

Data Mining

Dr. Raed Ibraheem Hamed

**University of Human Development,
College of Science and Technology
Department of Computer Science**






Department of CS - DM - UHD

Road map

- Cluster Analysis: Basic Concepts
- Partitioning Methods
- Hierarchical Methods
- What is Hierarchical Clustering
- General Steps Of Hierarchical Clustering
- Methods of Hierarchical Clustering
- Agglomerative (bottom up)
- Divisive (top down)
- Dendrogram
- Summary

Cluster Analysis: Basic Concepts and Methods

- Cluster Analysis: Basic Concepts
- Partitioning Methods
- Hierarchical Methods 
- Evaluation of Clustering
- Summary

What is Hierarchical Clustering

- ❖ In data mining and statistics, hierarchical clustering (also called **hierarchical cluster analysis** or **HCA**) is a method of cluster analysis which seeks to build a hierarchy of clusters.
- ❖ The idea is to build a binary tree of the data that successively merges similar groups of points

General Steps Of Hierarchical Clustering

Given a set of N items to be clustered, the basic process of hierarchical clustering is this:

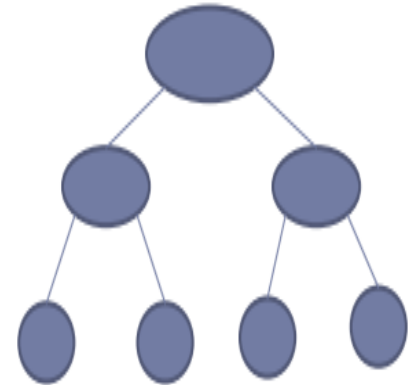
1. Start by assigning each item to a cluster, so that if you have N items, you now have N clusters, each containing just one item.
2. Find the closest (most similar) pair of clusters and merge them into a single cluster, so that now you have one cluster less.
3. Compute similarities between the new cluster and each of the old clusters.
4. Repeat steps 2 and 3 until all items are clustered into K number of clusters

Methods of Hierarchical Clustering

There are two main types of hierarchical clustering:
group data objects into a tree of clusters

Hierarchical methods can be

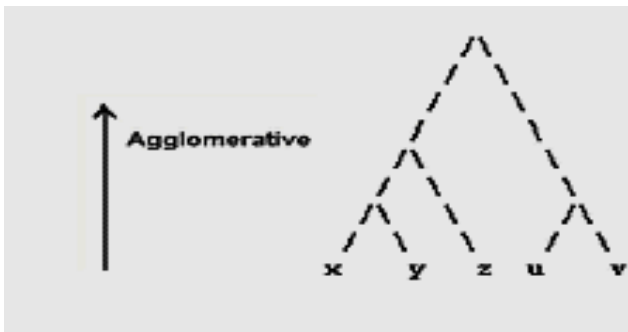
- ❖ **Agglomerative**: bottom-up approach
- ❖ **Divisive**: top-down approach



Hierarchical Clustering Agglomerative and Divisive Methods

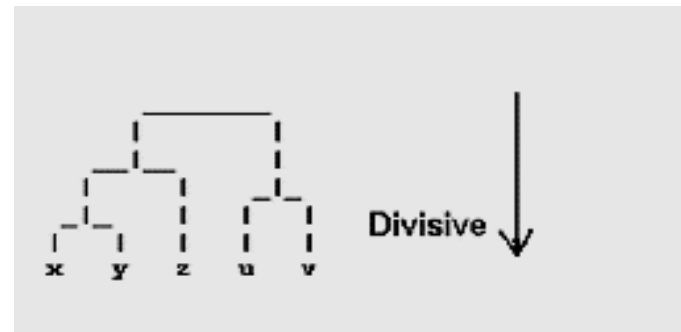
Agglomerative (bottom up)

1. Start with 1 point (singleton)
2. Recursively add two or more appropriate clusters
3. Stop when k number of clusters is achieved.



