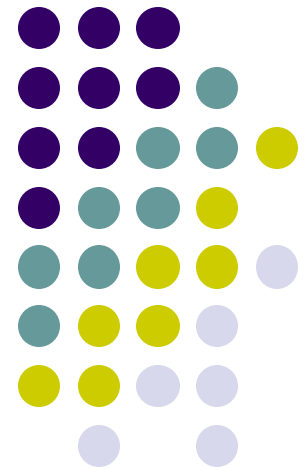


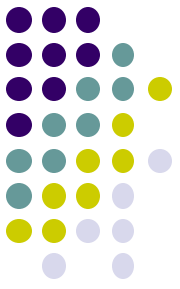
Hierarchical Clustering

Dr. R. Bhargavi

School of Computing Science & Engineering
VIT University, Chennai

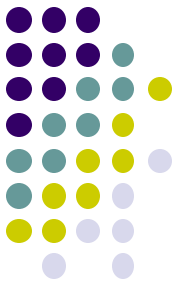


Introduction - Hierarchical Clustering

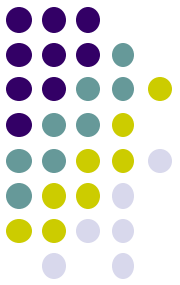


- Hierarchical Clustering is an unsupervised algorithm.
- K-means clustering requires us to pre-specify the number of clusters (K)
- In most of the real world applications it is difficult to pre-specify the value of K
- Hierarchical clustering does not require to commit for a particular K

Hierarchical Clustering - Approaches



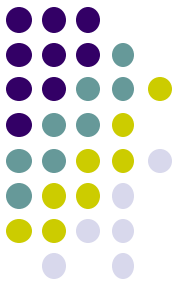
- Strategies for hierarchical clustering generally fall into two types
 - **Agglomerative** or **Bottom-up** clustering
 - **Divisive** or **Top-down** clustering



Agglomerative Clustering

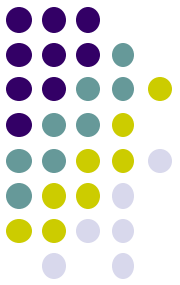
- Results in a Dendrogram – a tree (inverted) representation of the observations
- Dendrogram is built starting from the leaves and combining clusters up to the trunk.
- One single dendrogram can be used to obtain any number of clusters.

Agglomerative clustering – Working



- Start with each observation as an individual cluster
- Combine/merge two similar clusters based on some dissimilarity measure.
- Repeat till all the observations are merged into a single cluster

