CS 4104
APPLIED MACHINE LEARNING

Dr. Hashim Yasin

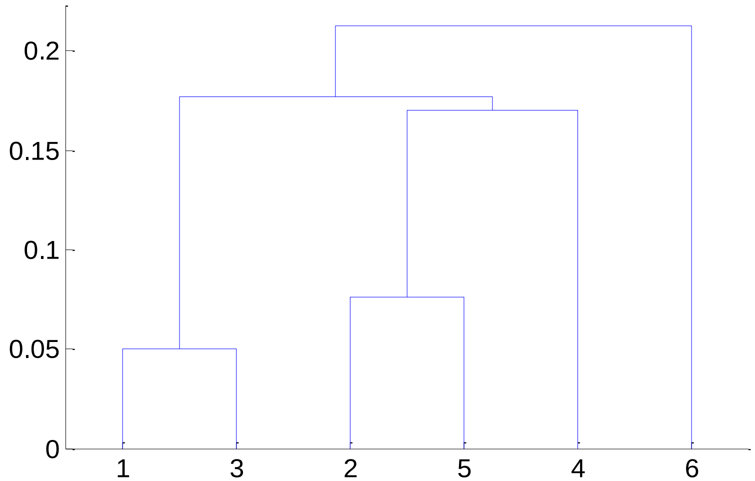
National University of Computer and Emerging Sciences,

Faisalabad, Pakistan.

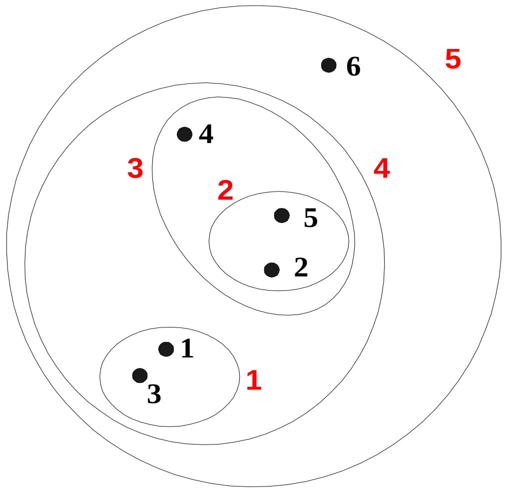
HIERARCHICAL CLUSTERING

Hierarchical Clustering

- Produces a set of nested clusters organized as a hierarchical tree
  - Can be visualized as a dendrogram
    - A tree like diagram that records the sequences of merges or splits



dendrogram



Venn diagram

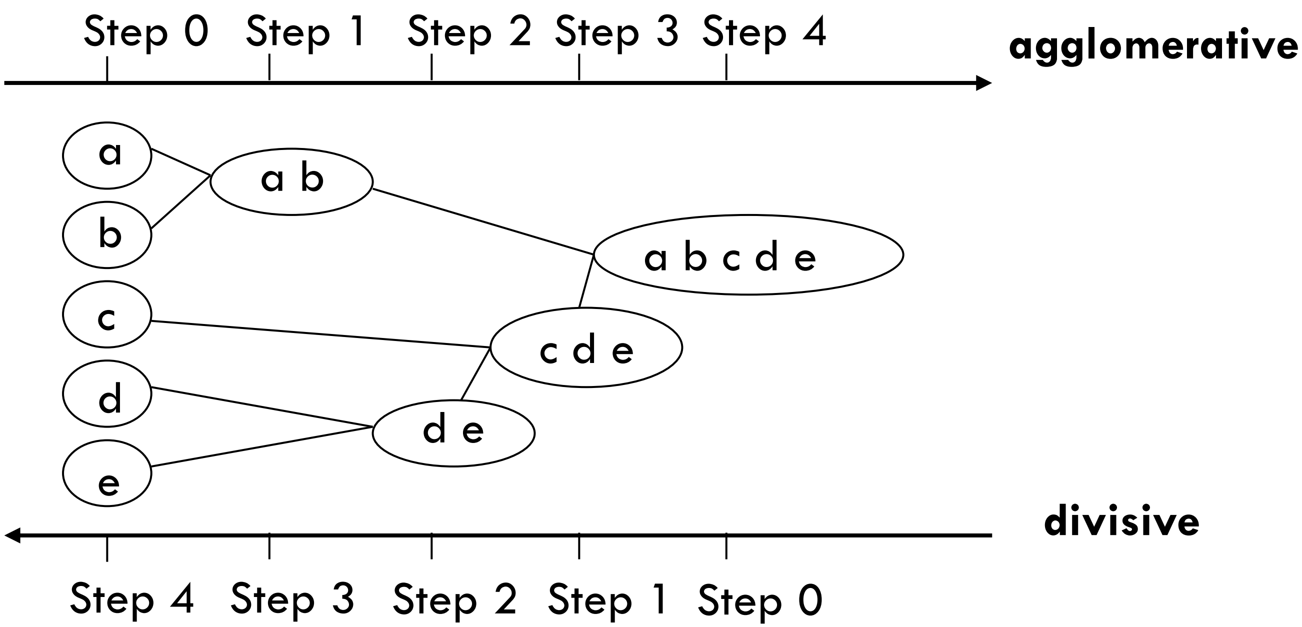
Hierarchical Clustering

- 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

Hierarchical Clustering

- Use **distance matrix** as clustering criteria. *This method does not require the number of clusters **k** as an input but needs a termination condition.*



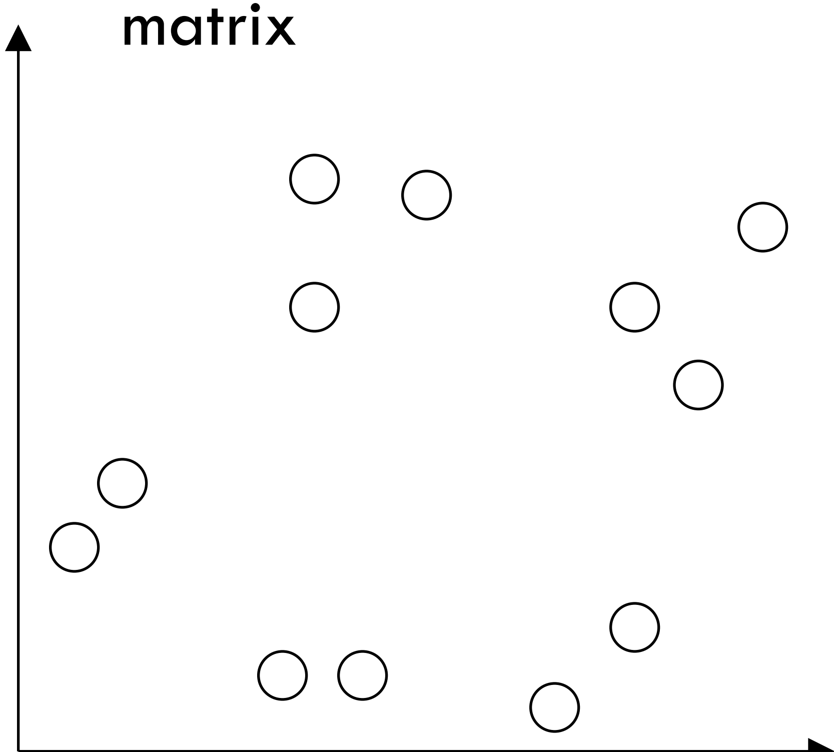
AGGLOMERATIVE HIERARCHICAL CLUSTERING

Agglomerative Hierarchical Clustering

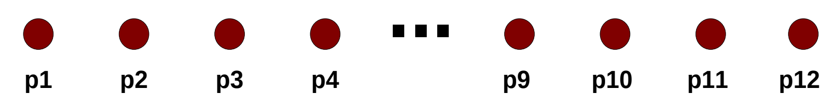
- More popular hierarchical clustering technique
- Basic algorithm is straightforward
  1. Compute the proximity matrix
  2. Let each data point be a cluster
  3. **Repeat**
  4. Merge the two closest clusters
  5. Update the proximity matrix
  6. **Until** only a single cluster remains
- *Key operation is the computation of the proximity of two clusters*
  - Different approaches to define the distance between clusters distinguish the different algorithms