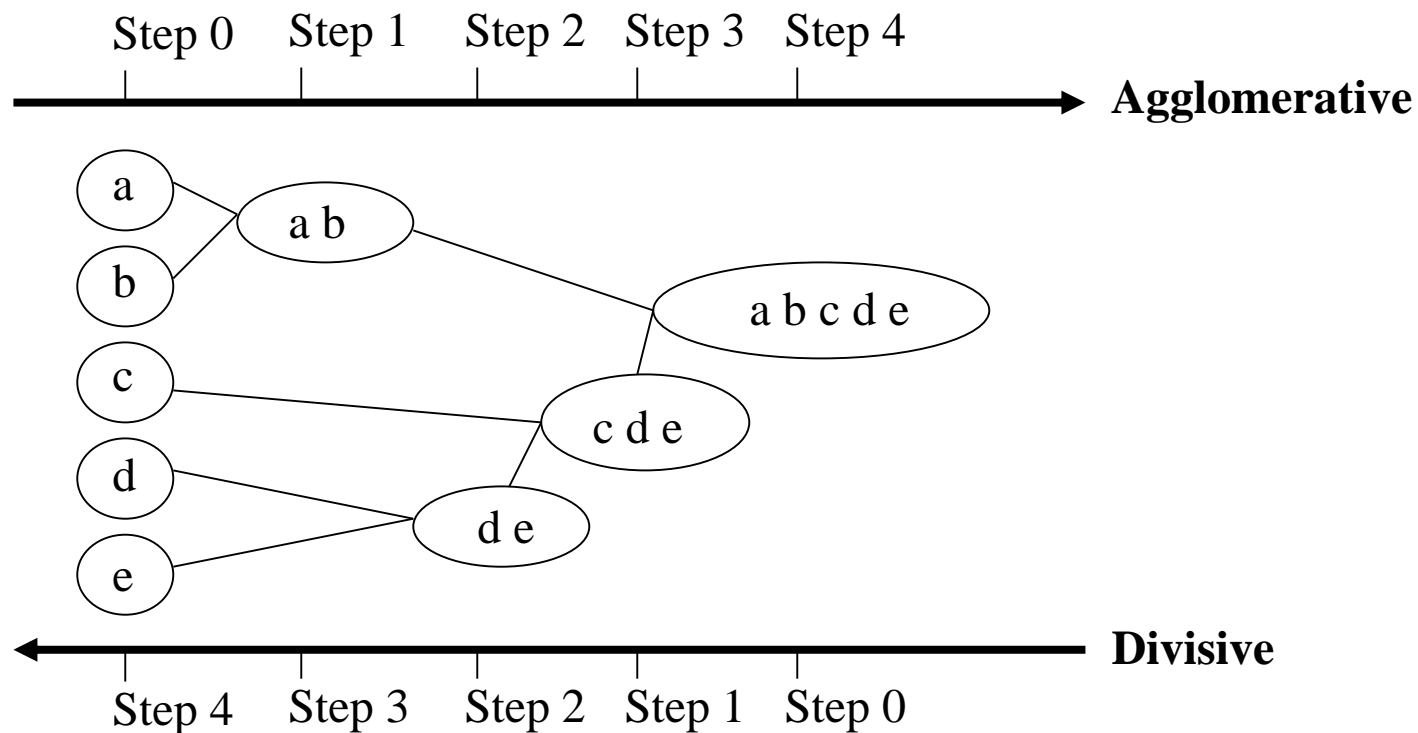
Divisive (top down)

1. Start with a big cluster
2. Recursively divide into smaller clusters
3. Stop when k number of clusters is achieved.

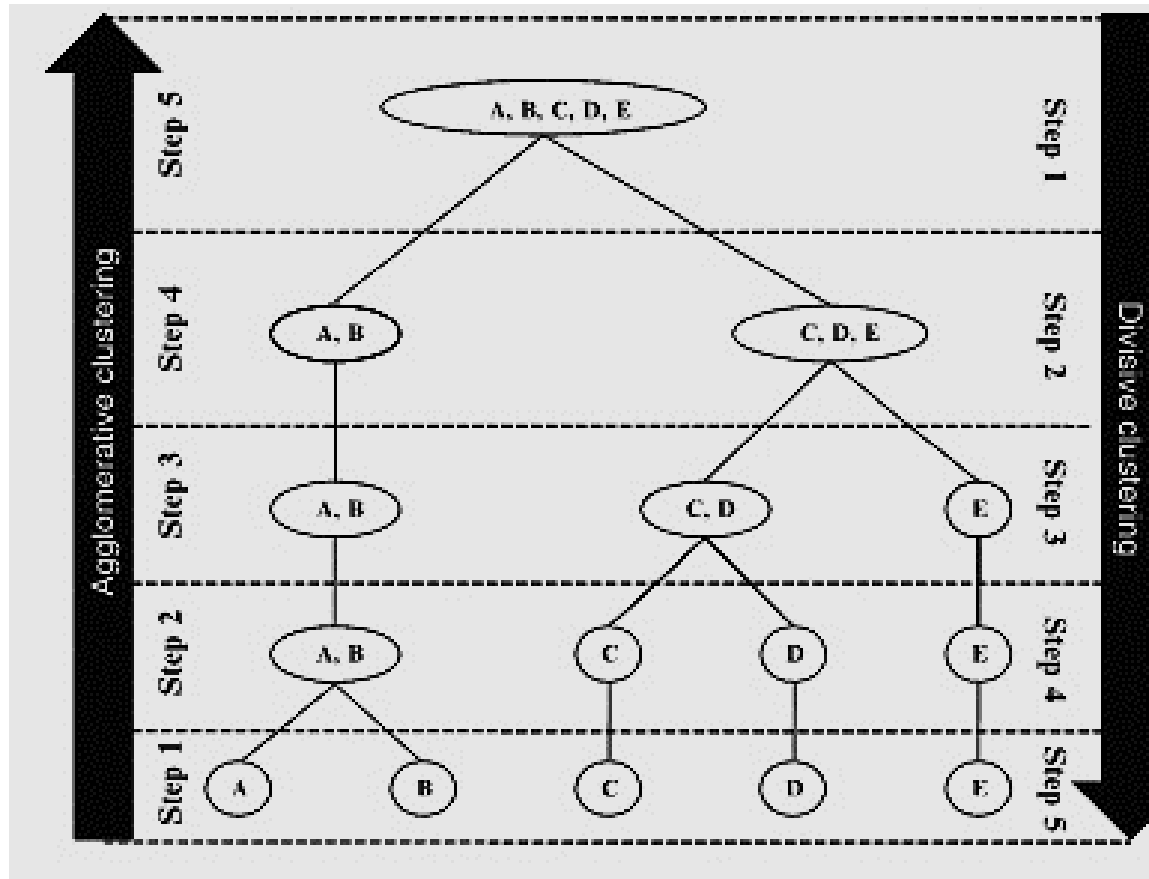


Hierarchical Clustering Agglomerative and Divisive Methods

- Use similarity matrix as clustering criteria. Stop when k number of clusters is achieved.



