CS6350 HW3

Problem 1

u_1 & u_2:

(a) Jaccard distance = 1 − Jaccard similarity $\left(\text{similarity} = \frac{A \cap B}{A \cup B} = \frac{\text{num}}{\text{num}}\right)$

$$= 1 - \frac{1+1+1+1}{8} = \boxed{\frac{1}{2}}$$

u_1 & u_3:

Jaccard distance = 1 − Jaccard similarity $= 1 - \frac{1+1+1+1}{8} = \boxed{\frac{1}{2}}$

u_2 & u_3:

Jaccard distance $= 1 - \frac{1+1+1+1}{8} = \boxed{\frac{1}{2}}$

(b) u_1 & u_2:

Cosine ~~Distance~~ Similarity $= \frac{5\times3+5\times3+1\times1+3\times1}{\sqrt{4^2+5^2+5^2+1^2+3^2+2^2}\cdot\sqrt{3^2+4^2+3^2+1^2+2^2+1^2}} = \frac{34}{\sqrt{80}\sqrt{40}}$

$$= 0.601$$

cosine distance $= 1 - 0.601 = \boxed{0.399}$

u_1 & u_3:

cosine similarity $= \frac{4\times2+5\times3+3\times5+2\times3}{\sqrt{4^2+5^2+5^2+1^2+3^2+2^2}\cdot\sqrt{2^2+1^2+3^2+4^2+5^2+3^2}} = \frac{44}{\sqrt{80}\cdot\sqrt{64}}$

$$= 0.614$$

cosine distance $= 1 - 0.614 = \boxed{0.386}$

u_2 & u_3:

$$\text{cosine similarity} = \frac{4\times1 + 3\times3 + 2\times4 + 1\times5}{\sqrt{3^2+4^2+3^2+1^2+2^2+1^2}\cdot\sqrt{2^2+1^2+3^2+4^2+5^2+3^2}} = \frac{26}{\sqrt{40}\cdot\sqrt{64}}$$

$$= 0.513$$

$$\text{Cosine distance} = 1 - 0.513 = \boxed{0.486}$$

(c) u_1 & u_2:

$$\text{cosine similarity} = \frac{1\times1 + 1\times1 + 1\times1 + 1\times1}{\sqrt{1^2+1^2+1^2+1^2+1^2+1^2}\sqrt{1^2+1^2+1^2+1^2+1^2+1^2}}$$

$$= \frac{4}{\sqrt{6}\cdot\sqrt{6}} = \frac{2}{3}$$

$$\text{cosine distance} = 1 - \text{cosine similarity} = 1 - \frac{2}{3} = \boxed{\frac{1}{3}}$$

u_1 & u_3:

$$\text{cosine similarity} = \frac{1\times1 + 1\times1 + 1\times1 + 1\times1}{\sqrt{1^2+1^2+1^2+1^2+1^2+1^2}\cdot\sqrt{1^2+1^2+1^2+1^2+1^2+1^2}} = \frac{2}{3}$$

$$\text{cosine distance} = 1 - \text{cosine similarity} = 1 - \frac{2}{3} = \boxed{\frac{1}{3}}$$

u_2 & u_3:

$$\text{cosine similarity} = \frac{1\times1 + 1\times1 + 1\times1 + 1\times1}{\sqrt{1^2+1^2+1^2+1^2+1^2+1^2}\cdot\sqrt{1^2+1^2+1^2+1^2+1^2+1^2}} = \frac{2}{3}$$

$$\text{cosine distance} = 1 - \text{cosine similarity} = 1 - \frac{2}{3} = \boxed{\frac{1}{3}}$$

(d)

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

u_1 & u_2:

Jaccard Similarity = $\frac{1+1}{5} = \frac{2}{5}$

Jaccard distance = $1 - \frac{2}{5} = \boxed{\frac{3}{5}}$

u_1 & u_3:

Jaccard similarity = $\frac{1+1}{6} = \frac{1}{3}$

Jaccard distance = $1 - \frac{1}{3} = \boxed{\frac{2}{3}}$

u_2 & u_3:

Jaccard similarity = $\frac{1}{6}$

Jaccard distance = $1 - \frac{1}{6} = \boxed{\frac{5}{6}}$

(e) u_1 & u_2:

cosine similarity = $\frac{1\times1+1\times1}{\sqrt{1^2+1^2+1^2+1^2}\cdot\sqrt{1^2+1^2+1^2}} = \frac{2}{\sqrt{4}\cdot\sqrt{3}} = 0.5774$

cosine distance = $1 - 0.5774 = \boxed{0.4226}$

~~u_1 & u_3:~~

~~cosine similarity = $\frac{1\times1+1\times1+1\times1+1\times1}{\sqrt{1^2+1^2+1^2+1^2}\cdot\sqrt{1^2+1^2+1^2+1^2+1^2+1^2}} = \frac{4}{\sqrt{4}\cdot\sqrt{6}} =$~~

u_1 & u_3:

$$\text{cosine similarity} = \frac{1\times1+1\times1}{\sqrt{1^2+1^2+1^2+1^2}\cdot\sqrt{1^2+1^2+1^2+1^2}} = \frac{2}{\sqrt{4}\cdot\sqrt{4}} = \frac{1}{2}$$

$$\text{cosine distance} = 1-\frac{1}{2} = \boxed{\frac{1}{2}}$$

u_2 & u_3:

$$\text{cosine similarity} = \frac{1}{\sqrt{1^2+1^2+1^2}\cdot\sqrt{1^2+1^2+1^2+1^2}} = \frac{1}{\sqrt{3}\cdot\sqrt{4}} = 0.2887$$

$$\text{cosine distance} = 1-0.2887 = \boxed{0.7113}$$

f) $\text{avg}(u_1) = \frac{4+5+5+1+3+2}{6} = \frac{10}{3}$

$\text{avg}(u_2) = \frac{3+4+3+1+2+1}{6} = \frac{7}{3}$

$\text{avg}(u_3) = \frac{2+1+3+4+5+3}{6} = \frac{18}{6} = 3$

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	$\frac{2}{3}$	$\frac{5}{3}$		$\frac{5}{3}$	$-\frac{7}{3}$		$-\frac{1}{3}$	$-\frac{4}{3}$
u_2		$\frac{2}{3}$	$\frac{5}{3}$	$\frac{2}{3}$	$-\frac{4}{3}$	$-\frac{1}{3}$	$-\frac{4}{3}$	
u_3	1		-2	0		1	2	0