Agglomerative Hierarchical Clustering

- Start with clusters of individual points and a proximity matrix

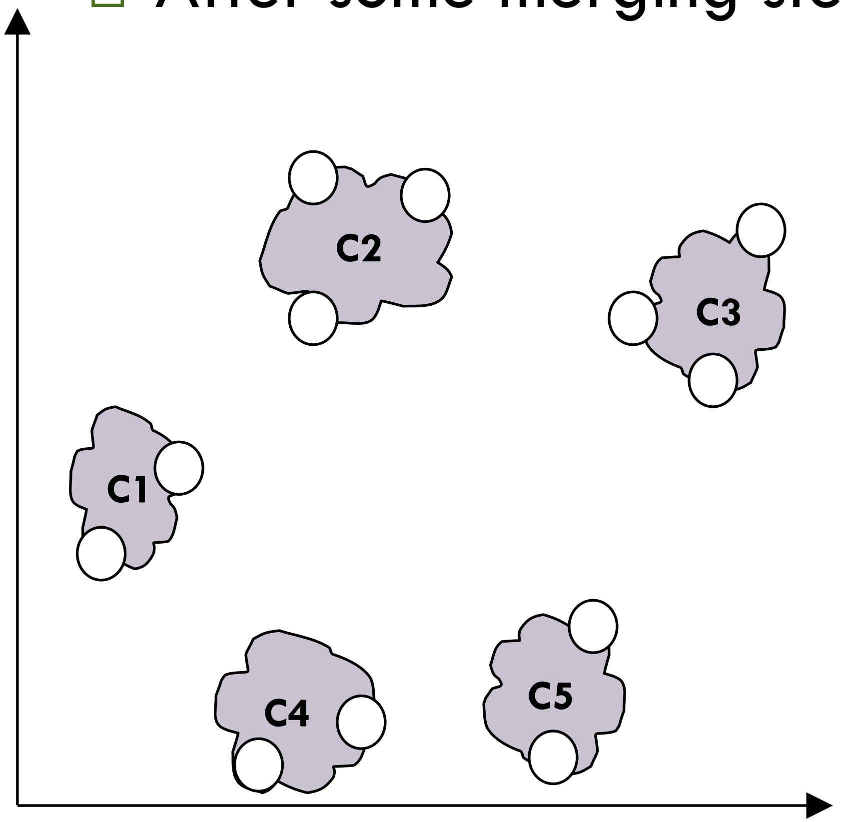


	p1	p2	p3	p4	p5	. . .
p1						
p2						
p3						
p4						
p5						
.						
.						
.						

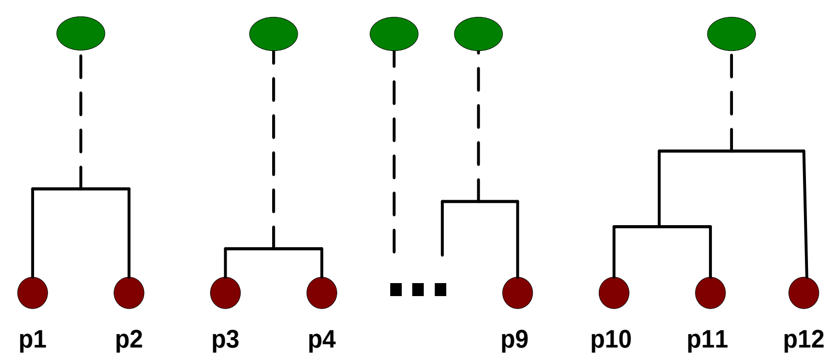


Agglomerative Hierarchical Clustering

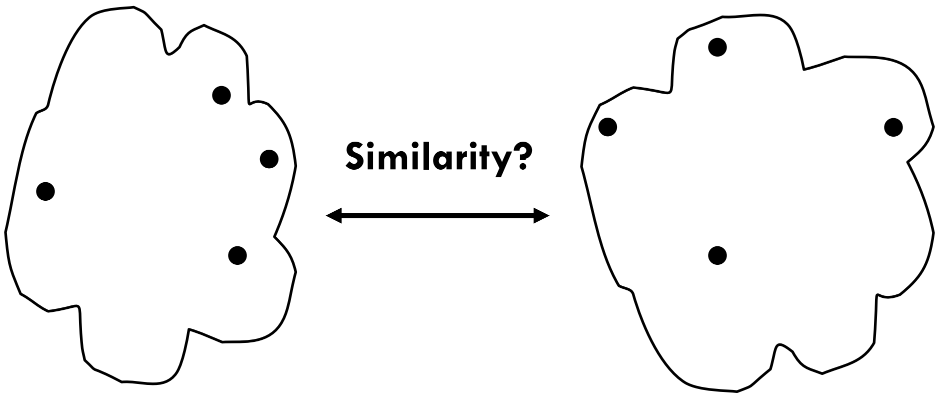
- After some merging steps, we have some clusters



	C1	C2	C3	C4	C5
C1					
C2					
C3					
C4					
C5					



Inter-Cluster Similarity

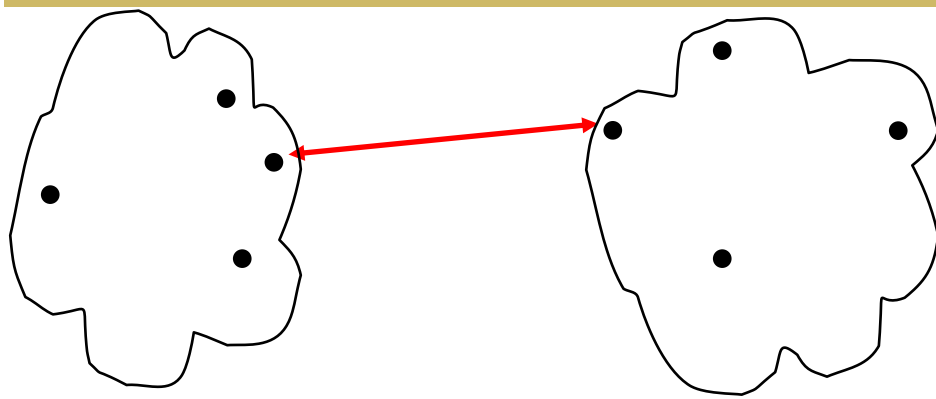


- MIN
- MAX
- Group Average
- Distance Between Centroids
- Other methods driven by an objective function
  - Ward's Method uses squared error

