contrário $S_v \subset S_u$.

□

2.2.1 n-Árvores Completamente Resolvidas

Dizemos que uma n-árvore Ψ é *completamente resolvida* se e somente se a inclusão em Ψ de qualquer subgrupo não vazio que não pertença a Ψ faz com que uma das quatro condições necessárias para que Ψ seja uma n-árvore deixe de ser satisfeita. Como a inclusão de um subgrupo não muda o fato de que qualquer subgrupo trivial pertencer a Ψ e a inclusão de um subgrupo vazio não é permitida, podemos simplificar a definição de uma n-árvore completamente resolvida Ψ como sendo uma n-árvore para a qual a inclusão de qualquer subgrupo não pertencente a Ψ e não vazio fere a condição de que para qualquer $A, B \in \Psi$, temos $A \cap B \in \{A, B, \emptyset\}$.

Teorema 2.2.3 *Uma n-árvore Ψ é completamente resolvida se e somente se para qualquer subgrupo $S \in \Psi$ com cardinalidade maior que um existirem dois subgrupos $A, B \in \Psi$ tais que $A \cup B = S$ e $A \cap B = \emptyset$.*

Prova ($\Rightarrow$)

Suponha que exista em Ψ ao menos o subgrupo S com cardinalidade maior que um que não possua os subgrupos $A, B \in \Psi$ tais que $A \cup B = S$ e $A \cap B = \emptyset$, mas que Ψ seja completamente resolvida mesmo assim. Seja $L_S \subseteq \Psi$ tal que $\bigcup_{R \in L_S} R = S$ com cardinalidade mínima. Note que, como para todo $x \in S$ existe o subgrupo trivial $\{x\} \in \Psi$, então L_S existe e $|L_S| \leq |S|$. Como ainda não existem $A, B \in \Psi$ tais que $A \cup B = S$ e $A \cap B = \emptyset$, temos $3 \leq |L_S| \leq |S|$.

Não é difícil perceber que qualquer elemento do conjunto $\{X \cup Y \mid X, Y \in L_S \text{ e } X \neq Y\}$ pode ser incluído em Ψ sem que a propriedade 4 deixe de ser verificada. Chegamos a uma contradição, pois se Ψ é completamente resolvida, então não deveria existir nenhum subgrupo que pudesse ser adicionado a Ψ sem que Ψ deixasse de ser uma n-árvore. Isso significa que, se Ψ é completamente resolvida, então para qualquer subgrupo $S \in \Psi$ existem subgrupos $A, B \in \Psi$ tais que $A \cup B = S$ e $A \cap B = \emptyset$.

($\Leftarrow$)

Suponha agora que Ψ seja uma n-árvore e não seja completamente resolvida, mas que seja verdade que para qualquer subgrupo $S \in \Psi$ com cardinalidade maior que um existam dois subgrupos $A, B \in \Psi$ tais que $A \cup B = S$ e $A \cap B = \emptyset$. Se Ψ realmente não é completamente resolvida, então podemos acrescentar ao menos um subgrupo a Ψ. Seja R um subgrupo que ainda pode ser acrescentado a Ψ sem que Ψ deixe de ser uma n-árvore. Seja S um subgrupo de cardinalidade mínima em Ψ tal que $R \subset S$. Por hipótese, existem dois subgrupos $A, B \in \Psi$ tais que $A \cup B = S$ e $A \cap B = \emptyset$.

Como R pode ser acrescentado a Ψ, $R \cap A \in \{R, A, \emptyset\}$. Assim, ao menos uma das condições abaixo deve ser verdadeira:

$R \cap A = \emptyset$: Esta situação não é possível, pois se a interseção entre R e A for vazia, então $R \subset B$ e S não é o menor subgrupo em Ψ para o qual $R \subset S$.

$R \cap A = R$, portanto $R \subset A$: Isto não pode ser verdade, pois $A = S \setminus B$ e B não é vazio, assim, $|A| < |S|$ e, neste caso, S não seria o menor subgrupo em Ψ para o qual $R \subset S$.

