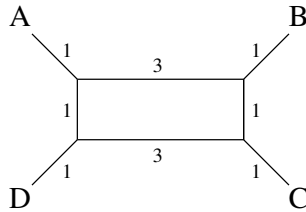


Definition 1 (Clade). A clade is a group of closely related organisms. As a clade directly involves a certain edge that separates the clade from the rest of the tree, it can also be represented by an X -split.

Ex. (FPC): Given a splitsgraph on $X = \{A, B, C, D\}$.



Let d_{AB} denote the length of the shortest path from A to B , then we get the following distance matrix d :

	A	B	C	D
A	0	5	6	3
B	5	0	3	6
C	6	3	0	5
D	3	6	5	0

Does the FPC (Four Point Condition) hold?

Presumably not, as there probably does not exist a tree that matches the distances.

Test of the FPC ($d_{ij} + d_{kl} \leq \max\{d_{ik} + d_{jl}, d_{il} + d_{jk}\} \forall i, j, k, l \in X$):

$$d_{AB} + d_{CD} = 10 \quad (1)$$

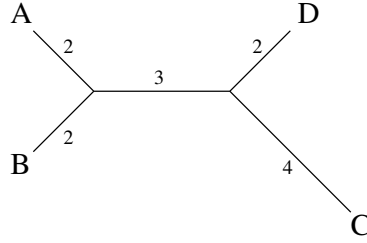
$$d_{AC} + d_{BD} = 12 \quad (2)$$

$$d_{AD} + d_{BC} = 6 \quad (3)$$

As $12 \not\leq \max\{10, 6\}$, the FPC does not hold.

Definition 2 (Weighted Splits). Given an X -Tree T with branch lengths. Assign a weight to each $S \in \Sigma(T)$, namely the branch length of the corresponding branch.

Ex. (Weighted Splits): T :



$\Sigma(T)$:

$$S_1 = \frac{\{A\}}{\{B, C, D\}}, \quad w(S_1) = 2 \quad (4)$$

$$S_2 = \frac{\{B\}}{\{A, C, D\}}, \quad w(S_2) = 2 \quad (5)$$

$$S_3 = \frac{\{C\}}{\{A, B, D\}}, \quad w(S_3) = 4 \quad (6)$$

$$S_4 = \frac{\{D\}}{\{A, B, C\}}, \quad w(S_4) = 2 \quad (7)$$

$$S_5 = \frac{\{A, B\}}{\{C, D\}}, \quad w(S_5) = 3 \quad (8)$$

Problem: Given a distance matrix D on X and a set Σ of compatible splits. We know that Σ corresponds to some X -tree T , i.e. $\Sigma = \Sigma(T)$. How to compute weights for Σ , i.e. branch lengths for T ?

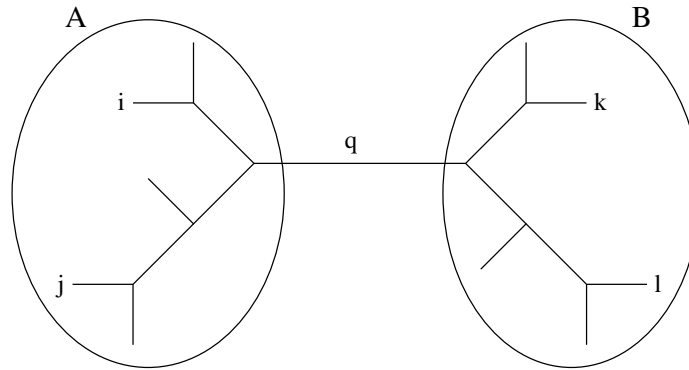
Answers: Given a distance matrix D on X and Σ .

1. Use the least squares method.
2. For any split $S \in \Sigma$ that "satisfies the FPC", we can use the FPC to assign a weight to S as follows.

