QUESTION - 1

GIVEN: UTILITY MATRIX :-

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
U_1	4	5	0	5	1	0	3	2
U_2	0	3	4	3	1	2	1	0
U_3	2	0	1	3	0	4	5	3

a) RTF :- Jacard distance between each pair of users

$$J_D(A,B) = 1 - J_S(A,B) = 1 - \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$

$$J_D(U_1,U_2) = 1 - \frac{(1+1+1+1)}{8} = \frac{4}{8} = 0.5$$

$$J_D(U_2,U_3) = 1 - \frac{(1+1+1+1)}{8} = \frac{4}{8} = 0.5$$

$$J_D(U_1,U_3) = 1 - \frac{(1+1+1+1)}{8} = \frac{4}{8} = 0.5$$

b) RTF :- Cosine Distance between each pair of users

$$C_D(A,B) = 1 - S_c(A,B) = 1 - \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

$$C_D(U_1,U_2) = 1 - \left[\frac{0+15+0+15+1+0+3+0}{(8.94427)(6.324)}\right]$$

$$= 0.3988$$

$$C_D(U_2,U_3) = 1 - \left[\frac{0+0+4+9+0+8+5+0}{(6.3245)(8)}\right]$$

$$= 0.48612$$

$$C_D(U_1,U_3) = 1 - \left[\frac{8+0+0+15+0+0+15+6}{(8.94427)(8)}\right]$$

$$= 0.38508$$

c) RTF: cosine distance between each pair of users by treating the utility matrix as boolean

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
U_1	1	1	0	1	1	0	1	1
U_2	0	1	1	1	1	1	1	0
U_3	1	0	1	1	0	1	1	1

$$C_D(U_1, U_2) = 1 - \left[\frac{1+1+1+1}{\sqrt{6}\ \sqrt{6}}\right] = \frac{1}{3}$$

$$C_D(U_2, U_3) = 1 - \left[\frac{1+1+1+1}{\sqrt{6}\ \sqrt{6}}\right] = \frac{1}{3}$$

$$C_D(U_1, U_3) = 1 - \left[\frac{1+1+1+1}{\sqrt{6}\ \sqrt{6}}\right] = \frac{1}{3}$$

d) RTF: Changing the ratings with 3/more → 1 and ratings with 2/less → 0. Jacard distance between each pair of users

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
U_1	1	1	0	1	0	0	1	0
U_2	0	1	1	1	0	0	0	0
U_3	0	0	0	1	0	1	1	1

$$J_D(U_1, U_2) = 1 - \frac{(1+1)}{5} = \frac{3}{5}$$

$$J_D(U_2, U_3) = 1 - \frac{(1)}{6} = \frac{5}{6}$$

$$J_D(U_1, U_3) = 1 - \frac{(1+1)}{6} = \frac{2}{3}$$

e) RTF :- cosine distance

$$C_D(U_1, U_2) = 1 - \left[\frac{1+1+0}{\sqrt{4}\sqrt{3}}\right] = \frac{\sqrt{3}-1}{\sqrt{3}}$$

$$C_D(U_2, U_3) = 1 - \left[\frac{1}{\sqrt{3}\sqrt{3}}\right] = \frac{2}{3}$$

$$C_D(U_1, U_3) = 1 - \left[\frac{1+1}{\sqrt{4}\sqrt{3}}\right] = \frac{\sqrt{3}-1}{\sqrt{3}}$$

f) Normalized Matrix

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
U_1	0.666	1.666	0	1.666	-2.333	0	-0.333	-0.333
U_2	0	0.666	1.666	0.666	-1.333	-0.33	-1.33	0
U_3	-1	0	-2	0	0	1	2	0

g) Cosine distances

$$C_D(U_1, U_2) = 1 - \left[\frac{1.1095 + 1.1095 + 3.1098 + 0.4438}{(3.6503)(2.7069)}\right]$$

$$= 1 - 0.3842$$

$$= 0.4157$$

$$C_D(U_2, U_3) = 1 - \left[\frac{-3.332 - 0.33 - 2.666}{(2.7069)(3.1622)}\right]$$