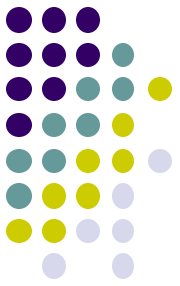
Similarity measure



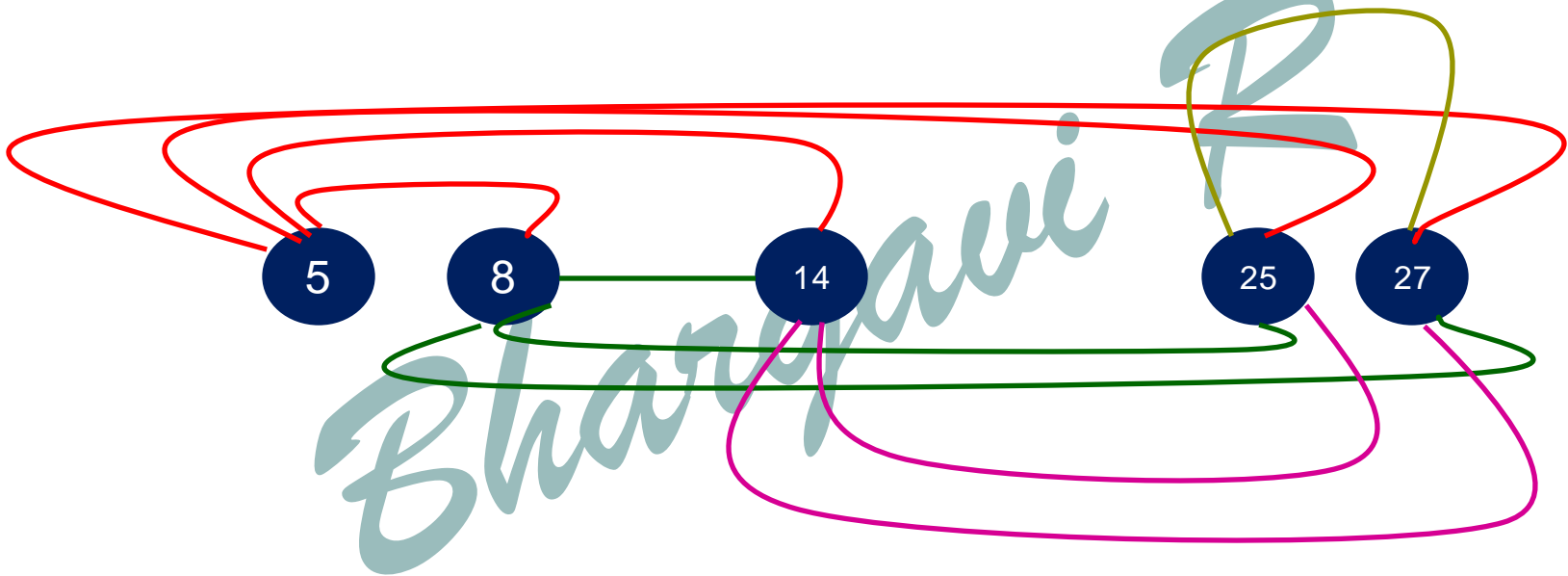
- Similarity measure between points – Euclidian distance, Manhattan distance, correlation etc.
- Similarity measure between clusters – **Linkage**

Linkage	Description
Complete	Maximal intercluster dissimilarity.
Single	Minimal intercluster dissimilarity.
Average	Mean intercluster dissimilarity.
Centroid	Dissimilarity between the centroids

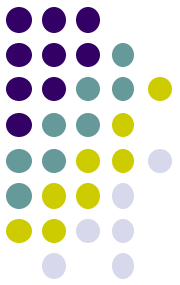
Example



- Data : 5, 8, 14, 25, 27
- Compute similarity among initial clusters



Example (cont...)

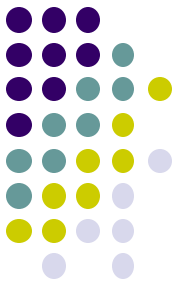


Cluster 1	Cluster 2	Distance
{5}	{8}	3
{5}	{14}	9
{5}	{25}	20
{5}	{27}	22
{8}	{14}	6
{8}	{25}	17
{8}	{27}	19
{14}	{25}	11
{14}	{27}	13
{25}	{27}	2

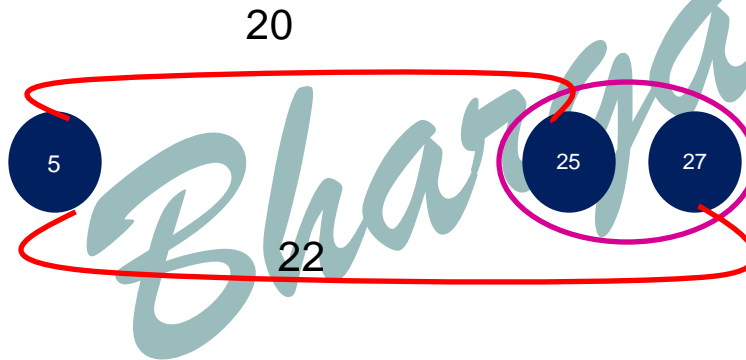
Min.
dissimilarity



Example (cont...)

