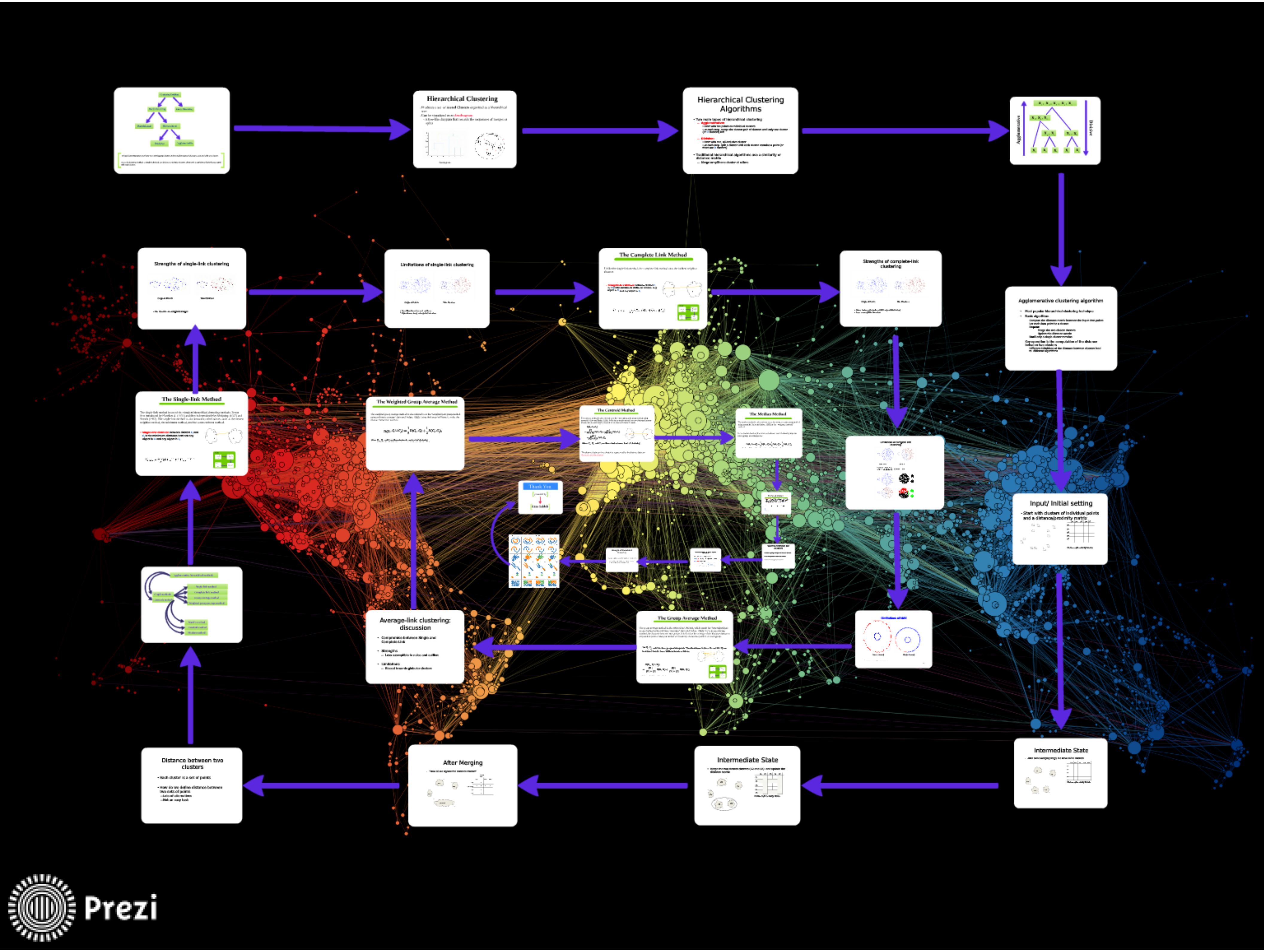
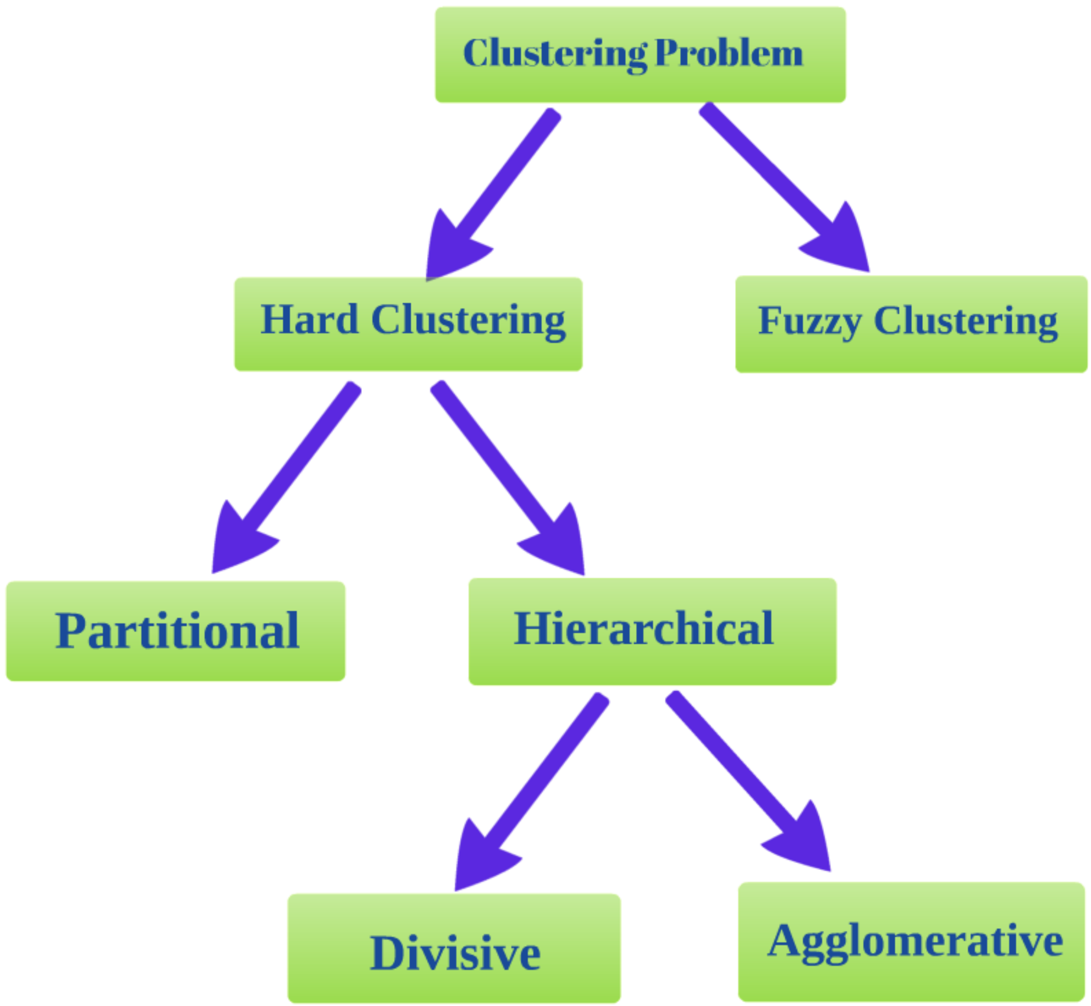

Hierarchical Clustering
Hierarchical Clustering Algorithms
Agglomerative
Divisive
Strengths of single-link clustering
Limitations of single-link clustering
The Complete Link Method
Strengths of complete-link clustering
Agglomerative clustering algorithm
The Single-link Method
The Weighted Group Average Method
The Centroid Method
The Median Method
Input/ Initial setting
Thank You
Average-link clustering: discussion
The Group Average Method
Distance between two clusters
After Merging
Intermediate State
Intermediate State
Prezi



A hard clustering means we have non-overlapping clusters, where each instance belongs to one and only one cluster.

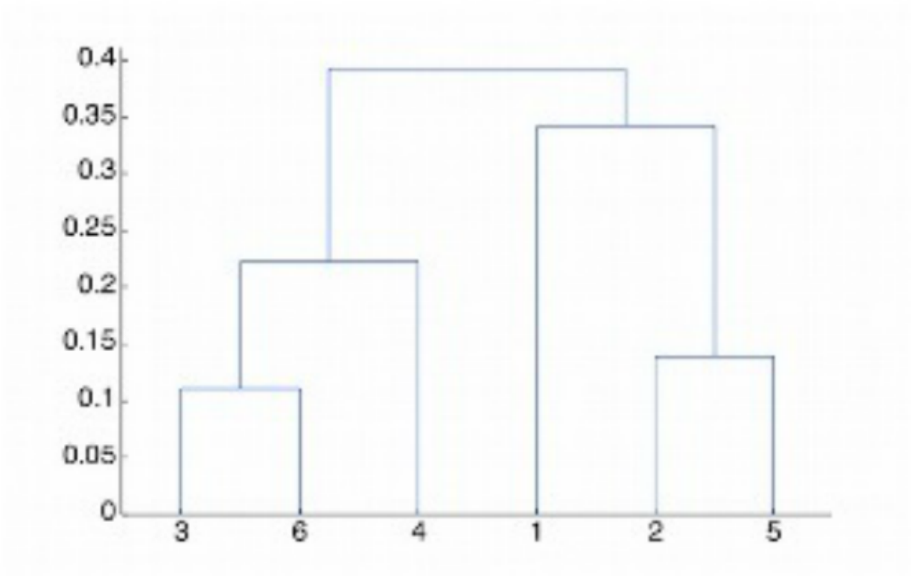
In a soft clustering method, a single individual can belong to multiple clusters, often with a confidence (belief) associated with each cluster.