$R \cap A = A$, portanto $A \subset R$: Como $R \subset S$, então $R \setminus A \subset S \setminus A$. Como supomos que $A \subset R$, então $R \setminus A \neq \emptyset$. Mas $S \setminus A = B$, logo, $R \setminus A \subset B$, o que significa que $R \cap B \neq \emptyset$. Restam então duas possibilidades para $R \cap S$, mas por semelhança com o caso anterior, já sabemos que $R \cap B \neq R$. Ficamos então com a hipótese $B \subset R$, mas como já supomos que $A \subset R$, então $A \cup B \subseteq R$, o que é impossível, pois $A \cup B = S$ e S foi escolhido de tal forma que $R \subset S$.

Chegamos à conclusão de que, se para qualquer subgrupo $S \in \Psi$ com cardinalidade maior que um existirem dois subgrupos $A, B \in \Psi$ tais que $A \cup B = S$ e $A \cap B = \emptyset$, então não existe nenhum subgrupo que possa ser acrescentado a Ψ, o que significa que Ψ é completamente resolvida.

□

Teorema 2.2.4 *O número de subgrupos de uma n-árvore completamente resolvida sobre L é $2|L| - 1$.*

Prova Provaremos este teorema por indução no tamanho de L.

Base: Se $|L| = 1$, só existe uma n-árvore possível, composta por apenas um subgrupo trivial. De fato, $2|L| - 1 = 2 - 1 = 1$

Hipótese de Indução: Se $1 \leq |L| < n$, então o número de subgrupos de uma n-árvore completamente resolvida sobre L é $2|L| - 1$.

Passo: Seja L um conjunto de taxa tal que $|L| = n$ e seja Ψ uma n-árvore completamente resolvida sobre L. Como $L \in \Psi$ por definição e Ψ é completamente resolvida, pelo Teorema 2.2.3 sabemos que existem dois subgrupos L' e L'' em Ψ tais que $L' \cup L'' = L$ e $L' \cap L'' = \emptyset$. Sejam $\Psi[L']$ e $\Psi[L'']$ as n-subárvores induzidas por L' e L'', respectivamente. É fácil notar que tanto $\Psi[L']$ quanto $\Psi[L'']$ são completamente resolvidas, pois qualquer subgrupo que possa ser adicionado em uma destas duas árvores também pode ser adicionado a Ψ e Ψ não seria completamente resolvida neste caso. Não é difícil perceber também que, pelo fato de $L' \cap L'' = \emptyset$, a igualdade

$$\Psi = \{L\} \cup \Psi[L'] \cup \Psi[L'']$$

é válida. Como $L' \neq \emptyset$ e $L'' \neq \emptyset$, $|L'| < n$ e $|L''| < n$, assim, por hipótese de indução, $|\Psi[L']| = 2\,|L'| - 1$ e $|\Psi[L'']| = 2\,|L''| - 1$ e podemos calcular o número de subgrupos em Ψ da seguinte forma:

$$\begin{aligned}
\Psi &= \{L\} \cup \Psi[L'] \cup \Psi[L''] \\
|\Psi| &= 1 + |\Psi[L']| + |\Psi[L'']| \\
&= 1 + 2\,|L'| - 1 + 2\,|L''| - 1 \\
&= 2(|L'| + |L''|) - 1 \\
&= 2\,|L| - 1
\end{aligned}$$

□

2.3 Relações entre Sistemas de Cortes e n-Árvores

Sistemas de cortes estão para árvores filogenéticas sem raiz assim como n-árvores estão para árvores filogenéticas com raiz. O mesmo se pode dizer dos conceitos de cortes e subgrupos. Em teoria dos grafos, certas relações entre árvores sem raiz e árvores com raiz são muito fáceis de serem percebidas:

- Com a remoção da raiz e arestas adjacentes e a inserção de uma aresta é possível transformar qualquer árvore filogenética completamente resolvida com raiz e mais de uma folha em uma árvore filogenética completamente resolvida sem raiz com o mesmo número de folhas.

- A remoção de uma aresta e a inserção de um vértice e duas arestas pode transformar uma árvore completamente resolvida sem raiz em uma de $2n - 3$ possíveis árvores completamente resolvidas com raiz, onde n é o número de folhas das árvores em questão.

