

In the name of the most high

Introduction to Bioinformatics

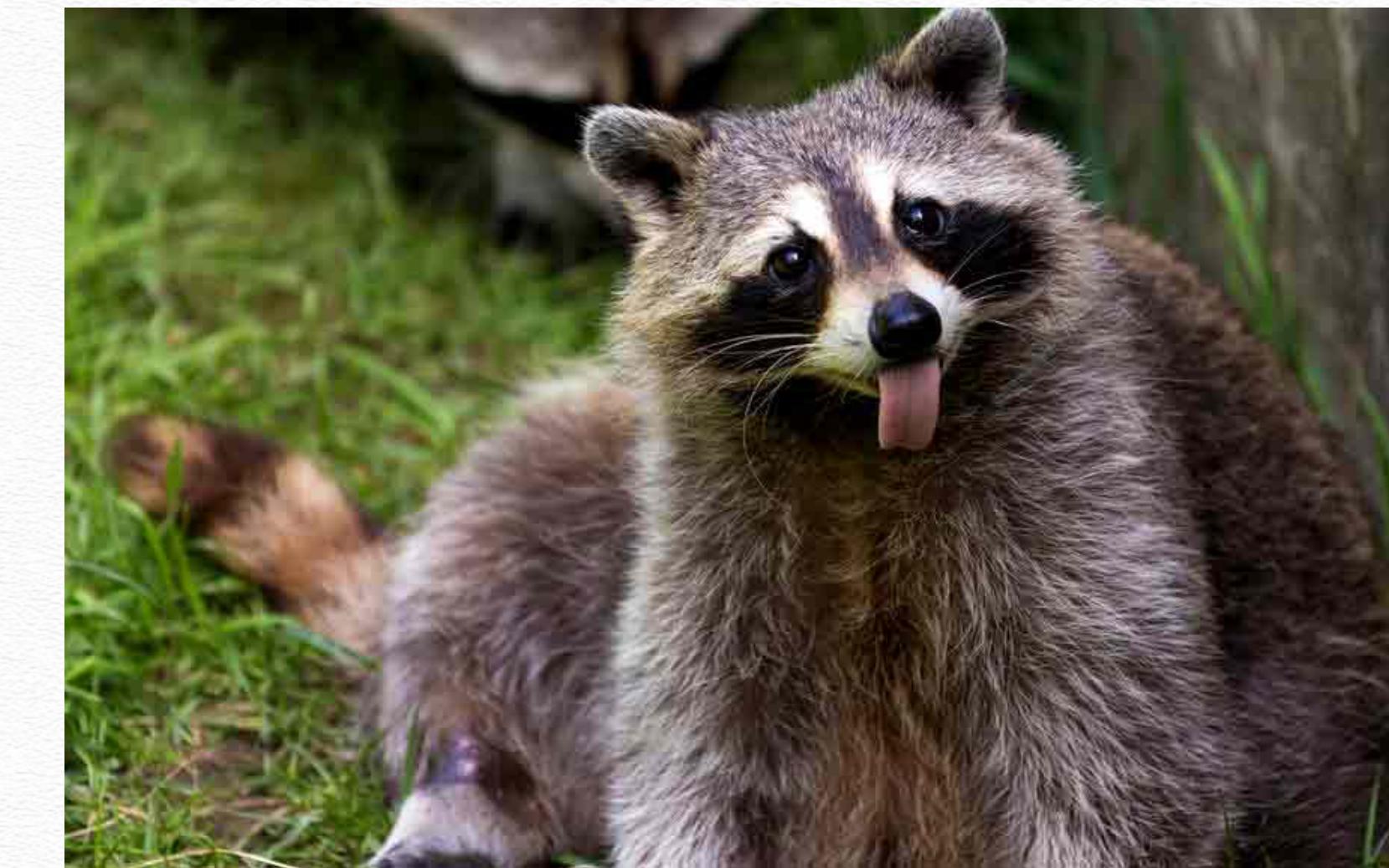
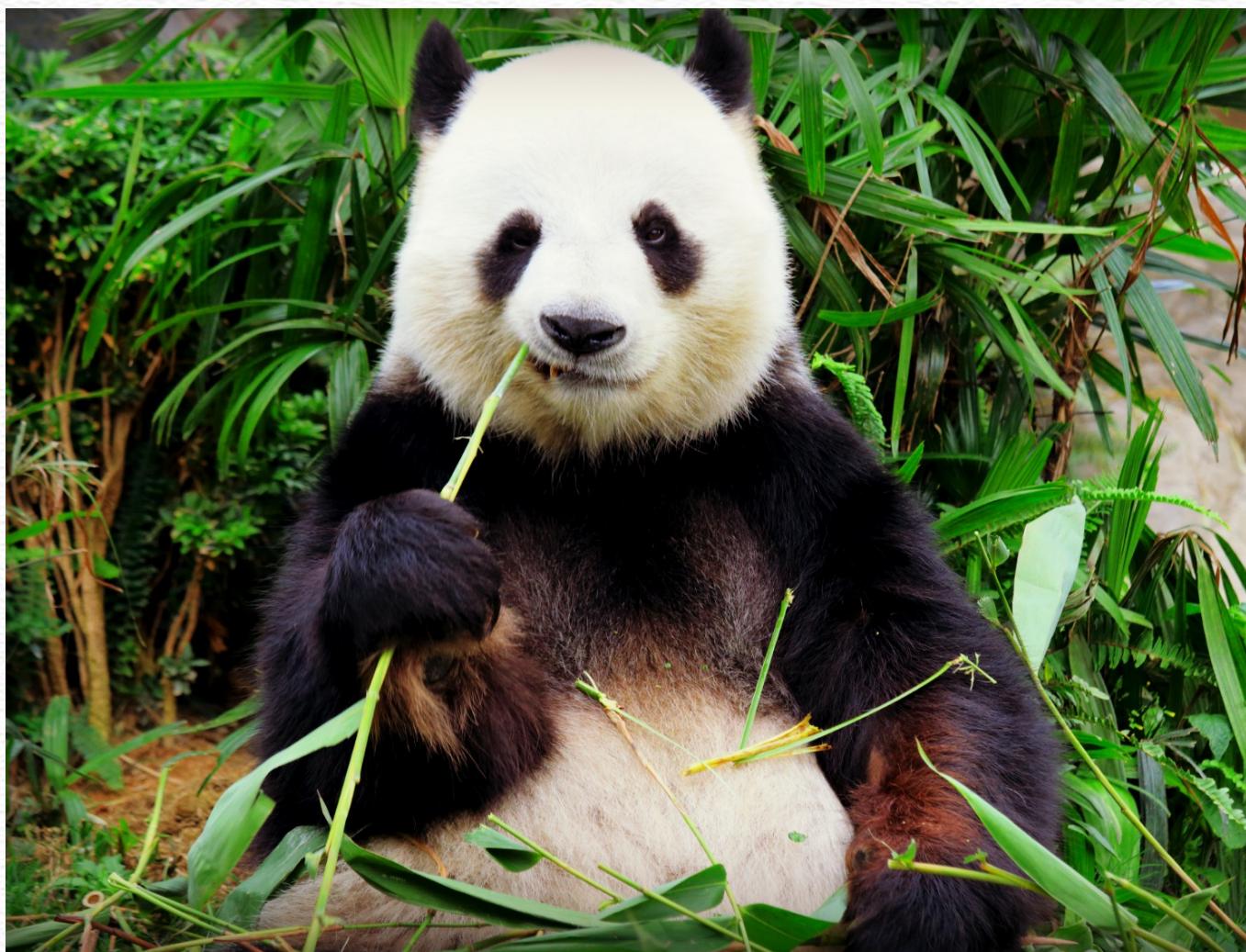
Phylogenetic Trees

