



After Mid

Phylogeny

১২০
 time ৩ ঘণ্টা
 দার
 ১২) আবার
 number of
 ২) mutations ৩
 ঘণ্টা দার

Distance in Trees:

tree distance

$d_{ij}(T) \rightarrow$ একটি tree কে T এর i ও j এর
 \downarrow node এর মার্কা- distance.

৭ আর ৮ আসলে leaf এর মার্কা- distance

Distance matrix: $n \times n$ matrix

$\rightarrow n$ সস্বত্বক species এর মার্কা distance

$D_{ij} \rightarrow$ Edit distance.

Fitting Distance Matrix:

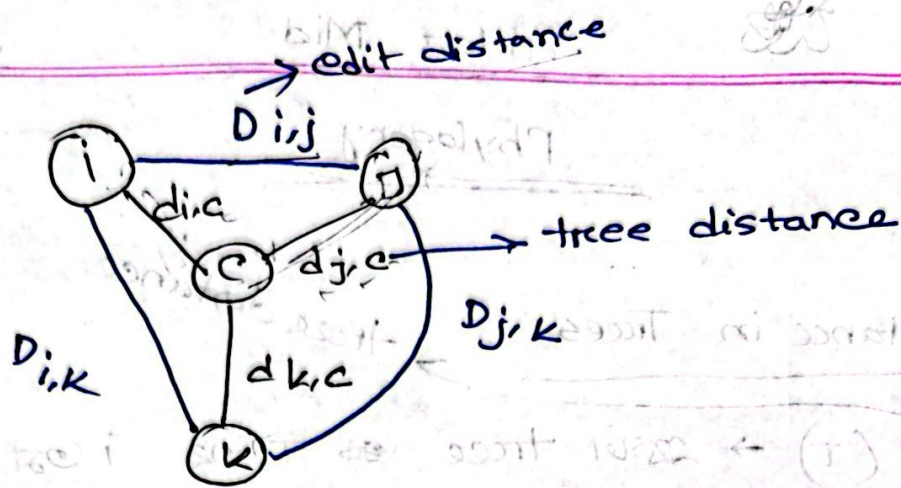
১) $D_{ij} = a_i b_j + a_j b_i$
 ২) $D_{ij} = a_i b_j + a_j b_i$
 ৩) $D_{ij} = a_i b_j + a_j b_i$

best fitting - এমনভাবে design করব যাতে

edit distance আর tree distance সমান

হয়।

$$D_{ij} - D_{ij} = a_i b_j + a_j b_i - a_i b_j - a_j b_i = 0$$



$$D = \begin{matrix} & \begin{matrix} i & j & k \end{matrix} \\ \begin{matrix} i \\ j \\ k \end{matrix} & \begin{bmatrix} 0 & d_{i,j} & d_{i,k} \\ d_{i,j} & 0 & d_{j,k} \\ d_{i,k} & d_{j,k} & 0 \end{bmatrix} \end{matrix}$$

$$d_{i,c} + d_{j,c} = D_{ij} \quad \text{--- (I) } \checkmark$$

$$d_{j,c} + d_{k,c} = D_{jk} \quad \text{--- (II) } \checkmark$$

$$d_{i,c} + d_{k,c} = D_{ik} \quad \text{--- (III) } \checkmark$$

$$\text{(I) - (II)}$$

$$\Rightarrow d_{i,c} - d_{k,c} = D_{ij} - D_{jk}$$

$$\text{(III) + (IV)}$$

$$2d_{i,c} = \frac{D_{ik} + D_{ij} - D_{jk}}{2}$$

Similarly,

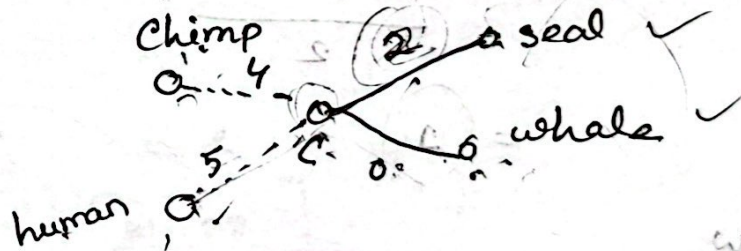
$$d_{ij} = \frac{D_{ij,k} + D_{j,k} - D_{i,k}}{2}$$

$$d_{kc} = \frac{D_{i,k} + D_{j,k} - D_{i,j}}{2}$$

Example

3 leaf theorem

Additive matrix
(- error = 1)