Definition 3 ("satisfies the FPC"). Let D be a distance matrix on X . Let $S = \frac{A}{B}$ be an X -split. We say S "satisfies the FPC", if $D_{ij} + D_{kl} \leq \max\{D_{ik} + D_{jl}, D_{il} + D_{jk}\} \forall i, j \in A, k, l \in B$. We say that S "strictly satisfies the FPC", if ' $<$ ' always holds in the formula above.

Lemma 1. Let D be a distance matrix on X and let S, S' be two X -splits. If both S and S' strictly satisfy the FPC, then S and S' are compatible.

Let $S = \frac{A}{B}$ be an X -split that strictly satisfies the FPC. S corresponds to some edge q in an X -tree.



We can obtain the weight of S , i.e. the length of q as follows:

$$w(S) = \beta(S) = \frac{1}{2} \min_{\substack{i, j \in A \\ \underbrace{k, l \in B}_{\text{(look at all possible configurations)}}}} (\min(D_{ik} + D_{jl}, D_{il} + D_{jk}) - (D_{ij} + D_{kl})) \quad (9)$$

$\beta(S)$ is called *Buneman-index* of S .

Problem: Given a distance matrix D on X . Find all X -splits with positive *Buneman-index*.

Theorem 1. Given a distance matrix D on X . The set of all X -splits with positive *Buneman-index* is compatible.

Buneman tree algorithm (1971)

Input: A distance matrix D on $X = \{x_1, \dots, x_n\}$.

Output: A set Σ of all splits with positive *Buneman-index*.

```

Init:  $\Sigma_0 := \emptyset, X_0 := \emptyset$ 
for  $i = 1$  to  $n$  do
   $X_i := X_{i-1} \cup \{x_i\}; \Sigma_i := \emptyset$ 
  if  $\beta\left(\frac{\{x_i\}}{X_{i-1}}\right) > 0$ , then add the split to  $\Sigma_i$ 
  for all  $S = \frac{A}{B} \in \Sigma_{i-1}$  do:
    if  $\beta\left(\frac{A \cup \{x_i\}}{B}\right) > 0$ , then add the split to  $\Sigma_i$ 
    else if  $\beta\left(\frac{A}{B \cup \{x_i\}}\right) > 0$ , then add the split to  $\Sigma_i$ 
return  $\Sigma_n$ 

```

Theorem 1 (restated). *Given a distance matrix D on X . The set of all X -splits generated by the Buneman tree algorithm is compatible.*

Proof (by contradiction): Assume the algorithm generates $S = \frac{A}{B}$ and $S' = \frac{A'}{B'}$ that are not compatible, then there exist $i, j, k, l \in X$ with $i \in A \cap A', j \in A \cap B', k \in B \cap A', l \in B \cap B'$.

Now $\beta(S) > 0$ implies $\underbrace{D_{ij}}_{\in A} + \underbrace{D_{kl}}_{\in B} < \min\{D_{ik} + D_{jl}, D_{il} + D_{jk}\},$

and $\beta(S') > 0$ implies $\underbrace{D_{ik}}_{\in A'} + \underbrace{D_{jl}}_{\in B'} < \min\{D_{ij} + D_{kl}, D_{il} + D_{jk}\},$

a contradiction!

□

Lemma 2. *Let D be a distance matrix on $X = \{a, b, c, d\}$. If D comes from a tree T , i.e. $D = D_T$, then of the three possible "non-trivial" splits: $\frac{\{a,b\}}{\{c,d\}}, \frac{\{a,c\}}{\{b,d\}}, \frac{\{a,d\}}{\{b,c\}}$ at most **one** can have positive Buneman-index.*

Problem: Can we define an "index" that allows at most **two** of the three "non-trivial" splits to have that "index" being positive?

Answer: *Bandelt & Dress (1992) introduced the isolation-index α .*

$$\alpha(S) = \frac{1}{2} \min_{\substack{i,j \in A \\ k,l \in B}} \left(\max(D_{ik} + D_{jl}, D_{il} + D_{jk}) - (D_{ij} + D_{kl}) \right) \quad (10)$$

(the only difference to the formula for the β -index)

The isolation-index α gives rise to non-compatible split sets.