Ali Sharifi-Zarchi

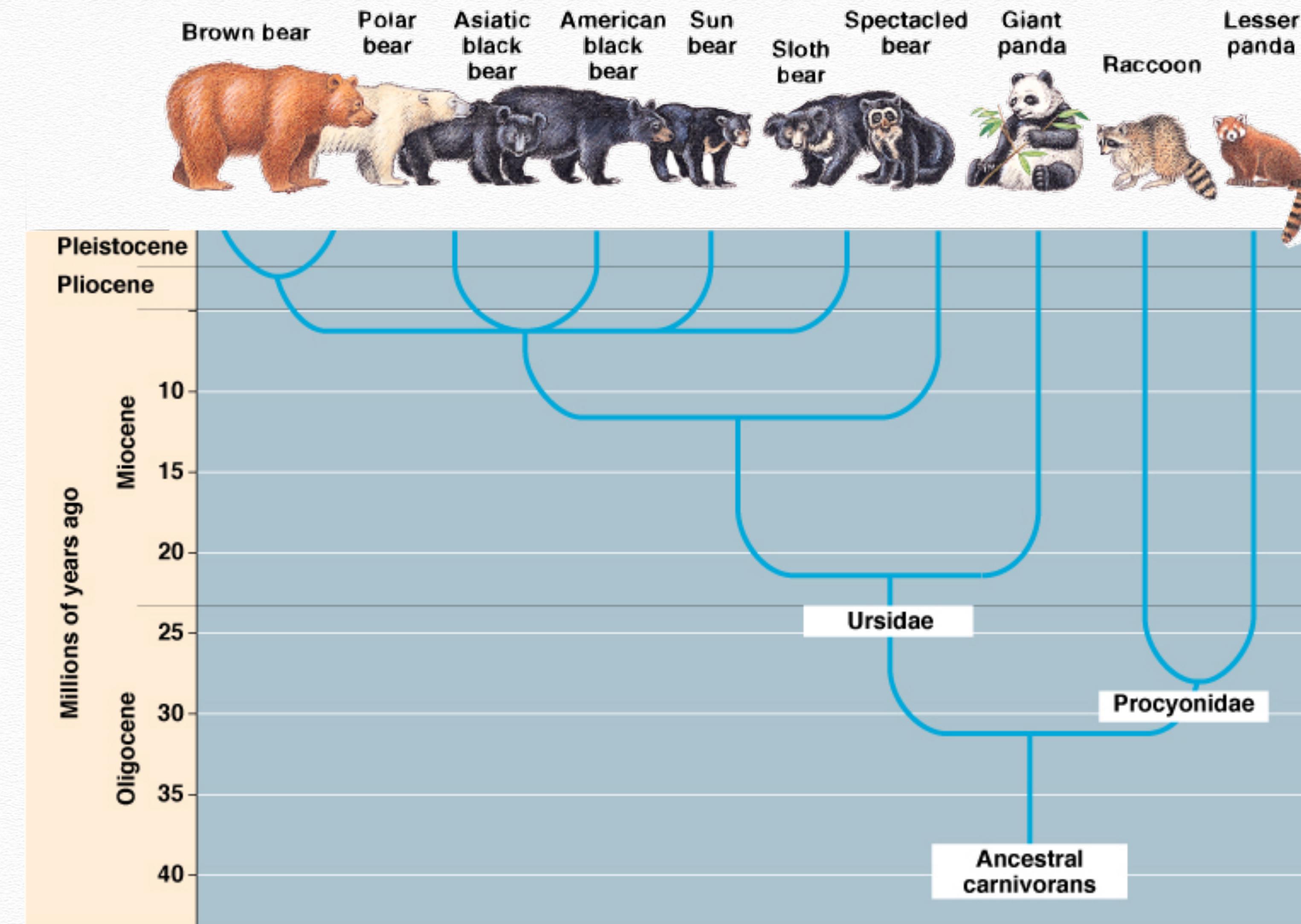
Department of Computer Engineering, Sharif University of Technology

