Thus, we obtain two clusters containing:

Cluster1 {R1, R2, R3} and Cluster2 {R4, R5, R6, R7}.

Their new centroids are:

C1 = (1.0+1.5+3.0)/3, (1.0+2.0+4.0)/3 C2 = (5.0+3.5+4.5+3.5)/4, (7+5+5+4.5)/4
= 5.5/3, 7.0/3 = 16.5/4, 21.5/4
= 1.83, 2.33 = 4.12, 5.37

Iteration2:

Record Number	**Close to C1(**1.83, 2.33**)**	**Close to C2(**4.12, 5.37**)**	**Assign to cluster**
R1(1.0,1.0)	dist(R1, C1)=1.57	dist(R1, C2)=5.37	Cluster1
R2(1.5,2.0)	dist(R2, C1)=0.47	dist(R2, C2)=4.27	Cluster1
R3(3.0,4.0)	dist(R3, C1)=2.04	dist(R3, C2)=1.77	Cluster2
R4(5.0,7.0)	dist(R4, C1)=5.64	dist(R4, C2)=1.85	Cluster2
R5(3.5,5.0)	dist(R5, C1)=3.15	dist(R5, C2)=0.72	Cluster2
R6(4.5,5.0)	dist(R6, C1)=3.78	dist(R6, C2)=0.53	Cluster2
R7(3.5,4.5)	dist(R7,C1)=2.74	dist(R7, C2)=1.07	Cluster2

Therefore, new clusters are:

Cluster1 {R1, R2} and Cluster2 {R3, R4, R5, R6, R7}.

Their new centroids are:

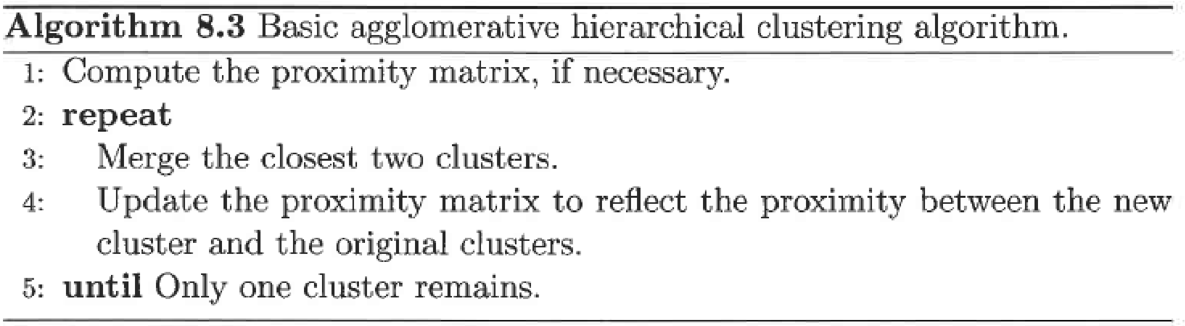
C1 = (1.0+1.5)/2, (1.0+2.0)/2 C2 = (3.0+5.0+3.5+4.5+3.5)/5, (4+7+5+5+4.5)/5
= 2.50/2, 3.0/2 = 19.5/5, 25.5/5
= 1.25, 1.5 = 3.9, 5.1

2. Use the **distance matrix** in Table1 to perform single link and complete link hierarchical clustering. Show your results by drawing a dendogram. The dendogram should clearly show the order in which the points are merged.

Table 1 Distance matrix

	P1	P2	P3	P4	P5
P1	0.00	0.10	0.41	0.55	0.35
P2	0.10	0.00	0.64	0.47	0.98
P3	0.41	0.64	0.00	0.44	0.85
P4	0.55	0.47	0.44	0.00	0.76
P5	0.35	0.98	0.85	0.76	0.00

Solution:



Algorithm 8.3 Basic agglomerative hierarchical clustering algorithm.

```
Compute the proximity matrix, if necessary.
repeat
    Merge the closest two clusters.
    Update the proximity matrix to reflect the proximity between the new
    cluster and the original clusters.
until Only one cluster remains.
```

For the single link or MIN version of hierarchical clustering, the proximity of two clusters is defined as the minimum of the distance (maximum of the similarity) between any two points in the two different clusters.