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

Proximity Matrix

Inter-Cluster Similarity

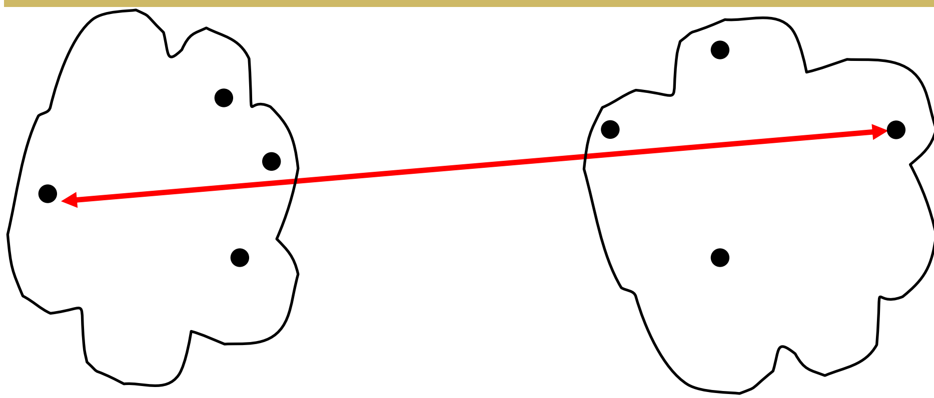


	p1	p2	p3	p4	p5	. . .
p1						
p2						
p3						
p4						
p5						
.						
.						
.						

Proximity Matrix

- MIN
- MAX
- Group Average
- Distance Between Centroids
- Other methods driven by an objective function
  - Ward's Method uses squared error

Inter-Cluster Similarity

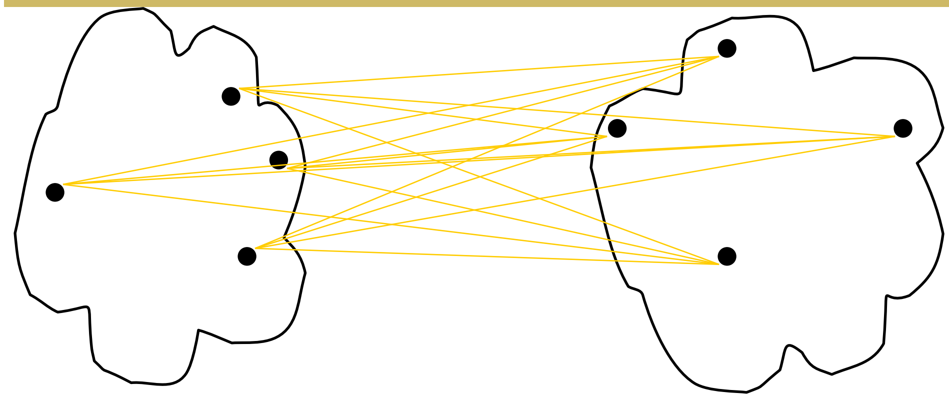


	p1	p2	p3	p4	p5	. . .
p1						
p2						
p3						
p4						
p5						
. . .						

Proximity Matrix

- MIN
- MAX
- Group Average
- Distance Between Centroids
- Other methods driven by an objective function
  - Ward's Method uses squared error

Inter-Cluster Similarity

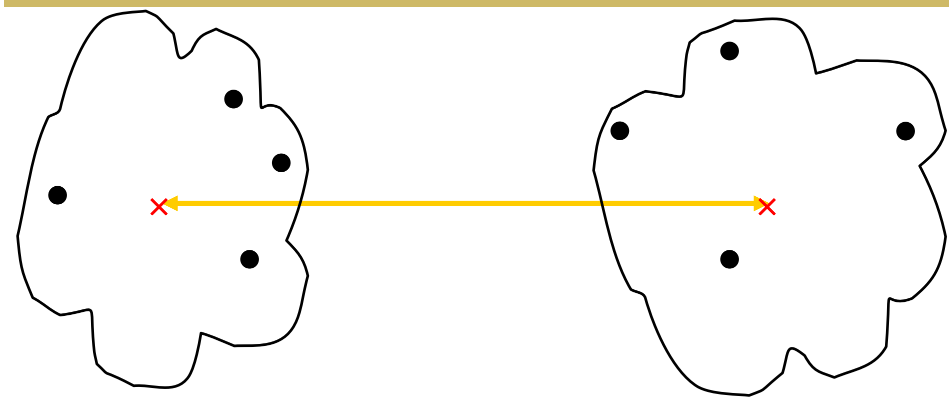


- MIN
- MAX
- Group Average
- Distance Between Centroids
- Other methods driven by an objective function
  - Ward's Method uses squared error

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

Proximity Matrix

Inter-Cluster Similarity

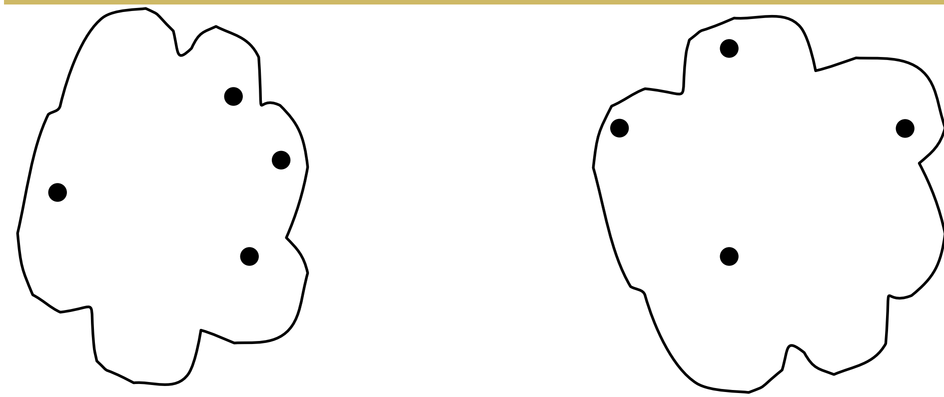


	p1	p2	p3	p4	p5	. . .
p1						
p2						
p3						
p4						
p5						
.						
.						
.						

Proximity Matrix

- MIN
- MAX
- Group Average
- Distance Between Centroids
- Other methods driven by an objective function
  - Ward's Method uses squared error

Inter-Cluster Similarity



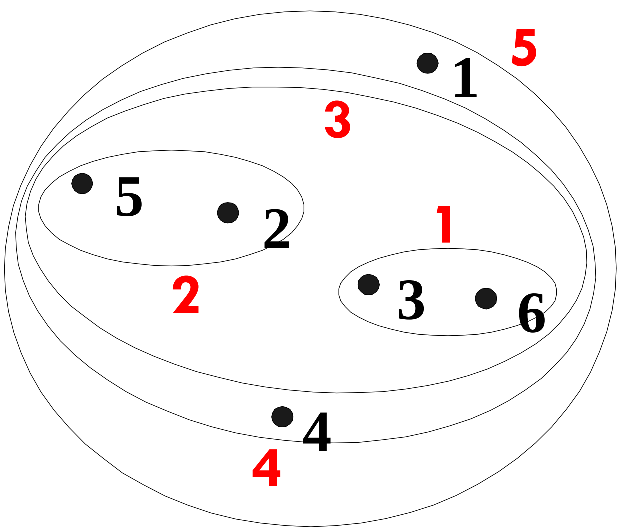
- MIN
- MAX
- Group Average
- Distance Between Centroids
- Other methods driven by an objective function
  - Ward's Method uses *squared error*