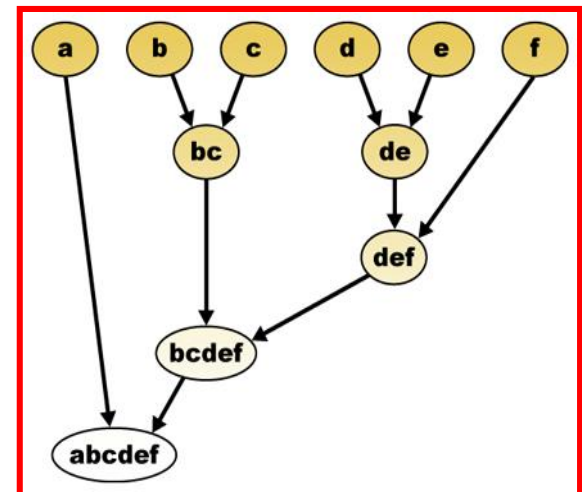
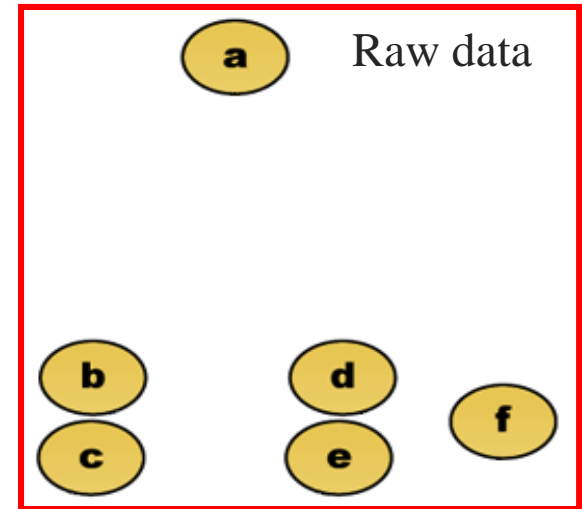
Hierarchical Clustering Agglomerative and Divisive Methods



Hierarchical Agglomerative General Algorithm

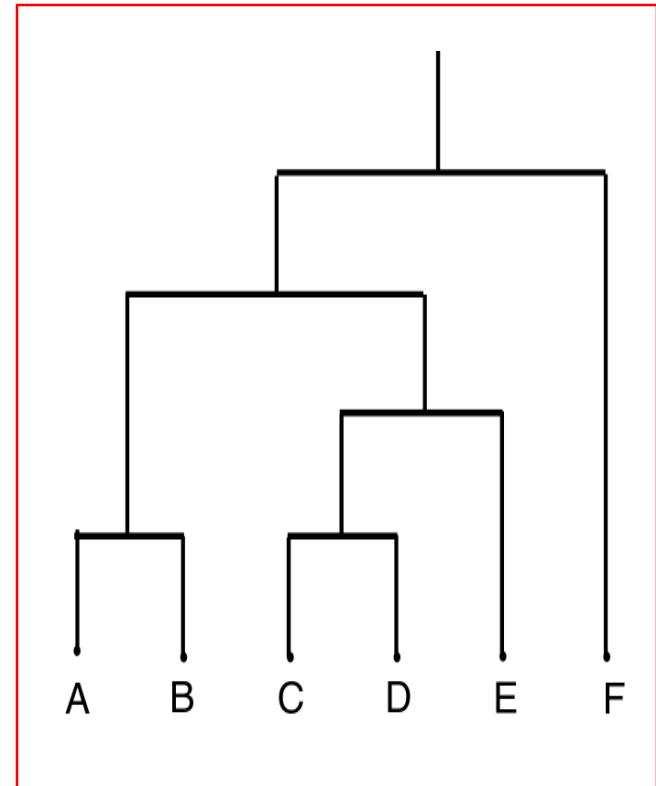
In this example, after the **second step** of the agglomerative algorithm will yield clusters:-
{a} {b c} {d e} {f}.