Steps:

- Using graph terminology, start with all points as singleton clusters.
- Add links between points one at a time (shortest links first).
- These single links combine the points into clusters.

	P1	P2	P3	P4	P5
P1	0.00	0.10	0.41	0.55	0.35
P2	0.10	0.00	0.64	0.47	0.98
P3	0.41	0.64	0.00	0.44	0.85
P4	0.55	0.47	0.44	0.00	0.76
P5	0.35	0.98	0.85	0.76	0.00

Combine P1 and P2:

$$\begin{aligned} dist(\{P1,P2\},\{P3\}) &= min\big(dist(P1,P3), dist(P2,P3)\big) \\ &= min(0.41, 0.64) \\ &= 0.41 \end{aligned}$$

$$\begin{aligned} dist(\{P1,P2\},\{P4\}) &= min\big(dist(P1,P4), dist(P2,P5)\big) \\ &= min(0.55, 0.98) \\ &= 0.55 \end{aligned}$$

$$\begin{aligned} dist(\{P1,P2\},\{P5\}) &= min\big(dist(P1,P5), dist(P2,P5)\big) \\ &= min(0.35, 0.98) \\ &= 0.35 \end{aligned}$$

	P12	P3	P4	P5
P12	0.00	0.41	0.55	0.35
P3	0.41	0.00	0.44	0.85
P4	0.55	0.44	0.00	0.76
P5	0.35	0.85	0.76	0.00

Combine P12 and P5:

$$dist(\{P12, P5\}, \{P3\}) = min\big(dist(P12, P3), dist(P5, P3)\big)$$
$$= min(0.41, 0.85)$$
$$= 0.41$$

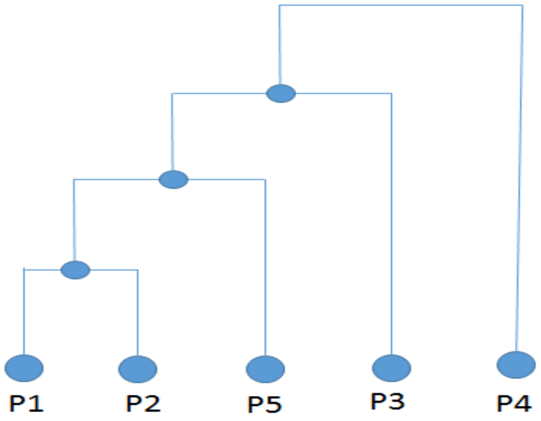
$$dist(\{P12, P5\}, \{P4\}) = min\big(dist(P12, P4), dist(P5, P4)\big)$$
$$= min(0.55, 0.76)$$
$$= 0.55$$

	P125	P3	P4
P125	0.00	0.41	0.55
P3	0.41	0.00	0.44
P4	0.55	0.44	0.00

Combine P125 and P3:

$$dist(\{P125, P3\}, \{P4\}) = min\big(dist(P125, P4), dist(P3, P4)\big)$$
$$= min(0.55, 0.44)$$
$$= 0.44$$

	P1235	P4
P1235	0.00	0.44
P4	0.44	0.00



Single Link Dendogram

For the complete link or MAX version of hierarchical clustering, the proximity of two clusters is defined as the maximum of the distance (minimum of the similarity) between any two points in the two different clusters.

Steps:

- Using graph terminology, start with all points as singleton clusters.
- Add links between points one at a time (shortest links first).

- Group points until all the points are completely linked, i.e., clique.