$$= 1 - [-0.7396]$$

$$= 1.7396$$

$$C_D(U_1, U_3) = 1 - \left[\frac{-0.666 - 0.666}{(3.6503)(3.1622)}\right]$$

$$= 1 - (-0.1153) = 1.1153$$

QUESTION 3:-

$P_1(2,2)$ $P_2(3,4)$ $P_3(4,7)$ $P_4(5,3)$
$P_5(6,7)$ $P_6(8,7)$ $P_7(8,1)$ $P_8(9,3)$

a) $d(P_1, P_5) = \sqrt{(6-2)^2 + (7-2)^2} = 6.4$
$d(P_2, P_5) = \sqrt{(6-3)^2 + (7-4)^2} = 4.24$
$d(P_3, P_5) = \sqrt{(6-4)^2 + (7-7)^2} = 2$
$d(P_4, P_5) = \sqrt{(6-5)^2 + (7-3)^2} = 4.123$
$d(P_6, P_5) = \sqrt{(8-8)^2 + (7-7)^2} = 2$
$d(P_7, P_5) = \sqrt{(6-8)^2 + (7-1)^2} = 6.32$
$d(P_8, P_5) = \sqrt{(6-9)^2 + (7-3)^2} = 5$

$d(P_1, P_6) = \sqrt{(8-2)^2 + (7-2)^2} = 7.81$
$d(P_2, P_6) = \sqrt{(8-3)^2 + (7-4)^2} = 5.83$
$d(P_3, P_6) = \sqrt{(8-4)^2 + (7-7)^2} = 4$
$d(P_4, P_6) = \sqrt{(8-5)^2 + (7-3)^2} = 5$
$d(P_5, P_6) = \sqrt{(8-6)^2 + (7-7)^2} = 2$
$d(P_7, P_6) = \sqrt{(8-8)^2 + (7-1)^2} = 6$
$d(P_8, P_6) = \sqrt{(9-8)^2 + (3-7)^2} = 4.12$

Cluster - 1: $\{P_1, P_2, P_3, P_4, \underline{P_5}\}$

Cluster - 2: $\{\underline{P_6}, P_7, P_8\}$

New Centroid: $M_1(4, 4.6)$
$M_2(8.33, 3.66)$

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
M_1	3.28	1.16	2.4	1.68	3.12	4.66	5.38	5.24
M_2	6.54	5.34	5.46	3.39	4.07	3.35	2.68	0.94

Cluster-1 : $\{M_1, P_1, P_2, P_3, P_4, P_5\}$
Cluster-2 : $\{M_2, P_6, P_7, P_8\}$

Final Cluster-1 :- $\{P_1, P_2, P_3, P_4, P_5\}$
Cluster-2 :- $\{P_6, P_7, P_8\}$

b) $P_1(2,2)$ $P_2(3,4)$ $P_3(4,7)$ $P_4(5,3)$
$P_5(6,7)$ $P_6(8,7)$ $P_7(8,1)$ $P_8(9,3)$

Initial centroids: $P_3(4,7)$ & $P_7(8,1)$

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
P_3	5.38	3.16	-	4.12	2	4	7.21	6.4
P_7	6.08	5.8	7.21	3.60	6.32	6	-	2.23

Cluster-1 : $\{P_3, P_1, P_2, P_5, P_6\}$
Cluster-2 : $\{P_7, P_4, P_8\}$

Now, centroids : $M_1 = (4.6, 5.4)$
$M_2 = (7.33, 2.33)$

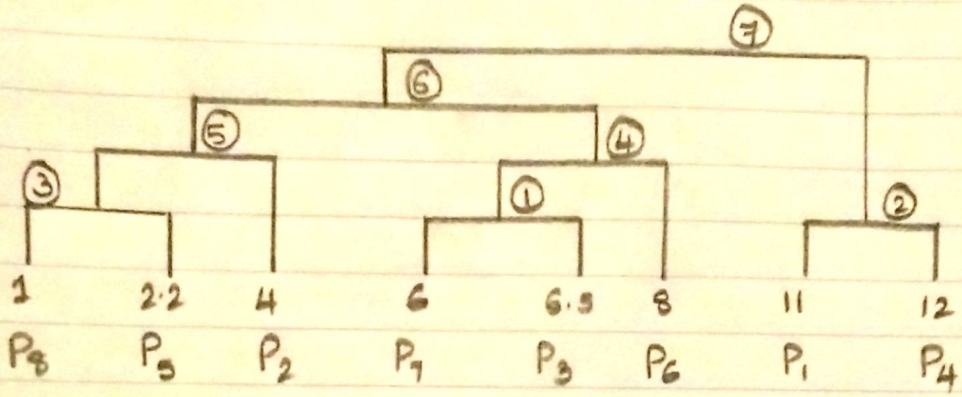
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
M_1	4.28	2.12	1.70	2.43	2.12	3.75	5.56	5.6
M_2	5.34	4.64	5.73	2.42	4.85	4.71	1.48	1.0

