## **UPGMA**

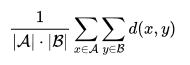
**UPGMA** (**unweighted pair group method with arithmetic mean**) is a simple agglomerative (bottom-up) [hierarchical clustering](https://en.wikipedia.org/wiki/Hierarchical_clustering) method. The method is generally attributed to Sokal and Michener.

The UPGMA method is similar to its weighted variant, the [WPGMA](https://en.wikipedia.org/wiki/WPGMA) method.

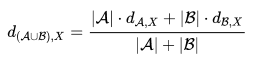
Note that the unweighted term indicates that all distances contribute equally to each average that is computed and does not refer to the math by which it is achieved. Thus the simple averaging in WPGMA produces a weighted result and the proportional averaging in UPGMA produces an unweighted result.

## **Algorithm**

The UPGMA algorithm constructs a rooted tree ([dendrogram](https://en.wikipedia.org/wiki/Dendrogram)) that reflects the structure present in a pairwise [similarity matrix](https://en.wikipedia.org/wiki/Similarity_matrix) (or a [dissimilarity matrix](https://en.wikipedia.org/wiki/Distance_matrix)). At each step, the nearest two clusters are combined into a higher-level cluster. The distance between any two clusters A and B , each of size (i.e., [cardinality](https://en.wikipedia.org/wiki/Cardinality)) |A | and |B|,is taken to be the average of all distance d(x,y) between pairs of objects x in A and y in B, that is, the mean distance between elements of each cluster:



In other words, at each clustering step, the updated distance between the joined clusters A U B and and a new cluster X is given by the propportional averaging of the dA,X and dB,X distances:



The UPGMA algorithm produces rooted dendrograms and requires a constant-rate assumption - that is, it assumes an [ultrametric](https://en.wikipedia.org/wiki/Ultrametric) tree in which the distances from the root to every branch tip are equal. When the tips are molecular data (i.e., [DNA](https://en.wikipedia.org/wiki/DNA), [RNA](https://en.wikipedia.org/wiki/RNA) and [protein](https://en.wikipedia.org/wiki/Protein)), the [ultrametricity](https://en.wikipedia.org/wiki/Ultrametricity) assumption is called the [molecular clock](https://en.wikipedia.org/wiki/Molecular_clock).