Hierarchical Clustering

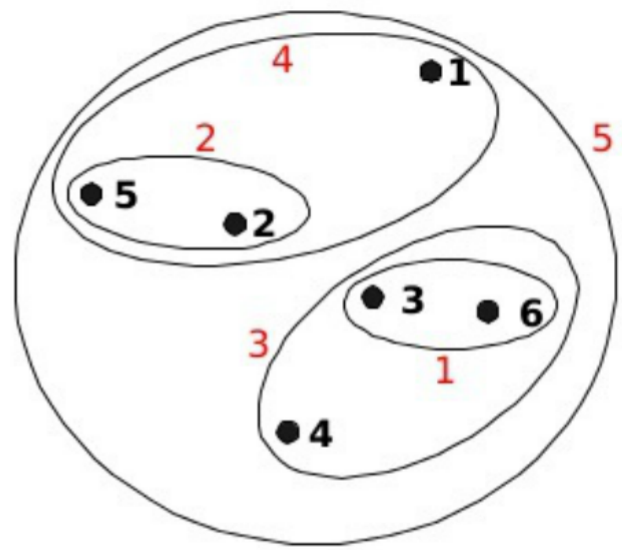
. Produces a set of **nested Clusters** organized as a hierarchical tree

. Can be visualized as as **dendrogram**

- A tree-like diagram that records the sequences of merges or splits

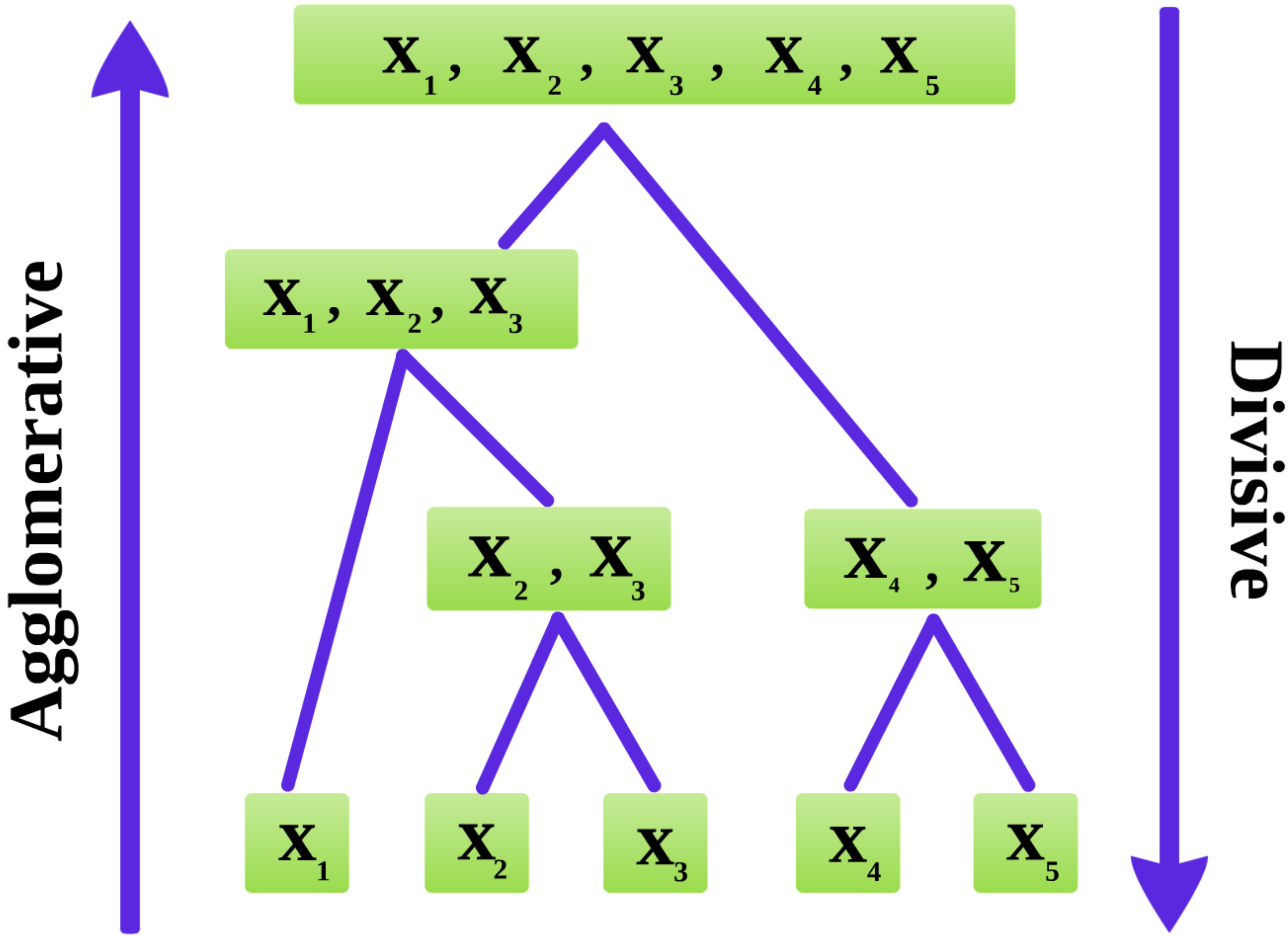


dendogram



Hierarchical Clustering Algorithms

- Two main types of hierarchical clustering
  - **Agglomerative:**
    - Start with the points as individual clusters
    - At each step, merge the closest pair of clusters until only one cluster (or **k** clusters) left
  - **Divisive:**
    - Start with one, all-inclusive cluster
    - At each step, split a cluster until each cluster contains a point (or there are **k** clusters)
- Traditional hierarchical algorithms use a similarity or distance matrix
  - Merge or split one cluster at a time



Agglomerative
x1 , x2 , x3 , x4 , x5
x1 , x2 , x3
x2 , x3
x4 , x5
x1
x2
x3
x4
x5
Divisive

Agglomerative clustering algorithm

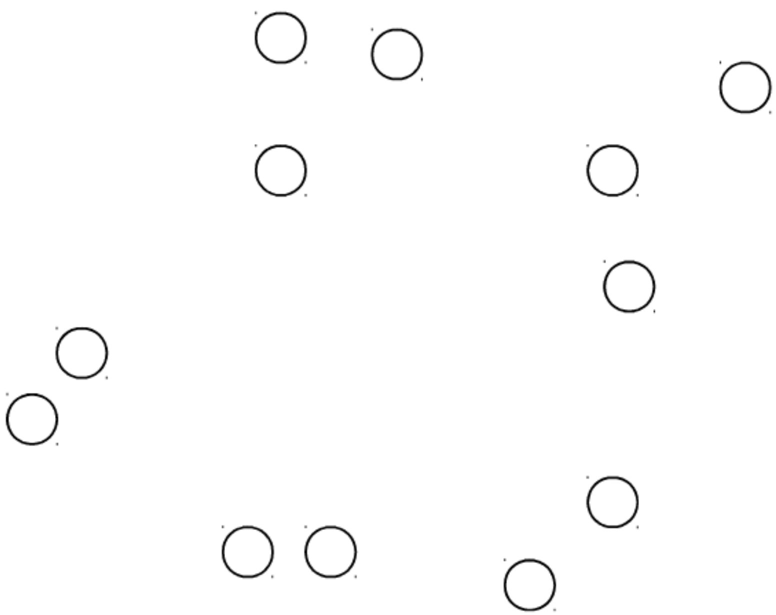
- Most popular hierarchical clustering technique
- Basic algorithm:
  - Compute the distance matrix between the input data points
  - Let each data point be a cluster
  - **Repeat**
    - Merge the two closest clusters
    - Update the distance matrix
  - **Until** only a single cluster remains

Key operation is the computation of the distance between two clusters

- Different definitions of the distance between clusters lead to different algorithms

Input/ Initial setting

- Start with clusters of individual points and a distance/proximity matrix

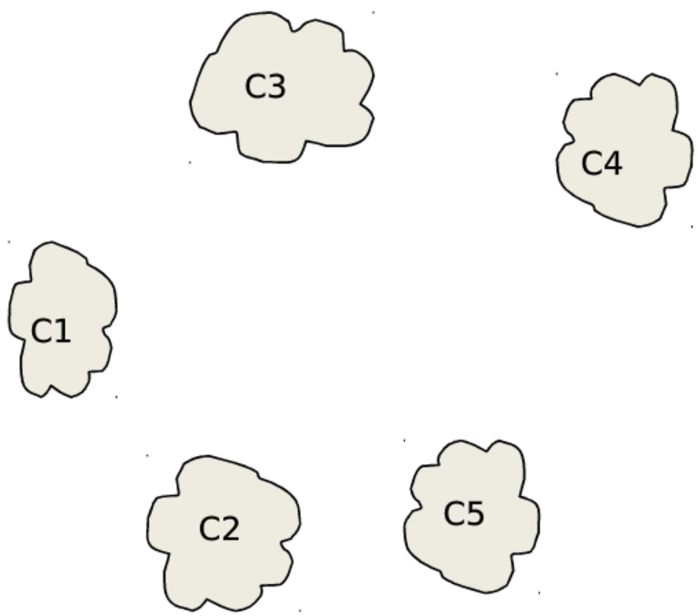


	p1	p2	p3	p4	p5	. . .
p1						
p2						
p3						
p4						
p5						
. . .						

Distance/Proximity Matrix

Intermediate State

- After some merging steps, we have some clusters

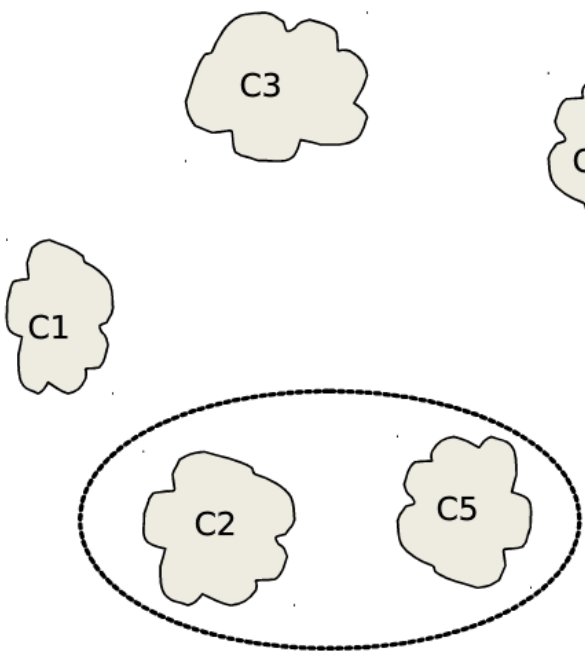


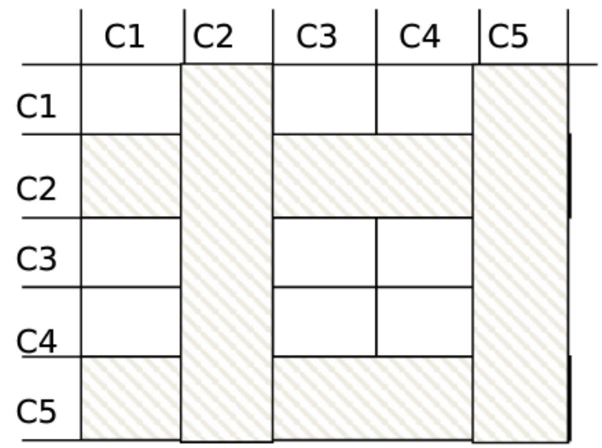
	C1	C2	C3	C4	C5
C1					
C2					
C3					
C4					
C5					

Distance/Proximity Matrix

Intermediate State

- Merge the two closest clusters (C2 and C5) and update the distance matrix.

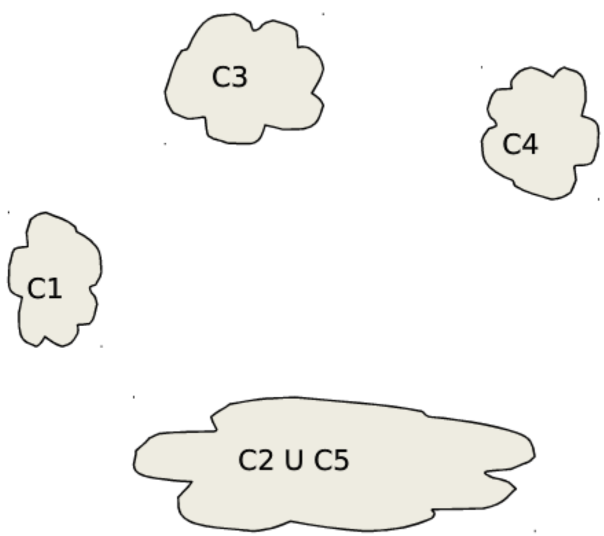




Distance/Proximity Matrix

After Merging

- “How do we update the distance matrix?”