Cluster-1 : $\{P_1, P_2, P_3, P_5, P_6\}$
Cluster-2 : $\{P_4, P_7, P_8\}$

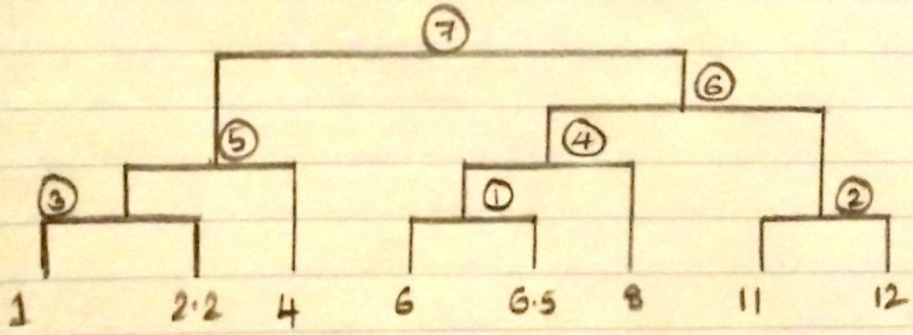
The clustering assignment has changed.

QUESTION-4 :-

a) SINGLE LINK :-

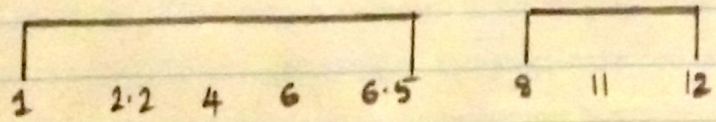


COMPLETE LINK :



b) clusters $\{P_2, P_3, P_5, P_7, P_8\}$
$\{P_1, P_4, P_6\}$

we have,



single link distance - $8 - 6.5 = 1.5$
complete link distance = $12 - 1 = 11$
Average link distance - 6.3933

QUESTION 2:

a)

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	1	1	0	1	0	0	1	0
u_2	0	1	1	1	0	0	0	0
u_3	0	0	0	1	0	1	1	1

$i_1 = \{1\}$ $i_2 = \{1,2\}$ $i_3 = \{2\}$ $i_4 = \{1,2,3\}$ $i_5 = \{\}$
$i_6 = \{3\}$ $i_7 = \{1,3\}$ $i_8 = \{3\}$

$S(i_1, i_2) = 0.5$ $S(i_1, i_5) = 0$ $S(i_1, i_8) = 0$
$S(i_1, i_3) = 0$ $S(i_1, i_6) = 0$
$S(i_1, i_4) = 0.33$ $S(i, i_7) = 0.5$

$S(i_2, i_3) = 0.5$ $S(i_2, i_6) = 0$
$S(i_2, i_4) = 0.66$ $S(i_2, i_7) = 0.33$
$S(i_2, i_5) = 0$ $S(i_2, i_8) = 0$

$S(i_3, i_4) = 0.33$ $S(i_3, i_7) = 0$
$S(i_3, i_5) = 0$ $S(i_3, i_8) = 0$
$S(i_3, i_6) = 0$

$S(i_4, i_5) = 0$ $S(i_4, i_8) = 0.33$
$S(i_4, i_6) = 0.33$
$S(i_4, i_7) = 0.66$

$S(i_5, i_6) = 0$
$S(i_5, i_7) = 0$
$S(i_5, i_8) = 0$

$S(i_6, i_7) = 0.5$ $S(i_7, i_8) = 0.5$
$S(i_6, i_8) = 1$

$D(i_1, i_2) = 0.5$ $D(i_2, i_3) = 0.5$ $D(i_3, i_4) = 0.66$
$D(i_1, i_3) = 1$ $D(i_2, i_4) = 0.33$ $D(i_3, i_5) = 1$
$D(i_1, i_4) = 0.66$ $D(i_2, i_5) = 1$ $D(i_3, i_6) = 1$
$D(i_1, i_5) = 1$ $D(i_2, i_6) = 1$ $D(i_3, i_7) = 1$
$D(i_1, i_6) = 1$ $D(i_2, i_7) = 0.66$ $D(i_3, i_8) = 1$
$D(i_1, i_7) = 0.5$ $D(i_2, i_8) = 1$
$D(i_1, i_8) = 1$

