D_1 is additive but not ultrametric

$D_{ij} \equiv d_{ij} - (r_i + r_j)$ $\quad r_i \equiv \frac{[\sum_{k \in L} d_{ik}]}{|L|-2}$ $\forall i \in L$ $\quad d_{mk} = \frac{1}{2}[d_{ik} + d_{jk} - d_{ij}]$ $\forall k \in L,\ k \neq i,j$

① ↳ use neighbor joining for unrooted phylogenetic tree

d_{ij}	X_1	X_2	X_3	X_4
X_1	0	0.3	0.7	0.7
X_2		0	0.6	0.6
X_3			0	0.6
X_4				0

r_i	X_1	X_2	X_3	X_4
	0.85	0.75	0.95	0.95

D_{ij}	X_1	X_2	X_3	X_4
X_1	0	-1.3	-1.1	-1.1
X_2		0	-1.1	-1.1
X_3			0	-1.3
X_4				0

$d_{1m} = \frac{1}{2}(d_{12} + r_1 - r_2) = \frac{0.3+0.85-0.75}{2} = 0.2$

$d_{m2} = \frac{1}{2}(d_{12} + r_2 - r_1) = \frac{0.3+0.75-0.85}{2} = 0.1$

②

d_{ij}	a	X_3	X_4
a	0	0.5	0.5
X_3		0	0.6
X_4			0

$d_{a3} = \frac{1}{2}[d_{13} + d_{23} - d_{12}] = \frac{0.7+0.6-0.3}{2} = 0.5$

$d_{a4} = \frac{1}{2}[d_{14} + d_{24} - d_{12}] = \frac{0.7+0.6-0.3}{2} = 0.5$

r_i	a	X_3	X_4
	1	1.1	1.1

D_{ij}	a	X_3	X_4
a	0	-1.6	-1.6
X_3		0	-1.6
X_4			0

choose any

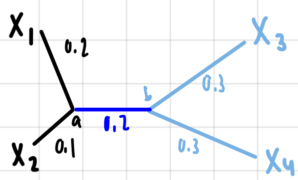
$r_3 = r_4 \rightarrow d_{3m} = d_{4m} = \frac{0.6+1.1-1.1}{2}$

$= 0.3$

③

d_{ij}	a	b
a	0	0.2
b		0

$d_{ba} = \frac{1}{2}(d_{3a} + d_{4a} - d_{34}) = 0.2$



D_2 is ultrametric and additive

① ultrametric → rooted phylogenetic tree

D_2	Y_1	Y_2	Y_3	Y_4	Y_5
Y_1	0	0.9	0.4	0.6	0.9
Y_2		0	0.9	0.9	0.4
Y_3			0	0.6	0.9
Y_4				0	0.9
Y_5					0

h = 0.2

②

D_2	Y_{13}	Y_2	Y_4	Y_5
Y_{13}	0	0.9	0.6	0.9
Y_2		0	0.9	0.4
Y_4			0	0.9
Y_5				0

$d(Y_{13}, Y_2) = \frac{0.9+0.9}{2} = 0.9$

$d(Y_{13}, Y_4) = \frac{0.6+0.6}{2} = 0.6$

$d(Y_{13}, Y_5) = \frac{0.9+0.9}{2} = 0.9$

h = 0.2

③

D_2	Y_{13}	Y_{25}	Y_4
Y_{13}	0	0.9	0.6
Y_{25}		0	0.9
Y_4			0

$d(Y_{25}, Y_{13}) = \frac{0.9+0.9}{2} = 0.9$

$d(Y_{25}, Y_4) = \frac{0.9+0.9}{2} = 0.9$

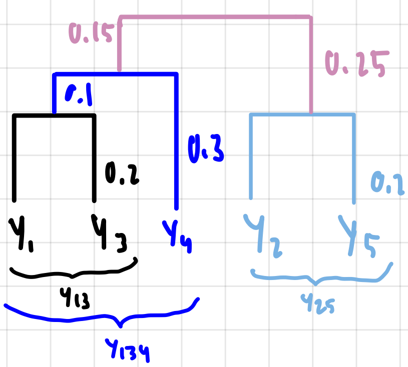
h = 0.3

④

D_2	Y_{134}	Y_{25}
Y_{134}	0	0.9
Y_{25}		0

$d(Y_{134}, Y_{25}) = \frac{2(0.9)+0.9}{2+1} = 0.9$

h = 0.45



① consider distances between pairs of clusters: pick smallest

② merge the clusters of least distance: set the height of merge point to be ½ that distance

③ recompute distance between new (merged) cluster and the others remaining

more efficient strategy:

in step ②, merged $C_p + C_q \rightarrow C_k$

$d(C_k, C_s) = \frac{d(C_p, C_s) \cdot |C_p| + d(C_q, C_s) \cdot |C_q|}{|C_p| + |C_q|}$