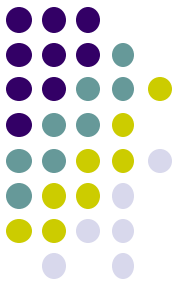


- Compute dissimilarity between clusters with two or more elements with complete linkage
- First compute distances from each element of one cluster to every element of the other cluster.



- Now take $\text{Max}(20, 22) = 22$

Example (cont...)



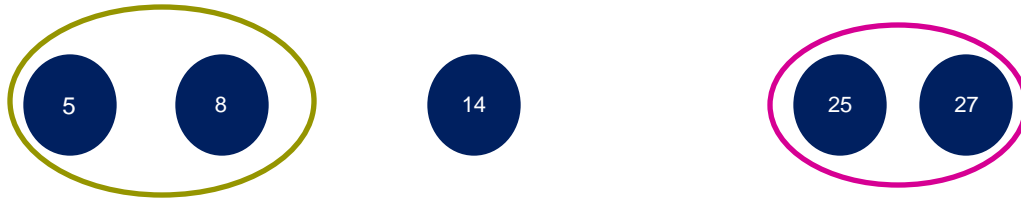
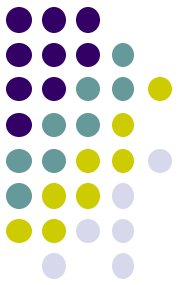
- Step 2

Cluster1	Cluster 2	Distance
{5}	{25, 27}	22
{8}	{25, 27}	19
{14}	{25, 27}	13
{5}	{8}	3
{5}	{14}	9
{8}	{14}	6

Min.
dissimilarity



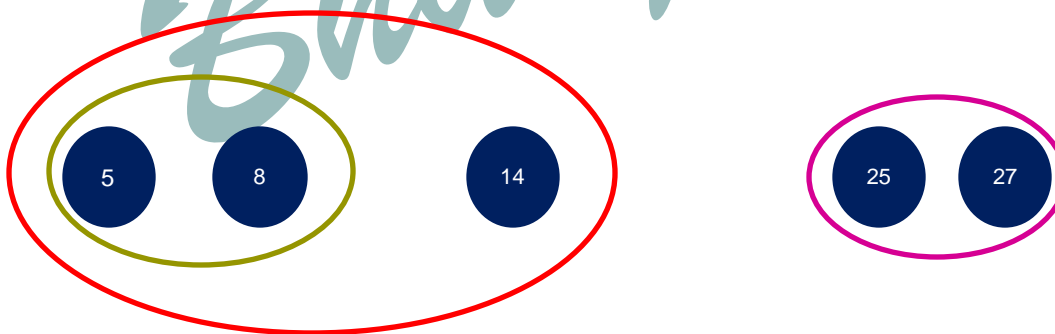
Example (cont...)



