

Phylogeny

Phylogeny:

Phylogeny is the study of the evolutionary history and relationships among species or group of organisms.

* Existing যত living being আছে তাদের বিষয়ে যা
তাৰা common ancestor share কৰে।

Evolutionary Tree:

An evolutionary tree, also called a phylogenetic tree, is a diagram that shows how different species, genes or organisms are related through evolution.

Out of Africa Hypothesis:

Around the time the giant panda riddle was solved, a DNA-based construction of the human evolutionary tree led to the 'Out of Africa Hypothesis' that claims our most ancient ancestor lived in Africa roughly 200,000 years ago.

- The Out of Africa Hypothesis is a scientific theory that proposes all modern humans (*Homo sapiens*) originated in Africa, and then migrated out to populated the rest of the world.

Rooted Tree & Unrooted Tree:

In the unrooted tree

Rooted tree has a root which represents the most recent common ancestor of all entities (species, genes etc) in the tree.

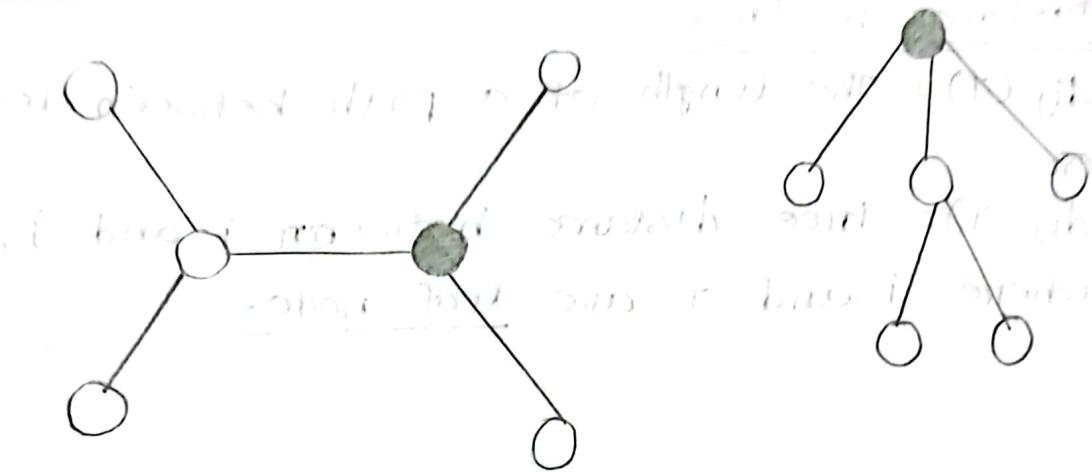


Fig: Rooted Tree

Unrooted tree does not show a root. It just shows relationships among entities without indicating the direction of evolution.

In the unrooted tree the position of the root ("oldest ancestor") is unknown. Otherwise, they are like rooted trees.