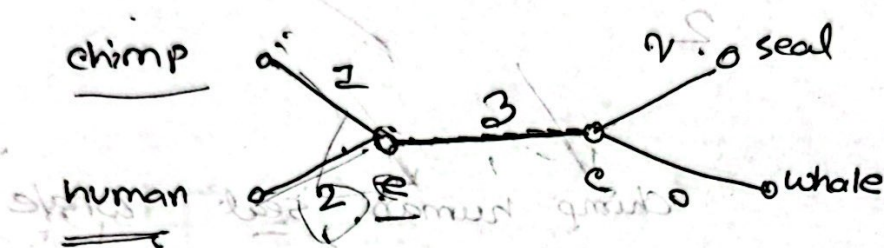
$$D = \begin{matrix} & \begin{matrix} \text{chimp} \\ \text{human} \\ \text{seal} \\ \text{whale} \end{matrix} \\ \begin{matrix} \text{chimp} \\ \text{human} \\ \text{seal} \\ \text{whale} \end{matrix} & \begin{bmatrix} 0 & 3 & 6 & 4 \\ 3 & 0 & 7 & 5 \\ 6 & 7 & 0 & 2 \\ 4 & 5 & 2 & 0 \end{bmatrix} \end{matrix}$$


using neighbouring leaves.

$$d_{\text{seal}, \text{human}} = \frac{D_{\text{seal}, \text{whale}} + D_{\text{seal}, \text{chimp}} + D_{\text{human}, \text{whale}}}{2}$$

$$= \frac{2 + 6 + 4}{2} = 2$$

$$D' = \begin{matrix} & \begin{matrix} \text{chimp} & \text{human} & e \end{matrix} \\ \begin{matrix} \text{chimp} \\ \text{human} \\ e \end{matrix} & \begin{bmatrix} 0 & 3 & 4 \\ 3 & 0 & 5 \\ 4 & 5 & 0 \end{bmatrix} \end{matrix}$$



$$d_{\text{chimp}, e} = \frac{D_{\text{chimp-human}} + D_{\text{chimp-e}} - D_{\text{human-e}}}{2}$$

$$= \frac{3 + 4 - 5}{2}$$

$$= 1$$

HW

example 2:

$$\begin{matrix} & \begin{matrix} i & j & k & l \end{matrix} \\ \begin{matrix} i \\ j \\ k \\ l \end{matrix} & \begin{bmatrix} 0 & 13 & 21 & 22 \\ 13 & 0 & 12 & 13 \\ 21 & 12 & 0 & 13 \\ 22 & 13 & 13 & 0 \end{bmatrix} \end{matrix}$$

non-additive
matrix
(negative
edge
error)

Limb length
theorem.



limb node - theorem:

- যে কোনো একটি leaf কে randomly choose করুন,
- যদি 3 টি leaf থেকে 2 টি নিয়ে 3 leaf theorem 302

$$\begin{aligned} (i, j, k) &\rightarrow d_j = \frac{D_{ij} + D_{jk} - D_{ik}}{2} = \frac{13 + 12 - 21}{2} = 2 \\ (i, j, l) &\rightarrow d_j = \frac{D_{ij} + D_{jl} - D_{il}}{2} = \frac{13 + 13 - 22}{2} = 2 \\ (j, k, l) &\rightarrow d_j = \frac{D_{jk} + D_{jl} - D_{kl}}{2} = \frac{12 + 13 - 13}{2} = 6 \end{aligned}$$

minimum value

Limblength (j) = 2



j তম row, j তম column থেকে limblength minus করতে হবে।

$D_{bald} =$

	i	j	k	l
i	0	11	21	22
j	11	0	10	11
k	21	10	0	13
l	22	11	13	0

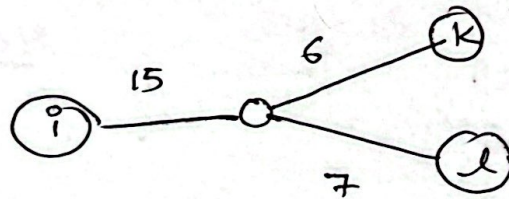
D_{bald}

$$D_{trim} = \begin{matrix} & i & k & l \\ \begin{matrix} i \\ k \\ l \end{matrix} & \begin{bmatrix} 0 & 21 & 22 \\ 21 & 0 & 13 \\ 22 & 13 & 0 \end{bmatrix} \end{matrix}$$