- A remoção de um nó interno de uma árvore filogenética sem raiz dá origem a três árvores filogenéticas com raiz.

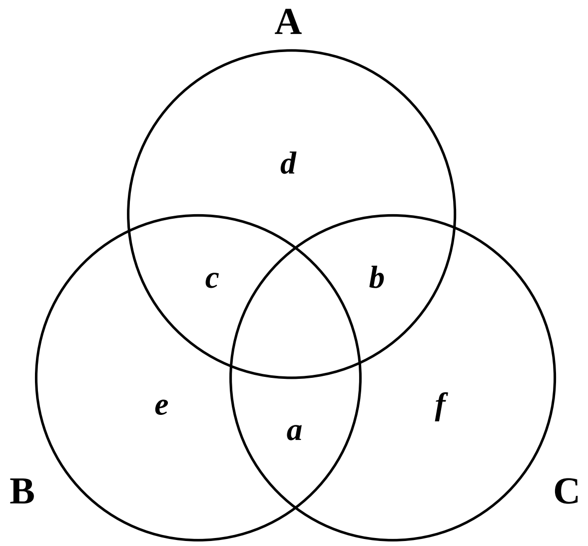

Figura 2.5: *Representação esquemática dos subgrupos A, B e C usados na prova do Lema 2.3.1. As variáveis em itálico representam o mínimo atingido pela função* φ *no subconjunto correspondente. Assim,* $a = \min(\varphi((B \cap C) \setminus A))$ *e* $d = \min(\varphi(A \setminus (B \cup C)))$*, por exemplo.*

Esta seção é dedicada ao estabelecimento de relações semelhantes entre subgrupos e cortes e entre n-árvores e sistemas de cortes. As propriedades apresentadas logo a seguir serão úteis para os algoritmos de construção de consenso com base em sistemas de cortes e n-árvores.

Definimos primeiramente o conjunto $\varphi(R)$, onde R é um subgrupo do conjunto L, da seguinte maneira:

$$\varphi(R) = \{\varphi(r) \mid r \in R\},$$

onde φ é a enumeração dos elementos de L definida na página 4.

Dados dois subgrupos distintos R e S de L, dizemos que R é *menor* que S, e denotamos esta relação por $R < S$, se e somente se $|R| < |S|$ ou $|R| = |S|$ e $\min(\varphi(R \setminus S)) < \min(\varphi(S \setminus R))$. Caso contrário, dizemos que R é *maior* que S e denotamos esta relação por $R > S$.

Lema 2.3.1 *Sejam* A, B *e* C *três subgrupos distintos de* L *tais que* $|A| = |B| = |C|$. *Se* $A < B$ *e* $B < C$, *então* $A < C$.

Prova A Figura 2.5 representa esquematicamente os subgrupos A, B e C de L. As variáveis em itálico representam os valores dos mínimos para os subconjuntos de A, B ou C onde elas se encontram. Por definição, assumiremos que $\min(\emptyset) = \infty$.

Note que, pela definição de subgrupos menores, provar que, se $A < B$ e $B < C$, então $A < C$ é equivalente a provar que, se

$$\min(\varphi(A \setminus B)) < \min(\varphi(B \setminus A))$$

e

$$\min(\varphi(B \setminus C)) < \min(\varphi(C \setminus B)),$$

então

$$\min(\varphi(A \setminus C)) < \min(\varphi(C \setminus A)).$$

Podemos utilizar os elementos ilustrados na Figura 2.5 para reescrever as sentenças acima da seguinte maneira: Se

$$\min(b, d) < \min(a, e)$$

e

$$\min(c, e) < \min(b, f),$$

então

$$\min(c, d) < \min(a, f).$$

Levando em consideração que, se $\min(b, d) < \min(a, e)$, então $\min(a, b, d, e) = \min(b, d)$, chegamos finalmente à conclusão de que provar que $A < C$ se $A < B$ e $B < C$ é equivalente a provar que, se

$$\min(a, b, d, e) < \min(b, d)$$

e

$$\min(b, c, e, f) < \min(c, e),$$

então

$$\min(a, c, d, f) < \min(c, d).$$

Para facilitar a notação, definimos $m_1 = \min(a, b, d, e) = \min(b, d)$, $m_2 = \min(b, c, e, f) = \min(c, e)$ e $m_3 = \min(a, c, d, f)$ e provaremos o lema mostrando que sempre que m_1 assume um dos valores b ou d e m_2 assume um dos valores c ou e, o valor assumido por m_3 é $\min(c, d)$.