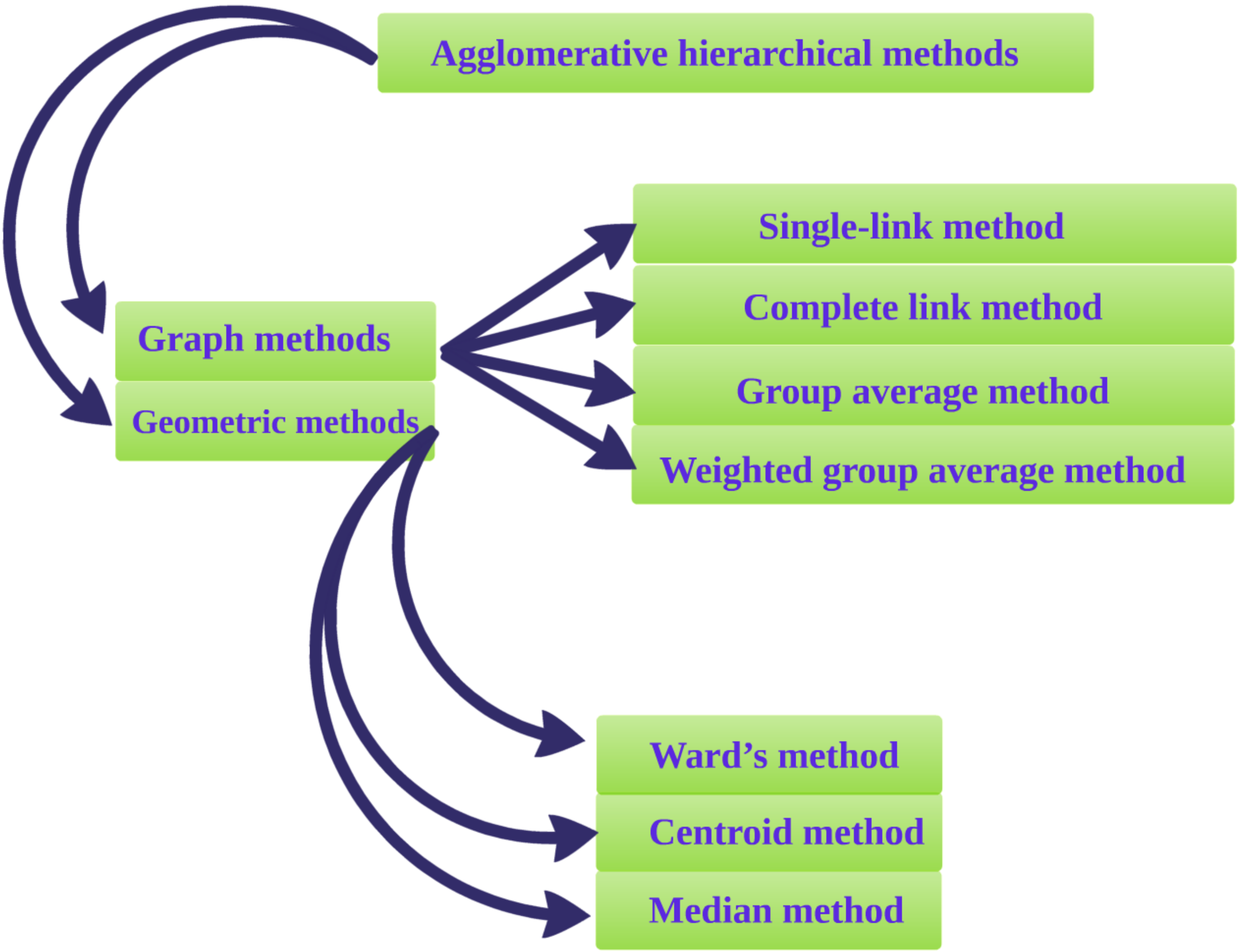


	C1	C2 U C5	C3	C4
C1		?		
C2 U C5	?	?	?	?
C3		?		
C4		?		

Distance between two clusters

- Each cluster is a set of points

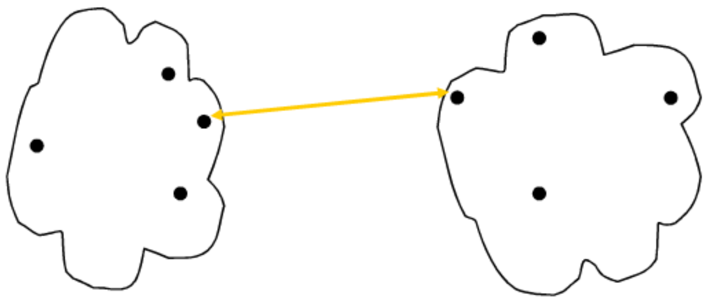
- How do we define distance between two sets of points
  - Lots of alternatives
  - Not an easy task


Agglomerative hierarchical methods
Graph methods
Geometric methods
Single-link method
Complete link method
Group average method
Weighted group average method
Ward's method
Centroid method
Median method

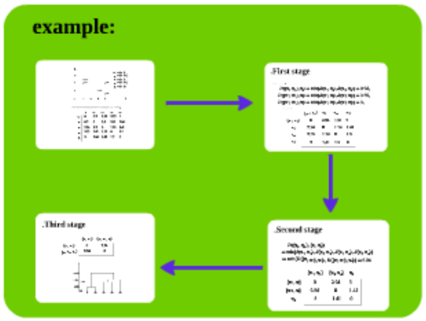
The Single-link Method

The single-link method is one of the simplest hierarchical clustering methods. It was first introduced by Florek et al. (1951) and then independently by McQuitty (1957) and Sneath (1957). The single-link method is also known by other names, such as the nearest neighbor method, the minimum method, and the connectedness method.

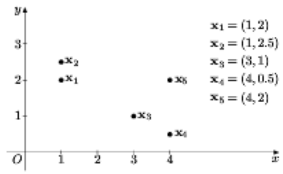
- **Single-link distance** between clusters C_i and C_j is the **minimum distance** between any object in C_i and any object in C_j



$$D_{\text{single}} = \min_{x,y}\{d(x,y) \mid x \in C_i, y \in C_j\}$$



example:



	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\mathbf{x}_1$	0	0.5	2.24	3.35	3
$\mathbf{x}_2$	0.5	0	2.5	3.61	3.04
$\mathbf{x}_3$	2.24	2.5	0	1.12	1.41
$\mathbf{x}_4$	3.35	3.61	1.12	0	1.5
$\mathbf{x}_5$	3	3.04	1.41	1.5	0

.First stage

$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3) = \min\{d(\mathbf{x}_1, \mathbf{x}_3), d(\mathbf{x}_2, \mathbf{x}_3)\} = 2.24,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4) = \min\{d(\mathbf{x}_1, \mathbf{x}_4), d(\mathbf{x}_2, \mathbf{x}_4)\} = 3.35,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5) = \min\{d(\mathbf{x}_1, \mathbf{x}_5), d(\mathbf{x}_2, \mathbf{x}_5)\} = 3,$$

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.24	3.35	3
$\mathbf{x}_3$	2.24	0	1.12	1.41
$\mathbf{x}_4$	3.35	1.12	0	1.5
$\mathbf{x}_5$	3	1.41	1.5	0

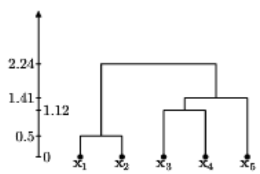
.Second stage

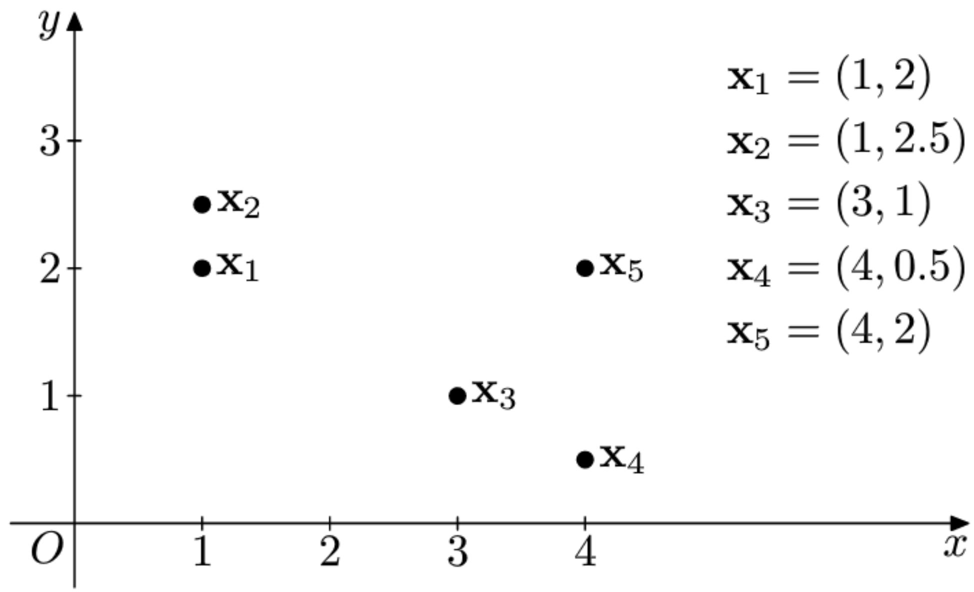
$$\begin{aligned} & D(\{\mathbf{x}_3, \mathbf{x}_4\}, \{\mathbf{x}_1, \mathbf{x}_2\}) \\ = & \min\{d(\mathbf{x}_1, \mathbf{x}_3), d(\mathbf{x}_2, \mathbf{x}_3), d(\mathbf{x}_1, \mathbf{x}_4), d(\mathbf{x}_2, \mathbf{x}_4)\} \\ = & \min\{D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3), D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4)\} = 2.24 \end{aligned}$$

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4\}$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.24	3
$\{\mathbf{x}_3, \mathbf{x}_4\}$	2.24	0	1.41
$\mathbf{x}_5$	3	1.41	0

.Third stage

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.24
$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$	2.24	0





	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\mathbf{x}_1$	0	0.5	2.24	3.35	3
$\mathbf{x}_2$	0.5	0	2.5	3.61	3.04
$\mathbf{x}_3$	2.24	2.5	0	1.12	1.41
$\mathbf{x}_4$	3.35	3.61	1.12	0	1.5
$\mathbf{x}_5$	3	3.04	1.41	1.5	0

.First stage

$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3) = \min\{d(\mathbf{x}_1, \mathbf{x}_3), d(\mathbf{x}_2, \mathbf{x}_3)\} = 2.24,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4) = \min\{d(\mathbf{x}_1, \mathbf{x}_4), d(\mathbf{x}_2, \mathbf{x}_4)\} = 3.35,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5) = \min\{d(\mathbf{x}_1, \mathbf{x}_5), d(\mathbf{x}_2, \mathbf{x}_5)\} = 3,$$