g) u_1 & u_2:

$$\text{cosine similarity} = \frac{\frac{5}{3}\times\frac{2}{3}+\frac{5}{3}\times\frac{2}{3}+(-\frac{7}{3})\times(-\frac{4}{3})+(-\frac{1}{3})\times(-\frac{4}{3})}{\sqrt{(\frac{2}{3})^2+(\frac{5}{3})^2+(\frac{5}{3})^2+(-\frac{7}{3})^2+(-\frac{1}{3})^2+(-\frac{4}{3})^2}\cdot\sqrt{(\frac{2}{3})^2+(\frac{5}{3})^2+(\frac{2}{3})^2+(-\frac{4}{3})^2+(-\frac{1}{3})^2+(-\frac{4}{3})^2}}$$

$$= \frac{52}{\sqrt{120}\cdot\sqrt{66}} = \boxed{0.5843}$$

cosine distance = $1 - 0.5843 = 0.4157$

U_1 & U_3:

$$\text{cosine similarity} = \frac{\frac{2}{3}\times(-1) + (-\frac{1}{3})\times 2}{\sqrt{(\frac{2}{3})^2+(\frac{5}{3})^2+(\frac{5}{3})^2+(-\frac{7}{3})^2+(-\frac{1}{3})^2+(-\frac{4}{3})^2}\cdot\sqrt{(\frac{2}{3})^2+(\frac{5}{3})^2+(\frac{2}{3})^2+(-\frac{4}{3})^2+(-\frac{1}{3})^2+(-\frac{4}{3})^2}}$$

$$= \frac{-4}{\sqrt{120}\cdot\sqrt{10}} = -0.1155$$

cosine distance = $1 + 0.1155 = \boxed{1.1155}$

U_2 & U_3:

$$\text{cosine similarity} = \frac{(\frac{5}{3})\times(-2) + (-\frac{1}{3})\times 1 + (-\frac{4}{3})\times 2}{\sqrt{(\frac{2}{3})^2+(\frac{5}{3})^2+(\frac{2}{3})^2+(-\frac{4}{3})^2+(-\frac{1}{3})^2+(-\frac{4}{3})^2}\cdot\sqrt{(-1)^2+(-2)^2+1^2+2^2}}$$

$$= \frac{-19}{\sqrt{66}\cdot\sqrt{10}} = -0.7396$$

cosine distance = $1 + 0.7396 = \boxed{1.7396}$

Problem 2

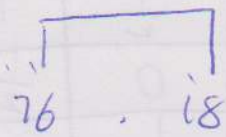
(a)

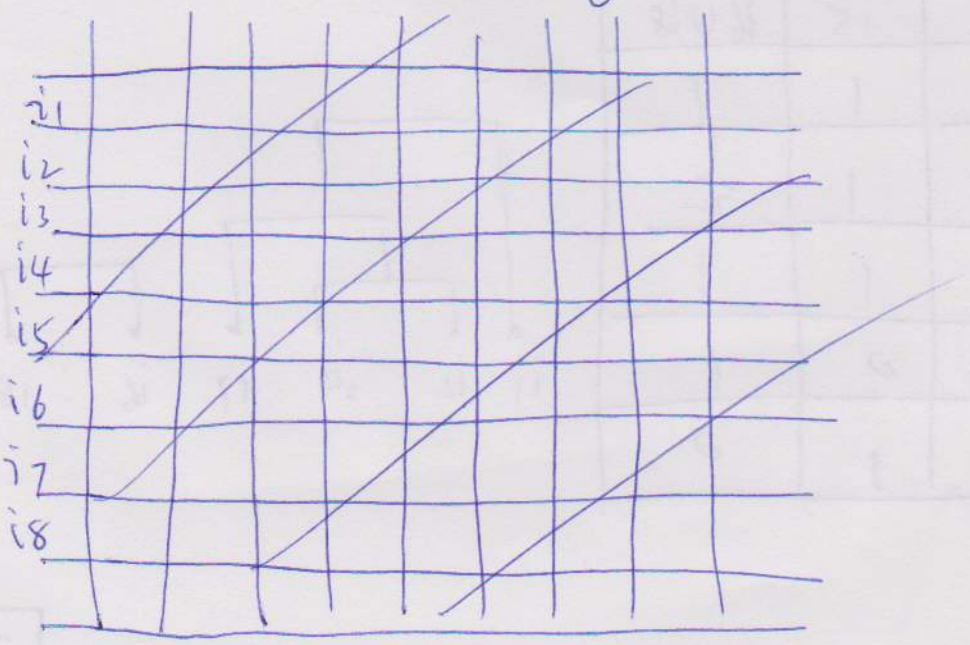
	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	1	1		1			1	
u_2		1	1	1				
u_3				1		1	1	1

1 − Jaccard similarity

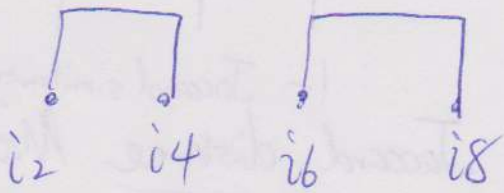
Jaccard distance Matrix: (~~de~~ column vector)

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
i_1	0	$\frac{1}{2}$	1	$\frac{2}{3}$	1	1	$\frac{1}{2}$	1
i_2	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{3}$	1	1	$\frac{2}{3}$	1
i_3	1	$\frac{1}{2}$	0	$\frac{2}{3}$	1	1	1	1
i_4	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	0	1	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$
i_5	1	1	1	1	0	1	1	1
i_6	1	1	1	$\frac{2}{3}$	1	0	$\frac{1}{2}$	0
i_7	$\frac{1}{2}$	$\frac{2}{3}$	1	$\frac{1}{3}$	1	$\frac{1}{2}$	0	$\frac{1}{2}$
i_8	1	1	1	$\frac{2}{3}$	1	0	$\frac{1}{2}$	0