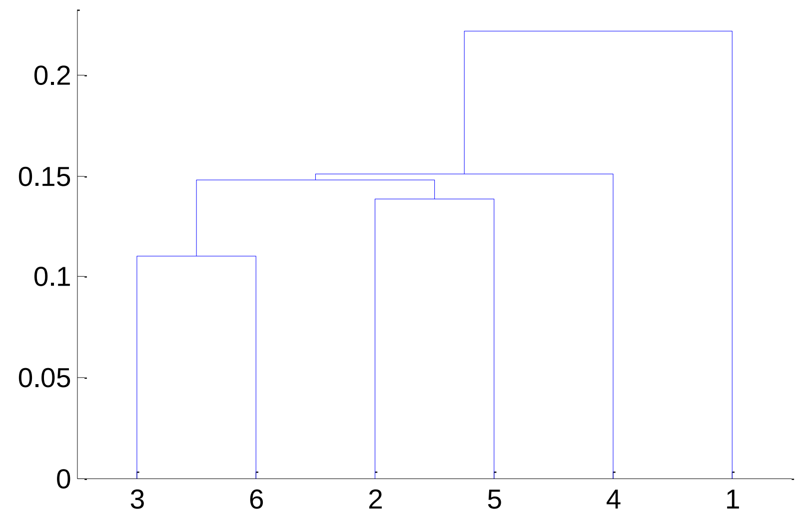
	p1	p2	p3	p4	p5	. . .
p1						
p2						
p3						
p4						
p5						
.						
.						
.						

Proximity Matrix

MIN or Single Link Linkage

- Similarity of two clusters is based on the **two most similar (closest) points** in the different clusters
  - Determined by one pair of points, i.e., by one link in the proximity graph.

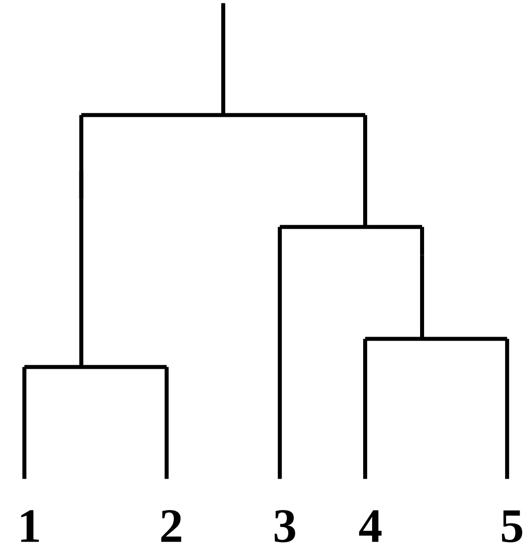




MAX or Complete Linkage

- Similarity of two clusters is based on the **two least similar (most distant) points** in the different clusters
  - Determined by the all pairs of points in the two clusters

	I1	I2	I3	I4	I5
I1	0.00	0.90	0.10	0.65	0.20
I2	0.90	0.00	0.70	0.60	0.50
I3	0.10	0.70	0.00	0.40	0.30
I4	0.65	0.60	0.40	0.00	0.80
I5	0.20	0.50	0.30	0.80	0.00



Group Average Linkage

- Proximity of two clusters is the **average of pairwise proximity between points** in the two clusters.

$$\text{proximity}(\text{Cluster}_i, \text{Cluster}_j) = \frac{\sum\limits_{\substack{p_i \in \text{Cluster}_i \\ p_j \in \text{Cluster}_j}} \text{proximity}(p_i, p_j)}{|\text{Cluster}_i| * |\text{Cluster}_j|}$$

- Need to use **average connectivity for scalability** since total proximity favors large clusters
- *Compromise between Single and Complete Link*
- Strengths --- Less susceptible to noise and outliers

Hierarchical Clustering

- All the algorithms are at least $O(n^2)$.
  - n is the number of data points.

- **Single link** can be done in $O(n^2)$.

- **Complete and average links** can be done in $O(n^2 \log n)$.

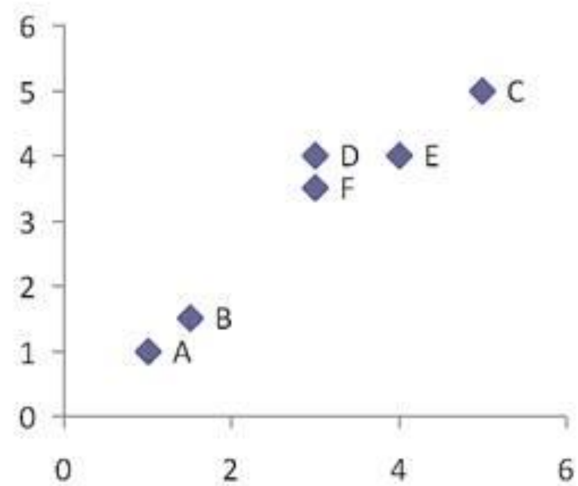
- Due the complexity, hard to use for large data sets.
  - Sampling may be the solution

SINGLE LINKAGE CLUSTERING EXAMPLE

Single Linkage Clustering: Example

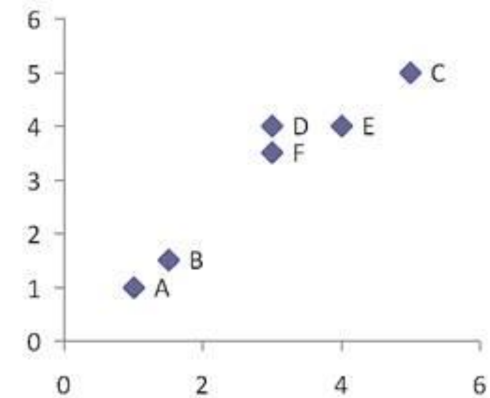
- Consider there are 6 objects (A, B, C, D, E, F).
- Our target is to group them into single one cluster at the end of the iteration.
- In each step of iteration, find the **closest pair cluster**.
- First, we must compute the distance matrix,

	X1	X2
A	1	1
B	1.5	1.5
C	5	5
D	3	4
E	4	4
F	3	3.5



Single Linkage Clustering: Example

	X1	X2
A	1	1
B	1.5	1.5
C	5	5
D	3	4
E	4	4
F	3	3.5



$$d_{ij} = \left(\sum_k \left(x_{ik} - x_{jk} \right)^2 \right)^{\frac{1}{2}}$$

$$d_{AB} = \left((1-1.5)^2 + (1-1.5)^2 \right)^{\frac{1}{2}} = \sqrt{\tfrac{1}{2}} = 0.7071$$

$$d_{DF} = \left((3-3)^2 + (4-3.5)^2 \right)^{\frac{1}{2}} = 0.5$$

Dist	A	B	C	D	E	F
A	0.00	0.71	5.66	3.61	4.24	3.20
B	0.71	0.00	4.95	2.92	3.54	2.50
C	5.66	4.95	0.00	2.24	1.41	2.50
D	3.61	2.92	2.24	0.00	1.00	0.50
E	4.24	3.54	1.41	1.00	0.00	1.12
F	3.20	2.50	2.50	0.50	1.12	0.00