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.24	3.35	3
$\mathbf{x}_3$	2.24	0	1.12	1.41
$\mathbf{x}_4$	3.35	1.12	0	1.5
$\mathbf{x}_5$	3	1.41	1.5	0

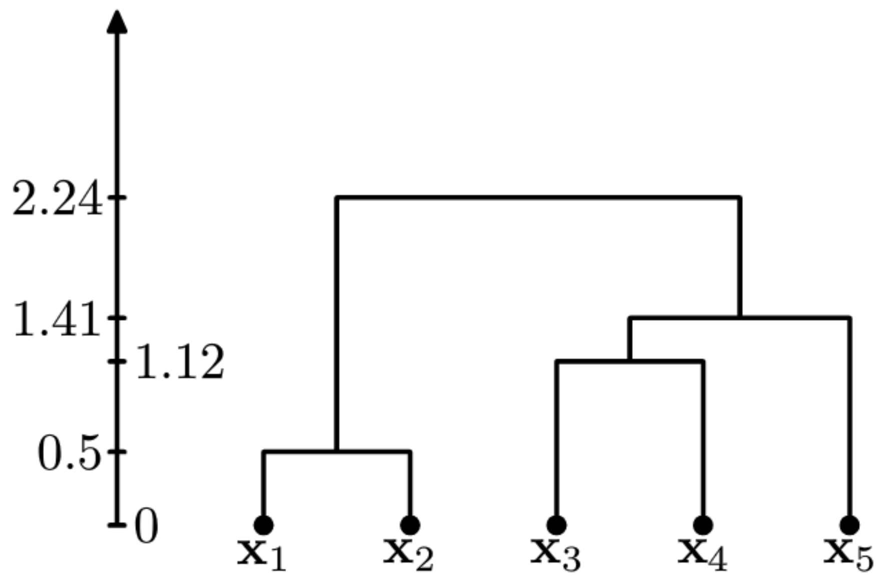
.Second stage

$$
\begin{aligned}
& D(\{\mathbf{x}_3, \mathbf{x}_4\}, \{\mathbf{x}_1, \mathbf{x}_2\}) \\
= & \min\{d(\mathbf{x}_1, \mathbf{x}_3), d(\mathbf{x}_2, \mathbf{x}_3), d(\mathbf{x}_1, \mathbf{x}_4), d(\mathbf{x}_2, \mathbf{x}_4)\} \\
= & \min\{D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3), D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4)\} = 2.24
\end{aligned}
$$

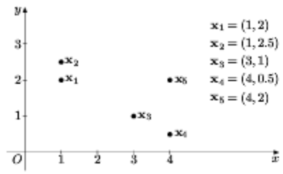
	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4\}$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.24	3
$\{\mathbf{x}_3, \mathbf{x}_4\}$	2.24	0	1.41
$\mathbf{x}_5$	3	1.41	0

.Third stage

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.24
$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$	2.24	0



example:



	x_1	x_2	x_3	x_4	x_5
x_1	0	0.5	2.24	3.35	3
x_2	0.5	0	2.5	3.61	3.04
x_3	2.24	2.5	0	1.12	1.41
x_4	3.35	3.61	1.12	0	1.5
x_5	3	3.04	1.41	1.5	0

.First stage

$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3) = \min\{d(\mathbf{x}_1, \mathbf{x}_3), d(\mathbf{x}_2, \mathbf{x}_3)\} = 2.24,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4) = \min\{d(\mathbf{x}_1, \mathbf{x}_4), d(\mathbf{x}_2, \mathbf{x}_4)\} = 3.35,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5) = \min\{d(\mathbf{x}_1, \mathbf{x}_5), d(\mathbf{x}_2, \mathbf{x}_5)\} = 3,$$

	$\{x_1, x_2\}$	x_3	x_4	x_5
$\{x_1, x_2\}$	0	2.24	3.35	3
x_3	2.24	0	1.12	1.41
x_4	3.35	1.12	0	1.5
x_5	3	1.41	1.5	0

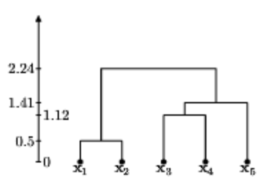
.Second stage

$$D(\{\mathbf{x}_3, \mathbf{x}_4\}, \{\mathbf{x}_1, \mathbf{x}_2\})$$
$$= \min\{d(\mathbf{x}_1, \mathbf{x}_3), d(\mathbf{x}_2, \mathbf{x}_3), d(\mathbf{x}_1, \mathbf{x}_4), d(\mathbf{x}_2, \mathbf{x}_4)\}$$
$$= \min\{D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3), D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4)\} = 2.24$$

	$\{x_1, x_2\}$	$\{x_3, x_4\}$	x_5
$\{x_1, x_2\}$	0	2.24	3
$\{x_3, x_4\}$	2.24	0	1.41
x_5	3	1.41	0

.Third stage

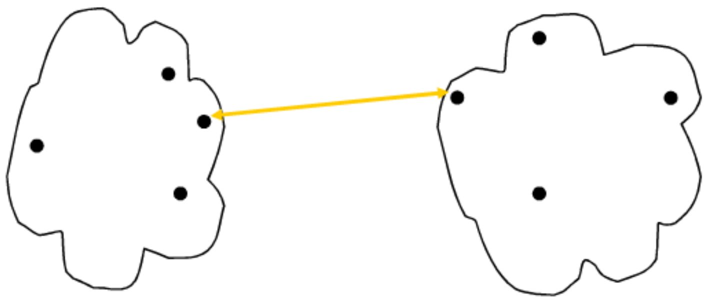
	$\{x_1, x_2\}$	$\{x_3, x_4, x_5\}$
$\{x_1, x_2\}$	0	2.24
$\{x_3, x_4, x_5\}$	2.24	0



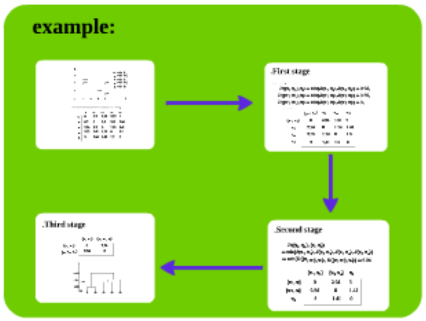
The Single-link Method

The single-link method is one of the simplest hierarchical clustering methods. It was first introduced by Florek et al. (1951) and then independently by McQuitty (1957) and Sneath (1957). The single-link method is also known by other names, such as the nearest neighbor method, the minimum method, and the connectedness method.

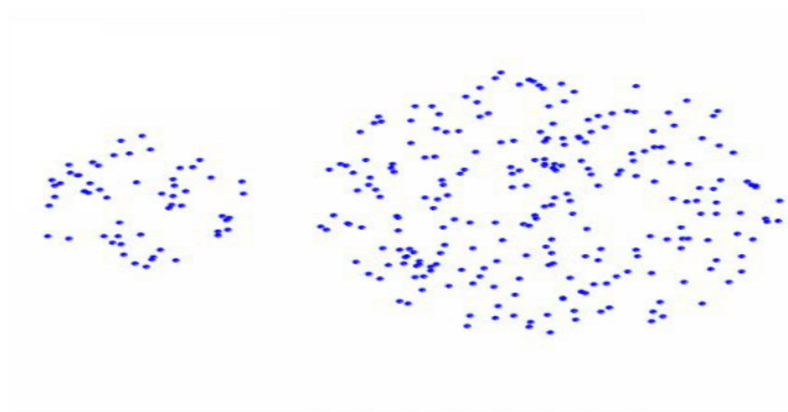
- **Single-link distance** between clusters C_i and C_j is the **minimum distance** between any object in C_i and any object in C_j



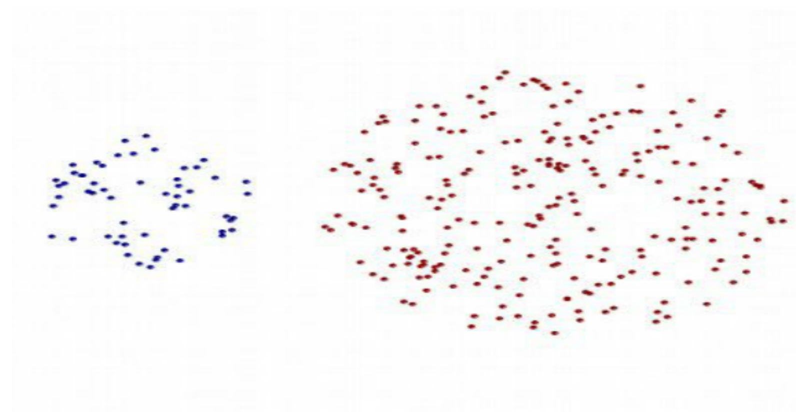
$$D_{\text{single}} = \min_{x,y}\{d(x,y) \mid x \in C_i, y \in C_j\}$$



Strengths of single-link clustering



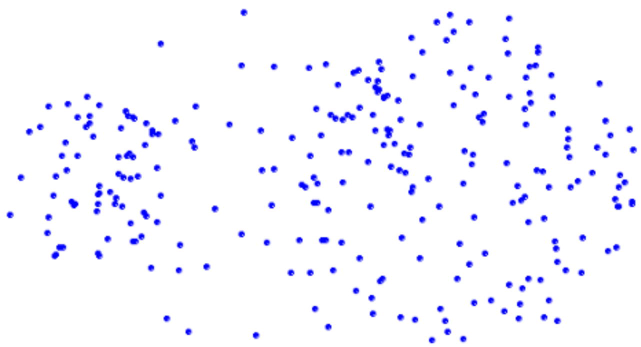
Original Points



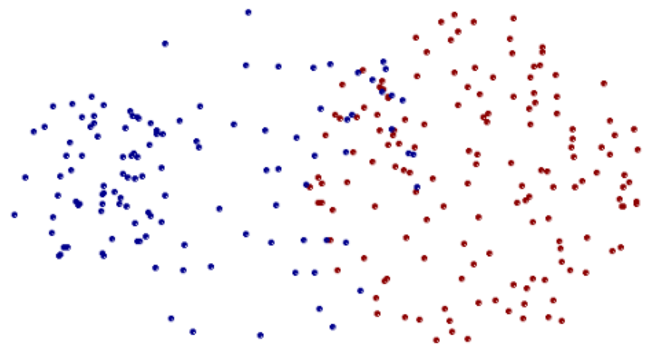
Two Clusters

- Can handle non-elliptical shapes

Limitations of single-link clustering



Original Points



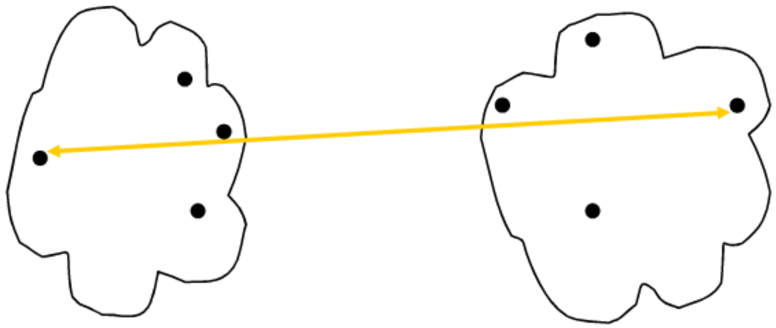
Two Clusters

- Sensitive to noise and outliers
- It produces long, elongated clusters

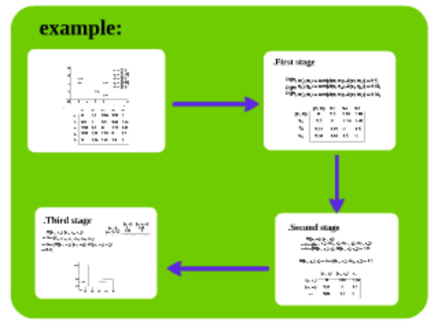
The Complete Link Method

Unlike the single-link method, the complete link method uses the farthest neighbor distance.

