**NAME: Ashish Bisht**

**ID: 801168390**

**Cloud Computing for Data Analysis**

**Exercise 10 : Hierarchical Clustering**

Use the similarity matrix in following table to perform single and complete link hierarchical clustering. Show your results by drawing a dendrogram. The dendrogram should clearly show the order in which the points are merged.

**Solution:**

**SINGLE LINK CLUSTERING:**

• Points p2 and p5 are more similar (0.98), combine p2 and p5.

Cluster p2 and p5.

Similarity from P2 U P5:

To P1: max(similarity from P2 and P5) = max(0.35, 0.10) = 0.35

To P3: max(similarity from P2 and P5) = max(0.64, 0.85) = 0.85

To P4: max(similarity from P2 and P5) = max(0.47, 0.76) = 0.76

P1

P2 U P5

P3

P4

P1

1

P2 U P5

0.35

1

P3

0.41

0.85

1

P4

0.55

0.76

0.44

1

Similarity Matrix 1

• Points p3 and (p2,p5) are more similar (0.85), combine p3 and (p2,p5).

Cluster p3 and p2, p5.

Similarity from P2 U P5 U P3:

To P1: max(similarity from P2,P3 and P5) = max(0.35, 0.41) = 0.41

To P4: max(similarity from P2,P3 and P5) = max(0.44, 0.76) = 0.76

P1

P2 U P5 U P3

P4

P1

1

P2 U P5 U P3

0.41

1

P4

0.55

0.76

1

Similarity Matrix 2

• Points p4 and (p2, p3, p5) are more similar (0.76), combine p4 and (p2, p3, p5).

Cluster p4 and p2, p3, p5.

Similarity from P2 U P5 U P3 U P4:

To P1: max(similarity from P2,P3,P4 and P5) = max(0.55) = 0.55

P1

P2 U P5 U P3 U P4

P1

1

P2 U P5 U P3 U P4

0.55

1

Similarity Matrix 3

**SINGLE LINK DENDROGRAM:**

**COMPLETE LINK HIERARCHICAL CLUSTERING:**

• Points p2 and p5 are more similar (0.98) combine p2 and p5.

Cluster p2 and p5.

Similarity from P2 U P5:

To P1: min(similarity from P2 and P5) = min(0.35, 0.10) = 0.10

To P3: min(similarity from P2 and P5) = min(0.64, 0.85) = 0.64

To P4: min(similarity from P2 and P5) = min(0.47, 0.76) = 0.47

P1

P2 U P5

P3

P4

P1

1

P2 U P5

0.10

1

P3

0.41

0.64

1

P4

0.55

0.47

0.44

1

Distance Matrix 1

• First cluster is formed like single link clustering (p2, p5).

Similarity from P2 U P5 U P3:

To P1: min(similarity from P2,P3 and P5) = min(0.10, 0.41) = 0.10

To P4: min(similarity from P2,P3 and P5) = min(0.44, 0.47) = 0.76

P1

P2 U P5 U P3

P4

P1

1

P2 U P5 U P3

0.10

1

P4

0.55

0.44

1

Distance Matrix 2

The similarity from P3 is more and cluster is formed.

P1 U P4

P2 U P5 U P3

P1 U P4

1

P2 U P5 U P3

0.10

1

Distance Matrix 3

P1 and P4 is have more similarity and Cluster is formed for (P1,P4)

Finally we combine both the clusters(P1,P4) and (P2,P3,P5).

**COMPLETE LINK HIERARCHICAL DENDOGRAM:**