In the **third step** will yield clusters **{a} {b c} {d e f}**, which is a clustering, in the **fourth step** will give a small number but larger clusters that are **{a} {b c d e f}** and **finally** will yield cluster of **{a b c d e f}**



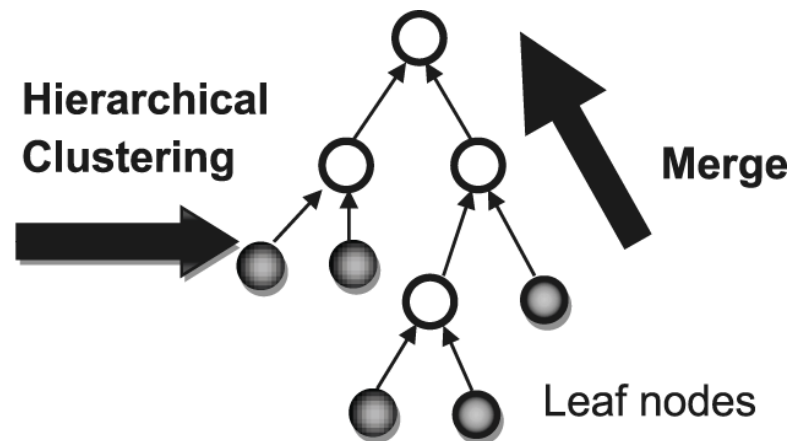
Dendrogram Hierarchical Clustering

- ***Dendrogram:*** a tree data structure which illustrates hierarchical clustering techniques.
- Each level shows clusters for that level.
 - Leaf – individual clusters
 - Root – one cluster