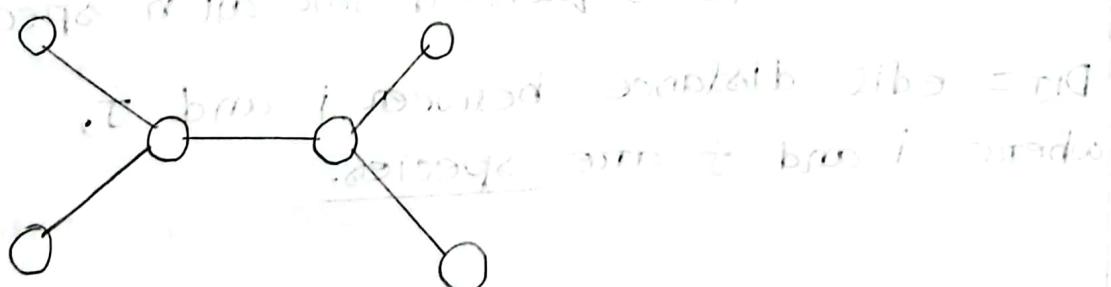


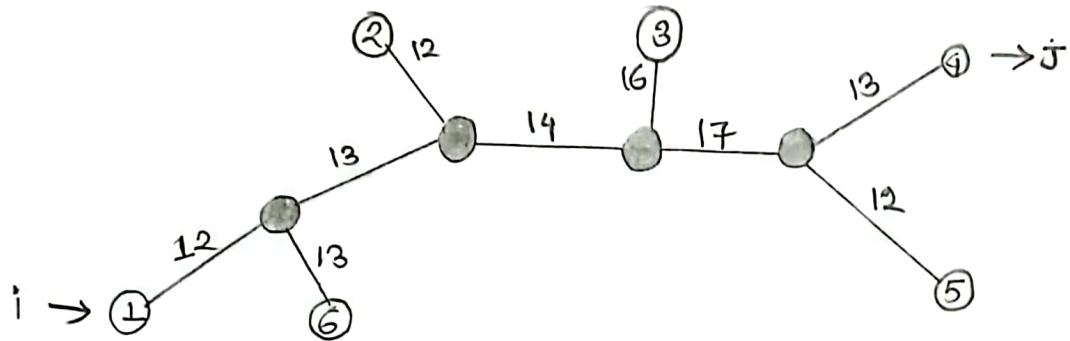
Fig: Unrooted Tree

Distance in tree:

$d_{ij}(T)$ = The length of a path between leaves i & j

Or,

$d_{ij}(T)$ = Tree distance between i and j ,
where i and j are leaf node.

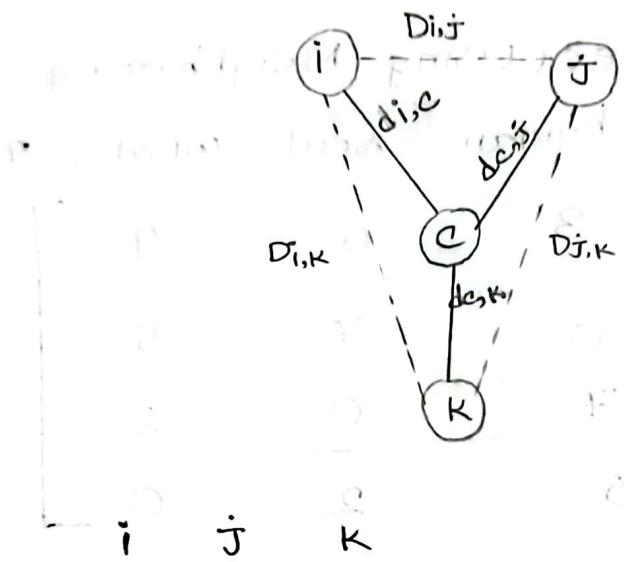


$d_{1,4} = 12 + 13 + 14 + 17 + 13$

Distance Matrix:

Dij may be defined as the edit distance between a gene in species i and species j, where the gene of interest is sequenced for all n species.

D_{ij} = edit distance between i and j ,
where i and j are species.



$$D = \begin{bmatrix} i & j & k \\ 0 & Dij & Dik \\ Dij & 0 & Djk \\ Dik & Djk & 0 \end{bmatrix}$$

Similarly,

$$djc = \frac{Dij + Djk - Dik}{2}$$

$$djk = \frac{Dik + Djk - Dij}{2}$$

Observe:

$$dic + djc = Dij \quad \text{--- (1)}$$

$$djc + dkc = Djk \quad \text{--- (II)}$$

$$dic + dkc = Dik \quad \text{--- (III)}$$

$$(I) - (II) \Rightarrow$$

$$dic - dkc = Dij - Djk \quad \text{--- (IV)}$$

$$(III) + (IV) \Rightarrow$$

$$2dic = Dik + Dij - Djk$$

$$\therefore dic = \frac{Dik + Dij - Djk}{2}$$

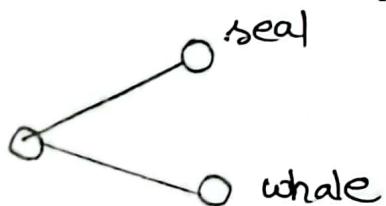
Example:

Distance matrix \rightarrow Tree (Using Neighboring Leaves):

	chimp	human	seal	whale	?
chimp	0	3	6	4	
human	3	0	7	5	
seal	6	7	0	2	
whale	4	5	2	0	

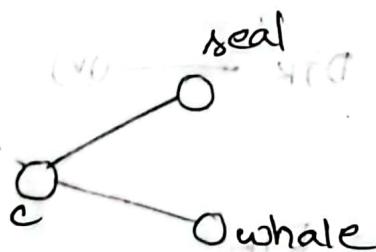
Solution:

- 1- Take two species that have the minimum value in the matrix. Which means, they share the same ancestor.
 $(\text{seal, whale}) = 3$



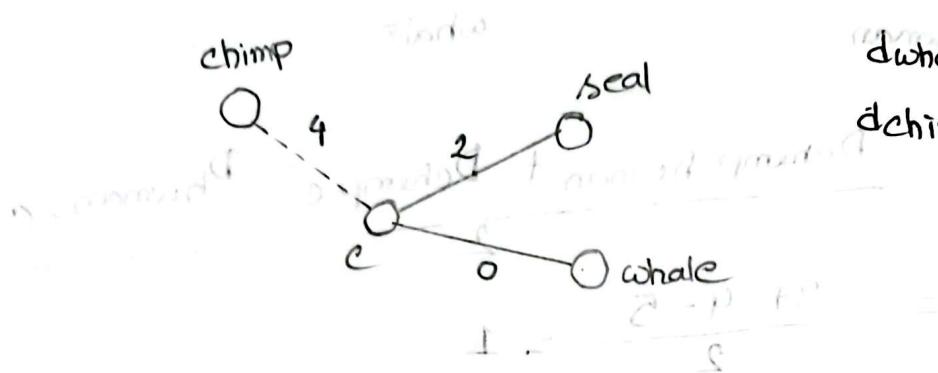
2. Then choose randomly choose one species.

As there is no proof that seal, whale and chimp share same ancestor, so we connect them with dot line.