These slides are available under the Creative Commons Attribution License.

Red Pandas Problem



Red Pandas

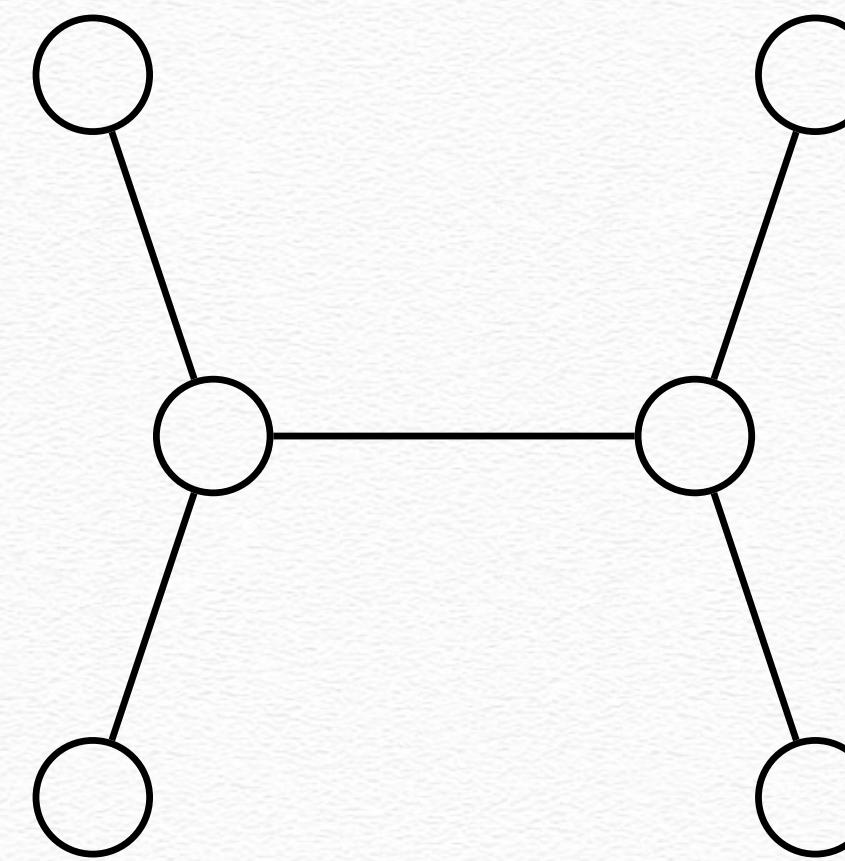


Burrows Wheeler Transform (BWT)