• j বাদ
দিক
trim
করা

• then 3 leaf apply

using 3 leaf calculation,



length
• limb_i এর value যেখান থেকে এসেছে উঠে

ডান দিক বের করার, প্রকারিতা min of

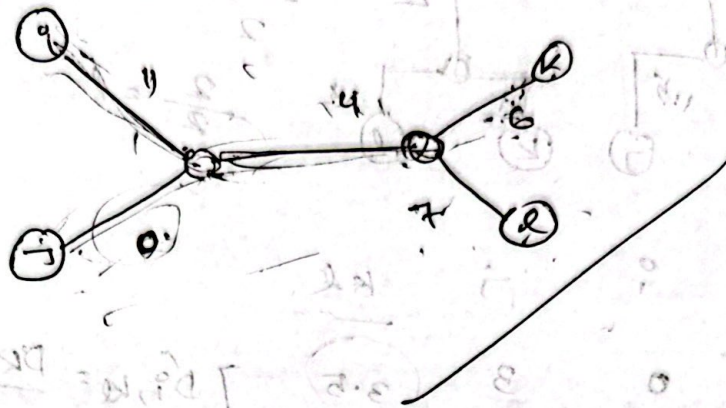
limb length থাকলে যেখানে অন্য consider

হবে,

$$D_{j \text{ bald}} = \frac{D_{ij \text{ bald}} + D_{jk \text{ bald}} - D_{ik \text{ bald}}}{2}$$

$$\Rightarrow D_{ij \text{ bald}} + D_{jk \text{ bald}} = D_{ik \text{ bald}}$$

Final tree,



চুটি leaf
মানের দেয়
যার condition
check

Four Point Condition: Matrix additive কিনা তা
বের করা যায়।

Slide

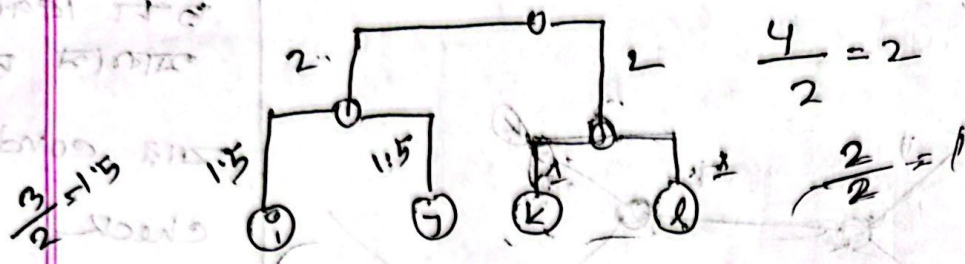
প্রত্যেকবার ৪ টা leaf নিয়ে calculation যদি প্রত্যেকবার
match হয় তাহলে additive matrix.

• চুটি path এর সোজান same হতে হবে তাই বারি
একটা ছোট্ট এটা

UPGMA : clustering এর tree construct করা হয়।

$$D = \begin{matrix} & \begin{matrix} 1 & j & k & l \end{matrix} \\ \begin{matrix} 1 \\ j \\ k \\ l \end{matrix} & \begin{bmatrix} 0 & 3 & 4 & 3 \\ 3 & 0 & 4 & 5 \\ 4 & 4 & 0 & 2 \\ 3 & 5 & 2 & 0 \end{bmatrix} \end{matrix}$$

• cluster - collection
of objects.



$$D' = \begin{matrix} & i & j & kl \\ \begin{matrix} i \\ j \\ kl \end{matrix} & \begin{bmatrix} 0 & 3 & 3.5 \\ 3 & 0 & 4.5 \\ 3.5 & 4.5 & 0 \end{bmatrix} \end{matrix}$$

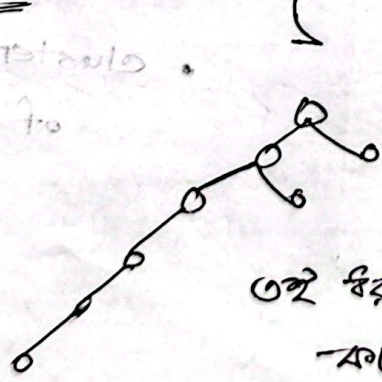
$$\hat{D}_{i,kl} = \frac{D_{ik} + D_{il}}{2} = \frac{4+3}{2} = 3.5$$

$$\hat{D}_{j,kl} = \frac{D_{jk} + D_{jl}}{2} = \frac{4+5}{2} = 4.5$$

$$D'' = \begin{matrix} & i & j & kl \\ \begin{matrix} i & j \\ kl \end{matrix} & \begin{bmatrix} 0 & 4 \\ 4 & 0 \end{bmatrix} \end{matrix}$$

$$D''_{j,kl} = \frac{D'_{i,kl} + D'_{j,kl}}{2} = \frac{3.5 + 4.5}{2} = 4$$

UPGMA - ultrametric tree



ଓଏ ଡିଟାନ୍ସ ଟ୍ରି ଫାଇ
- ଏଡ଼ା ।