$$d_{seal,c} = \frac{D_{seal, whale} + D_{seal, chimp} - D_{whale, chimp}}{2}$$

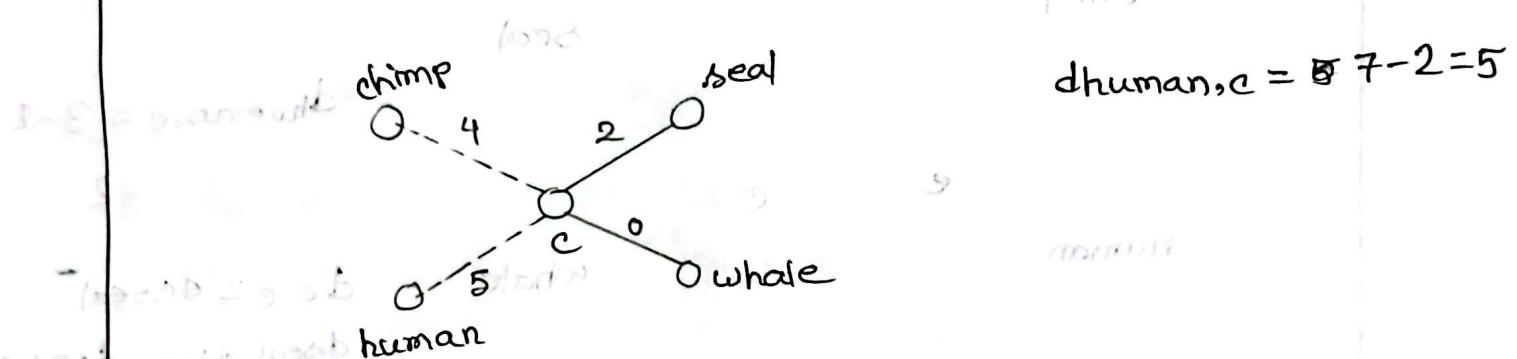
$$= \frac{2+6-4}{2} = 2$$



$$d_{whale,c} = 2-2=0$$

$$d_{chimp,c} = 6-2=4$$

3. Take human in the tree,



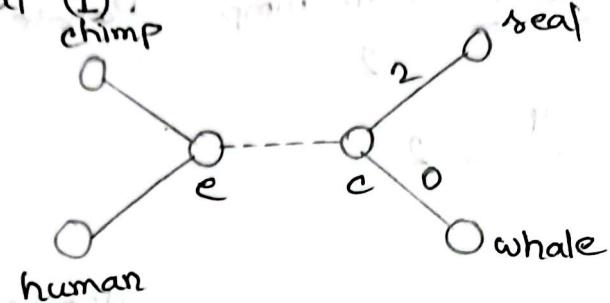
q.m.d

$$d_{human,c} = 7-2=5$$

4. Remove seal and whale from the matrix and put 'c' in the matrix.

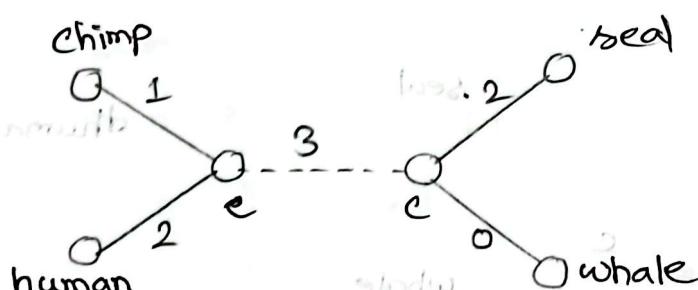
$$D' = \begin{bmatrix} \text{chimp} & \text{human} & c \\ \text{chimp} & 0 & 3 & 4 \\ \text{human} & 3 & 0 & 5 \\ c & 4 & 5 & 0 \end{bmatrix}$$

5. Repeat (1)



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

$$= \frac{3+4-5}{2} = 1$$

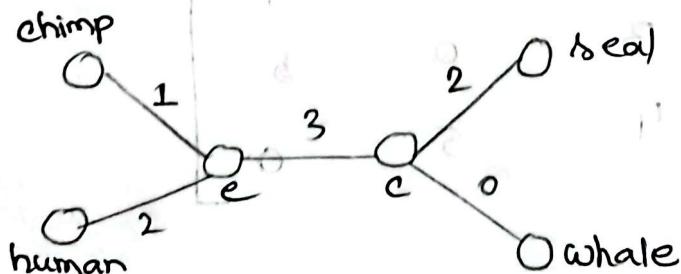


$$d_{human,e} = 3-1 = 2$$

$$d_{c,e} = d_{seal} = d_{seal,chimp} - d_{seal,c}$$

$$d_{chimp,e}$$

Since, chimp \rightarrow seal, chimp \rightarrow whale, human \rightarrow seal etc calculation are correct, so the connection between c and e are valid. So, the final tree is,



Limb Length Theorem:

Given an additive matrix D and a leaf i ,

$\text{LIMBLENGTH}(j)$ is equal to the minimum value of $(D_{ij} + D_{jk} - D_{ik})/2$ over all leaves i and k .

- It can be solved without any error.

Example:

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

1. choose any leaf randomly. choose j .

From the remaining 3 leaves (i, k, l), 3 leaves tree need to be constructed, keeping j common.

Possible leaf keeping j common:-

$$(i, j, k) \rightarrow d_j = \frac{D_{ij} + D_{jk} - D_{ik}}{2} = \frac{13 + 12 - 21}{2} = 2$$

$$(i, j, l) \rightarrow d_j = \frac{D_{ij} + D_{jl} - D_{il}}{2} = \frac{13 + 13 - 22}{2} = 0$$

$$(j, k, l) \rightarrow d_j = \frac{D_{jk} + D_{jl} - D_{kl}}{2} = \frac{12 + 13 - 13}{2} = 6$$

The one with the minimum value among all will be the weight of the edge of any leaf connected to j .