$m_1 = b$ e $m_2 = c$: Neste caso, sabemos que $c \leq b$ e $c \leq f$, pois $\min(b, c, e, f) = c$. Sabemos também que $b \leq a$, pois $\min(a, b, d, e) = b$. Como $c \leq b$, temos $c \leq a$. Além disso, $c \leq b$ e $b \leq d$ implicam em $c \leq d$. Assim, $m_3 = \min(a, c, d, f) = c = \min(c, d)$.

$m_1 = b$ e $m_2 = e$: Este caso na verdade nunca ocorre, pois se $\min(a, b, d, e) = b$, então $b \leq e$, mas $\min(b, c, e, f) = e$, o que indica que $e \leq b$. Daí conclui-se que $e = b = \infty$ e como $b \leq d$, temos $d = \infty$ mas isto significa que $A \subseteq B$, o que é uma contradição.

$m_1 = d$ **e** $m_2 = c$**:** Este caso é um pouco mais difícil de analizar, pois c e d nunca são comparados diretamente, mas apenas um de dois casos pode ocorrer:

Caso 1 ($c \leq d$): Como $\min(b, c, e, f) = c$, temos $c \leq b$ e $c \leq f$. Além disso, $b \leq a$, pois $\min(a, b, d, e) = b$. Assim, $c \leq a$ e $m_3 = \min(a, c, d, f) = c = \min(c, d)$.

Caso 2 ($d < c$): Desta vez, sabemos que $d < f$, pois $d < c$ e $c = min(b, c, e, f)$. Como d também é o mínimo dentre (a, b, d, e), sabemos que $d \leq a$. Assim, $m_3 = \min(a, c, d, f) = d = \min(c, d)$.

$m_1 = d$ **e** $m_2 = e$**:** Sabemos que $d \leq a$ e $d \leq e$ porque $\min(a, b, d, e) = d$. Como $e \leq f$ e $e \leq c$ porque $\min(b, c, e, f) = e$, então $d \leq c$ e $d \leq f$, assim, $m_3 = \min(a, c, d, f) = d = \min(c, d)$.

□

Teorema 2.3.2 *Sejam A, B e C três subgrupos distintos de L. Se $A < B$ e $B < C$, então $A < C$.*

Prova A comparação entre dois subgrupos depende da relação de suas cardinalidades, assim, dividimos a prova deste teorema em quatro casos:

$|A| < |B|$ **e** $|B| < |C|$**:** Neste caso, $|A|$ é claramente menor que $|C|$ e $A < C$ por definição.

$|A| = |B|$ **e** $|B| < |C|$**:** Assim como no caso anterior, $|A|$ é claramente menor que $|C|$ e $A < C$ por definição.

$|A| < |B|$ **e** $|B| = |C|$**:** Novamente, $|A|$ é menor que $|C|$ e $A < C$ por definição.

$|A| = |B|$ **e** $|B| = |C|$**:** $A < C$ pelo Lema 2.3.1.

□

Seja $S = \{A, B\}$ um corte de L. Não é difícil notar que A e B são subgrupos distintos de L, assim, $A > B$ ou $A < B$, de tal forma que podemos definir S_m como sendo o menor dos subgrupos de S. Chamamos S_m de *subgrupo menor* de S. Seja R um corte de L diferente de S. Dizemos que R é *menor* que S, e denotamos esta relação por $R < S$, se e somente se $R_m < S_m$, caso contrário, dizemos que R é *maior* que S, e denotamos esta relação por $R > S$.

Teorema 2.3.3 *Dois cortes distintos R e S de um mesmo conjunto L são compatíveis se e somente se os subgrupos menores R_m e S_m forem compatíveis.*

Prova ($\Rightarrow$)

Para facilitar a notação, denotaremos o subgrupo maior de um corte S por S_M, assim, $S = \{S_m, S_M\}$, onde S_m é o subgrupo menor do corte e $S_M = L \setminus S_m$.

