Hierarchical Methods

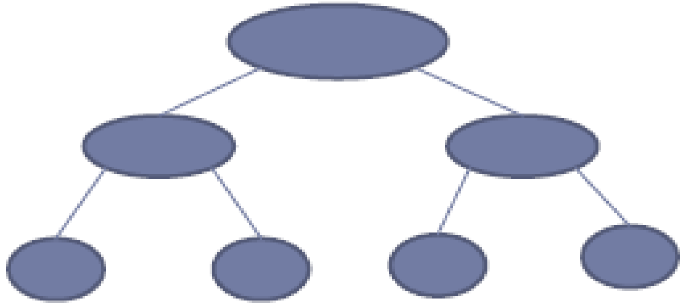
P.Mirunalini

Hierarchical Clustering

- Works by grouping data objects at different levels as hierarchy or “trees” of clusters
- Used in data summarization and visualization
- Eg: organize employees into major groups such as executives, managers and staff. Further divided into subgroups as seniors officers, officers and trainees.
- All these groups form a hierarchy.

Hierarchical Clustering

- Build a tree-based hierarchical taxonomy (*dendrogram*) from a set of documents.



Data points are represented as nodes of a graph

Edges linking the nodes

Each cluster are represented as sub graph.

Hierarchical Clustering Methods

- Hierarchical clustering methods can be further classified as
  - **Agglomerative (Bottom-up)**
  - **Divisive (Top-Down)**
- Needs a termination condition
- Does not require the number of clusters ***k*** in advance

Agglomerative hierarchical clustering

- This bottom-up strategy starts by placing each object form its own cluster
- Merges these atomic clusters into larger and larger clusters
- Until all of the objects are in a single cluster or until certain termination conditions are satisfied.
- The single cluster become's hierarchy's root.
- Merges two cluster that are closest to each other (based on similarity measure)
- Two clusters are merged per iteration each cluster contains one object requires at most n iterations.
- uses Single linkage Approach

Divisive Hierarchical Clustering

- This is top-down strategy does the reverse of agglomerative hierarchical clustering by starting with all objects in one cluster.
- It subdivides the clusters into smaller and smaller pieces
- Partitioning continues until each cluster is coherent enough
- Termination conditions :desired number of cluster or the diameter of each cluster is within a certain threshold.
- Challenge:
  - How to partition the large cluster into smaller ones.
  - Since $2^{n-1}-1$ possible ways to partition a set of n objects into two exclusive sets.
  - When “n” is large computationally prohibitive to examine so heuristics methods are used leads to errors

Agglomerative Vs Divisive Hierarchical clustering

- **AGNES Agglomerative NESting:** Places each object into cluster of its own.

- Clusters are merged step-by-step according to some criterion.

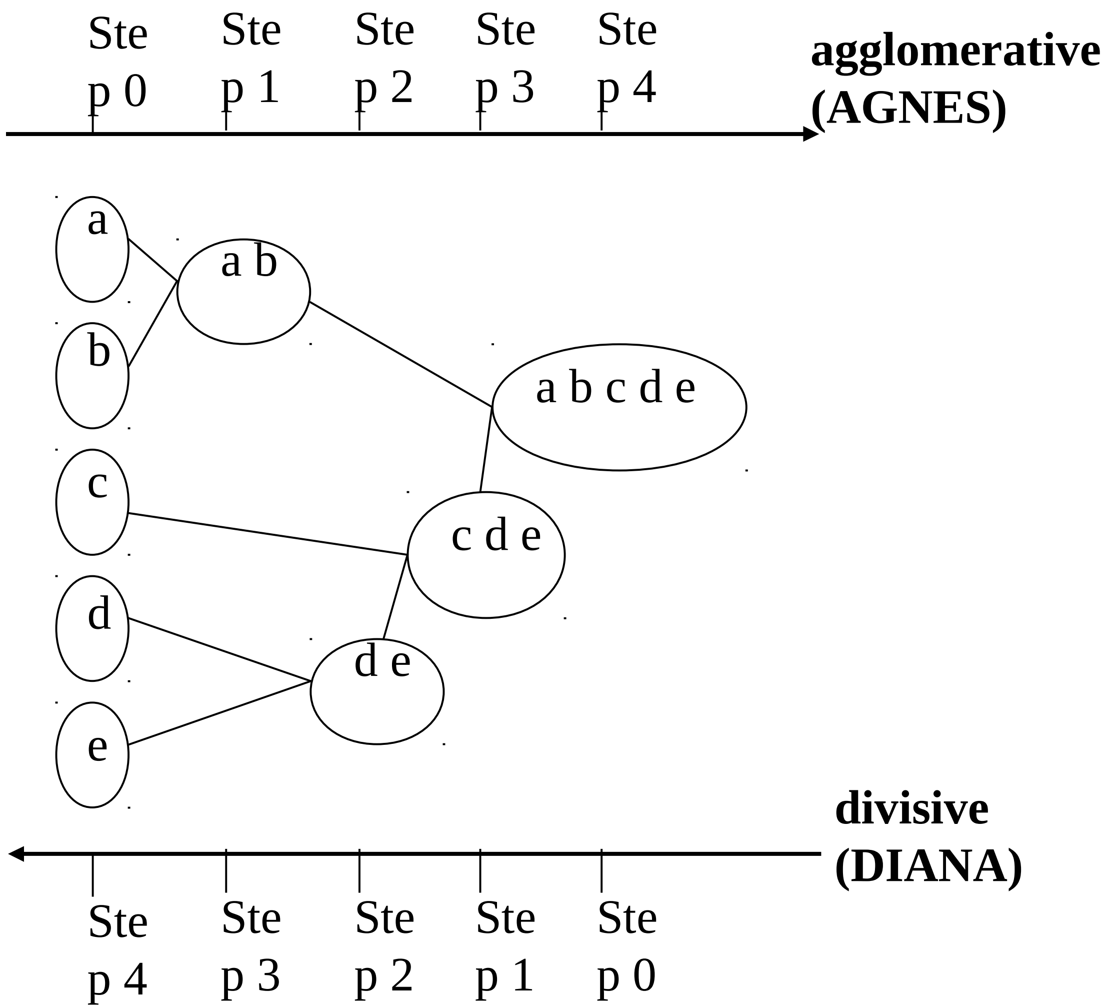
- Clusters C1 AND C2 are merged if the object in C1 and C2 form the minimum Euclidean distance between any two objects from different clusters.