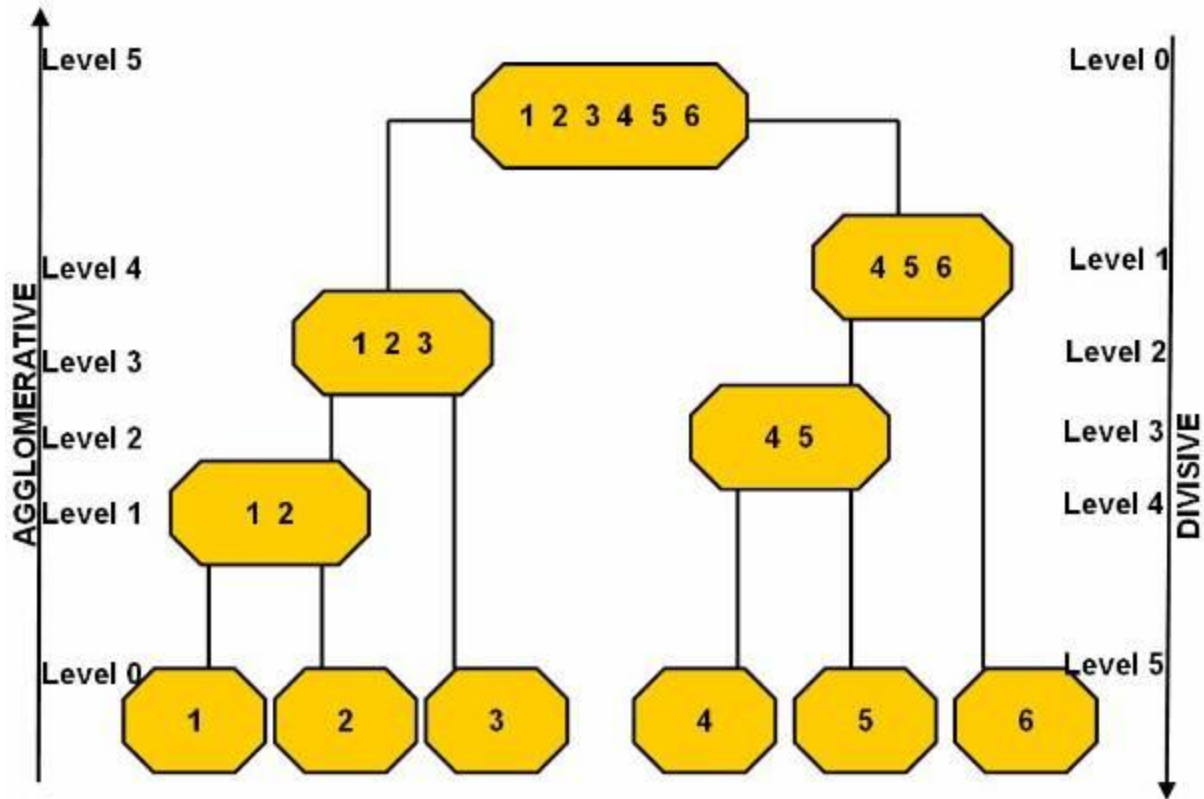


Dendrogram: Shows How Clusters are Merged

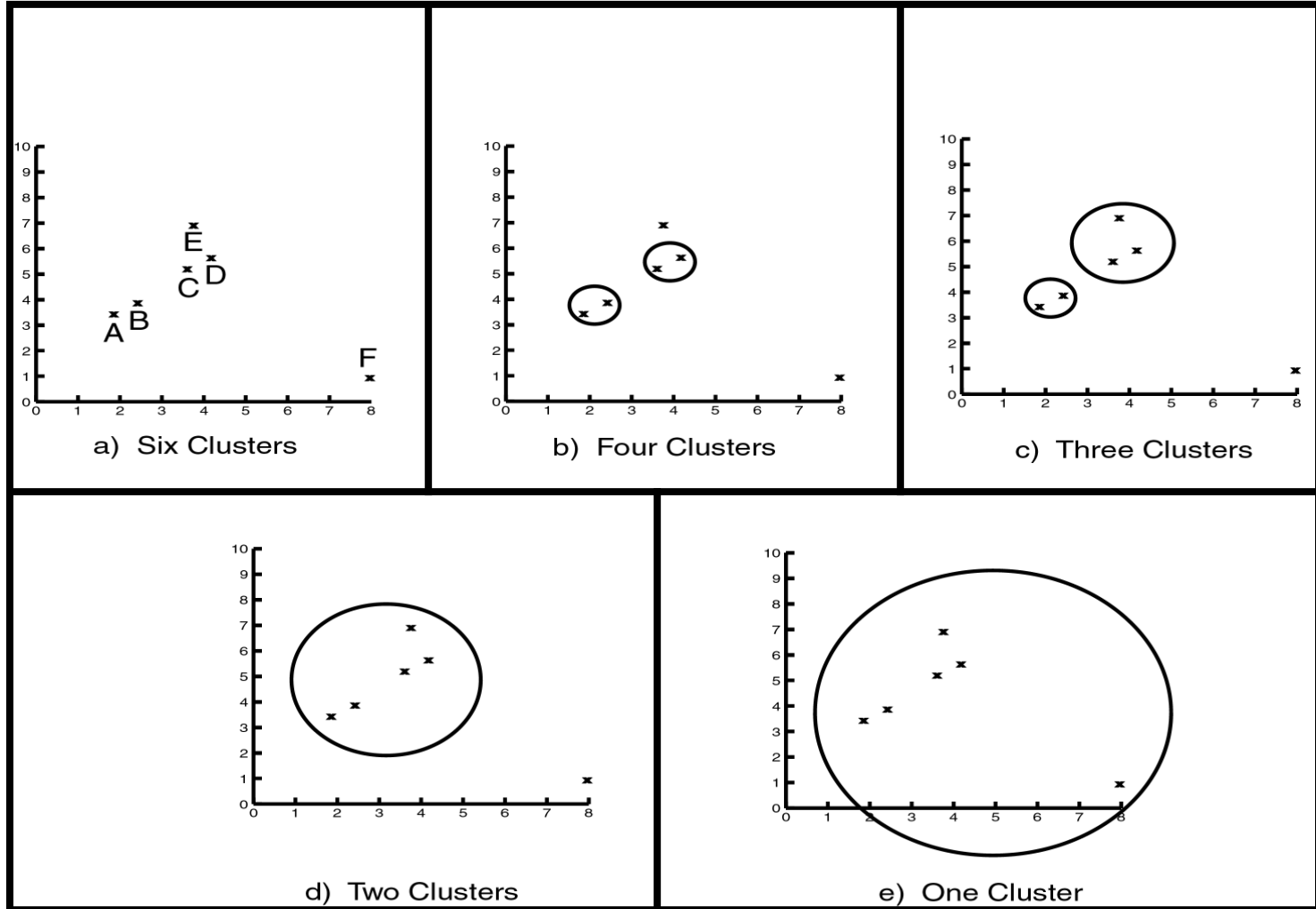
- Show how to merge clusters hierarchically
- Decompose data objects into a multi-level of a tree of clusters
- A clustering of the data objects: giving the dendrogram at the desired level
 - Each connected component forms a cluster



Levels of Clustering



Levels of Traditional Clustering

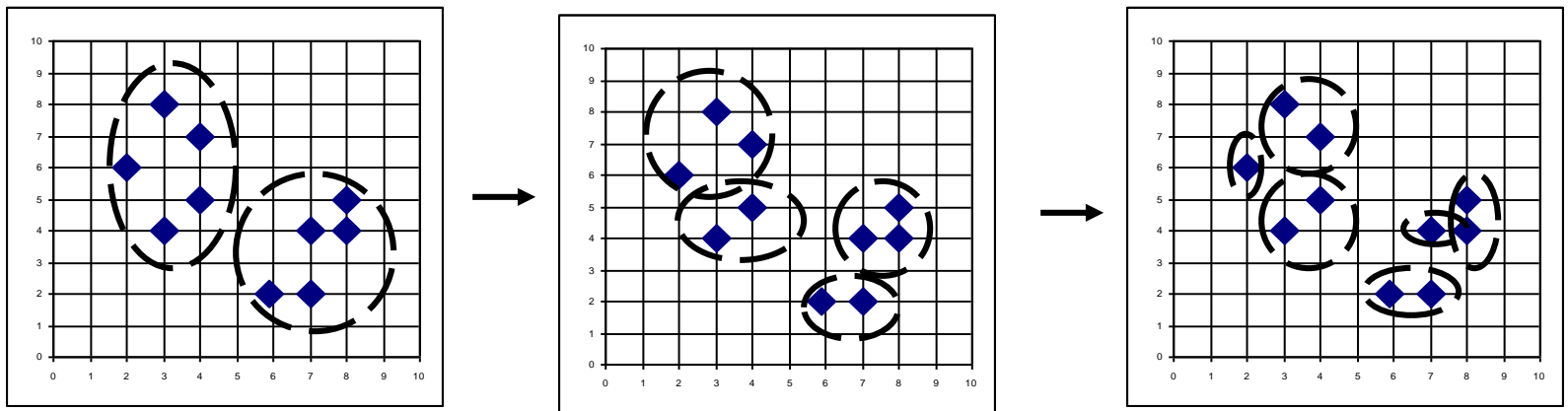


Traditional Hierarchical Clustering

Divisive Method

Divisive (top down)