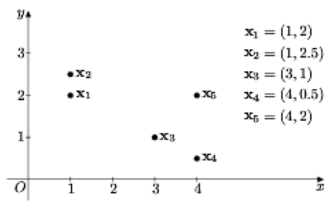
- **Complete-link distance** between clusters C_i and C_j is the **maximum distance** between any object in C_i and any object in C_j



$$D_{\text{complete}} = \max_{x,y}\{d(x,y) \mid x \in C_i, y \in C_j\}$$



example:



	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\mathbf{x}_1$	0	0.5	2.24	3.35	3
$\mathbf{x}_2$	0.5	0	2.5	3.61	3.04
$\mathbf{x}_3$	2.24	2.5	0	1.12	1.41
$\mathbf{x}_4$	3.35	3.61	1.12	0	1.5
$\mathbf{x}_5$	3	3.04	1.41	1.5	0

.First stage

$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3) = \max\{d(\mathbf{x}_1, \mathbf{x}_3), d(\mathbf{x}_2, \mathbf{x}_3)\} = 2.5,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4) = \max\{d(\mathbf{x}_1, \mathbf{x}_4), d(\mathbf{x}_2, \mathbf{x}_4)\} = 3.61,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5) = \max\{d(\mathbf{x}_1, \mathbf{x}_5), d(\mathbf{x}_2, \mathbf{x}_5)\} = 3.04,$$

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.5	3.61	3.04
$\mathbf{x}_3$	2.5	0	1.12	1.41
$\mathbf{x}_4$	3.61	1.12	0	1.5
$\mathbf{x}_5$	3.04	1.41	1.5	0

.Second stage

$$\begin{aligned} &D(\{\mathbf{x}_3, \mathbf{x}_4\}, \{\mathbf{x}_1, \mathbf{x}_2\}) \\ &= \max\{d(\mathbf{x}_1, \mathbf{x}_3), d(\mathbf{x}_2, \mathbf{x}_3), d(\mathbf{x}_1, \mathbf{x}_4), d(\mathbf{x}_2, \mathbf{x}_4)\} \\ &= \max\{D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3), D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4)\} = 3.61 \end{aligned}$$

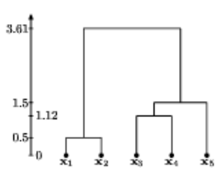
$$D(\{\mathbf{x}_3, \mathbf{x}_4\}, \mathbf{x}_5) = \max\{d(\mathbf{x}_3, \mathbf{x}_5), d(\mathbf{x}_4, \mathbf{x}_5)\} = 1.5.$$

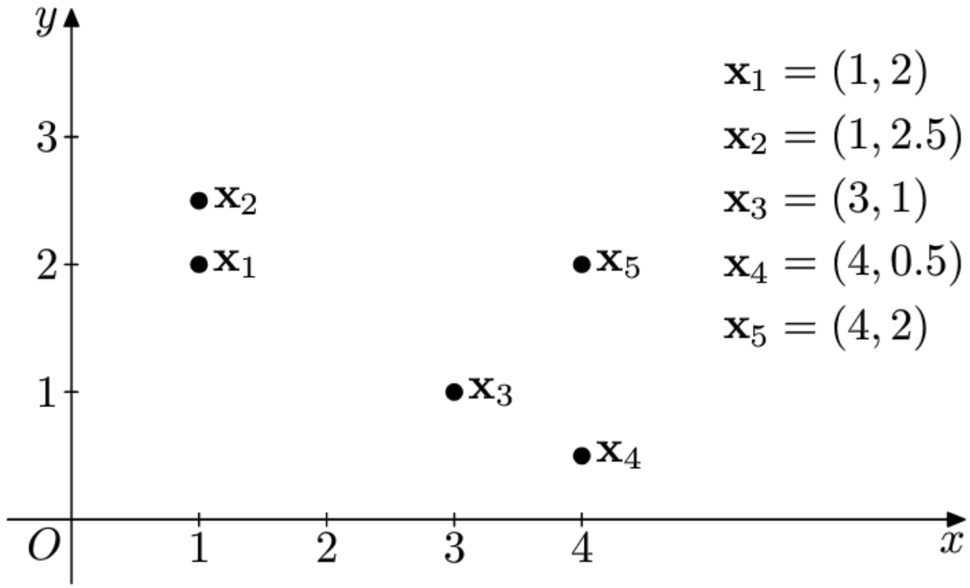
	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4\}$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	3.61	3.04
$\{\mathbf{x}_3, \mathbf{x}_4\}$	3.61	0	1.5
$\mathbf{x}_5$	3.04	1.5	0

.Third stage

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	3.61
$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$	3.61	0

$$\begin{aligned} &D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}) \\ &= \max\{d_{13}, d_{14}, d_{15}, d_{23}, d_{24}, d_{25}\} \\ &= \max\{D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4\}), D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5)\} \\ &= 3.61, \end{aligned}$$





	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\mathbf{x}_1$	0	0.5	2.24	3.35	3
$\mathbf{x}_2$	0.5	0	2.5	3.61	3.04
$\mathbf{x}_3$	2.24	2.5	0	1.12	1.41
$\mathbf{x}_4$	3.35	3.61	1.12	0	1.5
$\mathbf{x}_5$	3	3.04	1.41	1.5	0

.First stage

$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3) = \max\{d(\mathbf{x}_1, \mathbf{x}_3), d(\mathbf{x}_2, \mathbf{x}_3)\} = 2.5,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4) = \max\{d(\mathbf{x}_1, \mathbf{x}_4), d(\mathbf{x}_2, \mathbf{x}_4)\} = 3.61,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5) = \max\{d(\mathbf{x}_1, \mathbf{x}_5), d(\mathbf{x}_2, \mathbf{x}_5)\} = 3.04,$$

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.5	3.61	3.04
$\mathbf{x}_3$	2.5	0	1.12	1.41
$\mathbf{x}_4$	3.61	1.12	0	1.5
$\mathbf{x}_5$	3.04	1.41	1.5	0

.Second stage

$$\begin{aligned} &D(\{\mathbf{x}_3, \mathbf{x}_4\}, \{\mathbf{x}_1, \mathbf{x}_2\}) \\ &= \max\{d(\mathbf{x}_1, \mathbf{x}_3), d(\mathbf{x}_2, \mathbf{x}_3), d(\mathbf{x}_1, \mathbf{x}_4), d(\mathbf{x}_2, \mathbf{x}_4)\} \\ &= \max\{D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3), D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4)\} = 3.61 \end{aligned}$$

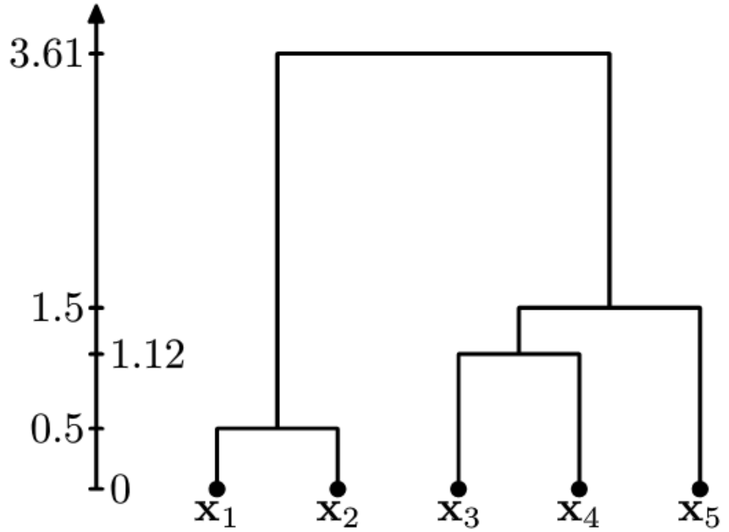
$$D(\{\mathbf{x}_3, \mathbf{x}_4\}, \mathbf{x}_5) = \max\{d(\mathbf{x}_3, \mathbf{x}_5), d(\mathbf{x}_4, \mathbf{x}_5)\} = 1.5.$$

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4\}$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	3.61	3.04
$\{\mathbf{x}_3, \mathbf{x}_4\}$	3.61	0	1.5
$\mathbf{x}_5$	3.04	1.5	0

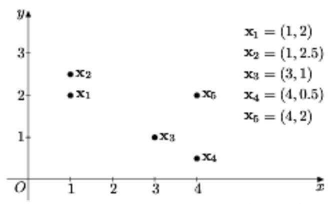
.Third stage

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	3.61
$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$	3.61	0

$$
\begin{aligned}
& D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}) \\
&= \max\{d_{13}, d_{14}, d_{15}, d_{23}, d_{24}, d_{25}\} \\
&= \max\{D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4\}), D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5)\} \\
&= 3.61,
\end{aligned}
$$



example:



	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\mathbf{x}_1$	0	0.5	2.24	3.35	3
$\mathbf{x}_2$	0.5	0	2.5	3.61	3.04
$\mathbf{x}_3$	2.24	2.5	0	1.12	1.41
$\mathbf{x}_4$	3.35	3.61	1.12	0	1.5
$\mathbf{x}_5$	3	3.04	1.41	1.5	0

.First stage

$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3) = \max\{d(\mathbf{x}_1, \mathbf{x}_3), d(\mathbf{x}_2, \mathbf{x}_3)\} = 2.5,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4) = \max\{d(\mathbf{x}_1, \mathbf{x}_4), d(\mathbf{x}_2, \mathbf{x}_4)\} = 3.61,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5) = \max\{d(\mathbf{x}_1, \mathbf{x}_5), d(\mathbf{x}_2, \mathbf{x}_5)\} = 3.04,$$

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.5	3.61	3.04
$\mathbf{x}_3$	2.5	0	1.12	1.41
$\mathbf{x}_4$	3.61	1.12	0	1.5
$\mathbf{x}_5$	3.04	1.41	1.5	0

.Second stage

$$\begin{aligned} &D(\{\mathbf{x}_3, \mathbf{x}_4\}, \{\mathbf{x}_1, \mathbf{x}_2\}) \\ &= \max\{d(\mathbf{x}_1, \mathbf{x}_3), d(\mathbf{x}_2, \mathbf{x}_3), d(\mathbf{x}_1, \mathbf{x}_4), d(\mathbf{x}_2, \mathbf{x}_4)\} \\ &= \max\{D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3), D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4)\} = 3.61 \end{aligned}$$