babaabdad\$ →

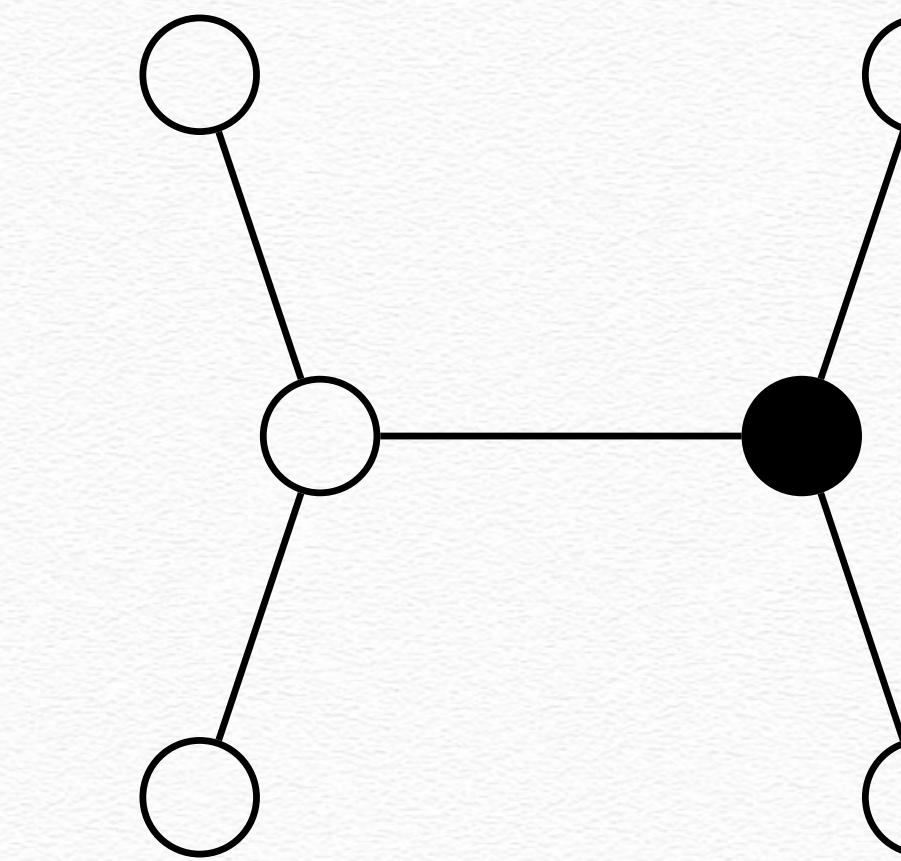
\$babaa**d**
aabdad\$b**a**
abaabdad\$b**a**
abdad\$b**a**
ad\$babaab**d**
baabdad\$b**a**
babaabdad\$b**a**
bdad\$baba**a**
d\$babaab**d**
dad\$babaab**b**