0.71	5.66	3.61	4.24	3.20	4.95	2.92	3.54	2.50	2.24	1.41	2.50	1.00	0.50	1.12

Single Linkage Clustering: Example

Dist	A	B	C	D	E	F
A	0.00	0.71	5.66	3.61	4.24	3.20
B	0.71	0.00	4.95	2.92	3.54	2.50
C	5.66	4.95	0.00	2.24	1.41	2.50
D	3.61	2.92	2.24	0.00	1.00	0.50
E	4.24	3.54	1.41	1.00	0.00	1.12
F	3.20	2.50	2.50	0.50	1.12	0.00

- *The shortest (min) distance is between pair F and D which is 0.50.* Then we update the distance matrix like,

Min Distance (Single Linkage)

Dist	A	B	C	D, F	E
A	0.00	0.71	5.66	?	4.24
B	0.71	0.00	4.95	?	3.54
C	5.66	4.95	0.00	?	1.41
D, F	?	?	?	0.00	?
E	4.24	3.54	1.41	?	0.00

Single Linkage Clustering: Example

The distance between cluster {D,F} and other cluster is:

Dist	A	B	C	D	E	F
A	0.00	0.71	5.66	3.61	4.24	3.20
B	0.71	0.00	4.95	2.92	3.54	2.50
C	5.66	4.95	0.00	2.24	1.41	2.50
D	3.61	2.92	2.24	0.00	1.00	0.50
E	4.24	3.54	1.41	1.00	0.00	1.12
F	3.20	2.50	2.50	0.50	1.12	0.00

Min Distance (Single Linkage)

Dist	A	B	C	D, F	E
A	0.00	0.71	5.66	?	4.24
B	0.71	0.00	4.95	?	3.54
C	5.66	4.95	0.00	?	1.41
D, F	?	?	?	0.00	?
E	4.24	3.54	1.41	?	0.00

$$d_{(D,F)\mapsto A} = \min\left(d_{DA}, d_{FA}\right) = \min\left(3.61, 3.20\right) = 3.20$$

$$d_{(D,F)\mapsto B} = \min\left(d_{DB}, d_{FB}\right) = \min\left(2.92, 2.50\right) = 2.50$$

$$d_{(D,F)\mapsto C} = \min\left(d_{DC}, d_{FC}\right) = \min\left(2.24, 2.50\right) = 2.24$$

$$d_{E\to(D,F)} = \min\left(d_{ED}, d_{EF}\right) = \min\left(1.00, 1.12\right) = 1.00$$

Single Linkage Clustering: Example

- The updated distance matrix would be:

Min Distance (Single Linkage)

Dist	A	B	C	D, F	E
A	0.00	0.71	5.66	?	4.24
B	0.71	0.00	4.95	?	3.54
C	5.66	4.95	0.00	?	1.41
D, F	?	?	?	0.00	?
E	4.24	3.54	1.41	?	0.00

$$d_{(D,F)\mapsto A} = \min\left(d_{DA}, d_{FA}\right) = \min\left(3.61, 3.20\right) = 3.20$$

$$d_{(D,F)\mapsto B} = \min\left(d_{DB}, d_{FB}\right) = \min\left(2.92, 2.50\right) = 2.50$$

$$d_{(D,F)\mapsto C} = \min\left(d_{DC}, d_{FC}\right) = \min\left(2.24, 2.50\right) = 2.24$$

$$d_{E\to(D,F)} = \min\left(d_{ED}, d_{EF}\right) = \min\left(1.00, 1.12\right) = 1.00$$

Min Distance (Single Linkage)

Dist	A	B	C	D, F	E
A	0.00	0.71	5.66	3.20	4.24
B	0.71	0.00	4.95	2.50	3.54
C	5.66	4.95	0.00	2.24	1.41
D, F	**3.20**	**2.50**	**2.24**	0.00	1.00
E	4.24	3.54	1.41	**1.00**	0.00

The next closest distance is between A and B is 0.71.

Single Linkage Clustering: Example

The distance between cluster {A,B} and other cluster is:

Dist	A,B	C	(D, F)	E
A,B	0	?	?	?
C	?	0	2.24	1.41
(D, F)	?	2.24	0	1.00
E	?	1.41	1.00	0

Dist	A	B	C	D	E	F
A	0.00	0.71	5.66	3.61	4.24	3.20
B	0.71	0.00	4.95	2.92	3.54	2.50
C	5.66	4.95	0.00	2.24	1.41	2.50
D	3.61	2.92	2.24	0.00	1.00	0.50
E	4.24	3.54	1.41	1.00	0.00	1.12
F	3.20	2.50	2.50	0.50	1.12	0.00

$$d_{C\to(A,B)} = \min\left(d_{CA}, d_{CB}\right) = \min\left(5.66, 4.95\right) = 4.95$$

$$d_{(D,F)\to(A,B)} = \min\left(d_{DA}, d_{DB}, d_{FA}, d_{FB}\right) = \min\left(3.61, 2.92, 3.20, 2.50\right) = 2.50$$

$$d_{E\to(A,B)} = \min\left(d_{EA}, d_{EB}\right) = \min\left(4.24, 3.54\right) = 3.54$$

Single Linkage Clustering: Example

- The updated distance matrix would be:

Dist	A,B	C	(D, F)	E
A,B	0	?	?	?
C	?	0	2.24	1.41
(D, F)	?	2.24	0	1.00
E	?	1.41	1.00	0

Min Distance (Single Linkage)

Dist	A,B	C	(D, F)	E
A,B	0	4.95	2.50	3.54
C	**4.95**	0	2.24	1.41
(D, F)	**2.50**	2.24	0	1.00
E	**3.54**	1.41	1.00	0

The next closest distance is between {D,F} and E which is 1.00.

Single Linkage Clustering: Example

The distance computations would be:

Min Distance (Single Linkage)

Dist	(A,B)	C	(D, F), E
(A,B)	0.00	4.95	?
C	4.95	0.00	?
(D, F), E	?	?	0.00

Dist	A	B	C	D	E	F
A	0.00	0.71	5.66	3.61	4.24	3.20
B	0.71	0.00	4.95	2.92	3.54	2.50
C	5.66	4.95	0.00	2.24	1.41	2.50
D	3.61	2.92	2.24	0.00	1.00	0.50
E	4.24	3.54	1.41	1.00	0.00	1.12
F	3.20	2.50	2.50	0.50	1.12	0.00

$$d_{((D,F),E)\mapsto(A,B)} = \min\left(d_{DA}, d_{DB}, d_{FA}, d_{FB}, d_{EA}, d_{EB}\right) = \min\left(3.61, 2.92, 3.20, 2.50, 4.24, 3.54\right) = 2.50$$

$$d_{((D,F),E)\mapsto C} = \min\left(d_{DC}, d_{FC}, d_{EC}\right) = \min\left(2.24, 2.50, 1.41\right) = 1.41$$

Single Linkage Clustering: Example

The updated distance matrix would be:

Min Distance (Single Linkage)

Dist	(A,B)	C	(D, F), E
(A,B)	0.00	4.95	?
C	4.95	0.00	?
(D, F), E	?	?	0.00

Min Distance (Single Linkage)

Dist	(A,B)	C	(D, F), E
(A,B)	0.00	4.95	2.50
C	4.95	0.00	1.41
(D, F), E	2.50	1.41	0.00

The next closest distance is between {(D,F),E} and C.