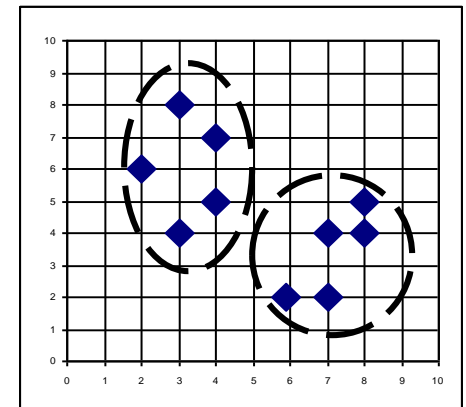
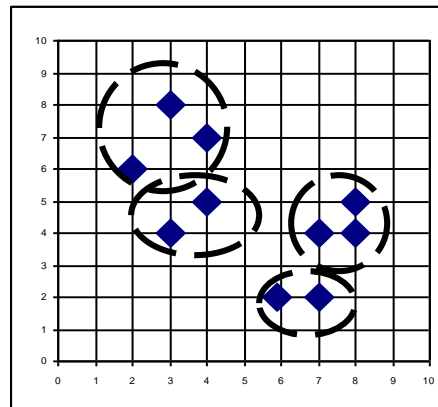
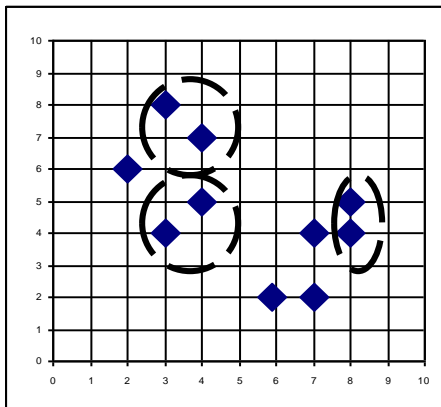
1. Start with a big cluster
2. Recursively divide into smaller clusters
3. Stop when k number of clusters is achieved.



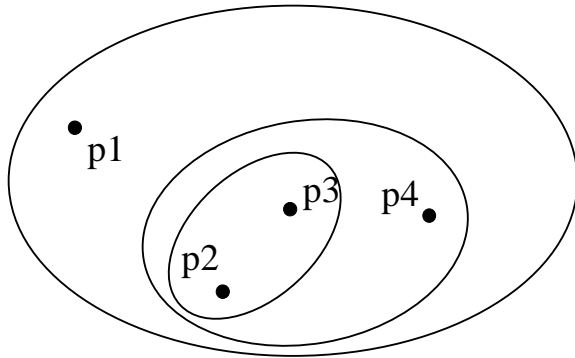
Traditional Hierarchical Clustering Agglomerative Method

Agglomerative (bottom up)

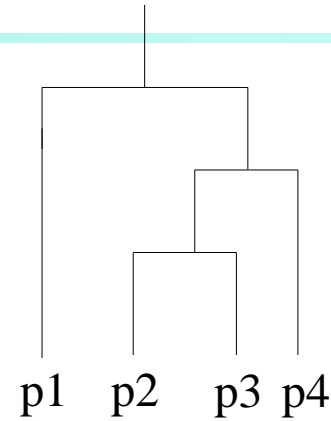
1. Start with 1 point (singleton)
2. Recursively add two or more appropriate clusters
3. Stop when k number of clusters is achieved.