$D(i_4, i_5) = 1$ $D(i_5, i_6) = 1$ $D(i_6, i_7) = 0.5$ $D(i_7, i_8) = 0.5$
$D(i_4, i_6) = 0.66$ $D(i_5, i_7) = 1$ $D(i_6, i_8) = 0$
$D(i_4, i_7) = 0.33$ $D(i_5, i_8) = 1$
$D(i_4, i_8) = 0.66$

Distance Matrix :

	i_1	i_2	i_3	i_4	i_5	i_6	i i_7	i_8
i_1	0	0.5	1	0.66	1	1	0.5	1
i_2	0.5	0	0.5	0.33	1	1	0.66	1
i_3	1	0.5	0	0.66	1	1	1	1
i_4	0.66	0.33	0.66	0	1	0.66	0.33	0.66
i_5	1	1	1	1	0	1	1	1
i_6	1	1	1	0.66	1	0	0.5	0
i_7	0.5	0.66	1	0.33	1	0.5	0	0.5
i_8	1	1	1	0.66	1	0	0.5	0

least distance = 0 between i_6, i_8

∴ Merge i_6, i_8.

	i_1	i_2	i_3	i_4	i_5	i_6, i_8	i_7
i_1	0	0.5	1	0.66	1	1	0.5
i_2	0.5	0	0.5	0.33	1	1	0.66
i_3	1	0.5	0	0.66	1	1	1
i_4	0.66	0.33	0.66	0	1	0.66	0.33
i_5	1	1	1	1	0	1	1
i_6, i_8	1	1	1	0.66	1	0	0.5
i_7	0.5	0.66	1	0.33	1	5	0

Least distance = 0.33
Merge i_2, i_4 first

	i_1	(i_2, i_4)	i_3	i_5	$(i_6 i_8)$	i_7
i_1	0	0.5	1	1	1	0.5
i_2, i_4	0.5	0	0.5	1	0.66	0.33
i_3	1	0	0	1	1	1
i_5	1	0.5	1	0	1	1
i_6, i_8	1	1	1	1	0	0.5
i_7	0.5	0.66	1	1	0.5	0

Merge i_2, i_4, i_7

	i_1	(i_2, i_4, i_7)	i_3	i_5	(i_6, i_8)
i_1	0	0.5	1	1	1
i_2, i_4, i_7	0.5	0	0.5	1	0
i_3	1	0.5	0	1	1
i_5	1	1	1	0	1
i_6, i_8	1	0.5	1	1	0

Merge i_1 and $i_2 i_4 i_7$

	$(i_1\ i_2\ i_4\ i_7)$	i_3	i_5	$(i_6\ i_8)$
$(i_1\ i_2\ i_4\ i_7)$	0	0.5	1	0.5
i_3	0.5	0	1	1
i_5	1	1	0	1
$(i_6\ i_8)$	0.5	1	1	0

∴ The 4 clusters are: $\{i_1, i_2, i_4, i_7\}$, i_3, i_5, $\{i_6, i_8\}$

b) Utility Matrix

	i_1, i_2, i_4, i_7	i_3	i_5	$i_6\ i_8$
U_1	4.25	2	0	1
U_2	2.33	2	4	1
U_3	3.33	3.5	1	0

c) Cosine distance = 1 − Cosine similarity

$$= 1 - \frac{\sum A \times B}{|A||B|}$$

$U_{1\,avg} = 4.80234$ $U_{2\,avg} = 5.140904$ $U_{3\,avg} = 4.933348$

$$D_c(U_1, U_2) = 1 - \frac{(4.25 \times 2.33 + 2 \times 2 + 0 \times 4 + 1 \times 1)}{4.80234 \times 5.140904}$$

$$= 0.39637$$

$$D_c(U_1, U_3) = 1 - \frac{(4.25 \times 3.33 + 2 \times 3.5)}{4.80234 \times 4.933348}$$

$$= 0.10719$$

$$D_c(U_2, U_3) = 1 - \frac{(2.33 \times 3.33) + 2 \times 3.5 + 4 \times 1}{5.140904 \times 4.9338}$$

$$= 0.26037$$