Single Linkage Clustering: Example

The distance computations would be:

Min Distance (Single Linkage)

Dist	(A,B)	(D, F), E),C
(A,B)	0.00	?
((D, F), E),C	?	0.00

Dist	A	B	C	D	E	F
A	0.00	0.71	5.66	3.61	4.24	3.20
B	0.71	0.00	4.95	2.92	3.54	2.50
C	5.66	4.95	0.00	2.24	1.41	2.50
D	3.61	2.92	2.24	0.00	1.00	0.50
E	4.24	3.54	1.41	1.00	0.00	1.12
F	3.20	2.50	2.50	0.50	1.12	0.00

$$d_{(((D,F),E),C)\rightarrow(A,B)} = \min\left(d_{DA}, d_{DB}, d_{FA}, d_{FB}, d_{EA}, d_{EB}, d_{CA}, d_{CB}\right)$$

$$d_{(((D,F),E),C)\rightarrow(A,B)} = \min\left(3.61, 2.92, 3.20, 2.50, 4.24, 3.54, 5.66, 4.95\right) = 2.50$$

Single Linkage Clustering: Example

- The updated distance matrix would be:

Min Distance (Single Linkage)

Dist	(A,B)	(D, F), E),C
(A,B)	0.00	?
((D, F), E),C	?	0.00

Min Distance (Single Linkage)

Dist	(A,B)	(D, F), E),C
(A,B)	0.00	2.50
((D, F), E),C	2.50	0.00

Single Linkage Clustering: Example

- In the beginning we have 6 clusters:

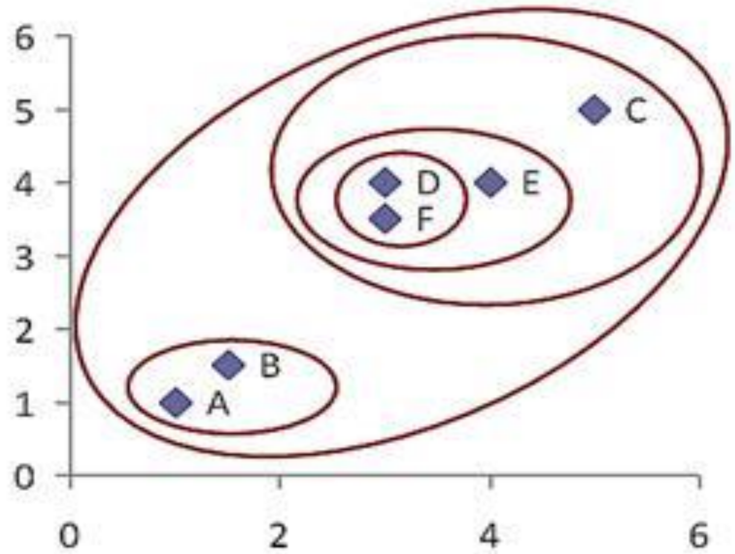
  A, B, C, D, E and F

- We merge cluster **D** and **F** into cluster **(D, F)** at distance **0.50**
- We merge cluster **A** and cluster **B** into **(A, B)** at distance **0.71**
- We merge cluster **E** and **(D, F)** into **((D, F), E)** at distance **1.00**

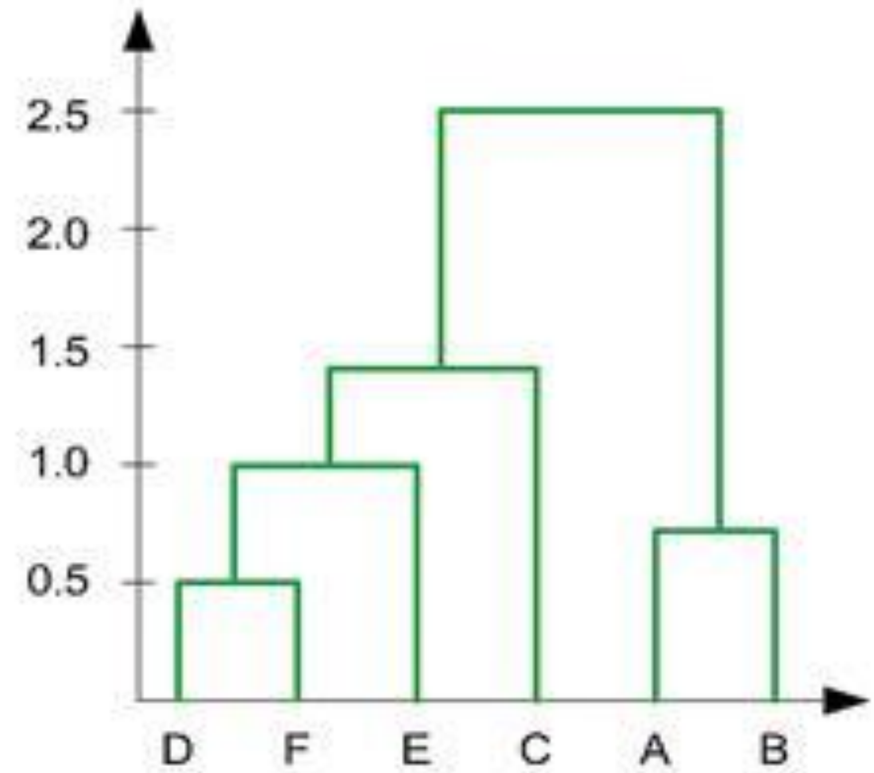
Single Linkage Clustering: Example

- We merge cluster **((D, F), E)** and C into **(((D, F), E), C)** at distance **1.41**
- We merge cluster **(((D, F), E), C)** and **(A, B)** into **((((D, F), E), C), (A, B))** at distance **2.50**
- The last cluster contain all the objects, thus conclude the computation.

Single Linkage Clustering: Example

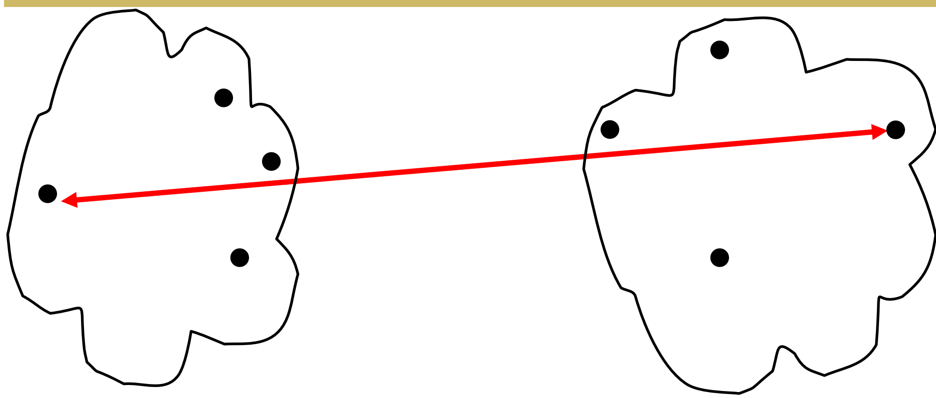


	X1	X2
A	1	1
B	1.5	1.5
C	5	5
D	3	4
E	4	4
F	3	3.5



COMPLETE LINKAGE CLUSTERING

Inter-Cluster Similarity



- MIN
- MAX
- Group Average
- Distance Between Centroids
- Other methods driven by an objective function
  - Ward's Method uses squared error