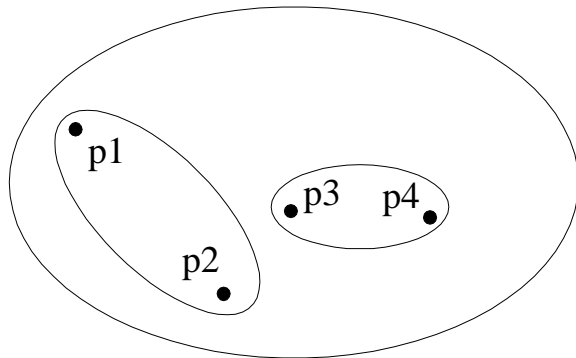
Hierarchical Clustering



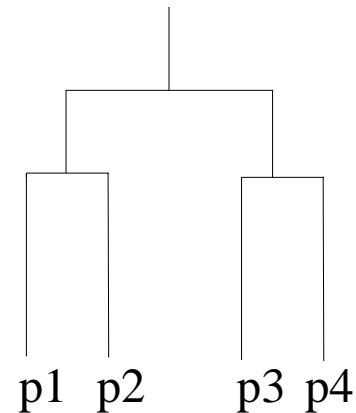
Traditional Hierarchical Clustering



Traditional Dendrogram

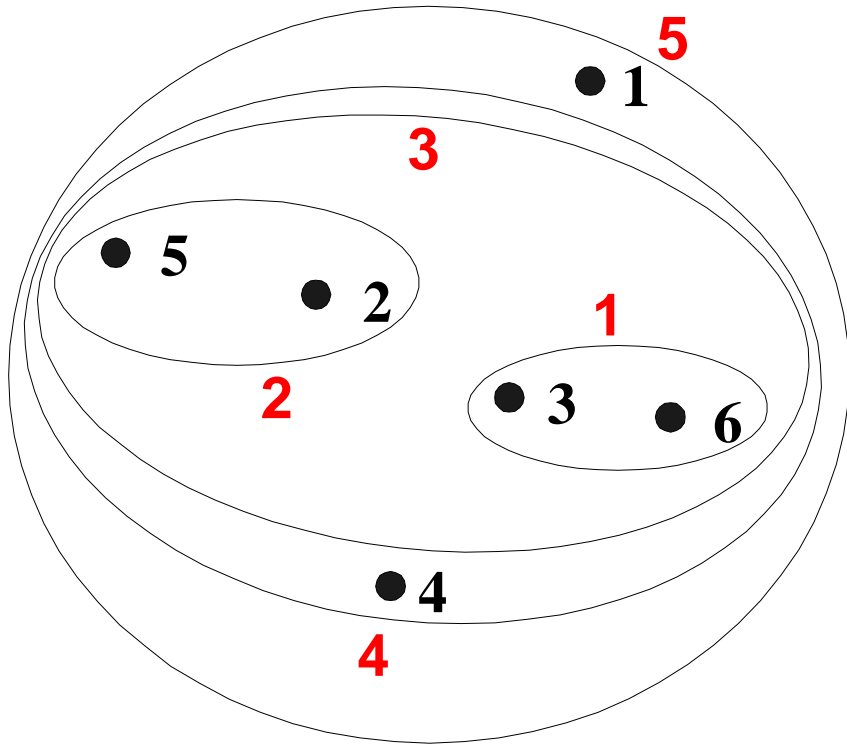


Non-traditional Hierarchical Clustering

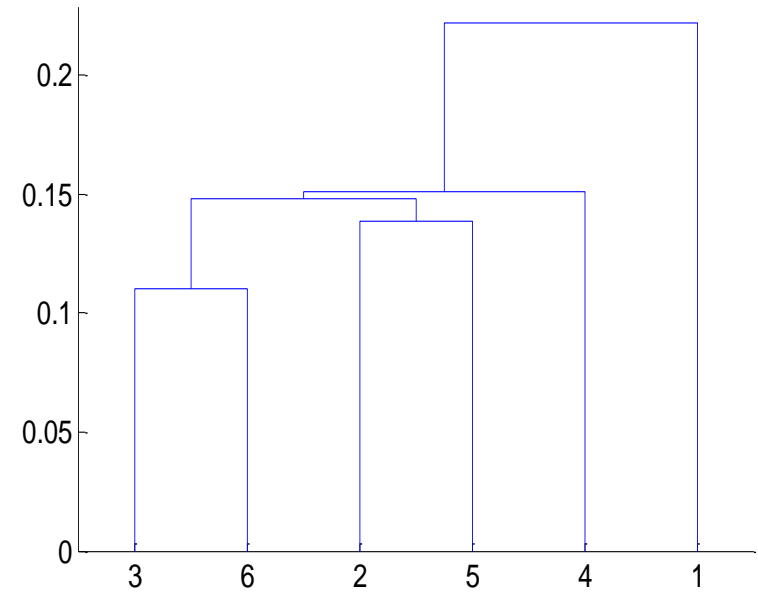


Non-traditional Dendrogram