- Uses single linkage approach:

Agglomerative hierarchical clustering

- **Single linkage Approach:**
  - Each cluster is represented by all objects in the cluster
  - Similarity between objects are measured by the similarity of closet pair of data points different clusters

Agglomerative Vs Divisive Hierarchical clustering

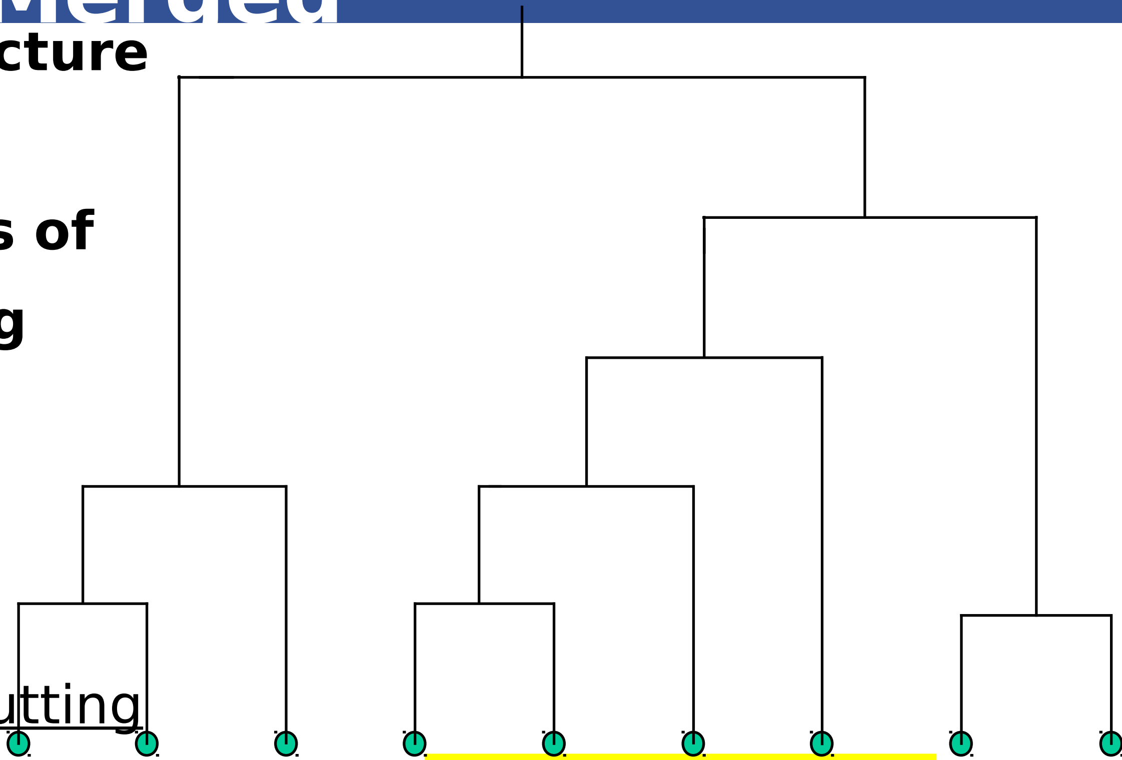


Agglomerative Vs Divisive Hierarchical clustering

- **DIANA(Divisive ANAlysis)**, the divisive method, proceeds in the contrasting way.
- All the objects are used to form one initial cluster.
- The cluster is split according to some principle such as the **maximum Euclidean distance** between the closest neighboring objects in the cluster.
- The cluster-splitting process repeats until, eventually, each new cluster contains only a single object.

Dendrogram: Shows How Clusters are Merged

- Dendrogram: **Tree structure commonly used to represent the process of hierarchical clustering**
- Shows how objects are grouped or partitioned
- A clustering of the data objects is obtained by cutting the dendrogram at the desired level, then each connected component forms a cluster



Hierarchical clustering generates a dendrogram (a hierarchy of clusters)

Distance Measures in Algorithmic methods

- Four widely used measures: let (p,p') be the points, mi is the mean for cluster Ci and ni is the number of objects in Ci.
- Measures are called as linkage measures

Minimum distance: $$dist_{min}(C_i, C_j) = \min_{p \in C_i, p' \in C_j} \{|p - p'|\}$$

Maximum distance: $$dist_{max}(C_i, C_j) = \max_{p \in C_i, p' \in C_j} \{|p - p'|\}$$

Mean distance: $$dist_{mean}(C_i, C_j) = |m_i - m_j|$$

Average distance: $$dist_{avg}(C_i, C_j) = \frac{1}{n_i n_j} \sum_{p \in C_i, p' \in C_j} |p - p'|$$

Hierarchical clustering

- Input: a pairwise matrix involved all instances in S

- Algorithm