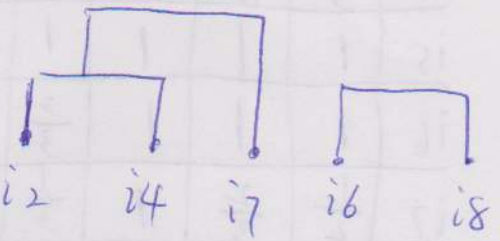
	i1	i2	i3	i4	i5	i6Ui8	i7
i1	0	$\frac{1}{2}$	1	$\frac{2}{3}$	1	1	$\frac{1}{2}$
i2	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{3}$	1	1	$\frac{2}{3}$
i3	1	$\frac{1}{2}$	0	$\frac{2}{3}$	1	1	1
i4	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	0	1	$\frac{2}{3}$	$\frac{1}{3}$
i5	1	1	1	1	0	1	1
i6Ui8	1	1	1	$\frac{2}{3}$	1	0	$\frac{1}{2}$
i7	$\frac{1}{2}$	$\frac{2}{3}$	1	$\frac{1}{3}$	1	$\frac{1}{2}$	0



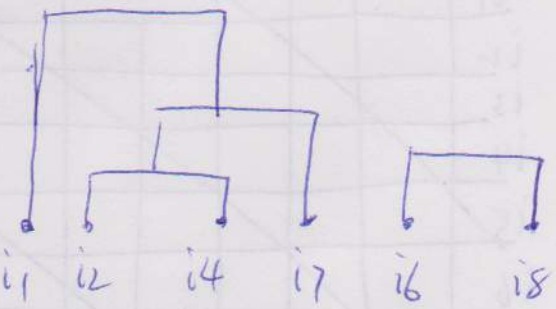
	i1	i2Ui4	i3	i5	i6Ui8	i7
i1	0	$\frac{1}{2}$	1	1	1	$\frac{1}{2}$
i2Ui4	$\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{2}{3}$	$\frac{1}{3}$
i3	1	$\frac{1}{2}$	0	1	1	1
i5	1	1	1	0	1	1
i6Ui8	1	$\frac{2}{3}$	1	1	0	$\frac{1}{2}$
i7	$\frac{1}{2}$	$\frac{1}{3}$	1	1	$\frac{1}{2}$	0

Choose distance smaller

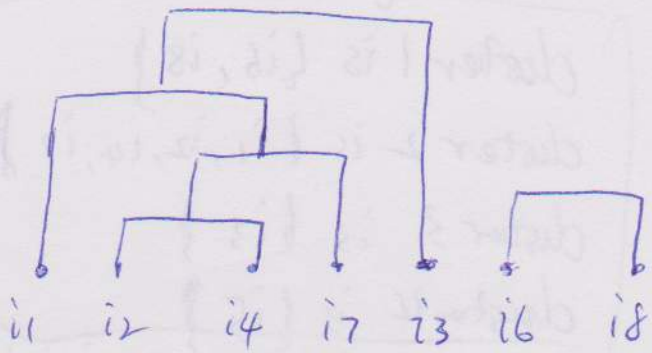
$\frac{1}{2} < \frac{2}{3}$



	i1	i2Ui4Ui7	i3	i5	i6Ui8
i1	0	$\frac{1}{2}$	1	1	1
i2Ui4Ui7	$\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{1}{2}$
i3	1	$\frac{1}{2}$	0	1	1
i5	1	1	1	0	1
i6Ui8	1	$\frac{1}{2}$	1	1	0

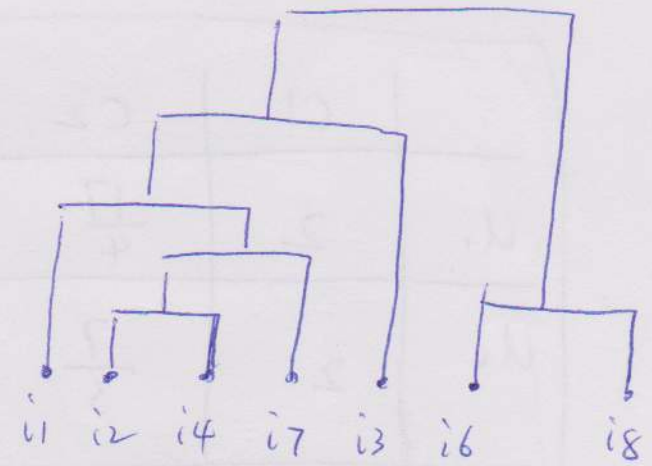


	$i_1 \cup i_2 \cup i_4 \cup i_7$	i_3	i_5	$i_6 \cup i_8$
$i_1 \cup i_2 \cup i_4 \cup i_7$	0	$\frac{1}{2}$	1	$\frac{1}{2}$
i_3	$\frac{1}{2}$ ✓	0	1	1
i_5	1	1	0	1
$i_6 \cup i_8$	$\frac{1}{2}$	1	1	0

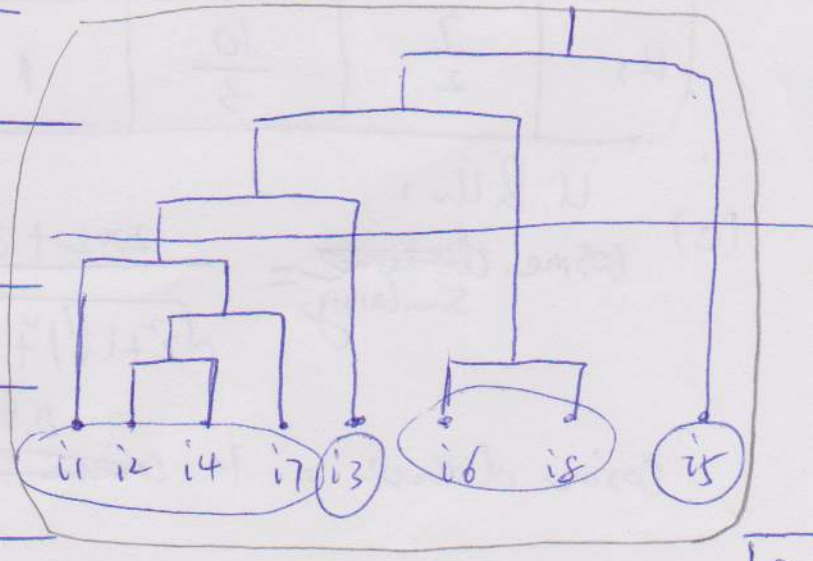


	$i_1 \cup i_2 \cup i_3 \cup i_4 \cup i_7$	i_5	$i_6 \cup i_8$
i_5	0		
$i_6 \cup i_8$			

