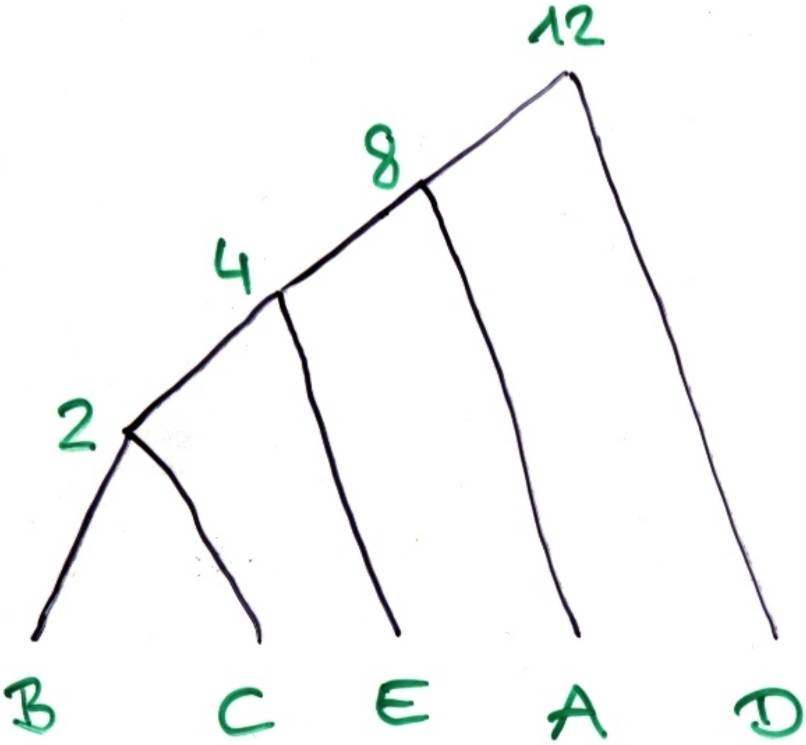
| A | B | C | D | E | |
|---|---|---|---|---|---|
| 0 | ~~8~~ 10 | 8 | ~~12~~ 10 | 8 | A |
| | 0 | 2 | 12 | ~~4~~ 6 | B |
| | | 0 | 12 | 4 | C |
| | | | 0 | 12 | D |
| | | | | 0 | E |



# Agglomeratives Clustern

WHILE mehr als ein Cluster in $\mathcal{C}$ DO

Finde Clusterpaar $C_i, C_j \in \mathcal{C}$ mit minimaler Distanz

Neuer Cluster $C_u = C_i \cup C_j$

Im Baum neuer Knoten $u$

- neue Kanten $(u,i)$ und $(u,j)$
- Datierung $t(u) = D(i,j)$

Bestimme Distanzen $D(u,k)$ für alle

$k \neq i,j$ $\longrightarrow$ single/complete Linkage, UPGMA, WPGMA

Lösche $C_i, C_j$ aus $\mathcal{C}$, füge $C_u$ ein

END WHILE

# Neighbor Joining (Saitou + Nei, 1987)

$X = \{1, 2, \dots, n\}$

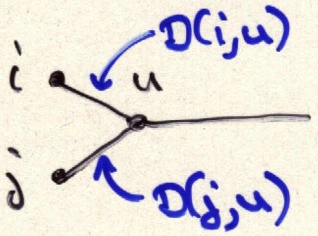
1. Bestimme $r_i = \sum_{j=1}^{n} D(i,j)$ für alle $i$

2. Bestimme $M_{ij} = D(i,j) - \frac{1}{n-2}(r_i + r_j)$

   für alle $i,j$

3. Finde minimales $M_{ij} \to i$ & $j$ Geschwister

4. Neues Objekt $u$




5. Setze

   $$D(i,u) = \frac{1}{2} D(i,j) + \frac{1}{2(n-2)}(r_i - r_j)$$

   $$D(j,u) = D(i,j) - D(i,u)$$

6. Setze

   $$D(u,k) = \frac{1}{2}\left(D(i,k) + D(j,k) - D(i,j)\right)$$

   für alle $k \neq i,j \to$ neue Distanzmatrix

7. Lösche $i,j$ aus Objekten, füge $u = (ij)$ hinzu

8. Falls noch 3 oder mehr Objekte $\to$ Schritt 1

9. Füge Kante $\{u,v\}$ zwischen den letzten beiden Obj. $u,v$ mit Länge $D(u,v)$ ein

# Voraussetzungen für die Anwendung von Neighbor Joining

- $D(i,i) = 0$ für alle $i$
- $D(i,j) = D(j,i)$ für alle $i,j$
  → sonst setze $D^*(i,j) = \frac{1}{2}\left(D(i,j) + D(j,i)\right)$
- $D$ muss <u>keine</u> Metrik sein
  → es können dabei aber Kanten mit negativem Gewicht entstehen
  → passiert bei biologischen Daten eher nicht
  → falls doch: post-processing, bspw. Kante zusammenziehen