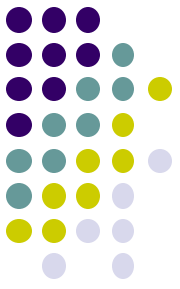
• Step 3

Cluster1	Cluster 2	Distance
{5, 8}	{14}	9
{5, 8}	{25, 27}	22
{25, 27}	{14}	13

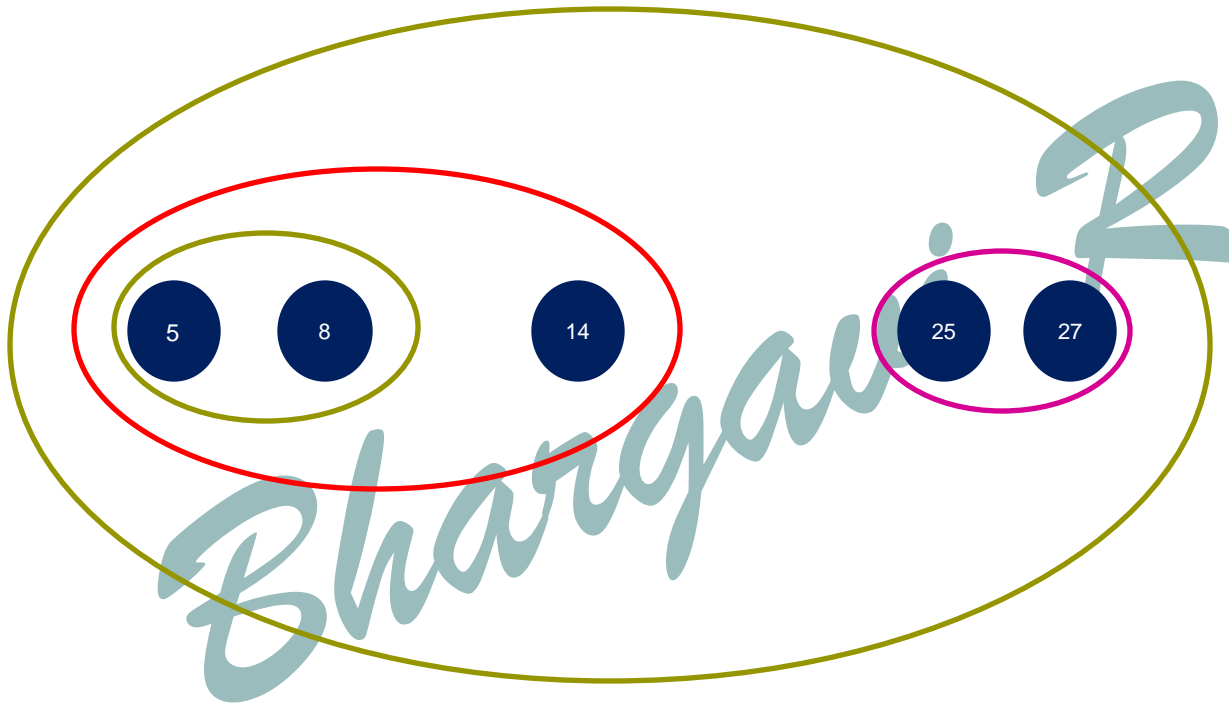
Min.
dissimilarity

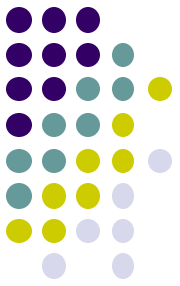


Example (cont...)