Hierarchical Clustering

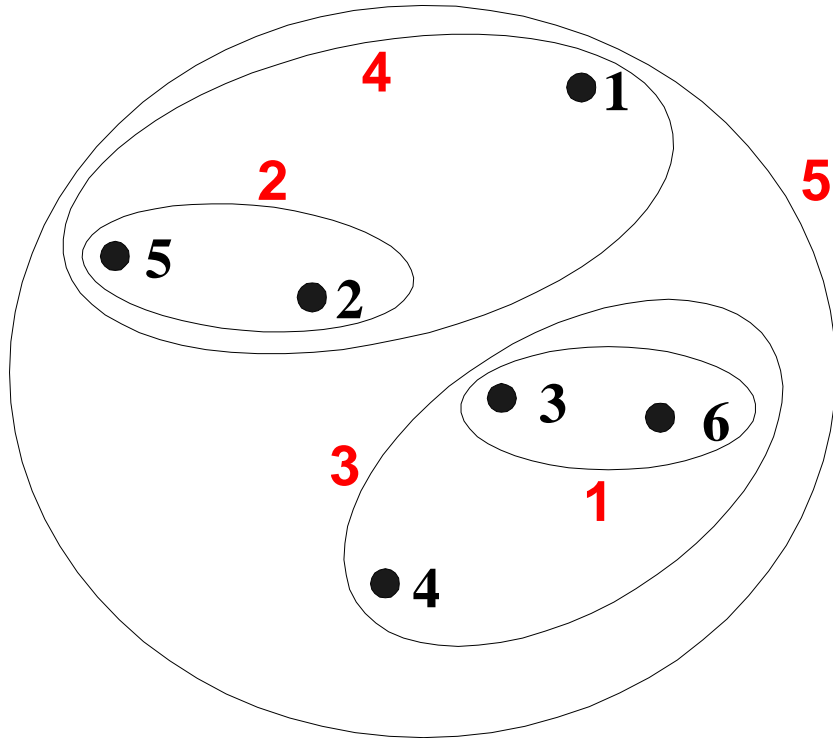


Nested Clusters

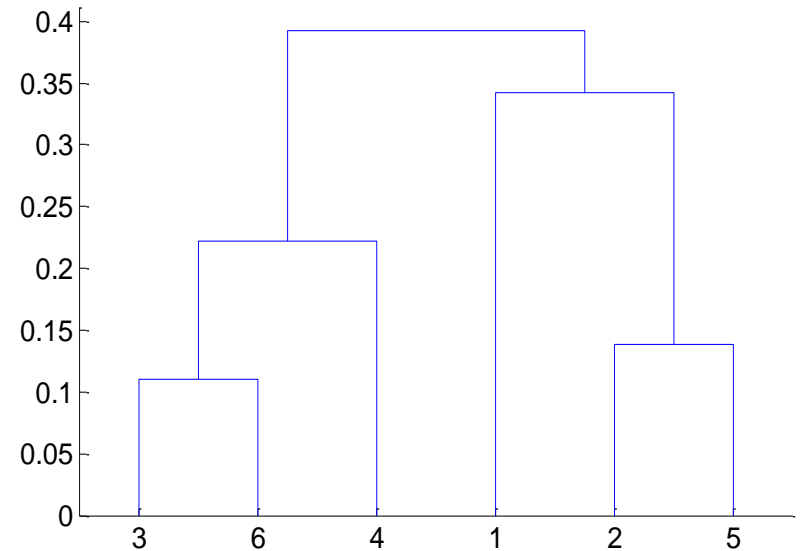


Dendrogram

Hierarchical Clustering

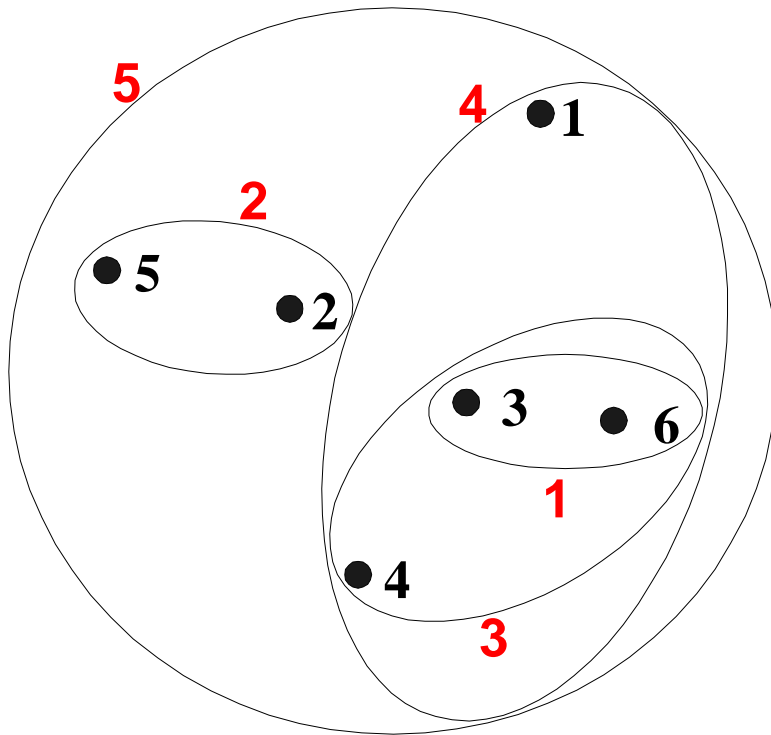


Nested Clusters

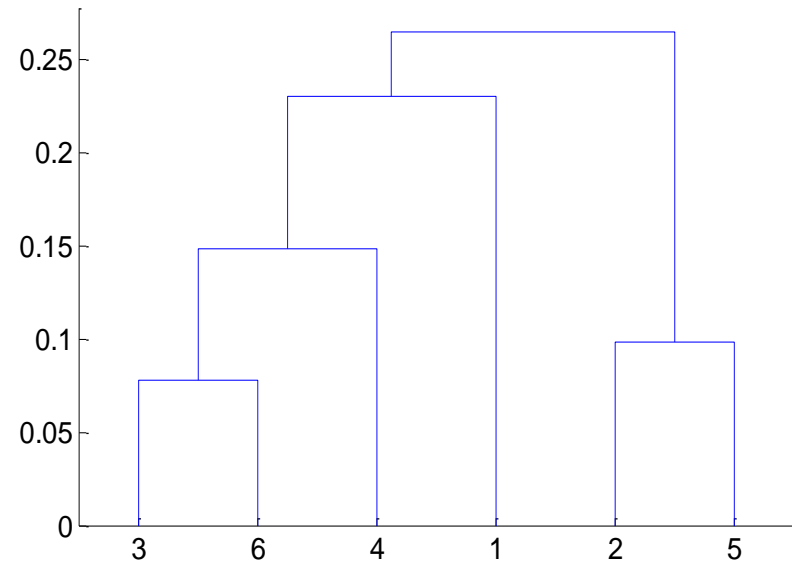


Dendrogram

Hierarchical Clustering



Nested Clusters



Dendrogram

Thank
you

