Step 2: Collapse joined taxa into new group and make collapsed new distance table $\rightsquigarrow$ identical to UPGMA

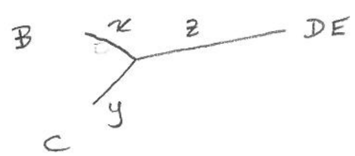
	DE	B	C
DE		1.75	1.3
B			1.1

$d(B,DE) = \frac{1}{2}(1.9+1.6) = \frac{3.5}{2} = 1.75$

$d(C,DE) = \frac{1}{2}(1.6+1.0) = 1.3$

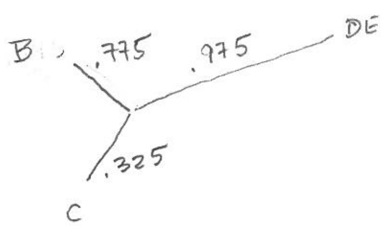
REPEAT ...

BC smallest ... only 3-groups $\Rightarrow$ use 3-point formula



$x = \frac{1}{2}\left(d(B,C) + d(B,DE) - d(C,DE)\right)$

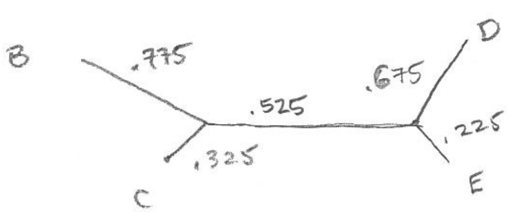
$= \frac{1}{2}\left[1.1 + 1.75 + 1.3\right] = \frac{1.55}{2} = .775$



$y = d(B,C) - .775 = 1.1 - .775 =$

$z = d(C,DE) - .325 = .975$

LAST STEP: Combine again using average



middle edge =

$.975 -$ (average dist. to DE)

$= .975 - .45$

FITCH - MARGOLIASH:

* not used in practice, but a means to ... NJ = neighbor joining

Same distance table used for UPGMA.

Step 1: Choose closest 2 taxa to join A,B

Different!

Collapse all other remaining taxa into group G $= \{C, D, E\}$

and form temporary distance table for A, B, G

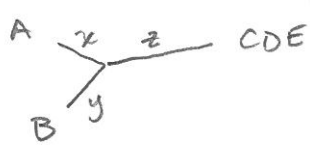
	B	CDE
A	.4	.6
B		1.8

$$d(A, CDE) = \frac{d(A,E) + d(A,D) + d(A,C)}{3}$$

$$= \frac{.5 + .6 + .7}{3} = \frac{1.8}{3} = .6$$

$$d(B, CDE) = \frac{1.1 + 1.9 + 1.6}{3} = \frac{4.6}{3} = 1.8$$

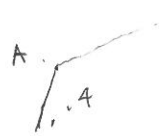
Use 3-point formula to fit chosen taxa (A,B) to tree



$$\frac{1}{2}\left(d(A,CDE) + d(A,B) - d(B,CDE)\right) = x$$

$$\frac{1}{2}\left[.6 + .4 - 1.8\right] < 0$$

Reset to zero!



Toss away temporary table

Step 2: Join AB into group for rest of analysis, Create new current distance table from original table just as in UPGMA.

REPEAT ...

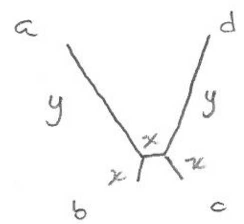
Notice: Fitch-Margoliash

1) constructs a non-ultrametric (also unrooted) tree which

2) has the same topology as the unrooted version of the UPGMA tree

since both methods use the same collapsed distance tables to choose which 2 taxa to join in the next step.

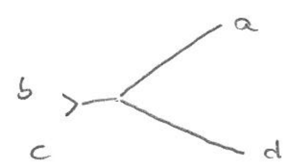
The "problem" with Fitch-Margoliash is evident in metric trees like the following:



metric tree with $y >> x$

$y > 2x$ sufficient

Both UPGMA, FM will choose to join b, c first, since they are closest leading to



the wrong tree.

These algorithms make a mistake in the very first step.

Defn: Two leaves of a tree are NEIGHBORS or form a CHERRY if the graph-theoretical distance between them is 2.