	$i_{1247} \cup i_3$	i_5	i_{68}
$i_{1247} \cup i_3$	0	1	$\frac{1}{2}$
i_5	1	0	1
i_{68}	$\frac{1}{2}$ ✓	1	0



	$i_{12347} \cup i_{68}$	i_5
$i_{12347} \cup i_{68}$	0	1
i_5	1	0



So we get 4 clusters.

cluster 1 is $\{i6, i8\}$
cluster 2 is $\{i1, i2, i4, i7\}$
cluster 3 is $\{i3\}$.
cluster 4 is $\{i5\}$

(b)

	i6 i8 cluster 1	i1 i2 i4 i7 cluster 2	i3 cluster 3	i5 cluster 4
u_1	~~$\frac{2+2}{2}$~~ 2	$\frac{4+5+5+3}{4}$		1
u_2	~~2~~	$\frac{3+3+1}{3}$	4	1
u_3	$\frac{4+3}{2}$	$\frac{2+3+5}{3}$	1	

	C1	C2	C3	C4
u_1	2	$\frac{17}{4}$		1
u_2	2	$\frac{7}{3}$	4	1
u_3	$\frac{7}{2}$	$\frac{10}{3}$	1	

(c) u_1 & u_2:

$$\text{cosine similarity} = \frac{2\times 2 + \frac{17}{4}\times\frac{7}{3} + 1\times 1}{\sqrt{2^2+(\frac{17}{4})^2+1^2}\cdot\sqrt{2^2+(\frac{7}{3})^2+4^2+1^2}} = \frac{179}{\sqrt{369}\sqrt{258}} = 0.604$$

$$\text{Cosine distance} = 1 - \text{~~cosine Similarity~~} \overset{0.604}{=} \boxed{0.396}$$

u_1 & u_3:

~~cosine similarity~~ $= \dfrac{2\times\frac{7}{2} + \frac{7}{3}\times\frac{10}{3} + 4\times 1}{\sqrt{2^2+(\frac{7}{3})^2+4^2+1^2}\cdot\sqrt{(\frac{7}{2})^2+(\frac{10}{3})^2+1^2}}$

cosine similarity $= \dfrac{2\times\frac{7}{2} + \frac{17}{4}\times\frac{10}{3}}{\sqrt{2^2+(\frac{17}{4})^2+1^2}\cdot\sqrt{(\frac{7}{2})^2+(\frac{10}{3})^2+1^2}} = 0.8930$

cosine distance $= 1 - 0.8930 =$ **0.107**

u_2 & u_3:

cosine similarity $= \dfrac{2\times\frac{7}{2} + \frac{7}{3}\times\frac{10}{3} + 4\times 1}{\sqrt{2^2+(\frac{7}{3})^2+4^2+1^2}\cdot\sqrt{(\frac{7}{2})^2+(\frac{10}{3})^2+1^2}} = 0.7398$

cosine distance $= 1 - 0.7398 =$ **0.2602**

Problem 3

(a) $dis(P_1, P_5) = \sqrt{(6-2)^2+(7-2)^2} = \sqrt{41}$ ✓

$dis(P_1, P_6) = \sqrt{(8-2)^2+(7-2)^2} = \sqrt{61}$

choose $\sqrt{41}$

$dis(P_2, P_5) = \sqrt{(6-3)^2+(7-4)^2} = \sqrt{18}$ ✓

$dis(P_2, P_6) = \sqrt{(8-3)^2+(7-4)^2} = \sqrt{34}$

choose $\sqrt{18}$

$dis(P_3, P_5) = \sqrt{(6-4)^2+(7-7)^2} = 2$ ✓

$dis(P_3, P_6) = \sqrt{(8-4)^2+(7-7)^2} = 4$

choose 2

$dis(P_4, P_5) = \sqrt{(6-5)^2+(7-3)^2} = \sqrt{17}$ ✓

$dis(P_4, P_6) = \sqrt{(8-5)^2+(7-3)^2} = \sqrt{25}$

choose $\sqrt{17}$

$dis(P_7, P_5) = \sqrt{(6-8)^2+(7-1)^2} = \sqrt{40}$

$dis(P_7, P_6) = \sqrt{(8-8)^2+(7-1)^2} = \sqrt{36}$ ✓

choose $\sqrt{36}$

$dis(P_8, P_5) = \sqrt{(6-9)^2+(7-3)^2} = \sqrt{25}$

$dis(P_8, P_6) = \sqrt{(8-9)^2+(7-3)^2} = \sqrt{17}$ ✓ chose $\sqrt{17}$

11

cluster 1 is $\{P_5(6,7), P_1(2,2), P_2(3,4), P_3(4,7), P_4(5,3)\}$

cluster 2 is $\{P_6(8,7), P_7(8,1), P_8(5,3)\}$

cluster 1's centroid $\left(\frac{6+2+3+4+5}{5}, \frac{7+2+4+7+3}{5}\right) = (4, 4.6)$

cluster 2's centroid $\left(\frac{8+8+5}{3}, \frac{7+1+3}{3}\right) = (8.33, 3.67)$

The first iteration produces 2 clusters whose centroids are (4, 4.6) and (8.33, 3.67) respectively.

