CS 4371 HW5

1. $d(A,B) = \sqrt{(x_B - x_A)^2 - (Y_B - Y_A)^2}$ $d(1,2) = \sqrt{(2-2)^2 + (5-10)^2} = \sqrt{25}$

A	1	2	3	4	5	6	7	8
1	0	$\sqrt{25}$	$\sqrt{36}$	$\sqrt{13}$	$\sqrt{50}$	$\sqrt{52}$	$\sqrt{65}$	$\sqrt{5}$
2		0	$\sqrt{17}$	$\sqrt{18}$	$\sqrt{25}$	$\sqrt{17}$	$\sqrt{10}$	$\sqrt{20}$
3			0	$\sqrt{15}$	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{53}$	$\sqrt{41}$
4				0	$\sqrt{13}$	$\sqrt{17}$	$\sqrt{52}$	$\sqrt{2}$
5					0	$\sqrt{2}$	$\sqrt{45}$	$\sqrt{25}$
6						0	$\sqrt{29}$	$\sqrt{29}$
7							0	$\sqrt{58}$
8								0

Cluster 1 (S1): A2 = (2,5)
Cluster 2 (S2): A4 = (5,8)
Cluster 3 (S3): A8 = (4,9)

seed ⇒ S

A1: $d(A1,S1) = \sqrt{25}$ $d(A1,S2) = \sqrt{13}$ $d(A1,S3) = \sqrt{5}$
 Cluster 2 (S2)
A2: $d(A2,S1) = 0$ (S1=A2) $d(A2,S2) = \sqrt{18}$ $d(A2,S3) = \sqrt{20}$
 Cluster 1 (S1)
A3: $d(A3,S1) = \sqrt{37}$ $d(A3,S2) = \sqrt{25}$ $d(A3,S3) = \sqrt{41}$
 Cluster 2 (S2)
A4: $d(A4,S1) = \sqrt{18}$ $d(A4,S2) = \sqrt{13}$ (S2=A4) $d(A4,S3) = \sqrt{2}$
 Cluster 3 (S3)
A5: $d(A5,S1) = \sqrt{25}$ $d(A5,S2) = \sqrt{13}$ $d(A5,S3) = \sqrt{25}$
 Cluster 2 (S2)
A6: $d(A6,S1) = \sqrt{17}$ $d(A6,S2) = \sqrt{17}$ $d(A6,S3) = \sqrt{29}$
 Cluster 1 or 2 (S1 and S2 close)
A7: $d(A7,S1) = \sqrt{10}$ $d(A7,S2) = \sqrt{52}$ $d(A7,S3) = \sqrt{58}$
 Cluster 1
A8: $d(A8,S1) = \sqrt{20}$ $d(A8,S2) = \sqrt{2}$ $d(A8,S3) = 0$ (S3=A8)
 Cluster 3 (S3)

Cluster 1 (C1): {(A2, A6, A7)}
Cluster 2 (C2): {(A3, A4, A5)}
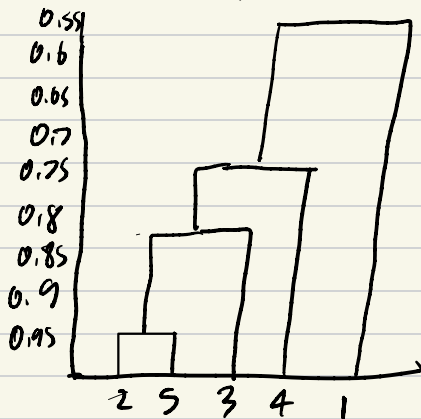Cluster 3 (C3): {(A1, A8)}
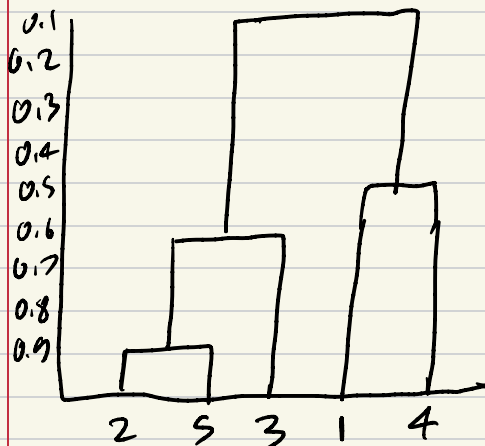
C1 centr: $((2+6+1)/3, (5+4+2)/3) = (3, 3.66)$
C2 centr: $((8+5+7)/3, (4+8+5)/3) = (6.66, 5.66)$
C3 centr: $((2+4)/2, (10+9)/2) = (3, 9.5)$

2. Single Link

Complete Link

3. $r_{S1} = \dfrac{\text{Sum}(xY) - \text{Sun}(x)\text{Sum}(\frac{y}{n})}{\sqrt{\text{Sum}(x^2) - \frac{\text{Sum}(x)^2}{n}}\sqrt{\text{Sum}(y^2) - \frac{\text{Sum}(y^2)}{n}}}$

$= \dfrac{(45) - (32)\frac{24}{5}}{\sqrt{(89) - \frac{(32)^2}{5}}\sqrt{29 - \frac{(24)^2}{5}}} = 0.948$

User 1: $r = 0.948$
User 6: $r = 0.923$
User 4: $r = 0.399$

$N = \{1, 6, 4\}$

$pred(5,2) = \dfrac{r_{S1} \times 1 + r_{S6} \times 6 + r_{S4} \times 4}{|r_{S1}| + |r_{S6}| + |r_{S4}|}$

$= \dfrac{0.948 + 0.923 \cdot 6 + 0.399 \cdot 5}{0.948 + 0.923 + 0.399} = 7$

$Sim(2,5) = \dfrac{65}{\sqrt{55} \times \sqrt{85}} = .904$

Product 1: 0.904
Product 2: 0.866 $N = \{5, 1\}$

$pred(5,2) = \dfrac{\sum sim(2,i) r_{ij}}{2} (|sim(2,i)|) = \dfrac{0.904 \cdot 9 + 0.866 \times 5}{|0.904| + |0.866|} = 7$