1. Place each instance of S in its own cluster (singleton), creating the list of clusters L (initially, the leaves of T):
   $L= S_1, S_2, S_3, ..., S_{n-1}, S_n$.

2. Compute a **merging cost function** between every pair of elements in L to find the two closest clusters $\{S_i, S_j\}$ which will be the cheapest couple to merge.

1. Remove S_i and S_j from L.

1. Merge S_i and S_j to create a new internal node S_{ij} in T which will be the parent of S_i and S_j in the resulting tree.

1. Go to **Step 2** until there is only one set remaining.

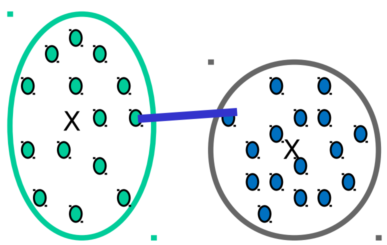
Hierarchical clustering

- **Step 2** can be done in different ways,
- **single-linkage clustering** (also called the connectedness or minimum method):
  - considers the distance between one cluster and another cluster to be equal to the **shortest** distance from any member of one cluster to any member of the other cluster.
- **Complete-linkage clustering** (also called the diameter or maximum method):
  - considers the distance between one cluster and another cluster to be equal to the **greatest** distance from any member of one cluster to any member of the other cluster.
- **Average-linkage clustering:**
  - Considers the distance between one cluster and another cluster to be equal to the **average** distance from any member of one cluster to any member of the other cluster.

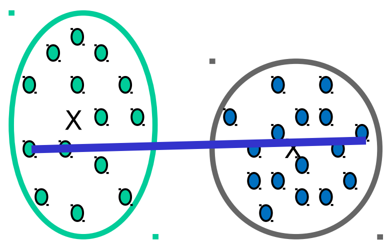
Single Link vs. Complete Link in Hierarchical Clustering

Nearest-Neighbouring Clustering Algorithm:

- Uses minimum distance dmin to measure the distance between clusters.
- IF the clustering process is terminated when the distance between nearest clusters exceeds a user defined threshold and is called **single-linkage algorithm**



- – Local similarity-based: Emphasizing more on close regions, ignoring the overall structure of the cluster
- – Capable of clustering non-elliptical shaped group of objects
- – Sensitive to noise and outliers