Then enter the second iteration:

$dis(P_1, C_1) = \sqrt{(4-2)^2+(4.6-2)^2} = 3.28$ ✓

$dis(P_1, C_2) = \sqrt{(8.33-2)^2+(3.67-2)^2} = 6.55$

choose 3.28

$dis(P_2, C_1) = \sqrt{(4-3)^2+(4.6-4)^2} = 1.17$ ✓

$dis(P_2, C_2) = \sqrt{(8.33-3)^2+(3.67-4)^2} = 5.34$

choose 1.17

$dis(P_3, C_1) = \sqrt{(4-4)^2+(4.6-7)^2} = 2.4$ ✓

$dis(P_3, C_2) = \sqrt{(8.33-4)^2+(3.67-7)^2} = 5.46$ choose 2.4

$dis(P_4, C_1) = \sqrt{(4-5)^2+(4.6-3)^2} = 1.89$ ✓

$dis(P_4, C_2) = \sqrt{(8.33-5)^2+(3.67-3)^2} = 3.4$

choose 1.89

$dis(P_5, C_1) = \sqrt{(4-6)^2+(4.6-7)^2} = 3.12$ ✓

$dis(P_5, C_2) = \sqrt{(8.33-6)^2+(3.67-7)^2} = 4.06$

choose 3.12

$dis(P_6, C_1) = \sqrt{(4-8)^2+(4.6-7)^2} = 4.66$

$dis(P_6, C_2) = \sqrt{(8.33-8)^2+(3.67-7)^2} = 3.35$ ✓

choose 3.35

$dis(P_7, C_1) = \sqrt{(4-8)^2+(4.6-1)^2} = 5.38$

$dis(P_7, C_2) = \sqrt{(8.33-8)^2+(3.67-1)^2} = 2.69$ ✓

choose 2.69

$dis(P_8, C_1) = \sqrt{(4-9)^2+(4.6-3)^2} = 5.25$

$dis(P_8, C_2) = \sqrt{(8.33-9)^2+(3.67-3)^2} = 0.95$ ✓

choose 0.95

cluster 1 is $\{P_1(2,2), P_2(3,4), P_3(4,7), P_4(5,3), P_5(6,7)\}$

13

cluster 2 is $\{P_6(8,7), P_7(8,1), P_8(9,3)\}$

cluster 1's centroid $\left(\frac{2+3+4+5+6}{5}, \frac{2+4+7+3+7}{5}\right) = (4, 4.6)$

cluster 2's centroid $\left(\frac{8+8+9}{3}, \frac{7+1+3}{3}\right) = (8.33, 3.67)$

The second iteration produces 2 clusters whose centroids are (4, 4.6) and (8.33, 3.67) respectively.

Because after first iteration and second iteration, the 2 clusters' centroids are at the same location, so the iterations will be end. So the final result is cluster 1 $\{P_1, P_2, P_3, P_4, P_5\}$ and $\{P_6, P_7, P_8\}$ with centroids (4, 4.6) and (8.33, 3.67) respectively.

(b) The first iteration

$dis(P_1, P_3) = \sqrt{(4-2)^2+(7-2)^2} = \sqrt{29}$ ✓

$dis(P_1, P_7) = \sqrt{(8-2)^2+(1-2)^2} = \sqrt{37}$

choose $\sqrt{29}$

$dis(P_2, P_3) = \sqrt{(4-3)^2+(7-4)^2} = \sqrt{10}$ ✓

choose $\sqrt{10}$

$dis(P_2, P_7) = \sqrt{(8-3)^2+(1-4)^2} = \sqrt{34}$

$dis(P_4, P_3) = \sqrt{(4-5)^2+(7-3)^2} = \sqrt{17}$

$dis(P_4, P_7) = \sqrt{(8-5)^2+(1-3)^2} = \sqrt{13}$ ✓

choose $\sqrt{13}$

$dis(P_5, P_3) = \sqrt{(4-6)^2+(7-7)^2} = \sqrt{4}$ ✓

$dis(P_5, P_7) = \sqrt{(8-6)^2+(1-7)^2} = \sqrt{40}$

choose $\sqrt{4}$

$dis(P_6, P_3) = \sqrt{(4-8)^2+(7-7)^2} = \sqrt{16}$ ✓

$dis(P_6, P_7) = \sqrt{(8-8)^2+(1-7)^2} = \sqrt{36}$

choose $\sqrt{16}$

$dis(P_8, P_3) = \sqrt{(4-9)^2+(7-3)^2} = \sqrt{41}$

$dis(P_8, P_7) = \sqrt{(8-9)^2+(1-3)^2} = \sqrt{5}$ ✓

choose $\sqrt{5}$

cluster 1 is $\{P_3(4,7), P_1(2,2), P_2(3,4), P_5(6,7), P_6(8,7)\}$

cluster 2 is $\{P_7(8,1), P_4(5,3), P_8(9,3)\}$

cluster 1's centroid $\left(\frac{4+2+3+6+8}{5}, \frac{7+2+4+7+7}{5}\right) = (4.6, 5.4)$

cluster 2's centroid $\left(\frac{8+5+9}{3}, \frac{1+3+3}{3}\right) = (7.33, 2.33)$

The first iteration produces 2 clusters whose centroids are (4.6, 5.4) & (7.33, 2.33) respectively.

15

Then enter 2nd iteration:

$dis(P_1, C_1) = \sqrt{(4.6-2)^2+(5.4-2)^2} = 4.28$ ✓

$dis(P_1, C_2) = \sqrt{(7.33-2)^2+(2.33-2)^2} = 5.34$

chose 4.28

~~Cluster~~

$dis(P_2, C_1) = \sqrt{(4.6-3)^2+(5.4-4)^2} = 2.13$ ✓

$dis(P_2, C_2) = \sqrt{(7.33-3)^2+(2.33-4)^2} = 4.64$

chose 2.13

$dis(P_3, C_1) = \sqrt{(4.6-4)^2+(5.4-7)^2} = 1.71$ ✓

$dis(P_3, C_2) = \sqrt{(7.33-4)^2+(2.33-7)^2} = 5.74$

chose ~~EF~~ 1.71

$dis(P_4, C_1) = \sqrt{(4.6-5)^2+(5.4-3)^2} = 2.43$

$dis(P_4, C_2) = \sqrt{(7.33-5)^2+(2.33-3)^2} = 2.42$ ✓

chose 2.42

$dis(P_5, C_1) = \sqrt{(4.6-6)^2+(5.4-7)^2} = 2.13$ ✓

$dis(P_5, C_1) = \sqrt{(7.33-6)^2+(2.33-7)^2} = $ ~~4.6~~ 4.86

chose 2.13

$dis(P_6, C_1) = \sqrt{(4.6-8)^2+(5.4-7)^2} = 3.76$ ✓

$dis(P_6, C_2) = \sqrt{(7.33-8)^2+(2.33-7)^2} = 4.72$

chose 3.76 ~~7.16~~

$dis(P_7, C_1) = \sqrt{(4.6-8)^2 + (5.4-11)^2} = 5.56$

$dis(P_7, C_2) = \sqrt{(7.33-8)^2 + (2.33-1)^2} = 1.49$ ✓

choose 1.49

$dis(P_8, C_1) = \sqrt{(4.6-9)^2 + (5.4-3)^2} = 5.01$

$dis(P_8, C_2) = \sqrt{(7.33-9)^2 + (2.33-3)^2} = 1.8$ ✓

choose 1.8

cluster 1 is $\{P_1(2,2), P_2(3,4), P_3(4,7), P_5(6,7), P_6(8,7)\}$

cluster 2 is $\{P_4(5,3), P_7(8,1), P_8(9,3)\}$

cluster 1's centroid is $\left(\frac{2+3+4+6+8}{5}, \frac{2+4+7+7+7}{5}\right)$

$= (4.6, 5.4)$

cluster 2's centroid is $\left(\frac{5+8+9}{3}, \frac{3+1+3}{3}\right) = (7.33, 2.33)$

The second iteration produces 2 clusters whose centroids ave $(4.6, 5.4)$ & $(7.33, 2.33)$ respectively.

Because after first iteration & seand iteration, the 2 clusters' centroids are at the same location, so the iterations will be end. So the final result is cluster 1 $\{P_1, P_2, P_3, P_5, P_6\}$ and $\{P_4, P_7, P_8\}$ with centroids $(4.6, 5.4)$ and $(7.33, 2.33)$ respectively.

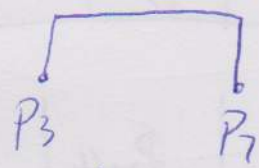
Yes. The clustering assignments have changed.

Problem 4:

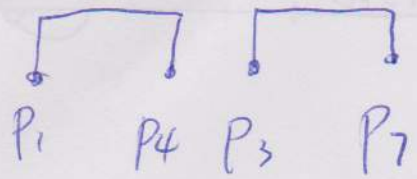
(a)

(i) Dendrogram (Euclidean Distance)

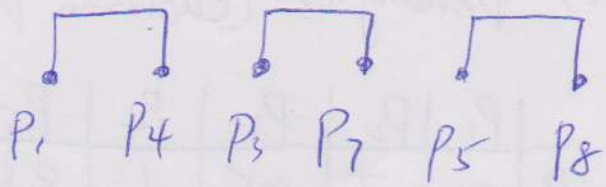
	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
P_1	0	7	4.5	1	8.8	3	5	10
P_2	7	0	2.5	8	1.8	4	2	3
P_3	4.5	2.5	0	5.5	4.3	1.5	0.5	5.5
P_4	1	8	5.5	0	9.8	4	6	11
P_5	8.8	1.8	4.3	9.8	0	5.8	3.8	1.2
P_6	3	4	1.5	4	5.8	0	2	7
P_7	5	2	0.5	6	3.8	2	0	5
P_8	10	3	5.5	11	1.2	7	5	0



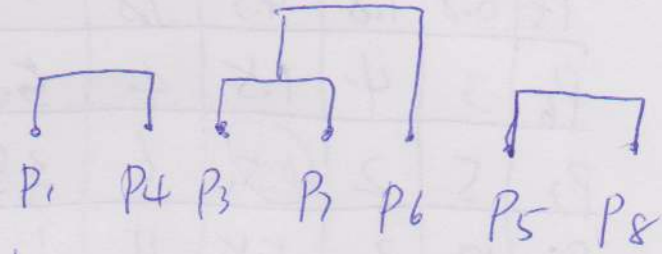
	P_1	P_2	$P_3 \cup P_7$	P_4	P_5	P_6	P_8
P_1	0	7	4.5	1	8.8	3	10
P_2	7	0	2	8	1.8	4	3
$P_3 \cup P_7$	4.5	2	0	5.5	3.8	1.5	5
P_4	1	8	5.5	0	9.8	4	11
P_5	8.8	1.8	3.8	9.8	0	5.8	1.2
P_6	3	4	1.5	4	5.8	0	7
P_8	10	3	5	11	1.2	7	0



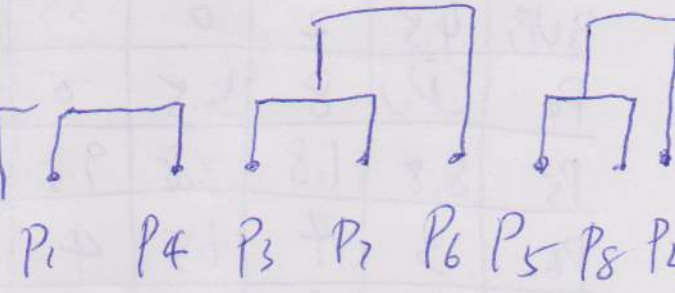
	$P_1 \cup P_4$	P_2	$P_{3,7}$	P_5	P_6	P_8
$P_1 \cup P_4$	0	7	4.5	8.8	3	10
P_2	7	0	2	1.8	4	3
$P_{3,7}$	4.5	2	0	3.8	1.5	5
P_5	8.8	1.8	3.8	0	5.8	1.2
P_6	3	4	1.5	5.8	0	7
P_8	10	3	5	1.2	7	0