Trees

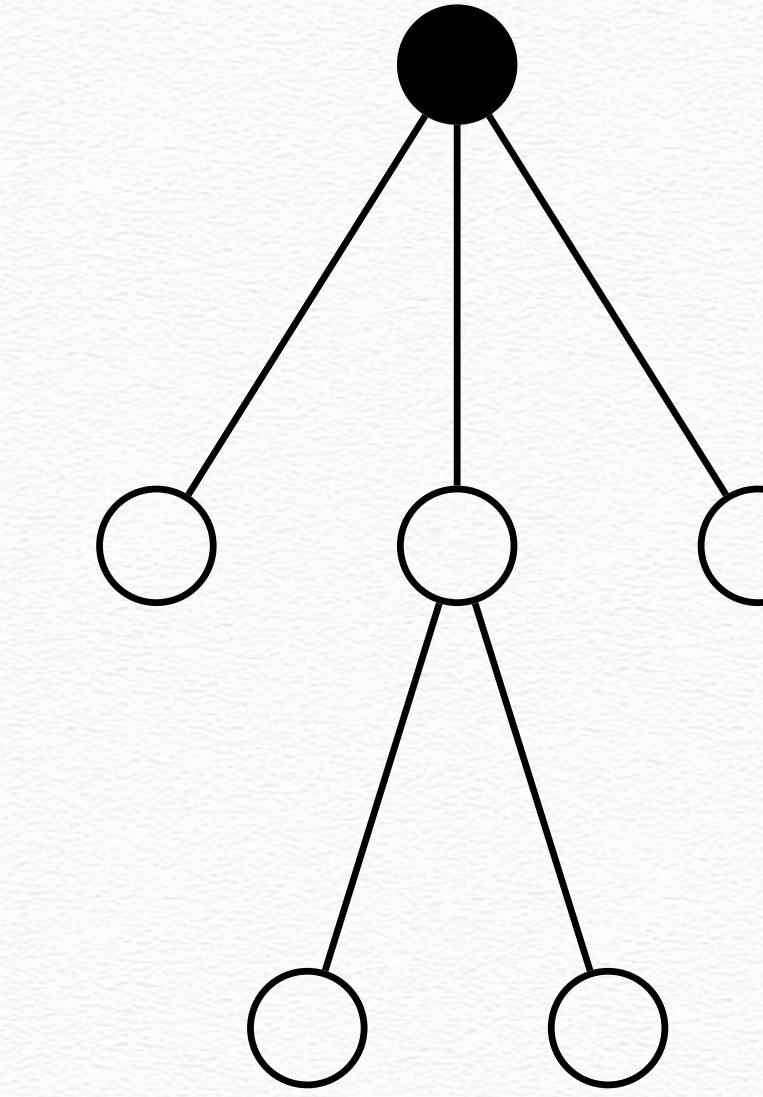


(a)
tree

Unrooted

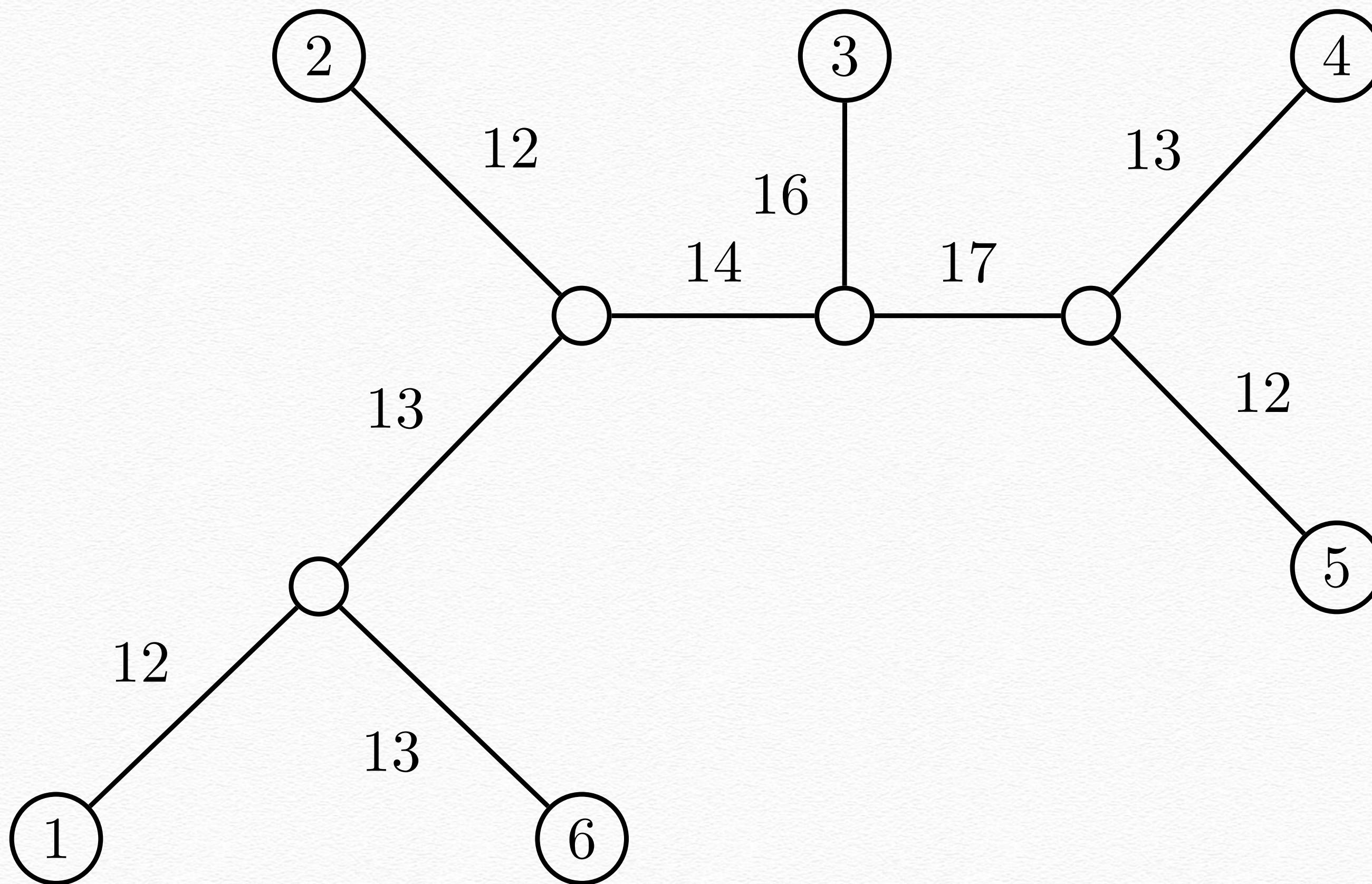


(b) Rooted tree

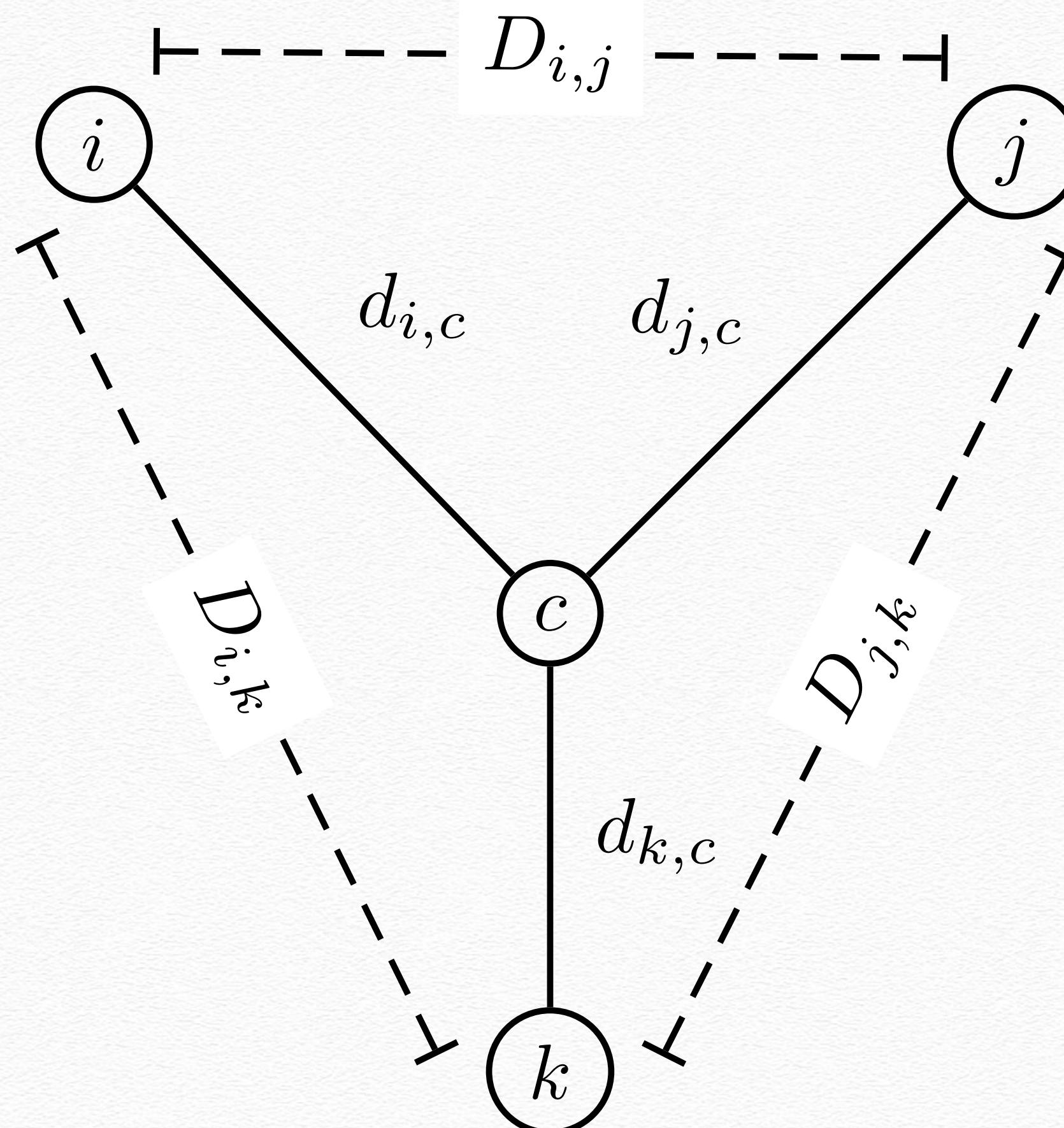


(c) The
same
rooted tree

Weighted Trees



Distance Based Tree Reconstruction