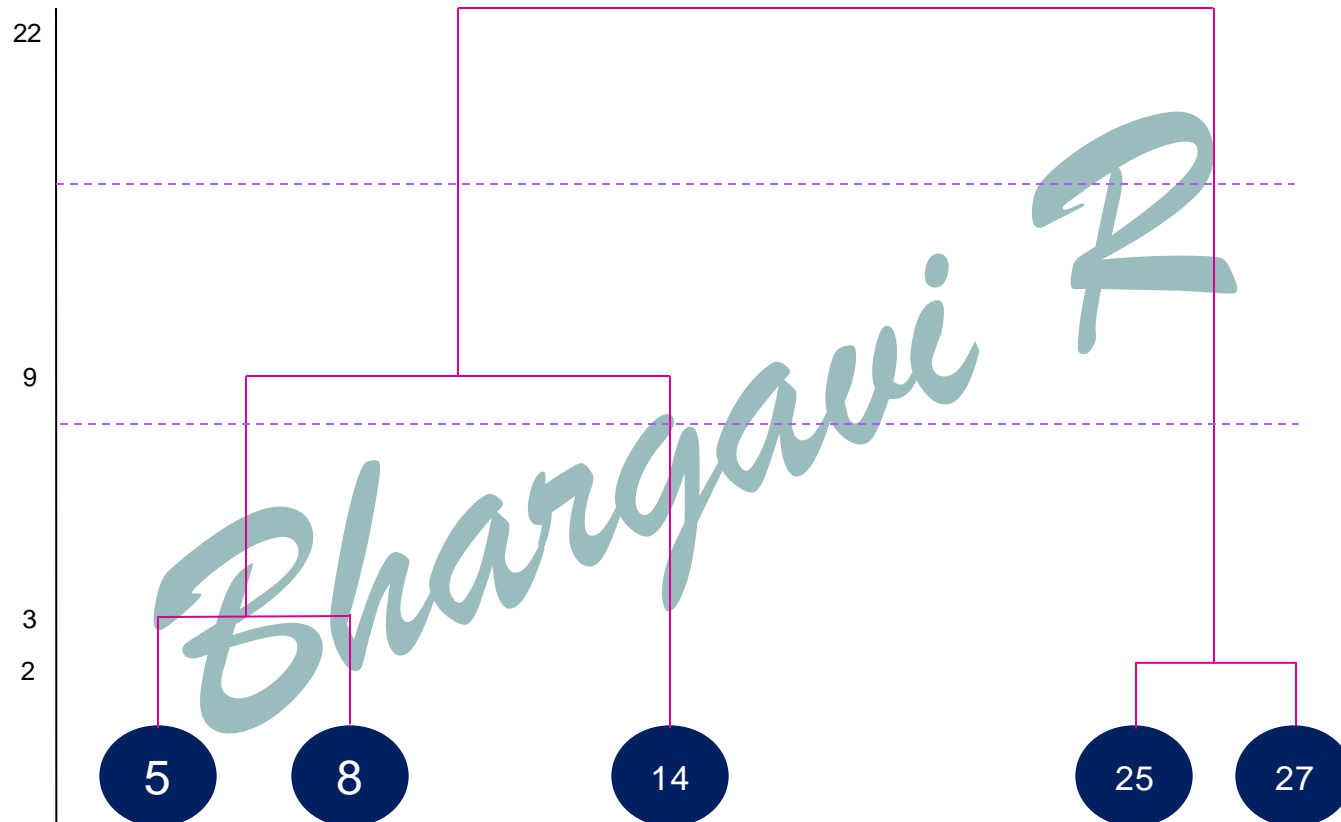


- Step 4

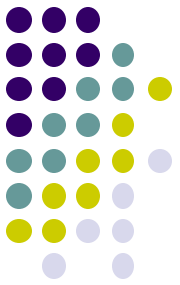




Dendrogram interpretation

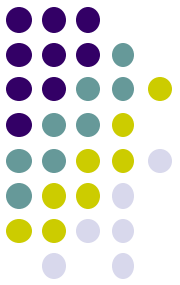


Dendrogram interpretation (cont...)



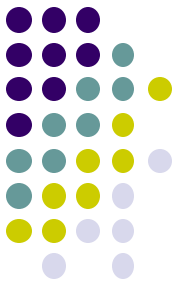
- Higher the similarity of the clusters, sooner the fusion (lower in the tree) occurs in the construction of the dendrogram.
- Clusters that get fused near the top of the tree are most dissimilar.
- The height of the fusion, as measured on the vertical axis, indicates how different the two observations are.

Algorithm



- Begin with n observations and a measure (ex: Euclidean distance) of all the $nC2 = n(n-1)/2$ pairwise dissimilarities. Treat each observation as its own cluster.
- For $i = n, n-1, \dots, 2$:
 - Examine all pairwise inter-cluster dissimilarities among the i clusters and identify the pair of clusters that are least dissimilar (that is, most similar). Fuse these two clusters.
 - Compute the new pairwise inter-cluster dissimilarities among the $i-1$ remaining clusters.

Practical Issues in Hierarchical Clustering



- In order to perform clustering, some decisions must be made like
- Should the observations or features first be standardized in some way?
- What dissimilarity measure should be used?
- What type of linkage should be used?
- Where should the dendrogram be cut in order to obtain clusters?