-	P1	P2	P3	P4	P5
P1	0.00	0.10	0.41	0.55	0.35
P2	0.10	0.00	0.64	0.47	0.98
P3	0.41	0.64	0.00	0.44	0.85
P4	0.55	0.47	0.44	0.00	0.76
P5	0.35	0.98	0.85	0.76	0.00

Combine P1 and P2:

$$\begin{aligned} dist(\{P1, P2\}, \{P3\}) &= max\big(dist(P1, P3), dist(P2, P3)\big) \\ &= max(0.41, 0.64) \\ &= 0.64 \end{aligned}$$

$$\begin{aligned} dist(\{P1, P2\}, \{P4\}) &= min\big(dist(P1, P4), dist(P2, P5)\big) \\ &= min(0.55, 0.98) \\ &= 0.98 \end{aligned}$$

$$\begin{aligned} dist(\{P1, P2\}, \{P5\}) &= min\big(dist(P1, P5), dist(P2, P5)\big) \\ &= min(0.35, 0.98) \\ &= 0.98 \end{aligned}$$

	P12	P3	P4	P5
P12	0.00	0.64	0.98	0.98
P3	0.64	0.00	0.44	0.85
P4	0.98	0.44	0.00	0.76
P5	0.98	0.85	0.76	0.00

Combine P3 and P4:

$$\begin{aligned} dist(\{P3, P4\}, \{P12\}) &= min\big(dist(P3, P12), dist(P4, P12)\big) \\ &= max(0.64, 0.98) \\ &= 0.98 \end{aligned}$$

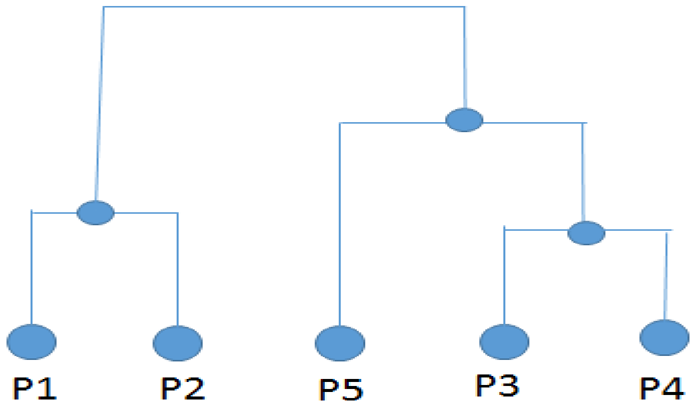
$$\begin{aligned} dist(\{P3, P4\}, \{P5\}) &= min\big(dist(P3, P5), dist(P4, P5)\big) \\ &= max(0.85, 0.76) \\ &= 0.85 \end{aligned}$$

	P12	P34	P5
P12	0.00	0.98	0.98
P34	0.98	0.00	0.85
P5	0.98	0.85	0.00

Combine P34 and P5:

$$dist(\{P34, P5\}, \{P12\}) = max\big(dist(P34, P12), dist(P5, P12)\big)$$
$$= max(0.98, 0.98)$$
$$= 0.98$$

	P12	P345
P12	0.00	0.98
P345	0.98	0.00



Complete Link Dendogram

Exercises

1. Find all well separated clusters in the set of points shown in Figure 1.

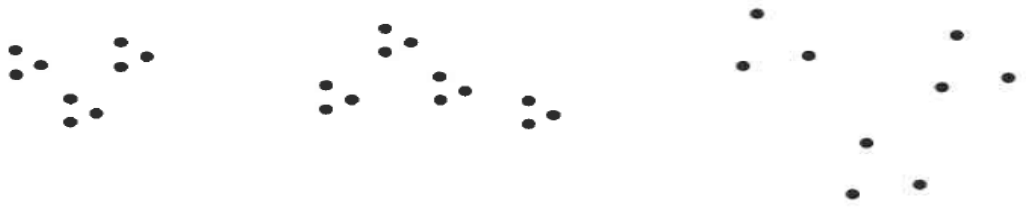


Figure 1: Points

2. Many partitional clustering algorithms that automatically determine the number of clusters claim that this is an advantage. List two situations in which this is not the case.