Distance Based Tree Reconstruction

$$d_{i,c} + d_{j,c} = D_{i,j}$$

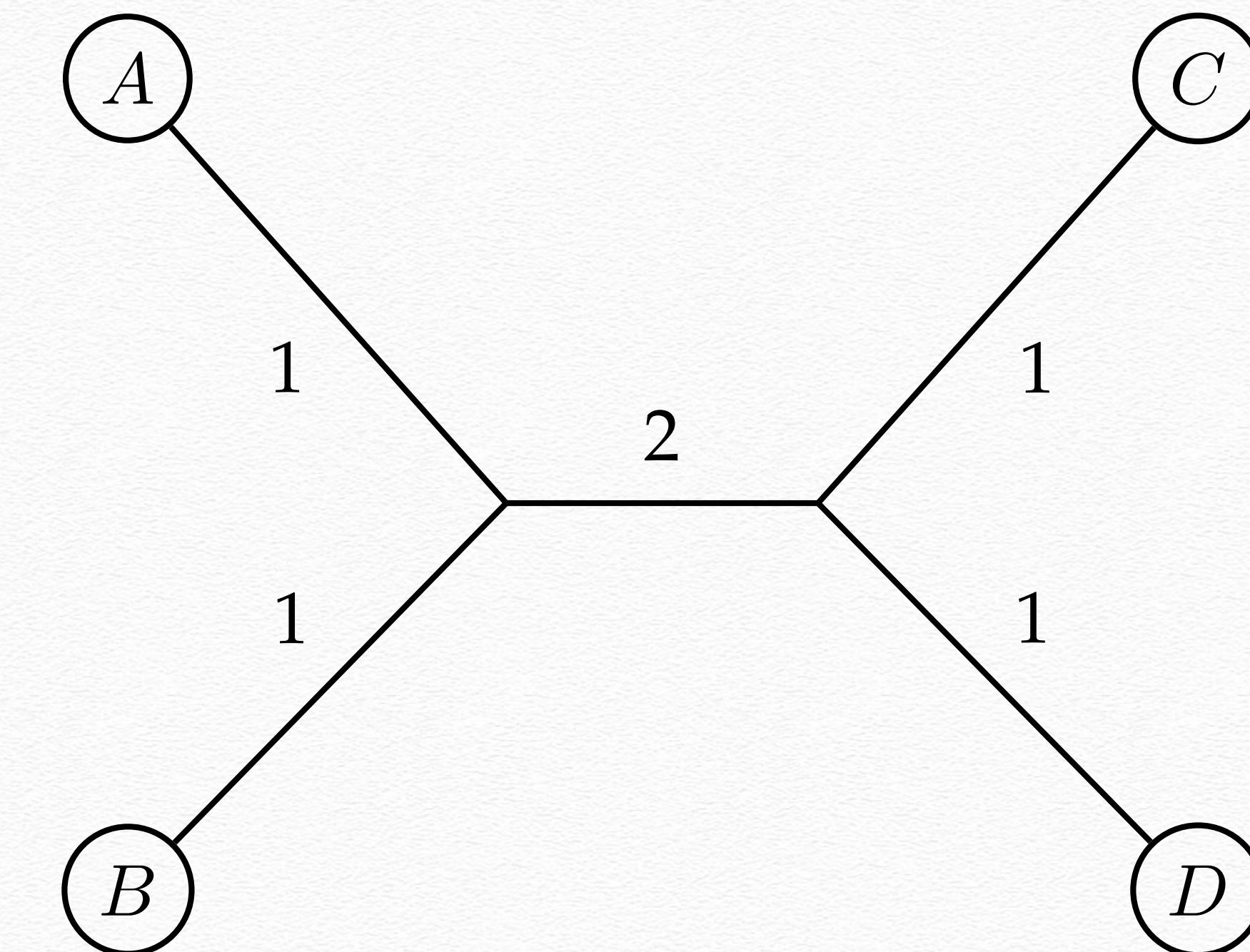
$$d_{i,c} + d_{k,c} = D_{i,k}$$

$$d_{j,c} + d_{k,c} = D_{j,k}.$$

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