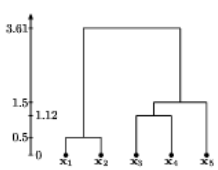
$$D(\{\mathbf{x}_3, \mathbf{x}_4\}, \mathbf{x}_5) = \max\{d(\mathbf{x}_3, \mathbf{x}_5), d(\mathbf{x}_4, \mathbf{x}_5)\} = 1.5.$$

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4\}$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	3.61	3.04
$\{\mathbf{x}_3, \mathbf{x}_4\}$	3.61	0	1.5
$\mathbf{x}_5$	3.04	1.5	0

.Third stage

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	3.61
$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$	3.61	0

$$\begin{aligned} &D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}) \\ &= \max\{d_{13}, d_{14}, d_{15}, d_{23}, d_{24}, d_{25}\} \\ &= \max\{D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4\}), D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5)\} \\ &= 3.61, \end{aligned}$$

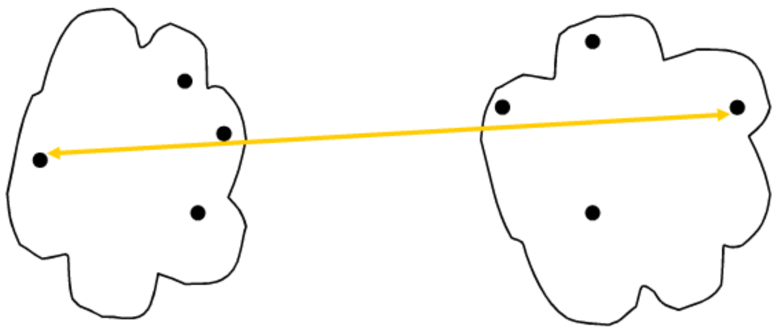


Prezi

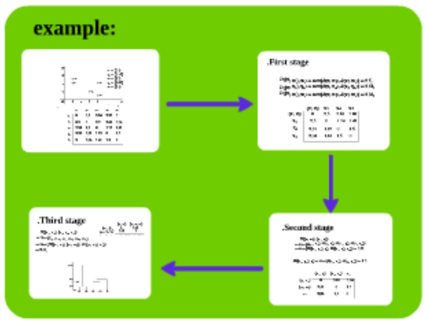
The Complete Link Method

Unlike the single-link method, the complete link method uses the farthest neighbor distance.

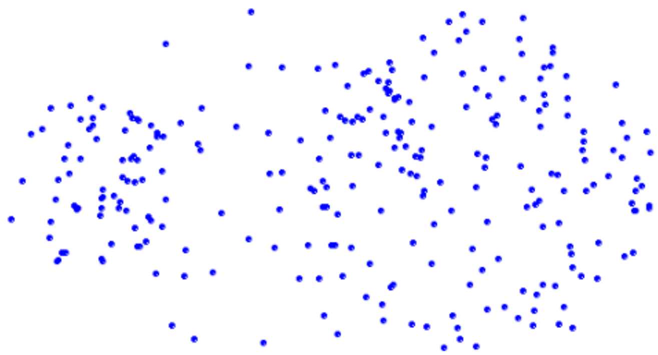
- **Complete-link distance** between clusters C_i and C_j is the **maximum distance** between any object in C_i and any object in C_j



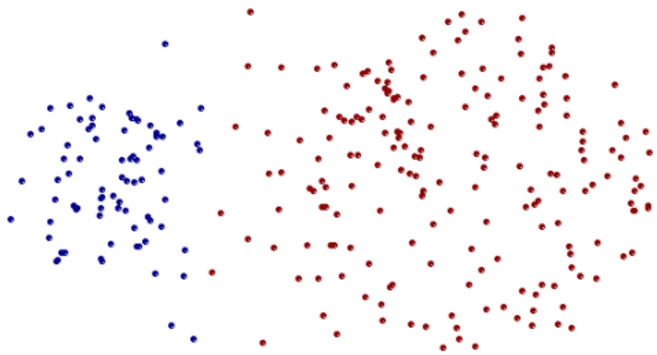
$$D_{\text{complete}} = \max_{x,y}\{d(x,y) \mid x \in C_i, y \in C_j\}$$



Strengths of complete-link clustering



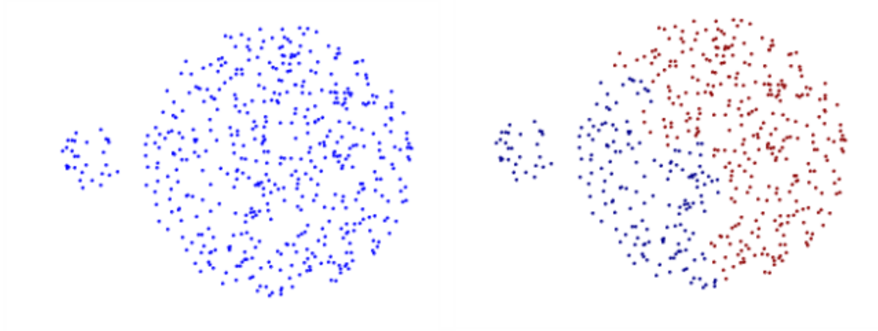
Original Points



Two Clusters

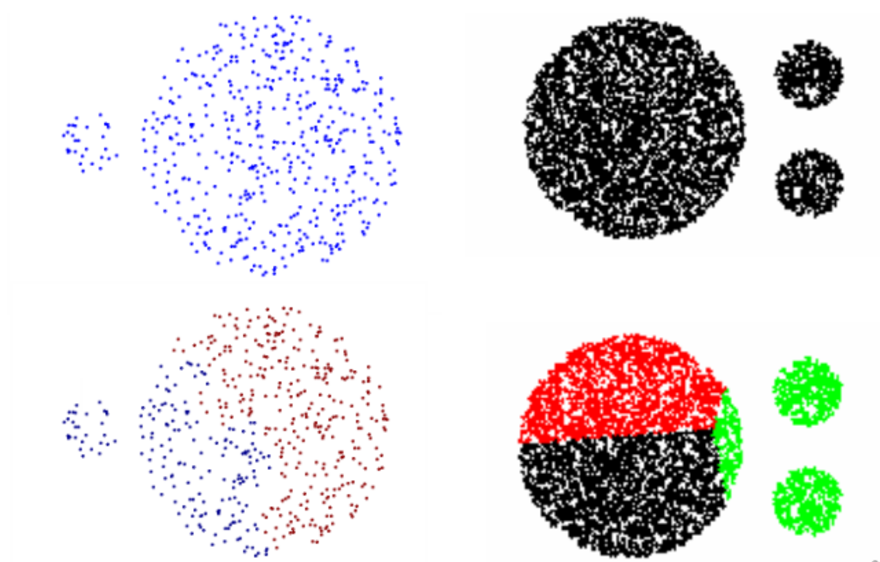
- More balanced clusters (with equal diameter)
- Less susceptible to noise

Limitations of complete-link clustering

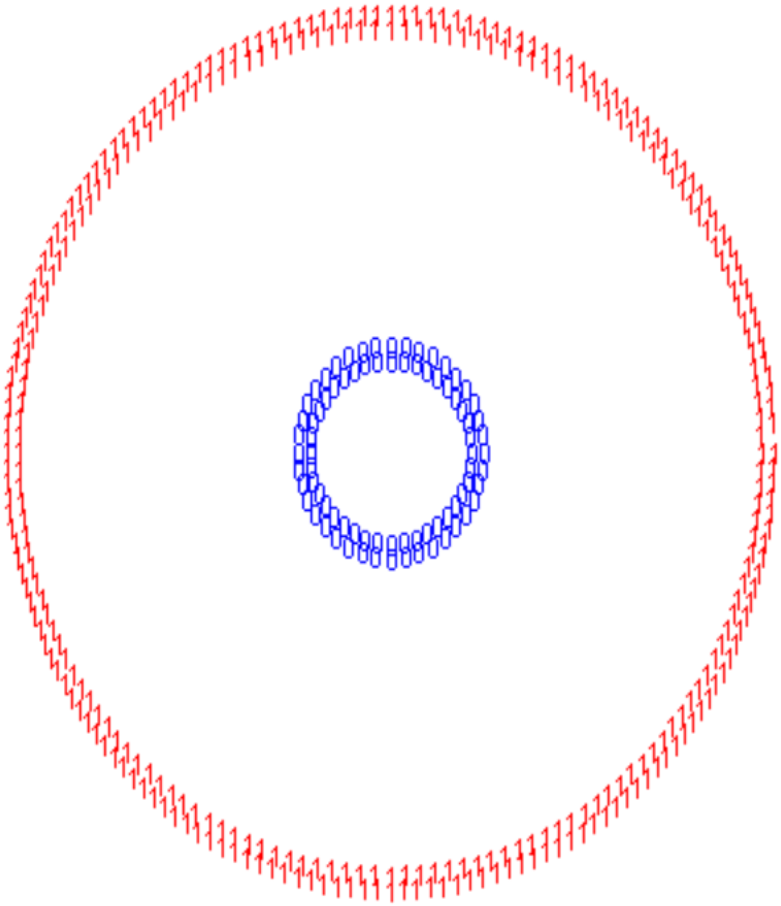


Original Points Two Clusters

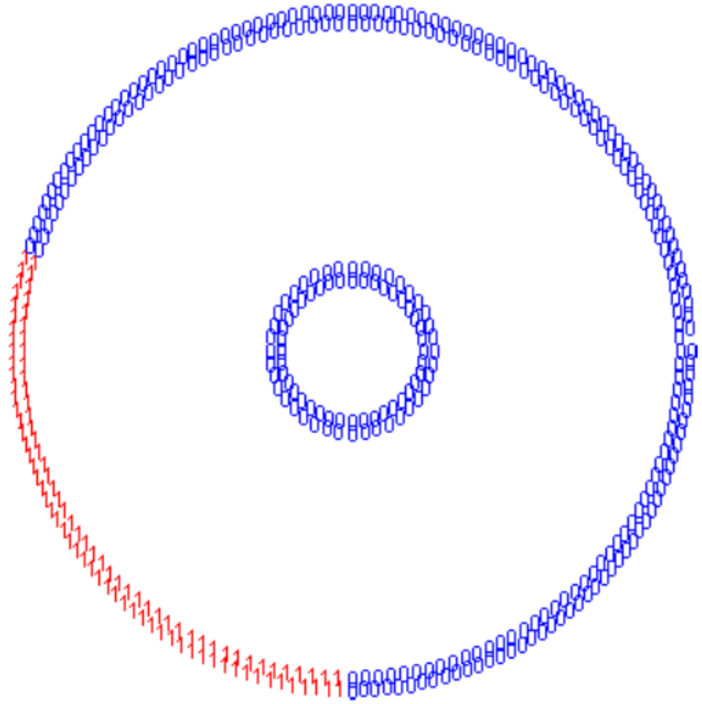
- Tends to break large clusters
- All clusters tend to have the same diameter – small clusters are merged with larger ones



Limitations of MAX



MIN (2 clusters)

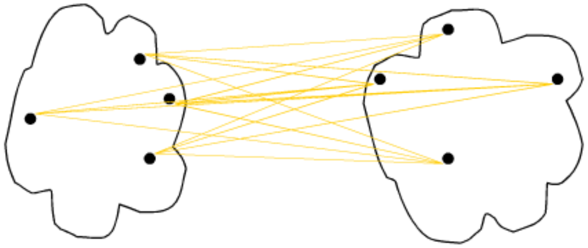


MAX (2 clusters)

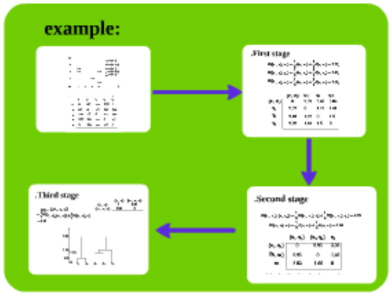
The Group Average Method

The group average method is also referred as UPGMA, which stands for "unweighted pair group method using arithmetic averages" (Jain and Dubes, 1988). In the group average method, the distance between two groups is defined as the average of the distances between all possible pairs of data points that are made up of one data point from each group.