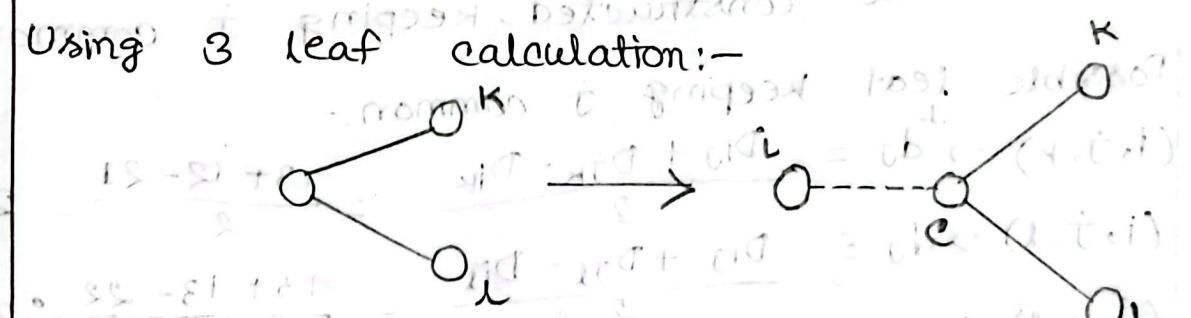
- So. limb length of $j = 2$. But the edge weight of $i-j$ is omitted because the edge weight of the leaf connected to j is set to 0.
1. First of all normal widths are given.
 2. The matrix needs to be updated by subtracting the limb length of $i-j$ from the j th row and j th column.

$$D^{\text{bold}} = \begin{bmatrix} i & 0 & 11 & 21 & 22 \\ j & 21 & 0 & 10 & 11 \\ k & 21 & 10 & 0 & 13 \\ l & 22 & 11 & 13 & 0 \end{bmatrix}$$

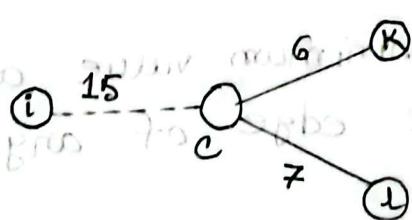
The limb length is omitted because the edge weight of the leaf connected to j is set to 0.

$$D^{\text{trim}} = \begin{bmatrix} i & 0 & 21 & 22 \\ j & 21 & 0 & 13 \\ k & 22 & 13 & 0 \end{bmatrix}$$

Using 3 leaf calculation:-

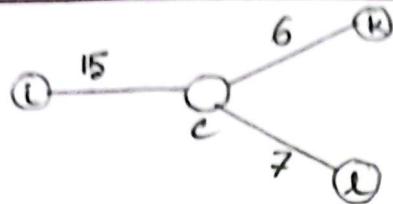


$$d_{k,c} = \frac{d_{k,l} + d_{k,i} - D_{i,j}}{2} = \frac{13 + 21 - 22}{2} = \frac{34 - 22}{2} = \frac{12}{2} = 6$$



$$d_{k,c} = 13 - 6 = 7$$

$$d_{i,c} = 21 - 6 = 15$$



~~3. Take any leaf (possible leaf from ①)~~

3. Take the leaf fore from, which the limb length was taken.

$$d_{ij}^{\text{bold}} = \frac{D_{ij}^{\text{bold}} + D_{jk}^{\text{bold}} - D_{ik}^{\text{bold}}}{2} \rightarrow \frac{11 + 10 - 21}{2} = \frac{21 - 21}{2} = 0$$