Provaremos que os subgrupos menores de cortes compatíveis são também compatíveis por contradição. Para isso, suponha que R e S são dois cortes compatíveis, mas que R_m e S_m não são subgrupos compatíveis. Deste modo, $R_m \cap S_m \neq \emptyset$, $R_m \not\subset S_m$ e $S_m \not\subset R_m$. Para que R e S sejam compatíveis, é preciso que R_m e S_M ou S_m e R_M ou S_M e R_M tenham a intersecção vazia, pela própria definição de compatibilidade entre cortes. Os três casos são analisados a seguir:

$R_m \cap S_M = \emptyset$: Se R_m e S_m não são compatíveis, então $R_m \setminus S_m \neq \emptyset$. Como $S_m \cup S_M = L$, $R_m \setminus S_m = R_m \cap S_M$, assim $R_m \cap S_M \neq \emptyset$.

$R_M \cap S_m = \emptyset$: Também não é verdade, por analogia ao caso anterior.

$R_M \cap S_M = \emptyset$: Não é verdade, pois ambos estão contidos no conjunto L e, se ambos são subgrupos maiores de cortes compatíveis distintos, então a cardinalidade de ambos é maior ou igual a $\frac{|L|}{2}$, sendo que a igualdade pode ser verificada para no máximo um deles pelo fato de se tratarem de cortes compatíveis e distintos, assim, se supusermos que a intersecção entre R_M e S_M é vazia, chegamos à conclusão que $|R_M \cup S_M| > |L|$, o que é uma contradição.

Analisando todos os casos, chegamos à conclusão que, se R_m e S_m não são subgrupos compatíveis, então R e S não são cortes compatíveis.

($\Leftarrow$)

Para provar que se os subgrupos menores R_m e S_m forem compatíveis os cortes R e S também são, provarei a contra-positiva, ou seja, provarei que, se dois cortes R e S não são compatíveis então os subgrupos menores R_m e S_m também não são.

Para tanto, basta lembrar que, pela definição de compatibilidade entre cortes, se R e S não são compatíveis, então nenhum par de subconjuntos formado por um dos subconjuntos de R e um dos subconjuntos de S possui interseção não-vazia. Como $R_m \in R$ e $S_m \in S$, então $R_m \cap S_m \neq \emptyset$.

□

Seja $T \in \mathcal{T}(L)$ uma árvore filogenética sem raiz. Denotamos por $\mathcal{F}(T)$ o conjunto formado pelos subgrupos menores de todos os cortes em $\mathcal{S}_T$.

Chamamos de *n-árvore maximal* de $\mathcal{F}(T)$ qualquer n-árvore contida em $\mathcal{F}(T)$ que não seja n-subárvore de nenhuma outra n-árvore contida em $\mathcal{F}(T)$.

Lema 2.3.4 *Seja L um conjunto de taxa de cardinalidade maior que dois. Se $T \in \mathcal{T}(L)$ é uma árvore filogenética sem raiz, então o conjunto $\mathcal{F}(T)$ é a união disjunta de um número finito de n-árvores definidas sobre subconjuntos de L.*

Prova Como arestas folhas sempre determinam cortes triviais em $\mathcal{S}_T$, e subgrupos menores de cortes triviais são subgrupos triviais, para qualquer elemento x de L existe um subgrupo $\{x\}$ em $\mathcal{F}(T)$, não importando qual seja a árvore T, pois L tem pelo menos três elementos.

Se considerarmos para cada grupo X em $\mathcal{F}(T)$ o conjunto $\mathcal{F}(T)[X]$ chegaremos rapidamente à conclusão de que este conjunto é uma n-árvore sobre X, pois:

- X pertence ao conjunto.
- Qualquer elemento x de X é também elemento de L, portanto $\{x\}$ pertence a $\mathcal{F}(T)$ e ao conjunto definido.
- Como $\mathcal{F}(T)$ é formado por subgrupos menores de cortes, $\emptyset$ não pertence a $\mathcal{F}(T)$ pela definição de corte.
- Como os subgrupos de $\mathcal{F}(T)$ são subgrupos menores de cortes compatíveis, todos eles são compatíveis entre si pelo Teorema 2.3.3.