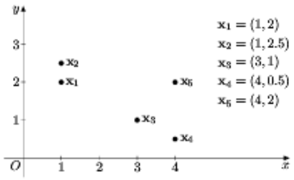
Let C_i, C_j, and C_k be three groups of data points. Then the distance between C_k and $C_i \cup C_j$ can be obtained from the Lance-Williams formula as follows:



$$D(C_k, C_i \cup C_j) = \frac{|C_i|}{|C_i| + |C_j|} D(C_k, C_i) + \frac{|C_j|}{|C_i| + |C_j|} D(C_k, C_j),$$



example:



$\mathbf{x}_1 = (1, 2)$
$\mathbf{x}_2 = (1, 2.5)$
$\mathbf{x}_3 = (3, 1)$
$\mathbf{x}_4 = (4, 0.5)$
$\mathbf{x}_5 = (4, 2)$

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\mathbf{x}_1$	0	0.5	2.24	3.35	3
$\mathbf{x}_2$	0.5	0	2.5	3.61	3.04
$\mathbf{x}_3$	2.24	2.5	0	1.12	1.41
$\mathbf{x}_4$	3.35	3.61	1.12	0	1.5
$\mathbf{x}_5$	3	3.04	1.41	1.5	0

.First stage

$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3) = \frac{1}{2}d(\mathbf{x}_1, \mathbf{x}_3) + \frac{1}{2}d(\mathbf{x}_2, \mathbf{x}_3) = 2.37,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4) = \frac{1}{2}d(\mathbf{x}_1, \mathbf{x}_4) + \frac{1}{2}d(\mathbf{x}_2, \mathbf{x}_4) = 3.48,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5) = \frac{1}{2}d(\mathbf{x}_1, \mathbf{x}_5) + \frac{1}{2}d(\mathbf{x}_2, \mathbf{x}_5) = 3.02,$$

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.37	3.48	3.02
$\mathbf{x}_3$	2.37	0	1.12	1.41
$\mathbf{x}_4$	3.48	1.12	0	1.5
$\mathbf{x}_5$	3.02	1.41	1.5	0

.Second stage

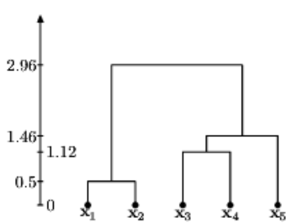
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4\}) = \frac{1}{2}D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4) + \frac{1}{2}D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3) = 2.93,$$
$$D(\{\mathbf{x}_3, \mathbf{x}_4\}, \mathbf{x}_5) = \frac{1}{2}d(\mathbf{x}_3, \mathbf{x}_5) + \frac{1}{2}d(\mathbf{x}_4, \mathbf{x}_5) = 1.46.$$

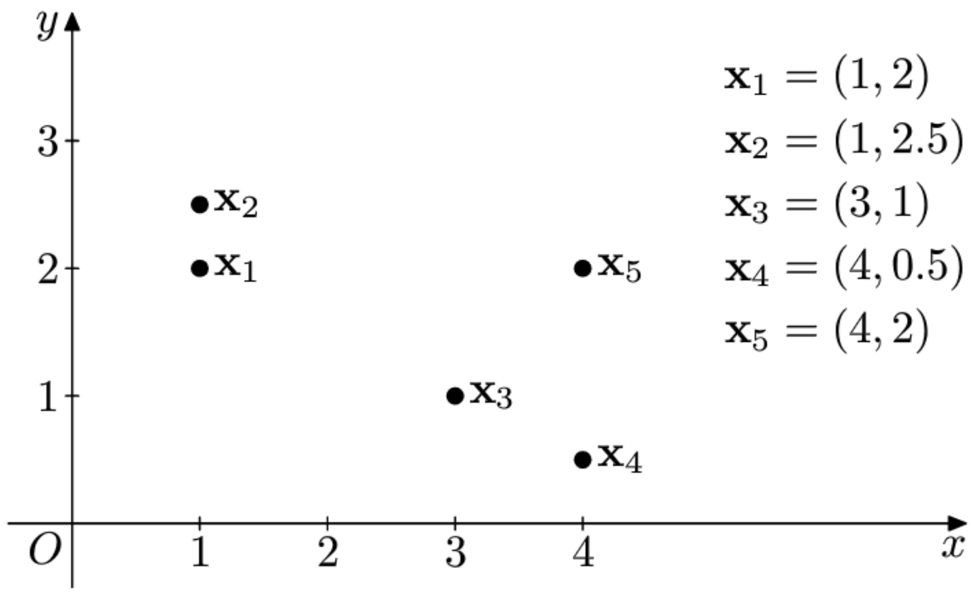
	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4\}$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.93	3.02
$\{\mathbf{x}_3, \mathbf{x}_4\}$	2.93	0	1.46
$\mathbf{x}_5$	3.02	1.46	0

.Third stage

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.96
$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$	2.96	0

$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\})$$
$$= \frac{2}{3}D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4\}) + \frac{1}{3}D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5)$$
$$= 2.96.$$





	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\mathbf{x}_1$	0	0.5	2.24	3.35	3
$\mathbf{x}_2$	0.5	0	2.5	3.61	3.04
$\mathbf{x}_3$	2.24	2.5	0	1.12	1.41
$\mathbf{x}_4$	3.35	3.61	1.12	0	1.5
$\mathbf{x}_5$	3	3.04	1.41	1.5	0

.First stage

$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3) = \frac{1}{2}d(\mathbf{x}_1, \mathbf{x}_3) + \frac{1}{2}d(\mathbf{x}_2, \mathbf{x}_3) = 2.37,$$

$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4) = \frac{1}{2}d(\mathbf{x}_1, \mathbf{x}_4) + \frac{1}{2}d(\mathbf{x}_2, \mathbf{x}_4) = 3.48,$$

$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5) = \frac{1}{2}d(\mathbf{x}_1, \mathbf{x}_5) + \frac{1}{2}d(\mathbf{x}_2, \mathbf{x}_5) = 3.02,$$

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.37	3.48	3.02
$\mathbf{x}_3$	2.37	0	1.12	1.41
$\mathbf{x}_4$	3.48	1.12	0	1.5
$\mathbf{x}_5$	3.02	1.41	1.5	0

.Second stage

$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4\}) = \frac{1}{2}D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4) + \frac{1}{2}D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3) = 2.93,$$

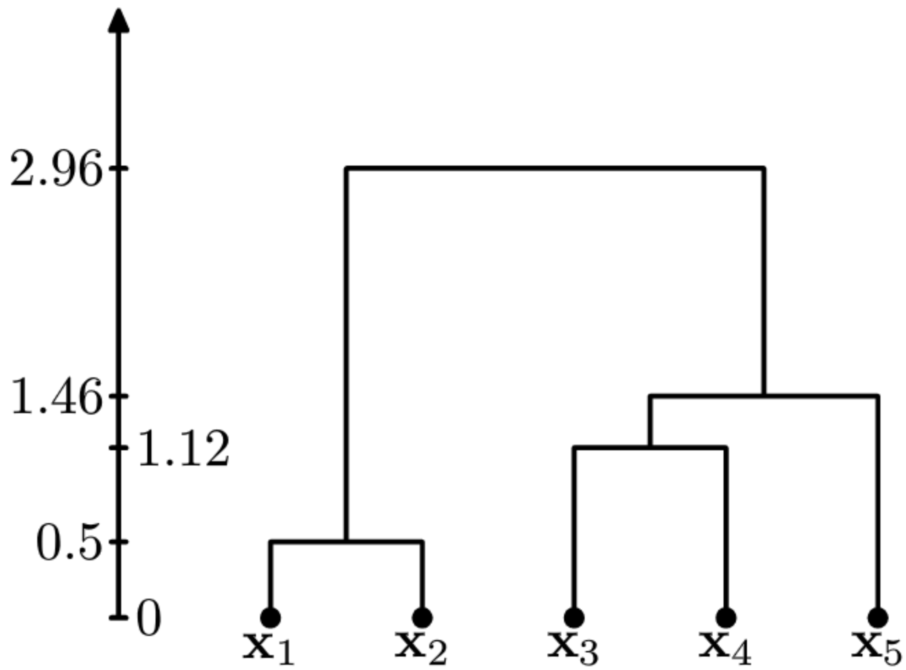
$$D(\{\mathbf{x}_3, \mathbf{x}_4\}, \mathbf{x}_5) = \frac{1}{2}d(\mathbf{x}_3, \mathbf{x}_5) + \frac{1}{2}d(\mathbf{x}_4, \mathbf{x}_5) = 1.46.$$

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4\}$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.93	3.02
$\{\mathbf{x}_3, \mathbf{x}_4\}$	2.93	0	1.46
$\mathbf{x}_5$	3.02	1.46	0

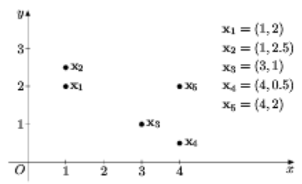
.Third stage

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.96
$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$	2.96	0

$$
\begin{aligned}
&D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}) \\
&= \frac{2}{3} D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4\}) + \frac{1}{3} D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5) \\
&= 2.96.
\end{aligned}
$$



example:



	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\mathbf{x}_1$	0	0.5	2.24	3.35	3
$\mathbf{x}_2$	0.5	0	2.5	3.61	3.04
$\mathbf{x}_3$	2.24	2.5	0	1.12	1.41
$\mathbf{x}_4$	3.35	3.61	1.12	0	1.5
$\mathbf{x}_5$	3	3.04	1.41	1.5	0

.First stage

$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3) = \frac{1}{2} d(\mathbf{x}_1, \mathbf{x}_3) + \frac{1}{2} d(\mathbf{x}_2, \mathbf{x}_3) = 2.37,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4) = \frac{1}{2} d(\mathbf{x}_1, \mathbf{x}_4) + \frac{1}{2} d(\mathbf{x}_2, \mathbf{x}_4) = 3.48,$$
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5) = \frac{1}{2} d(\mathbf{x}_1, \mathbf{x}_5) + \frac{1}{2} d(\mathbf{x}_2, \mathbf{x}_5) = 3.02,$$

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.37	3.48	3.02
$\mathbf{x}_3$	2.37	0	1.12	1.41
$\mathbf{x}_4$	3.48	1.12	0	1.5
$\mathbf{x}_5$	3.02	1.41	1.5	0