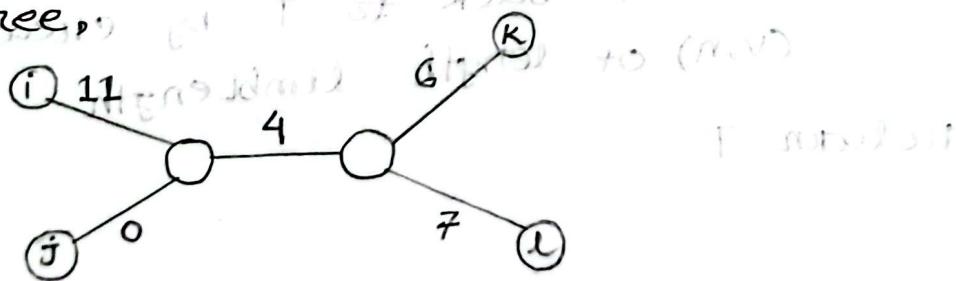
$$\Rightarrow \frac{D_{ij}^{\text{bold}} + D_{jk}^{\text{bold}} - D_{ik}^{\text{bold}}}{2} = 0$$

$$\Rightarrow D_{ij}^{\text{bold}} + D_{jk}^{\text{bold}} - D_{ik}^{\text{bold}} = 0$$

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

$$11 + 10 = 21$$

Final Tree,



Limb Length Algorithm:

AdditivePhylogeny (D, n)

if $n = 2$

return the tree consisting of a single edge of length $D_{1,2}$

limbLength $\leftarrow \text{Limb}(D, n)$

for $j \leftarrow 1$ to $n-1$

$D_{j,n} \leftarrow D_{j,n} - \text{limbLength}$

$D_{n,j} \leftarrow D_{j,n}$

$(i, n, k) \leftarrow$ three leaves such that $D_{i,k} = D_{i,n} + D_{n,k}$

$z \leftarrow D_{i,n}$

remove row n and column n from D

$T \leftarrow \text{AdditivePhylogeny}(D, n-1)$

$v \leftarrow$ the (potentially new) node in T at distance n from i on the path between i and k and leaf n back to T by creating a limb (v, n) of length limbLength

return T

The Four Point Condition:

Let $1 \leq i, j, k, l \leq n$ be four distinct leaves in a tree

Compute: 1. $D_{ij} + D_{kl}$, 2. $D_{ik} + D_{jl}$, 3. $D_{il} + D_{jk}$

The four point condition for the quartet i, j, k, l is satisfied if two of these sums are the same, with the third sum smaller than these first two.

Theorem:

An $n \times n$ matrix D is additive if and only if the four point condition holds for every quartet

$1 \leq i, j, k, l \leq n$

Example: $i \ j \ k \ l$

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

Tree:

1. $D_{ij} + D_{kl}$

$$= 13 + 13$$

$$= 26$$

2. $D_{ik} + D_{jl}$

$$= 21 + 13$$

$$= 34$$

3. $D_{il} + D_{jk}$

$$= 22 + 12$$

$$= 34$$

$$D_{ij} + D_{kl} < D_{ik} + D_{jl} = D_{il} + D_{jk}$$

This satisfy the four point condition.
So the matrix is additive matrix.

Example: $i \ j \ k \ l \ m$

$$D = \begin{bmatrix} i & 0 & 3 & 4 & 5 & 6 \\ j & 3 & 0 & 5 & 6 & 7 \\ k & 4 & 5 & 0 & 7 & 8 \\ l & 5 & 6 & 7 & 0 & 9 \\ m & 6 & 7 & 8 & 9 & 0 \end{bmatrix}$$