É claro que todo subgrupo de $\mathcal{F}(T)$ está em no mínimo uma n-árvore, pois o conjunto definido anteriormente é definido para qualquer subgrupo em $\mathcal{F}(T)$. Como $\mathcal{F}(T)$ é finito, o conjunto de n-árvores definidas sobre subgrupos de $\mathcal{F}(T)$ também é finito. Assim, $\mathcal{F}(T)$ é a união de um conjunto finito de n-árvores. Também é fácil perceber que toda n-árvore maximal de $\mathcal{F}(T)$ é da forma $\mathcal{F}(T)[X]$ para certo $X \in \mathcal{F}(T)$.

Sejam duas n-árvores maximais de $\mathcal{F}(T)$ definidas sobre A e B respectivamente. Como A e B são subgrupos menores de cortes compatíveis, temos $A \cap B = \emptyset$, caso contrário teríamos $A \subset B$ ou $B \subset A$, o que significaria, respectivamente, que $\mathcal{F}(T)[A]$ é uma n-subárvore de $\mathcal{F}(T)[B]$ ou $\mathcal{F}(T)[B]$ é uma n-subárvore de $\mathcal{F}(T)[A]$. Assim, o conjunto das n árvores maximais $\mathcal{F}(T)$ é formado por n-árvores disjuntas, uma vez que os subgrupos nos quais duas n-árvores maximais estão definidos não têm intersecção.

Por outro lado, qualquer subgrupo em $\mathcal{F}(T)$ pertence a pelo menos uma n-árvore maximal, pois se existe um subgrupo X que não faz parte de nenhuma n-árvore maximal, então se $\mathcal{F}(T)[X]$ não for maximal, é n-subárvore de uma n-árvore maximal que contém X.

Chegamos à conclusão de que o conjunto de árvores maximais de $\mathcal{F}(T)$ cobre todos os subgrupos de $\mathcal{F}(T)$ e que todas as árvores maximais são disjuntas. Assim, $\mathcal{F}(T)$ é a união disjuntas das árvores maximais de $\mathcal{F}(T)$.

□

Teorema 2.3.5 *Seja L um conjunto de taxa de cardinalidade maior que dois e $T \in \mathcal{T}(L)$. Então T é completamente resolvida se e somente se $\mathcal{F}(T)$ tiver exatamente três n-árvores maximais e estas árvores forem completamente resolvidas.*

Prova (⇒)

Suponha que T seja uma árvore filogenética completamente resolvida e sem raiz, mas que $\mathcal{F}(T)$ não tenha as propriedades acima. Temos dois casos:

Caso 1: *O número de n-árvores maximais em $\mathcal{F}(T)$ é diferente de 3.*

Como todos os subgrupos em $\mathcal{F}(T)$ são compatíveis, as n-árvores maximais de $\mathcal{F}(T)$ são disjuntas, ou seja, elas não compartilham nenhum subgrupo. Assim, a união entre os maiores subgrupos das n-árvores maximais de $\mathcal{F}(T)$ é L e a interseção entre eles é vazia.

O número de árvores maximais não pode ser um, pois toda árvore maximal em $\mathcal{F}(T)$ é definida sobre um subgrupo contido em $\mathcal{F}(T)$ e sabemos que, como os subgrupos em $\mathcal{F}(T)$ são subgrupos menores de cortes definidos pelas arestas de T em L, para cada elemento $x \in L$ há um subgrupo unitário que contém x em $\mathcal{F}(T)$ e o maior subgrupo em $\mathcal{F}(T)$ não é maior que $\frac{|L|}{2}$, assim, não há subgrupo em $\mathcal{F}(T)$ cuja n-subárvore induzida seja o próprio conjunto $\mathcal{F}(T)$.

Suponha que o número de n-árvores maximais é 2. Assim, há dois subgrupos A e B em $\mathcal{F}(T)$ tais que $A \cup B = L$ e $A \cap B = \emptyset$. Se isto é verdade, então A e B pertencem ao mesmo corte de L, logo, não podem ambos estar ao mesmo tempo em $\mathcal{F}(T)$, uma vez que apenas um deles é o subgrupo menor do corte.