.Second stage

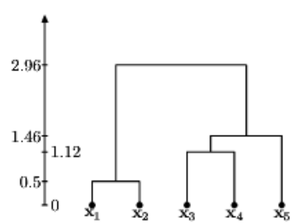
$$D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4\}) = \frac{1}{2} D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_4) + \frac{1}{2} D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_3) = 2.93,$$
$$D(\{\mathbf{x}_3, \mathbf{x}_4\}, \mathbf{x}_5) = \frac{1}{2} d(\mathbf{x}_3, \mathbf{x}_5) + \frac{1}{2} d(\mathbf{x}_4, \mathbf{x}_5) = 1.46.$$

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4\}$	$\mathbf{x}_5$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.93	3.02
$\{\mathbf{x}_3, \mathbf{x}_4\}$	2.93	0	1.46
$\mathbf{x}_5$	3.02	1.46	0

.Third stage

	$\{\mathbf{x}_1, \mathbf{x}_2\}$	$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$
$\{\mathbf{x}_1, \mathbf{x}_2\}$	0	2.96
$\{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$	2.96	0

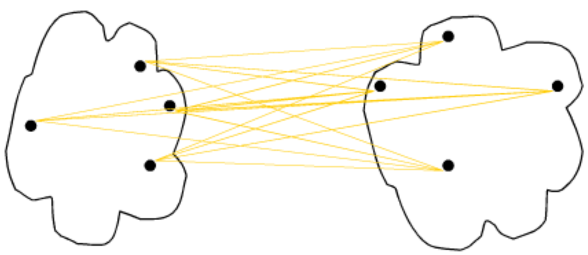
$$\begin{aligned}
&D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}) \\
&= \frac{2}{3} D(\{\mathbf{x}_1, \mathbf{x}_2\}, \{\mathbf{x}_3, \mathbf{x}_4\}) + \frac{1}{3} D(\{\mathbf{x}_1, \mathbf{x}_2\}, \mathbf{x}_5) \\
&= 2.96.
\end{aligned}$$



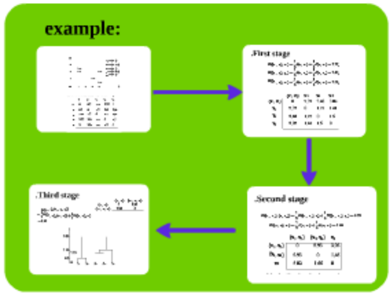
The Group Average Method

The group average method is also referred as UPGMA, which stands for “unweighted pair group method using arithmetic averages” (Jain and Dubes, 1988). In the group average method, the distance between two groups is defined as the average of the distances between all possible pairs of data points that are made up of one data point from each group.

Let C_i, C_j, and C_k be three groups of data points. Then the distance between C_k and $C_i \cup C_j$ can be obtained from the Lance-Williams formula as follows:



$$
\begin{aligned}
& D(C_k, C_i \cup C_j) \\
= & \frac{|C_i|}{|C_i| + |C_j|} D(C_k, C_i) + \frac{|C_j|}{|C_i| + |C_j|} D(C_k, C_j),
\end{aligned}
$$



Average-link clustering: discussion

- Compromise between Single and Complete Link
- Strengths
  - Less susceptible to noise and outliers
- Limitations
  - Biased towards globular clusters

The Weighted Group Average Method

The weighted group average method is also referred to as the "weighted pair group method using arithmetic average" (Jain and Dubes, 1988). Using the Lance-Williams formula, the distance between clusters is

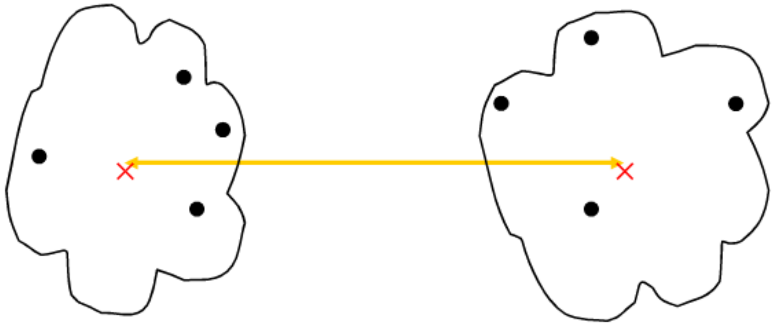
$$D(C_k, C_i \cup C_j) = \frac{1}{2} D(C_k, C_i) + \frac{1}{2} D(C_k, C_j),$$

where C_k, C_i, and C_j are three clusters in one level of clustering.

The Centroid Method

The centroid method is also referred to as the "unweighted pair group method using centroids" (Jain and Dubes, 1988). With the centroid method, the new distances between clusters can be calculated by the following Lance-Williams formula:

$$\begin{aligned} & D(C_k, C_i \cup C_j) \\ = & \frac{|C_i|}{|C_i| + |C_j|} D(C_k, C_i) + \frac{|C_j|}{|C_i| + |C_j|} D(C_k, C_j) \\ & - \frac{|C_i||C_j|}{(|C_i| + |C_j|)^2} D(C_i, C_j), \end{aligned}$$



where C_k, C_i, and C_j are three clusters in one level of clustering.

The distance between two clusters is represented by the distance between the centers of the clusters

The Median Method

The median method is also referred to as the “weighted pair group method using centroids”(Jain and Dubes, 1988) or the “weighted centroid” method.

In the median method, the distances between newly formed groups and other groups are computed as:

$$D(C_k, C_i \cup C_j) = \frac{1}{2}D(C_k, C_i) + \frac{1}{2}D(C_k, C_j) - \frac{1}{4}D(C_i, C_j),$$

The Ward Method

- Ward's distance between clusters C_i and C_j is the difference between the total within cluster sum of squares for the two clusters separately, and the within cluster sum of squares resulting from merging the two clusters in cluster C_{ij}

$$D_W(C_i, C_j) = \sum_{x \in C_i} (x - r_i)^2 + \sum_{x \in C_j} (x - r_j)^2 - \sum_{x \in C_{ij}} (x - r_{ij})^2$$

- $\mathbf{r_i}$: centroid of $\mathbf{C_i}$
- $\mathbf{r_j}$: centroid of $\mathbf{C_j}$
- $\mathbf{r_{ij}}$: centroid of $\mathbf{C_{ij}}$

Ward’s distance for clusters

- Similar to group average and centroid distance
- Less susceptible to noise and outliers

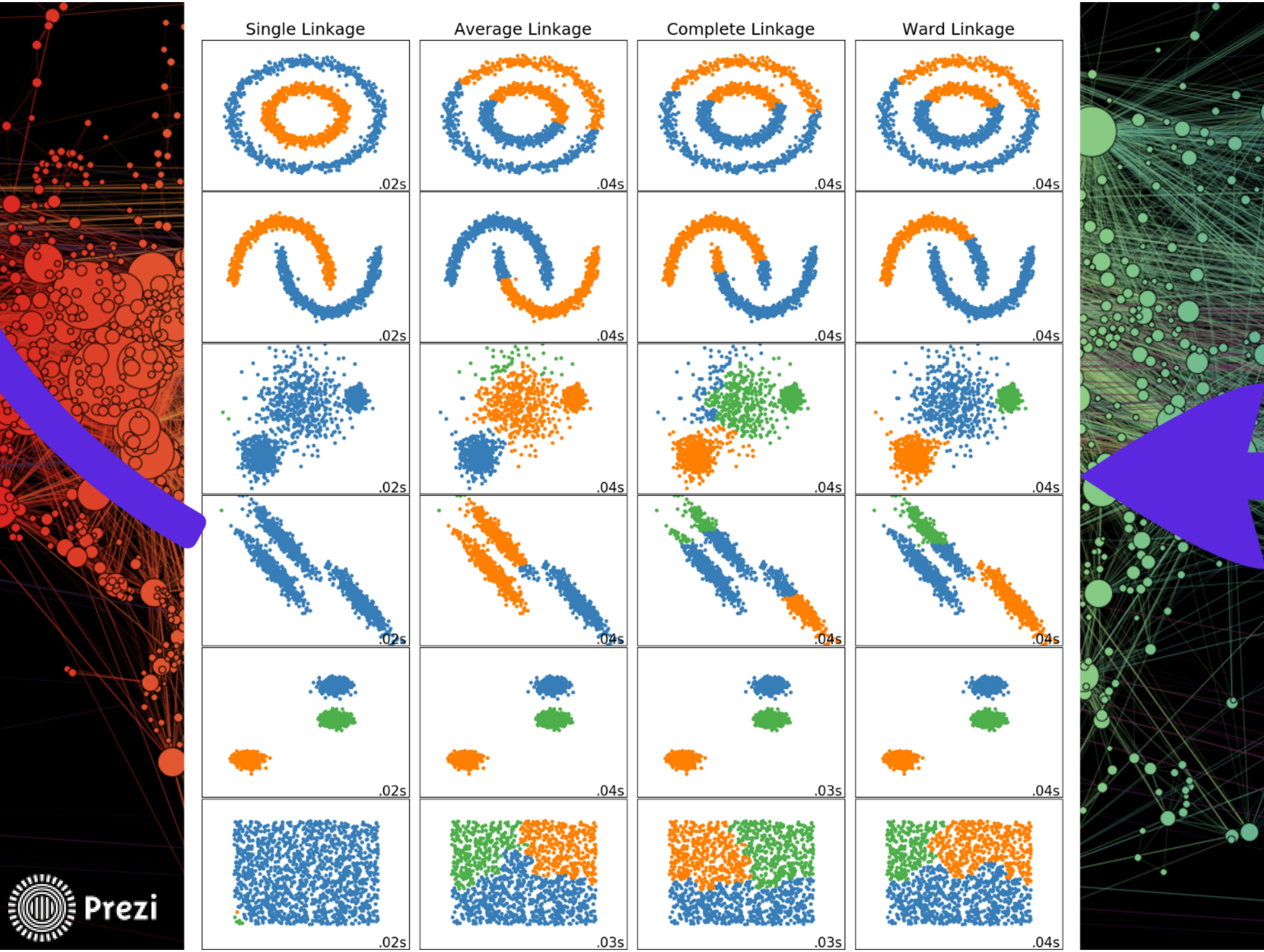
Hierarchical Clustering: Time and Space requirements

- For a dataset **X** consisting of **n** points
- **$O(n^2)$ space**; it requires storing the distance matrix
- **$O(n^3)$ time** in most of the cases
  - There are **n** steps and at each step the size **n^2** distance matrix must be updated and searched
  - Complexity can be reduced to **$O(n^2 \log(n))$** time for some approaches by using appropriate data structures

Strength of Hierachical Clustering

. No assumptions on the number of clusters

- Any desired number of clusters can be obtianed by 'cutting' the dendogram at the proper level


Single Linkage
Average Linkage
Complete Linkage
Ward Linkage
.02s
.04s
.04s
.04s
.02s
.04s
.04s
.04s
.02s
.04s
.04s
.04s
.02s
.04s
.04s
.04s
.02s
.04s
.03s
.04s
.02s
.03s
.03s
.04s
Prezi

Thank You

[**presented by**]

[**Reza Sahleh**]