The possible 4×4 matrix from the above matrix:

$$D_1 = \begin{bmatrix} i & j & k & l \\ i & 0 & 3 & 4 & 5 \\ j & 3 & 0 & 5 & 6 \\ k & 4 & 5 & 0 & 7 \\ l & 5 & 6 & 7 & 0 \end{bmatrix}$$

$$D_2 = \begin{bmatrix} i & j & k & m \\ i & 0 & 3 & 4 & 6 \\ j & 3 & 0 & 5 & 7 \\ k & 4 & 5 & 0 & 8 \\ m & 6 & 7 & 8 & 0 \end{bmatrix}$$

$$D_3 = \begin{bmatrix} i & j & l & m \\ i & 0 & 3 & 5 & 6 \\ j & 3 & 0 & 6 & 7 \\ l & 5 & 6 & 0 & 9 \\ m & 6 & 7 & 9 & 0 \end{bmatrix}$$

$$D_4 = \begin{bmatrix} i & k & l & m \\ i & 0 & 4 & 5 & 6 \\ k & 4 & 0 & 7 & 8 \\ l & 5 & 7 & 0 & 9 \\ m & 6 & 8 & 9 & 0 \end{bmatrix}$$

$$D_5 = \begin{bmatrix} j & k & l & m \\ j & 0 & 45 & 65 & 87 \\ k & 5 & 0 & 7 & 18 \\ l & 6 & 7 & 0 & 9 \\ m & 7 & 8 & 9 & 0 \end{bmatrix}$$

From D_1 ,

compute:

$$1. D_{ij} + D_{kl}$$

$$= 3 + 7$$

$$= 10$$

$$2. D_{ik} + D_{jl}$$

$$= 4 + 6$$

$$= 10$$

$$3. D_{il} + D_{jk}$$

$$= 5 + 5$$

$$= 10$$

From D_2 ,

compute:

$$1. D_{ij} + D_{km}$$

$$= 3 + 8$$

$$= 11$$

$$2. D_{ik} + D_{jm}$$

$$= 4 + 7$$

$$= 11$$

$$3. D_{im} + D_{jk}$$

$$= 6 + 5$$

$$= 11$$

From D_3 ,

$$1. D_{ij} + D_{lm}$$

$$= 3 + 9$$

$$= 12$$

$$2. D_{il} + D_{jm}$$

$$= 5 + 7$$

$$= 12$$

$$3. D_{im} + D_{jl}$$

$$= 6 + 6$$

$$= 12$$

From D_4 ,

$$1. D_{ik} + D_{lm}$$

$$= 4 + 9$$

$$= 13$$

$$2. D_{il} + D_{km}$$

$$= 5 + 8$$

$$= 13$$

$$3. D_{im} + D_{kl}$$

$$= 6 + 7$$

$$= 13$$

From D_5 ,

$$1. D_{jk} + D_{lm}$$

$$= 5 + 9$$

$$= 14$$

$$2. D_{jl} + D_{km}$$

$$= 6 + 8$$

$$= 14$$

$$3. D_{jm} + D_{kl}$$

$$= 7 + 7$$

$$= 14$$

UPGMA : Unweighted Pair Group Method with Arithmetic Mean:

- UPGMA is a clustering algorithm that:
 - computes the distance between clusters using average pairwise distance
 - assigns a height to every vertex in the tree, effectively assuming the presence of a molecular clock and dating every vertex.
- * the distance between clusters c_1 and c_2 is defined as the average pairwise distance between elements of c_1 and c_2

$$D_{c_1, c_2} = \frac{\sum_{i \in c_1} \sum_{j \in c_2} D_{i,j}}{|c_1| \cdot |c_2|}$$

Example:

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

$D_{i,j} = 3$
 $\frac{3}{2} = 1.5$

$$D_{ij,kl} = 4$$

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

$$D_{kl} = 2$$

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

$$\Rightarrow D' = \begin{bmatrix} i & j & kl \\ i & 0 & 3 & 3.5 \\ j & 3 & 0 & 4.5 \\ kl & 3.5 & 4.5 & 0 \end{bmatrix}; D'_{i,kl} = \frac{D_{ik} + D_{il}}{2} = \frac{4+3}{2} = 3.5$$

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

$$\Rightarrow D'' = \begin{bmatrix} ij & kl \\ ij & 0 & 4 \\ kl & 4 & 0 \end{bmatrix}; D''_{ij,kl} = \frac{D_{i,kl} + D_{j,kl}}{2} = \frac{3.5 + 4.5}{2} = 4$$