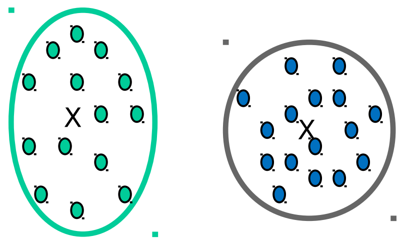


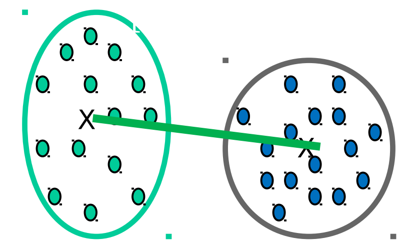
Single Link vs. Complete Link in Hierarchical Clustering

- **Farthest Neighbor clustering Algorithm:**
  - Algorithm uses the maximum distance dmax(Ci,Cj) to measure the distance between clusters.
  - When the clustering process is terminated when the maximum distance between nearest clusters exceeds a user-defined threshold, it is called **complete linkage algorithm**
  - The distance between two clusters is determined by the most distant nodes in the two clusters.
  - Nonlocal in behavior, obtaining compact shaped clusters
  - Sensitive to outliers

Agglomerative Clustering: Average vs. Centroid Links

- Agglomerative clustering with **average link**
  - **Average link**: The average distance between an element in one cluster and an element in the other (i.e., all pairs in two clusters)
    - Expensive to compute



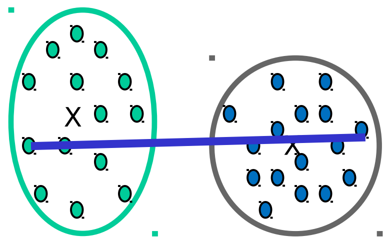


- Agglomerative clustering with **centroid link**
  - **Centroid link**: The distance between the centroids of two clusters

$$c_{a \cup b} = \frac{N_a c_a + N_b c_b}{N_a + N_b}$$

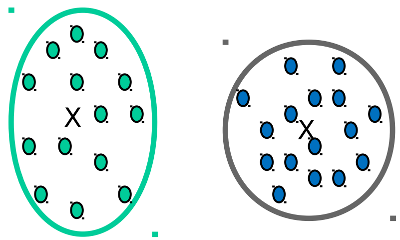
Single Link vs. Complete Link in Hierarchical Clustering

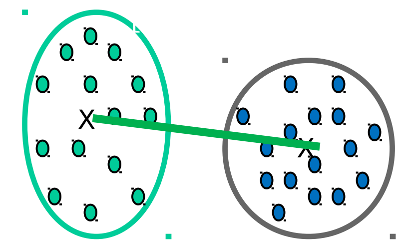
- Complete link (diameter)
  - The similarity between two clusters is the similarity between their most dissimilar members
  - Merge two clusters to form one with the smallest diameter
  - Nonlocal in behavior, obtaining compact shaped clusters
  - Sensitive to outliers



Agglomerative Clustering: Average vs. Centroid Links

- Agglomerative clustering with **average link**
  - **Average link**: The average distance between an element in one cluster and an element in the other (i.e., all pairs in two clusters)
    - Expensive to compute



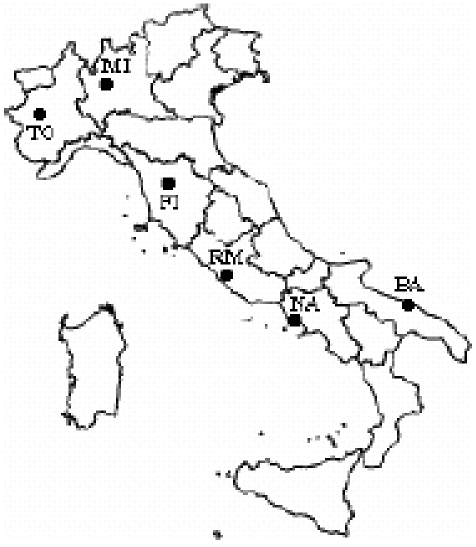


- Agglomerative clustering with **centroid link**
  - **Centroid link**: The distance between the centroids of two clusters

$$c_{a \cup b} = \frac{N_a c_a + N_b c_b}{N_a + N_b}$$

Hierarchical clustering: example

	BA	FI	MI	NA	RM	TO
BA	0	662	877	255	412	996
FI	662	0	295	468	268	400
MI	877	295	0	754	564	138
NA	255	468	754	0	219	869
RM	412	268	564	219	0	669
TO	996	400	138	869	669	0



Hierarchical Agglomerative clustering: example

	BA	FI	MI/TO	NA	RM
BA	0	662	877	255	412
FI	662	0	295	468	268
MI/TO	877	295	0	754	564
NA	255	468	754	0	219
RM	412	268	564	219	0