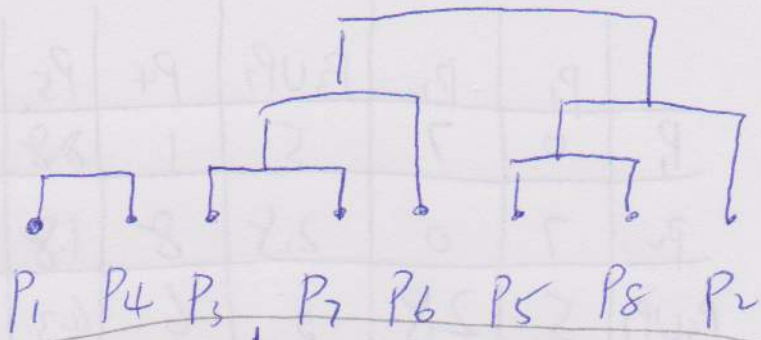
	$P_{1,4}$	P_2	$P_{3,7}$	$P_5 \cup P_8$	P_6
$P_{1,4}$	0	7	4.5	8.8	3
P_2	7	0	2	1.8	4
$P_{3,7}$	4.5	2	0	3.8	1.5
$P_5 \cup P_8$	8.8	1.8	3.8	0	5.8
P_6	3	4	1.5	5.8	0



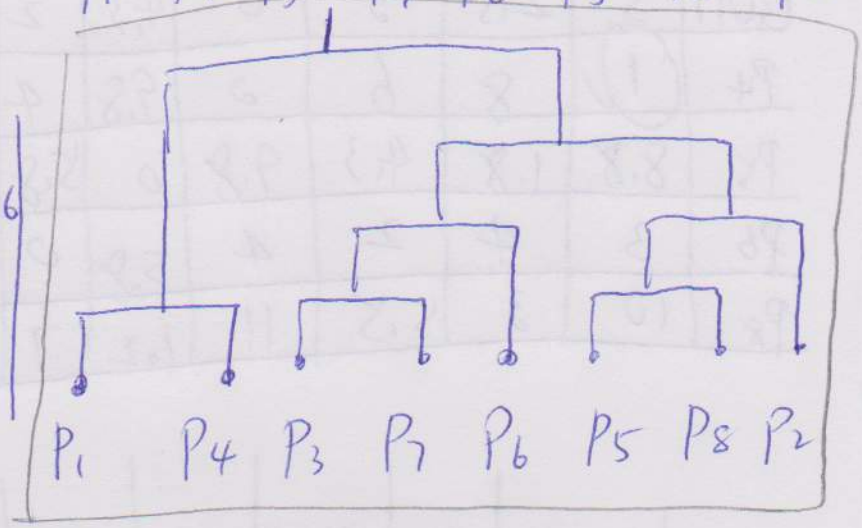
	$P_{1,4}$	P_2	$P_{3,7} \cup P_6$	$P_{5,8}$
$P_{1,4}$	0	7	3	8.8
P_2	7	0	2	1.8
$P_{3,7} \cup P_6$	3	2	0	3.8
$P_{5,8}$	8.8	1.8	3.8	0



	$P_{1,4}$	$P_2UP_{5,8}$	$P_{3,7,6}$
$P_{1,4}$	0	7	3
$P_2UP_{5,8}$	7	0	2
$P_{3,7,6}$	3	(2) ✓	0



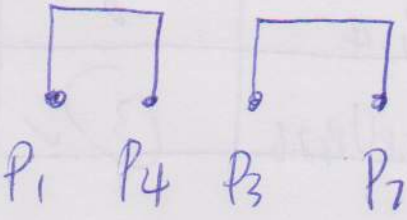
	$P_{1,4}$	$P_{2,5,8}UP_{3,7,6}$
$P_{1,4}$	0	3
$P_{2,5,8}UP_{3,7,6}$	(3) ✓	2



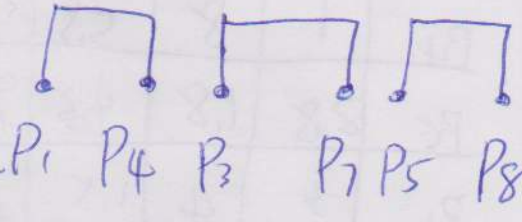
(ii)

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
P_1	0	7	4.5	1	8.8	3	5	10
P_2	7	0	2.5	8	1.8	4	2	3
P_3	4.5	2.5	0	~~5.5~~	4.3	1.5	0.5	5.5
P_4	1	8	5.5	0	9.8	4	6	11
P_5	8.8	1.8	4.3	9.8	0	5.8	3.8	1.2
P_6	3	4	1.5	4	5.8	0	2	7
P_7	5	2	(0.5) ✓	6	3.8	2	0	5
P_8	10	3	5.5	11	1.2	7	5	0

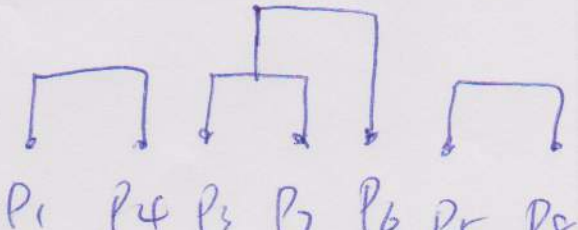
	P_1	P_2	$P_3 U P_7$	P_4	P_5	P_6	P_8
P_1	0	7	5	1	8.8	3	10
P_2	7	0	2.5	8	1.8	4	3
$P_3 U P_7$	5	2.5	0	6	4.3	2	5.5
P_4	1	8	6	0	9.8	4	11
P_5	8.8	1.8	4.3	9.8	0	5.8	1.2
P_6	3	4	2	4	5.8	0	7
P_8	10	3	5.5	11	1.2	7	0