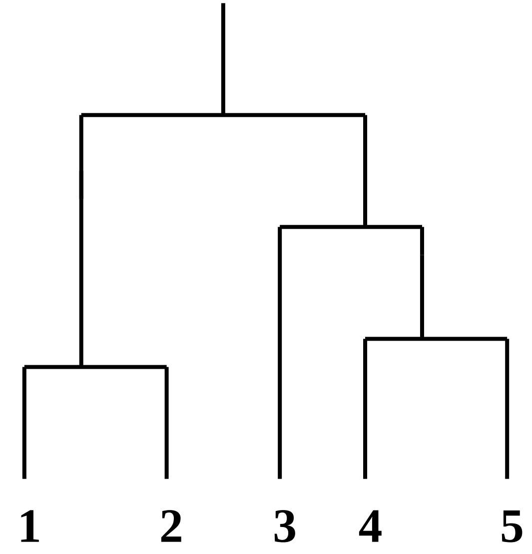
	p1	p2	p3	p4	p5	. . .
p1						
p2						
p3						
p4						
p5						
. . .						

Proximity Matrix

MAX or Complete Linkage

- Similarity of two clusters is based on the **two least similar (most distant) points** in the different clusters
  - Determined by the all pairs of points in the two clusters

	I1	I2	I3	I4	I5
I1	0.00	0.90	0.10	0.65	0.20
I2	0.90	0.00	0.70	0.60	0.50
I3	0.10	0.70	0.00	0.40	0.30
I4	0.65	0.60	0.40	0.00	0.80
I5	0.20	0.50	0.30	0.80	0.00



Complete Linkage Example

- Consider the following data points,
- Assume the **Manhattan distance** metric

Points	x-coordinate	y-coordinate
P1	3	5
P2	5	6
P3	1	7
P4	2	2
P5	8	3

Complete Linkage Example

Points	x-coordinate	y-coordinate
P1	3	5
P2	5	6
P3	1	7
P4	2	2
P5	8	3

	P1	P2	P3	P4	P5
P1	0	3	4	4	7
P2	3	0	5	7	6
P3	4	5	0	6	11
P4	4	7	6	0	7
P5	7	6	11	7	0

Complete Linkage Example

	P1	P2	P3	P4	P5
P1	0	3	4	4	7
P2	3	0	5	7	6
P3	4	5	0	6	11
P4	4	7	6	0	7
P5	7	6	11	7	0

	(P1, P2)	P3	P4	P5
(P1, P2)	0	5	7	7
P3	5	0	6	11
P4	7	6	0	7
P5	7	11	7	0

$$
\begin{aligned}
D_{(P1,P2)\to P3} &= \max(D_{P1\to P3}, D_{P2\to P3}) \\
&= \max(4,5) = 5 \\
D_{(P1,P2)\to P4} &= \max(D_{P1\to P4}, D_{P2\to P4}) \\
&= \max(4,7) = 7 \\
D_{(P1,P2)\to P5} &= \max(D_{P1\to P5}, D_{P2\to P5}) \\
&= \max(7,6) = 7
\end{aligned}
$$

Complete Linkage Example

	P1	P2	P3	P4	P5
P1	0	3	4	4	7
P2	3	0	5	7	6
P3	4	5	0	6	11
P4	4	7	6	0	7
P5	7	6	11	7	0

	(P1, P2, P3)	P4	P5
(P1, P2, P3)	0	7	11
P4	7	0	7
P5	11	7	0

$$\begin{aligned} D_{(P1,P2,P3)\rightarrow P4} &= \max(D_{P1\rightarrow P4}, D_{P2\rightarrow P4}, D_{P3\rightarrow P4}) \\ &= \max(4, 7, 6) = 7 \\ D_{(P1,P2,P3)\rightarrow P5} &= \max(D_{P1\rightarrow P5}, D_{P2\rightarrow P5}, D_{P3\rightarrow P5}) \\ &= \max(7, 6, 11) = 11 \end{aligned}$$

Complete Linkage Example

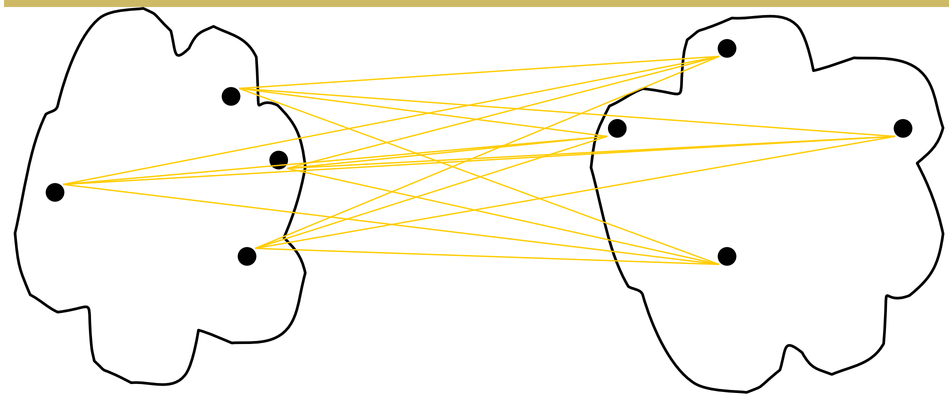
	P1	P2	P3	P4	P5
P1	0	3	4	4	7
P2	3	0	5	7	6
P3	4	5	0	6	11
P4	4	7	6	0	7
P5	7	6	11	7	0

	(P1, P2, P3, P4)	P5
(P1, P2,P3, P4)	0	11
P5	11	0

$$D_{(P1,P2,P3,P4)\rightarrow P5} =$$
$$= \max\left(D_{P1\rightarrow P5}, D_{P2\rightarrow P5}, D_{P3\rightarrow P5}, D_{P4\rightarrow P5}\right)$$
$$= \max(7,6,11,7) = 11$$

AVERAGE LINKAGE CLUSTERING EXAMPLE

Inter-Cluster Similarity



- MIN
- MAX
- Group Average
- Distance Between Centroids
- Other methods driven by an objective function
  - Ward's Method uses squared error

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

Proximity Matrix

Average Linkage Example

- Consider the following data points,
- Assume the **Manhattan distance** metric

Points	x-coordinate	y-coordinate
P1	3	5
P2	5	6
P3	1	7
P4	2	2
P5	8	3

Average Linkage Example

Points	x-coordinate	y-coordinate
P1	3	5
P2	5	6
P3	1	7
P4	2	2
P5	8	3

	P1	P2	P3	P4	P5
P1	0	3	4	4	7
P2	3	0	5	7	6
P3	4	5	0	6	11
P4	4	7	6	0	7
P5	7	6	11	7	0

Average Linkage Example

	P1	P2	P3	P4	P5
P1	0	3	4	4	7
P2	3	0	5	7	6
P3	4	5	0	6	11
P4	4	7	6	0	7
P5	7	6	11	7	0

	(P1, P2)	P3	P4	P5
(P1, P2)	0	?	?	?
P3	?	0	6	11
P4	?	6	0	7
P5	?	11	7	0

$$\text{proximity}(\text{Cluster}_i, \text{Cluster}_j) = \frac{\sum\limits_{\substack{p_i \in \text{Cluster}_i \\ p_j \in \text{Cluster}_j}} \text{proximity}(p_i, p_j)}{|\text{Cluster}_i| * |\text{Cluster}_j|}$$