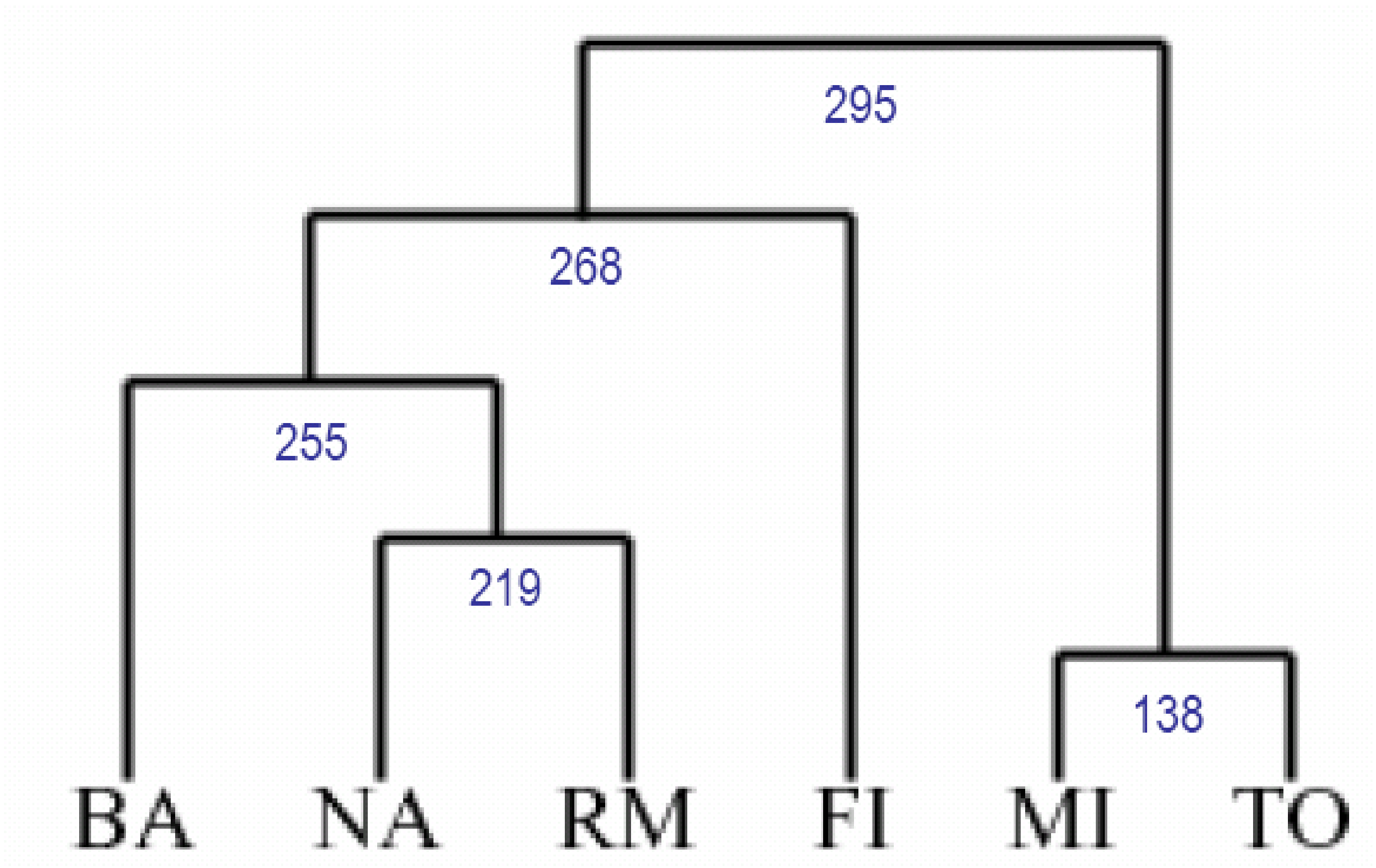
	BA	FI	MI/TO	NA/RM
BA	0	662	877	255
FI	662	0	295	268
MI/TO	877	295	0	564
NA/RM	255	268	564	0

Hierarchical Agglomerative clustering: example

	BA/NA/RM	FI	MI/TO
BA/NA/RM	0	268	564
FI	268	0	295
MI/TO	564	295	0

	BA/FI/NA/RM	MI/TO
BA/FI/NA/RM	0	295
MI/TO	295	0

Hierarchical clustering: example using single linkage



Challenges and Solutions

- It is **difficult to select merge or split points**
- No **backtracking**
- Hierarchical clustering **does not scale well: examines a good** number of objects before any decision of split or merge
- One promising directions to solve these problems is to combine hierarchical clustering with other clustering techniques: **multiple phase clustering**