Additive Matrix

	A	B	C	D
A	0	2	4	4
B	2	0	4	4
C	4	4	0	2
D	4	4	2	0



Non-Additive Matrix

	A	B	C	D
A	0	2	2	2
B	2	0	3	2
C	2	3	0	2
D	2	2	2	0

?

Distance Based Phylogeny

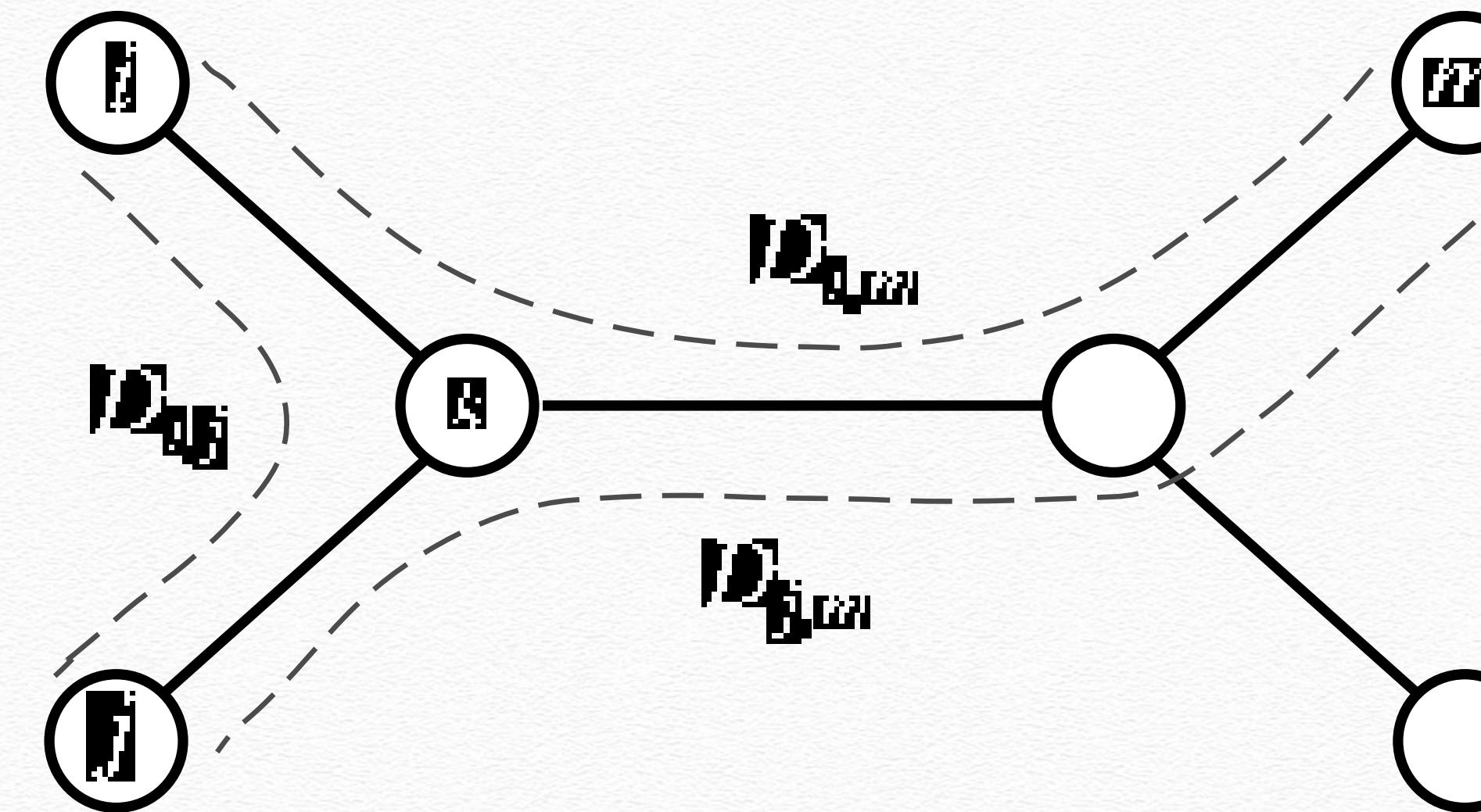
Distance-Based Phylogeny Problem:

Reconstruct an evolutionary tree from a distance matrix.

Input: An $n \times n$ distance matrix $(D_{i,j})$.

Output: A weighted unrooted tree T with n leaves fitting D , that is, a tree such that $d_{i,j}(T) = D_{i,j}$ for all $1 \leq i < j \leq n$ if $(D_{i,j})$ is additive.

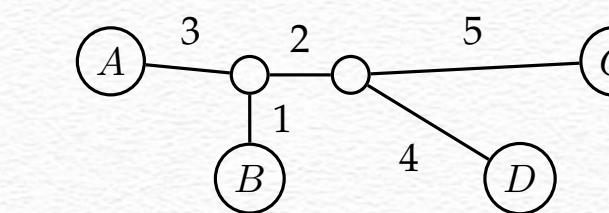
From Additive Matrix



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

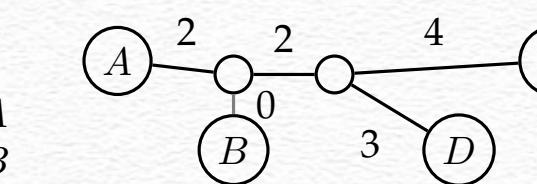
Strategy

	A	B	C	D
A	0	4	10	9
B	4	0	8	7
C	10	8	0	9
D	9	7	9	0



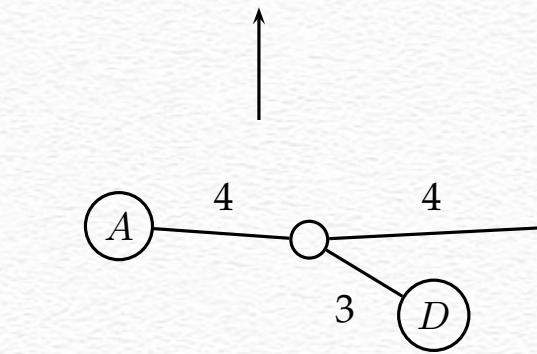
$\downarrow \delta = 1$

	A	B	C	D
A	0	2	8	7
B	2	0	6	5
C	8	6	0	7
D	7	5	7	0



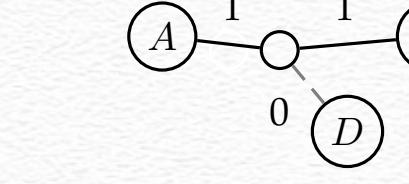
\downarrow

	A	C	D
A	0	8	7
C	8	0	7
D	7	7	0



$\downarrow \delta = 3$

	A	C	D
A	0	2	1
C	2	0	1
D	1	1	0



\downarrow

	A	C
A	0	2
C	2	0



Strategy

ADDITIVEPHYLOGENY(D)

```
1  if  $D$  is a  $2 \times 2$  matrix
2       $T \leftarrow$  the tree consisting of a single edge of length  $D_{1,2}$ .
3      return  $T$ 
4  if  $D$  is non-degenerate
5       $\delta \leftarrow$  trimming parameter of matrix  $D$ 
6      for all  $1 \leq i \neq j \leq n$ 
7           $D_{i,j} \leftarrow D_{i,j} - 2\delta$ 
8  else
9       $\delta \leftarrow 0$ 
10     Find a triple  $i, j, k$  in  $D$  such that  $D_{ij} + D_{jk} = D_{ik}$ 
11      $x \leftarrow D_{i,j}$ 
12     Remove  $j$ th row and  $j$ th column from  $D$ .
13      $T \leftarrow$  ADDITIVEPHYLOGENY( $D$ )
14     Add a new vertex  $v$  to  $T$  at distance  $x$  from  $i$  to  $k$ 
15     Add  $j$  back to  $T$  by creating an edge  $(v, j)$  of length 0
16     for every leaf  $l$  in  $T$ 
17         if distance from  $l$  to  $v$  in the tree  $T$  does not equal  $D_{l,j}$ 
18             output "Matrix  $D$  is not additive"
19             return
20     Extend hanging edges leading to all leaves by  $\delta$ 
21     return  $T$ 
```