3. Identify the clusters in Figure 2 using the center-, contiguity-, and density-based definitions. Also indicate the number of clusters for each case and give a brief indication of your reasoning. Note that darkness or the number of dots indicates density. If it helps, assume center-based means k-means, contiguity-based single link, and density-based means DBSCAN.

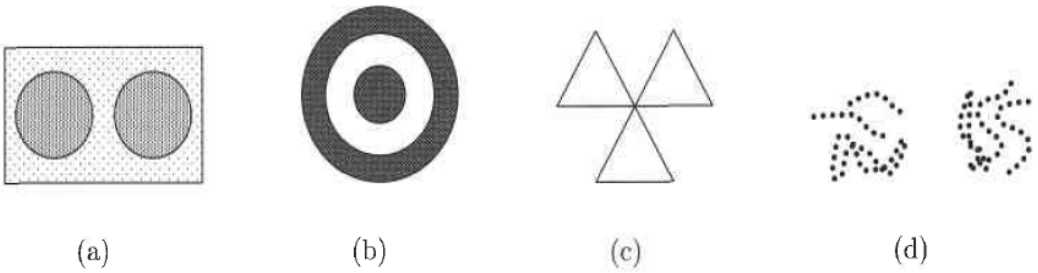
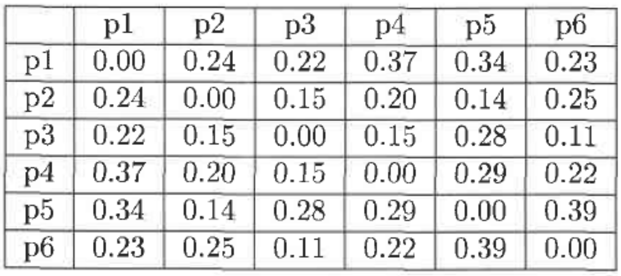


Figure 2: Clusters

4. Given the following points: $2, 4, 10, 12, 3, 20, 30, 11, 25$. Given $k = 3$, and the initial means, $\mu_1 = 2, \mu_2 = 4$ and $\mu_3 = 6$. Show the clusters obtained and new means after each iteration using the K-means algorithm.

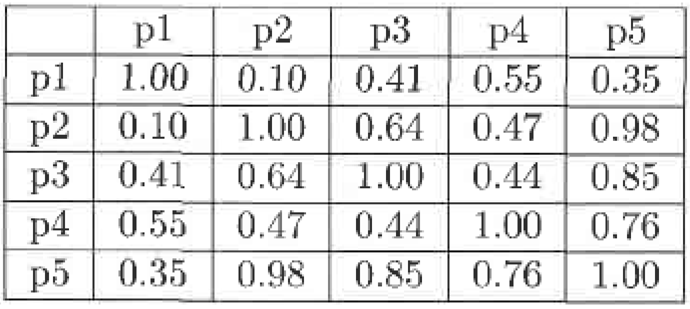
5. Use the **distance matrix** in Table2 to perform single link and complete link hierarchical clustering. Show your results by drawing a dendogram. The dendogram should clearly show the order in which the points are merged.
Table 2: Distance Matrix



	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

6. Use the **similarity matrix** in Table 3 to perform single link and complete link hierarchical clustering. Show your results by drawing a dendogram. The dendogram should clearly show the order in which the points are merged.

Table 3: Similarity matrix



	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00

7. Consider the mean of a cluster of objects from a binary transaction data set. What are the minimum and maximum values of the components of the mean? What is the interpretation of components of the cluster mean? Which component most accurately characterize the objects in the cluster?

8. Differentiate between agglomerative and divisive methods of hierarchical clustering with the help of a diagram.

9. Explain the following terms with reference to the DBSCAN clustering algorithm:

   (a) Core points
   (b) Noise points
   (c) Border points

10. Describe the following clustering algorithm in terms of:
   (a) Shape of clusters
   (b) Limitations:
      i. K-means
      ii. DBSCAN