Resta a hipótese de que o número de n-árvores maximais em $\mathcal{F}(T)$ é maior que 3. Sejam A e B dois subconjuntos de L sobre os quais duas n-árvores maximais distintas de $\mathcal{F}(T)$ estão definidas. Novamente temos duas possibilidades:

1.a) $|A \cup B| < \frac{|L|}{2}$: Se isto for verdade, então o subgrupo $A \cup B$, que não está em $\mathcal{F}(T)$, poderia ser inserido neste conjunto sem problemas, pois $A \cup B$ é claramente compatível tanto com A quanto com B quanto com qualquer outro subgrupo em $\mathcal{F}(T)$. Mas se o subgrupo $A \cup B$ é compatível com qualquer subgrupo em $\mathcal{F}(T)$, então o corte $\{A \cup B, L \setminus (A \cup B)\}$ é compatível com todos os cortes de $\mathcal{S}_T$, pelo Teorema 2.3.3, e pode ser adicionado a este sistema de cortes sem problemas, o que significa que $\mathcal{S}_T$ não é maximal, assim como T não é completamente resolvida.

1.b) $|A \cup B| \geq \frac{|L|}{2}$: Como $\mathcal{F}(T)$ tem mais de três n-árvores maximais, podemos encontrar mais dois subconjuntos C e D de L sobre os quais duas outras n-árvores maximais de $\mathcal{F}(T)$ estão definidas. Como $|A \cup B| \geq \frac{|L|}{2}$, $|C \cup D| < \frac{|L|}{2}$ e, usando o mesmo raciocínio do caso anterior, chegamos à conclusão que também neste caso T não é completamente resolvida.

Caso 2: *Ao menos uma das n-árvores maximais de $\mathcal{F}(T)$ não é completamente resolvida.*

Seja A o subconjunto de L sobre o qual uma das n-árvores maximais não completamente resolvidas de $\mathcal{F}(T)$ está definida. Pela definição de n-árvore, A pertence a esta n-árvore, logo $A \in \mathcal{F}(T)$. Então, A é o subgrupo menor do corte $\{A, L \setminus A\}$ e $|A| \leq \frac{|L|}{2}$. Se a n-árvore maximal de $\mathcal{F}(T)$ definida sobre A não é completamente resolvida, então ela admite a inclusão de ao menos um subgrupo B compatível com todos os demais subgrupos. Como B não pertence à n-árvore, mas pode ser incorporado a ela, $|B| < |A|$. Assim, B certamente é o subgrupo menor de $\{B, L \setminus B\}$ que, pelo Teorema 2.3.3, é compatível com todos os cortes do sistema $\mathcal{S}_T$ que, portanto, não é maximal, o que significa que T também não é completamente resolvida.

($\Leftarrow$)

Suponha agora que haja exatamente três n-árvores maximais completamente resolvidas em $\mathcal{F}(T)$.

Sejam A, B e $C \in \mathcal{F}(T)$ os três subgrupos de L sobre os quais as três n-árvores maximais de $\mathcal{F}(T)$ estão definidas. Quaisquer dois subgrupos em $\mathcal{F}(T)$ definem cortes compatíveis em $\mathcal{S}_T$ pela definição de $\mathcal{F}(T)$.

O número de subgrupos em $\mathcal{F}(T)$, dado que as três n-árvores maximais de $\mathcal{F}(T)$ são completamente resolvidas, é:

$$2\,|A| - 1 + 2\,|B| - 1 + 2\,|C| - 1 = 2(|A| + |B| + |C|) - 3 = 2\,|L| - 3$$

e este é o maior número de cortes compatíveis dois a dois que um sistema de cortes definido sobre L pode ter, como foi dito na página 5. Assim, se $\mathcal{F}(T)$ tem exatamente três n-árvores maximais e estas n-árvores são completamente resolvidas, então o sistema de cortes $\mathcal{S}_T$ é máximo, o que significa que T é completamente resolvida.

$\square$