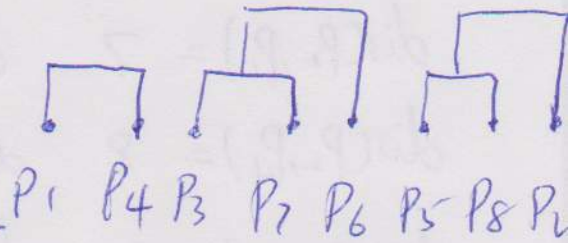
	$P_1 U P_4$	P_2	$P_{3,7}$	P_5	P_6	P_8
$P_1 U P_4$	0	8	6	9.8	4	11
P_2	8	0	2.5	1.8	4	3
$P_{3,7}$	6	2.5	0	4.3	2	5.5
P_5	9.8	1.8	4.3	0	5.8	1.2
P_6	4	4	2	5.8	0	7
P_8	11	3	5.5	1.2	7	0



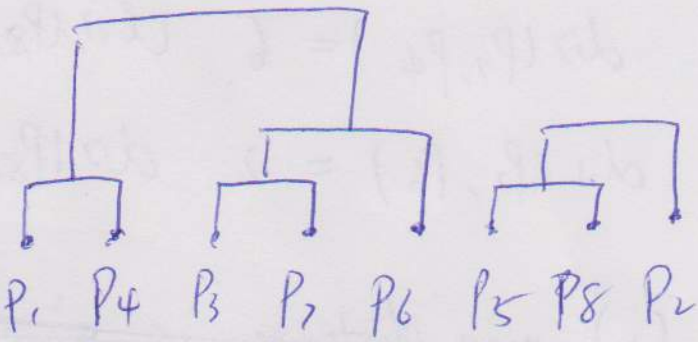
	$P_{1,4}$	P_2	$P_{3,7}$	$P_5 U P_8$	P_6
$P_{1,4}$	0	8	6	11	4
P_2	8	0	2.5	3	4
$P_{3,7}$	6	2.5	0	5.5	2
$P_5 U P_8$	11	3	5.5	0	7
P_6	4	4	2	7	0



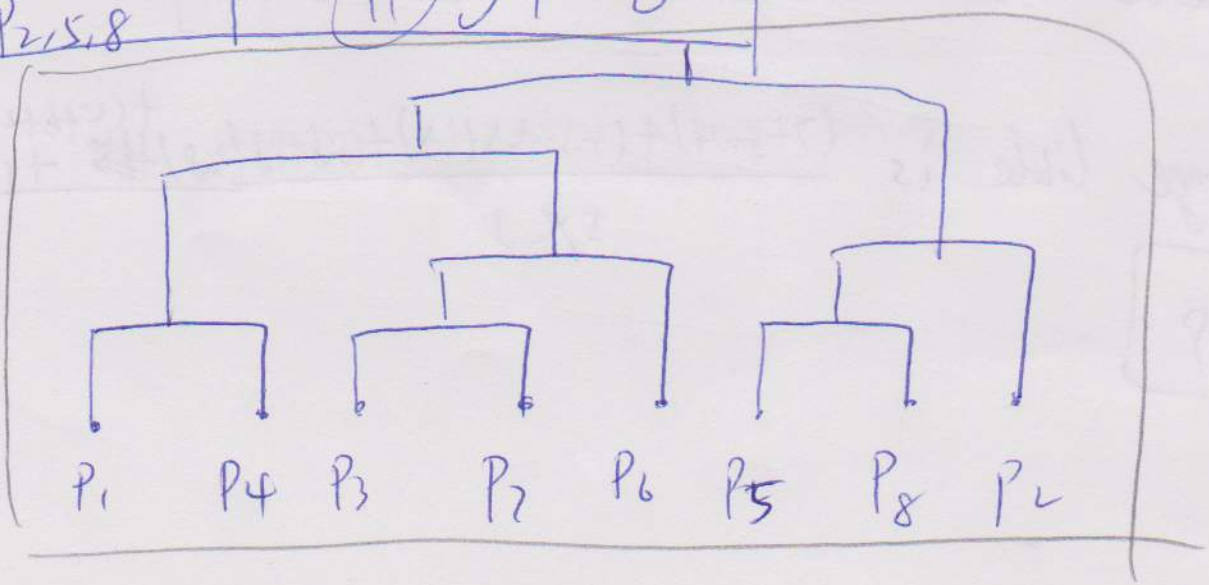
	$P_{1,4}$	P_2	$P_{3,7} \cup P_6$	$P_{5,8}$
$P_{1,4}$	0	8	6	11
P_2	8	0	4	3
$P_{3,7} \cup P_6$	6	4	0	7
$P_{5,8}$	11	(3) ✓	7	0



	$P_{1,4}$	$P_2 \cup P_{5,8}$	$P_{3,7,6}$
$P_{1,4}$	0	11	6
$P_2 \cup P_{5,8}$	11	0	7
$P_{3,7,6}$	(6) ✓	7	0



	$P_{1,4} \cup P_{3,7,6}$	$P_{2,5,8}$
$P_{1,4} \cup P_{3,7,6}$	0	11
$P_{2,5,8}$	(11) ✓	0



(b)

$P_2\ P_3\ P_5\ P_7\ P_8$ $\quad$ $P_1\ P_4\ P_6$

$dis(P_2, P_1) = 7$ $\quad dis(P_3, P_1) = 4.5$ $\quad dis(P_5, P_1) = 8.8$

$dis(P_2, P_4) = 8$ $\quad dis(P_3, P_4) = 5.5$ $\quad dis(P_5, P_4) = 9.8$

$dis(P_2, P_6) = 4$ $\quad dis(P_3, P_6) = 1.5$ ✓ $\quad dis(P_5, P_6) = 5.8$

$dis(P_7, P_1) = 5$ $\quad dis(P_8, P_1) = 10$

$dis(P_7, P_4) = 6$ $\quad dis(P_8, P_4) = 11$ ✓

$dis(P_7, P_6) = 2$ $\quad dis(P_8, P_6) = 7$

(i) ~~min distance is single link ; P_3 & P_6 (1.5)~~

Single link is minimum distance is $\boxed{1.5}$

(ii) Complete link is maximum distance is $\boxed{11}$

(iii) Average link is $\dfrac{(7+8+4)+(4.5+5.5+1.5)+(8.8+9.8+5.8)+(5+6+2)+(10+11+7)}{5\times 3}$

$= \boxed{6.39}$