Average Linkage Example

	P1	P2	P3	P4	P5
P1	0	3	4	4	7
P2	3	0	5	7	6
P3	4	5	0	6	11
P4	4	7	6	0	7
P5	7	6	11	7	0

	(P1, P2)	P3	P4	P5
(P1, P2)	0	4.5	5.5	6.5
P3	4.5	0	6	11
P4	5.5	6	0	7
P5	6.5	11	7	0

$$D_{(P1,P2)\to P3} = \frac{(D_{P1\to P3} + D_{P2\to P3})}{2 \times 1} = \frac{4+5}{2} = 4.5$$

$$D_{(P1,P2)\to P4} = \frac{(D_{P1\to P4} + D_{P2\to P4})}{2 \times 1} = \frac{4+7}{2} = 5.5$$

$$D_{(P1,P2)\to P5} = \frac{(D_{P1\to P5} + D_{P2\to P5})}{2 \times 1} = \frac{7+6}{2} = 6.5$$

Average Linkage Example

	P1	P2	P3	P4	P5
P1	0	3	4	4	7
P2	3	0	5	7	6
P3	4	5	0	6	11
P4	4	7	6	0	7
P5	7	6	11	7	0

	(P1, P2, P3)	P4	P5
(P1, P2, P3)	0	5.7	8
P4	5.7	0	7
P5	8	7	0

$$D_{(P1,P2,P3)\to P4} = \frac{(D_{P1\to P4} + D_{P2\to P4} + D_{P3\to P4})}{3 \times 1}$$
$$= \frac{4+7+6}{3 \times 1} = 5.7$$
$$D_{(P1,P2,P3)\to P5} = \frac{(D_{P1\to P5} + D_{P2\to P5} + D_{P3\to P5})}{3 \times 1}$$
$$= \frac{7+6+11}{3 \times 1} = 8$$

Average Linkage Example

	P1	P2	P3	P4	P5
P1	0	3	4	4	7
P2	3	0	5	7	6
P3	4	5	0	6	11
P4	4	7	6	0	7
P5	7	6	11	7	0

	(P1, P2, P3, P4)	P5
(P1, P2,P3, P4)	0	7.75
P5	7.75	0

$$D_{(P1,P2,P3,P4)\rightarrow P5} =$$

$$= \frac{(D_{P1\rightarrow P5} + D_{P2\rightarrow P5} + D_{P3\rightarrow P5} + D_{P4\rightarrow P5})}{4 \times 1}$$

$$= \frac{7+6+11+7}{4\times1} = \frac{31}{4} = 7.75$$

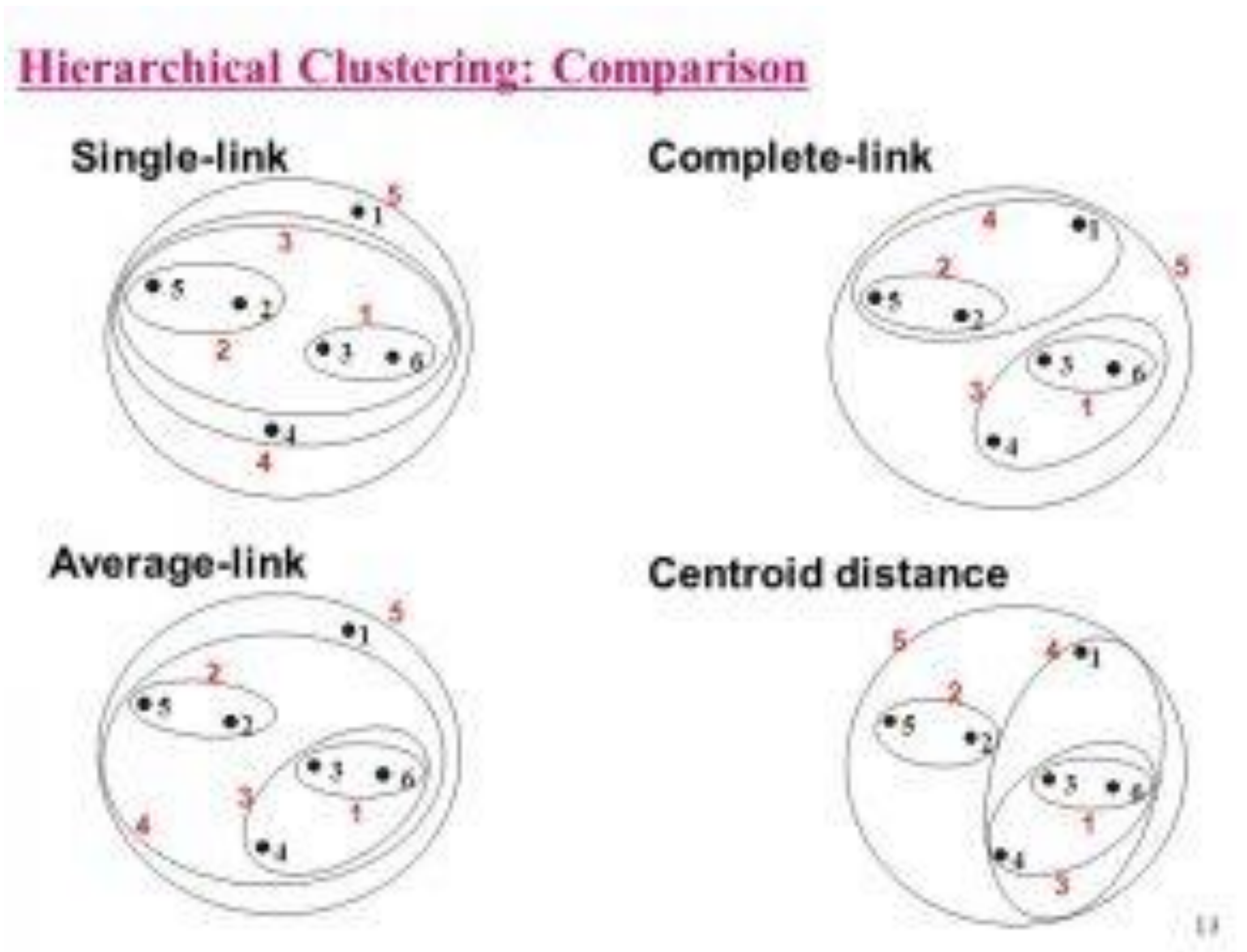
Group Average Linkage

- Proximity of two clusters is the **average of pairwise proximity between points** in the two clusters.

$$\mathbf{proximity}(\mathbf{Cluster_i}, \mathbf{Cluster_j}) = \frac{\sum\limits_{\substack{p_i \in Cluster_i \\ p_j \in Cluster_j}} \mathbf{proximity}(p_i, p_j)}{|\mathbf{Cluster_i}| * |\mathbf{Cluster_j}|}$$

- Need to use **average connectivity for scalability** since total proximity favors large clusters
- *Compromise between Single and Complete Link*
- Strengths --- Less susceptible to noise and outliers

Comparison



Hierarchical Clustering: Comparison
Single-link
Complete-link
Average-link
Centroid distance