UPGMA / WPGM

Motivation

- How close are different "things":

  - genes

  - family members

  - bacterias

- In general: how close some subjects are related

- UPGMA/WPGMA is just a clustering approach

Algorithm

- Every node is a single cluster

- Initialize the distiance dist(x,y) between every node x,y

- Repeat n - 1:

  - search minimal distance d_{min} between two clusters x, y

  - merge these two clusters to one cluster

  - define a new node, where e has distance d_{min} /2 to all leaves reachable from x and y

  - compute the distances to all other clusters for this new cluster

UPGMA

- New distance of cluster w and the new cluster x which was created of the two clusters y and z

$$dist(w, x) = \frac{dist(w, y) + dist(w, z)}{2}$$

WPGMA

- New distance of cluster w and the new cluster x which was created of the two clusters y and z

- Weighted by the size of the joined groups

$$dist(w, x) = \frac{|y| * dist(w, y) + |z| * dist(w, z)}{|y| * |z|}$$

Example

	A	B	C	D
A	0	1	2	3
B	1	0	2	3
C	2	2	0	3
D	3	3	3	0

$d_{min} = dist(a,b) = 1$

Example

	A	**B**	**C**	**D**
A	0	1	2	3
B	1	0	2	3
C	2	2	0	2
D	3	3	2	0

	(A B)	**C**	**D**
(A B)	0	2	3
C	2	0	2
D	3	2	0

dist(c, (a b)) = (2+2) / 2 = 2
dist(d, (a b)) = (3+3) / 2 = 3

0.5   0.5

A   B
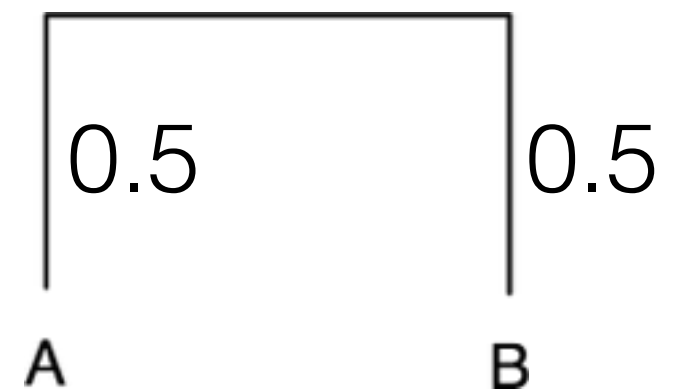
Example

	(A B)	C	D
(A B)	0	2	3
C	2	0	3
D	3	3	0

→

	(A B C)	D
(A B C)	0	3
D	3	0

$d_{min} = dist((a,b),c) = 2$

$dist(d, (a\ b\ c)) = (3+3) / 2 = 3$
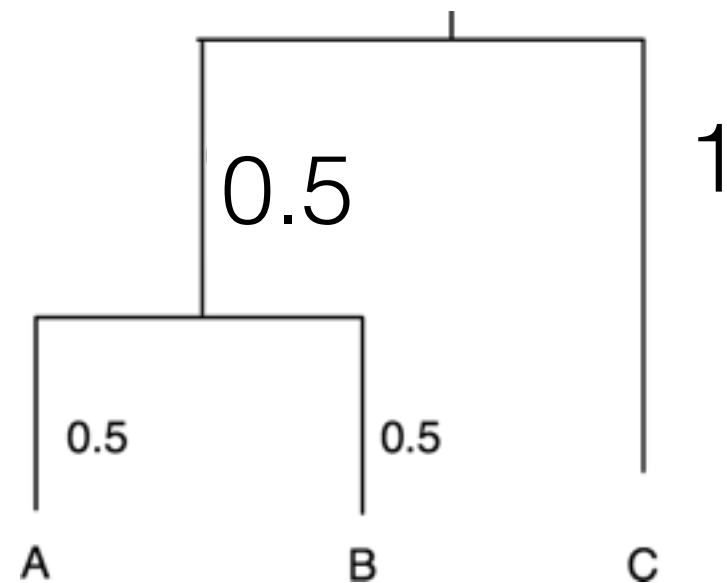
Example

	(A B C)	**D**
(A B C)	0	3
D	3	0

$d_{min} = dist((a,b,c),d) = 3$
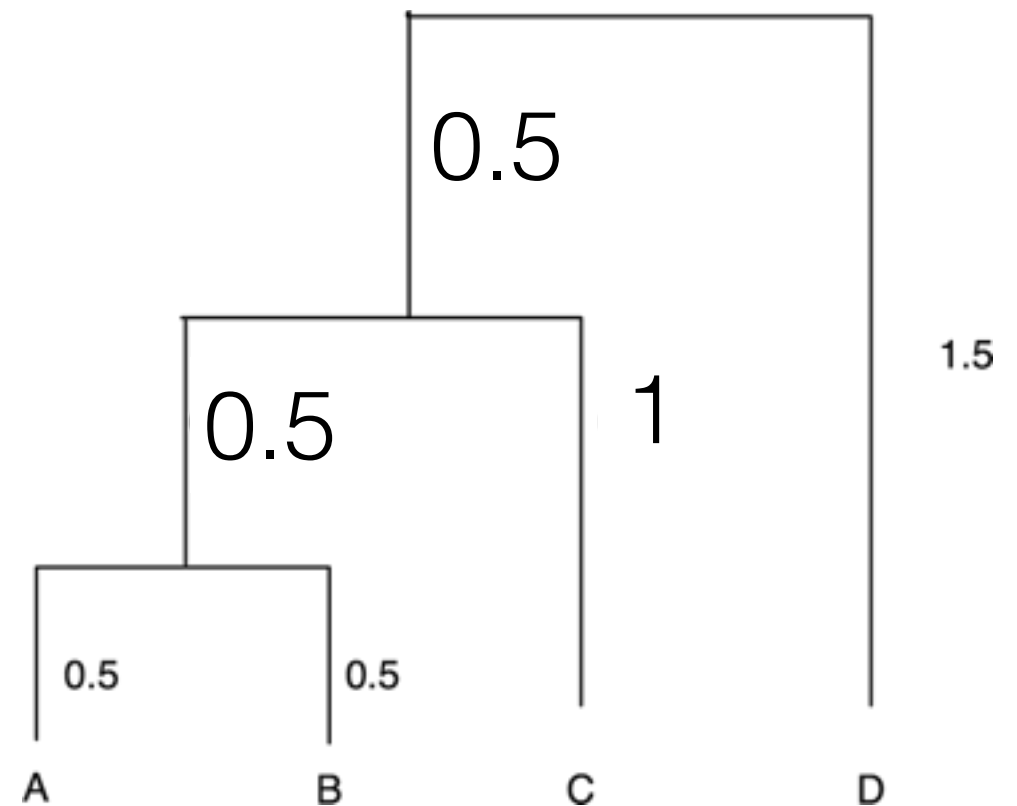
List of references

- Lecture "Phylogeny - Introduction and UPGMA" Bioinformatics I, Prof. Backofen, Chair of Bioinformatics, Albert-Ludwigs-Universität Freiburg

  URL: http://www.bioinf.uni-freiburg.de//Lehre/Courses/2014_SS/V_Bioinformatik_1/phylo-UPGMA.pdf, visited: 13/11/2014

- Exercise Sheet 10 - Phylogenetic Trees, Lecture Bioinformatics I SS 2014, Prof. Backofen, Chair of Bioinformatics, Albert-Ludwigs-Universität Freiburg

  http://www.bioinf.uni-freiburg.de/Lehre/Courses/2014_SS/U_Bioinformatik_1/ex10.pdf, visited: 13/11/2014