Biologia Computacional - 2004/2	02/01/05
Aula 8: Árvores e Distâncias Ultramétricas	
Instrutor: Berilhes Borges Garcia	Escriba: Rafael Gabeira Cola

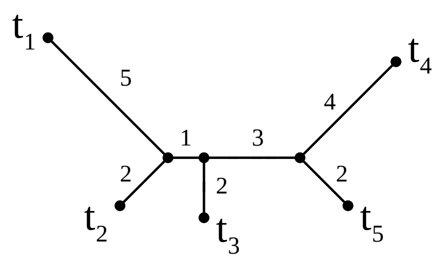
DRAFT

1 Introdução

Antes de discutir distâncias ultramétricas, vamos considerar distâncias aditivas.

Definição 1. *Distâncias aditivas são distâncias que podem ser ajustadas a uma árvore não enraizada tal que todos os pares de distâncias entre grupos taxonômicos (espécies) são iguais à soma do tamanho dos ramos conectando-as.*

t_1	0				
t_2	7	0			
t_3	8	5	0		
t_4	13	10	9	0	
t_5	11	8	7	6	0
	t_1	t_2	t_3	t_4	t_5



Distâncias ultramétricas são mais restritivas que distâncias aditivas.

Definição 2. *Distâncias ultramétricas são distâncias que:*

1. *Se ajustam a uma árvore tal que a distância entre quaisquer duas taxas é igual à soma dos ramos conectando-as.*

2. *Para quaisquer três taxas i, j e k, as duas maiores distâncias são iguais, isto é*

   Se $d_{ik} > d_{jk}$ *então* $d_{ik} = d_{ij}$

   senão *se* $d_{ik} > d_{ij}$ *então* $d_{ik} = d_{jk}$
   senão $d_{ij} = d_{jk}$

2 O que é uma árvore ultramétrica?

Uma árvore ultramétrica T para uma matriz de distâncias $n \times n$ simétrica D possui as seguintes propriedades:

1. T possui n folhas, uma para cada linha de D.

2. Os nós internos são rotulados por um elemento de D e têm dois filhos.

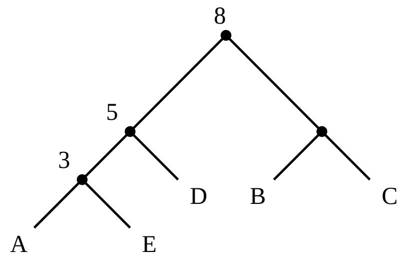
3. Os números rotulando nós internos estritamente decrescem ao longo do caminho da raiz para uma folha.

4. $D(i, j)$ denota o rótulo do ancestral mínimo comum entre as folhas i e j em T.

5. As distâncias em D devem ser ultramétricas.

Considere o seguinte exemplo de Gusfield 97.
Verifique a condição ultramétrica, isto é, para quaisquer três taxas, duas das distâncias serão iguais e maior que a terceira distância.

	A	B	C	D	E
A	0	8	8	5	3
B		0	3	8	8
C			0	8	8
D				0	5
E					0



3 Interpretação de árvores ultramétricas como árvores evolutivas

1. As folhas representam as d unidades taxonômicas existentes.

2. Os nós internos são os eventos divergentes.

Um evento divergente é um ponto onde a história evolutiva de duas espécies se separou.
Se as espécies A e B se divergiram no tempo t, que afirmações podem ser inferidas?

1. A é o ancestral de B

2. B é o ancestral de A

3. Nem A nem B são ancestrais um do outro

Teorema 1. *Uma matriz simétrica de distâncias D possui uma árvore ultramétrica T se e somente se D é uma matriz ultramétrica.*

Proof. $\Leftarrow$ (Se T é ultramétrica então D é ultramétrica)
Se T é ultramétrica

- Cada nó interno v é rotulado $D(i, j)$ onde i e j são folhas e v é o ancestral comum mínimo.

- Para quaisquer três folhas i, j e k em T, assumindo que u é o ancestral comum mínimo, então:

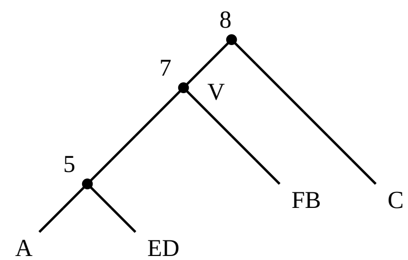
  - u é rotulado por dois de $\{D(i, j), D(i, k), D(j, k)\}$, isto é dois destes valores são iguais e devem ser maiores que o terceiro.

Portanto D é ultramétrica.
$\Rightarrow$ (Se D é ultramétrica então T é ultramétrica)
Se D é ultramétrica:

- O número de elementos distintos d em cada linha i define o número de nós da raiz para a folha
- Cada nó neste caminho é rotulado em ordem decrescente com um label distinto
- Qualquer nó v sob este caminho rotulado $D(i,j)$ deve ser o ancestral comum mínimo das folhas i e j
- O caminho para a folha i particiona as $n-1$ folhas restantes em $d-1$ classes
- Cada nó distinto sob o caminho para i é rotulado pelas distâncias de i para as folhas naquela partição

Example:

	A	B	C	D	E	F
A	0	7	8	5	5	7
B						
C						
D						
E						
F						



⊠

Nós queremos recursivamente encontrar a árvore ultramétrica para cada uma das $d-1$ partições e então combiná-las.
Considere a partição definida por um nó interno v. j é uma folha contida nesta partição.
l é uma outra folha. Existem três casos:

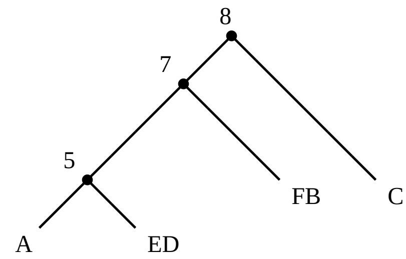
1. l está na mesma partição que j.
2. l está em uma partição entre i e o nó v.
3. l está em uma partição entre o nó v e a raiz.

Os três casos, assumindo $i = A$, $j = F$

1. l está na mesma partição que j, exemplo: $l = B$.
2. l está em uma partição entre i e v, exempo: $l = D$.
3. l está em uma partição entre v e a raiz, exemplo: $l = C$.

Caso 1: $(i = A,\ j = F,\ l = B)$
$D(i,j) = D(i,l)$ assim $D(j,l) \leq D(i,j)$. Por quê?
$\Rightarrow$Assim nós podemos adicionar a subárvore contendo j e l, sabendo que $D(j,l)$ está correto.
Caso 2: $(i = A,\ j = F,\ l = D)$

	A	B	C	D	E	F
A	0	7	8	5	5	7
B						
C						
D						
E						
F						



$D(i,l) < D(i,j)$ assim $D(i,j) = D(j,l)$

$\Rightarrow$Logo nós podemos adicionar a subárvore em v contendo j sabendo que $D(j,l)$ está correto, ou seja v está rotulado corretamente.

Caso 3: $(i = A,\ j = F,\ l = C)$

$D(i,l) > D(i,j)$ assim $D(i,l) = D(j,l)$

$\Rightarrow$Logo nós podemos adicionar a subárvore contendo j sabendo que $D(j,l)$ é correto, isto é este rotula o ancestral comum mínimo.

Em cada um dos três casos, a árvore ultramétrica definida por v pode ser corretamente conectada a v.

Dessa forma, nós podemos construir, utilizando recursão, a árvore ultramétrica T para D.

$\square$