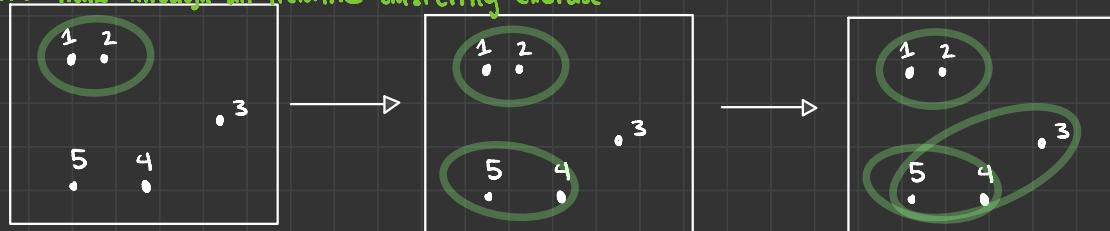


CHAPTER 3 ~ PHYLOGENIES

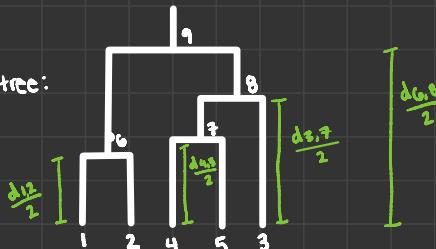
distance based methods

- sequence alignment could be interpreted as distance b/wn sequences
- Jukes-Cantor distance ~ another method that's explored on HW
- OTU: operational taxonomic units
- UPGMA alg ~ "unweighted pair group method using arithmetic averages"
 - a clustering method (for sequences)
- input: n points ~ x_1, x_2, \dots, x_n (or difference)
- Similarity matrix ~ S_{ij} : similarity between points x_i and x_j
- axioms of clustering: create groupings such that ...
 - points in the same cluster are highly similar (most clustering methods address this)
 - points in different clusters are very different

Let's walk through an iterative clustering exercise:

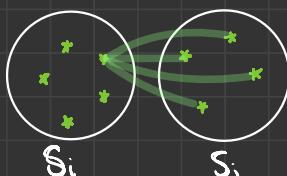


corresponding phylogenetic tree:



- How do we calculate the distance between clusters?

let S_i, S_j be 2 clusters of sequences. The distance d_{ij} between S_i, S_j is defined as follows:



$$d_{ij} = \frac{1}{|S_i| \cdot |S_j|} \sum_{\substack{\text{all pairs} \\ i \in S_i, j \in S_j}} d_{i,j} = \text{average distance between all pairs of sequences from each cluster}$$

of lines = 5*4

- Universal biological clock: we are assuming all species evolve at the same rate (not really accurate)

AI Overview
An operational taxonomic unit (OTU) is a group of closely related individuals that are grouped together based on the similarity of specific sequences. OTUs are a basic unit used in numerical taxonomy, and can refer to an individual, species, genus, or class.