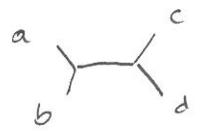
Eg. b, c are neighbors, a, d form a cherry.

To remedy the potential downfall of FM described above, notice

$$d(b,c) + d(a,d) > d(b,d) + d(a,c) = d(a,b) + d(c,d)$$

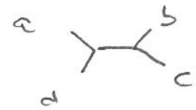
(assuming distances > 0)

Assume pairwise distances (dissimilarities) are positive. Then



T_1

T_2



T_3

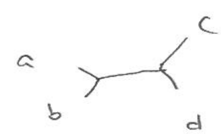
T_1: $d(a,b) + d(c,d) < d(a,c) + d(b,d) = d(a,d) + d(b,c)$

T_2: $d(a,c) + d(b,d) < d(a,b) + d(c,d) = d(a,d) + d(b,c)$

T_3: $d(a,d) + d(b,c) < d(a,b) + d(c,d) = d(a,c) + d(b,d)$

POINT: These inequalities and equalities can choose the tree in which (a,b) are neighbors.

Theorem: If T is a metric quartet tree with $X = \{a,b,c,d\}$ and all pairwise distances are positive, then (a,b) and (c,d) are neighbors



if, and only if,

$$d(a,b) + d(c,d) < d(a,c) + d(b,d) = d(a,d) + d(b,c).$$

How to use this to reconstruct trees?

Defn: Let $\delta : X \times X$ be a dissimilarity map on X. Then δ satisfies the 4-POINT CONDITION if for all $x,y,z,w \in X$

$$\delta(x,y) + \delta(z,w) \leq \max\{\delta(x,z) + \delta(y,w),\ \delta(x,w) + \delta(y,z)\}$$

i.e. 2 are equal and at least as big as the third.

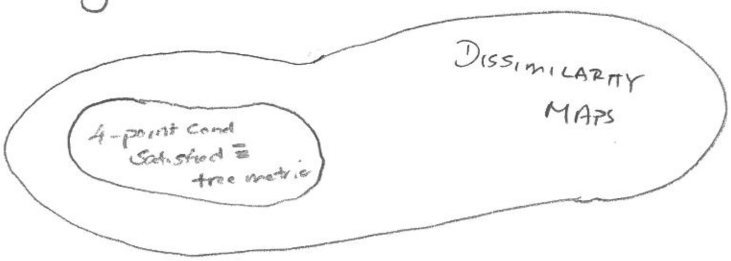
Note:

- used $\leq$ and not just $<$ This allows non-binary trees.
- the taxa names x, y, z, w are not necessarily distinct

Theorem: If (T, w) is a metric tree with positive edge weights, then its tree metric is a dissimilarity map satisfying the 4-point condition.

Remarkably, the converse is true

Theorem: If $|X| \geq 3$ and δ is a dissimilarity map with $\delta(x,y) \neq 0$ for $x \neq y$ and δ satisfies the 4-point condition, then there is a unique X-tree with positive edge weights whose tree metric agrees with δ

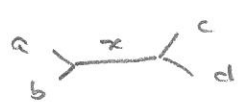


Proof (parts thereof). Induction on $n = |X|$.

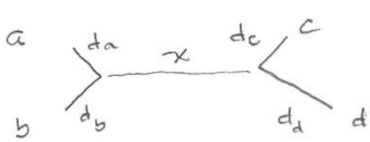
$n=3$: follows from 3-point formulas

$n=4$: Case 1: (Strict inequality) Suppose $X = \{a, b, c, d\}$ with

$d(a,b) + d(c,d) < d(a,c) + d(b,d) = d(a,d) + d(b,c)$

Then the topology is  with $x > 0$.

Use the 3-point formulas to find the distances to terminal branches

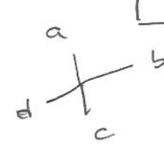


Solve for x via

$$x = \frac{d(a,c) + d(b,d) - \boxed{d(a,b) + d(c,d)}}{2}$$

$\boxed{\text{Check: . . . !}}$

Case 2: Equality: $d(a,b) + d(c,d) = d(a,c) + d(b,d)$
$= d(a,d) + d(b,c)$



Star tree.