Abdelrahman Adel Abdelfattah Sheet 3 — 17012296 NOTES

K-means Clustering :-

(1) 3 clusters, Center A1, A4, A7

D(A1, A1) = 0
D(A1, A4) = $\sqrt{13}$ ≈ 3.61
D(A1, A7) = $\sqrt{65}$ ≈ 8.06

D(A2, A1) = 5
D(A2, A4) = 4.24
D(A2, A7) = 3.16

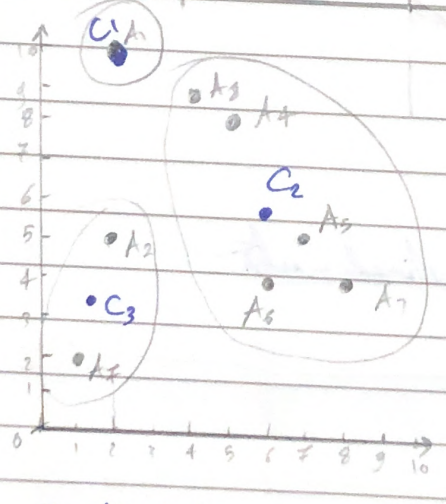
D(A3, A1) = 6
D(A3, A4) = 5
D(A3, A7) = 7.28

D(A5, A1) = 7.07
D(A5, A4) = 3.61
D(A5, A7) = 6.708

D(A6, A1) = 7.211
D(A6, A4) = 4.123
D(A6, A7) = 5.385

D(A8, A1) = 2.24
D(A8, A4) = 1.41
D(A8, A7) = 7.615

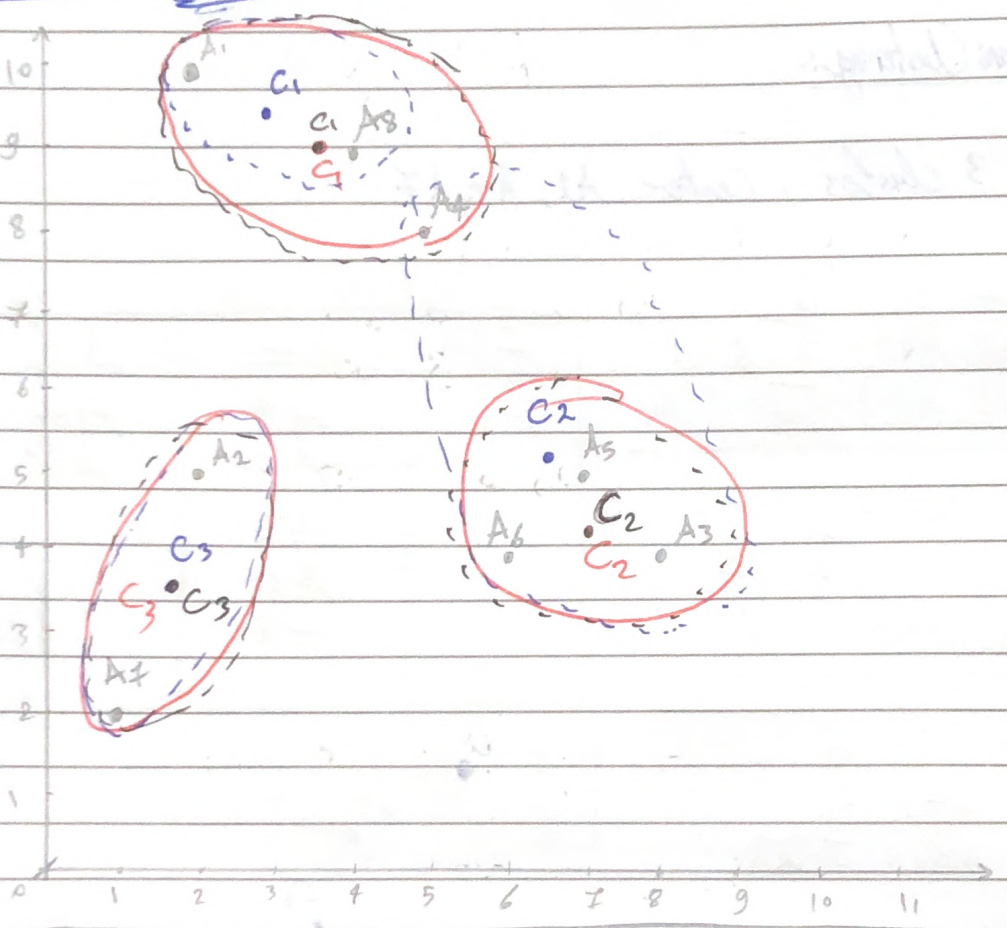
Clusters		
A1	A4	A7
A_1	A3	A2
	A4	A7
	A5	
	A6	
	A8	



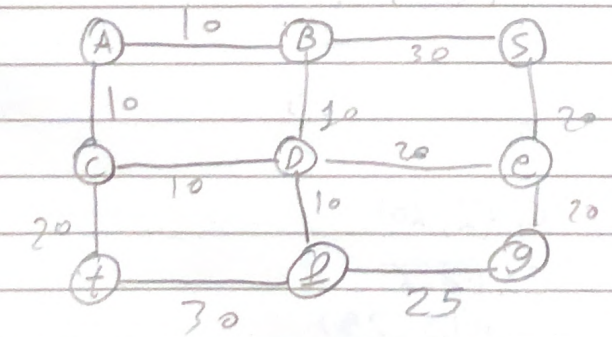
C_1 = (2, 10)
C_2 (6.6
C_3 = (1.5, 3.9)

2nd Epoch :- & 3rd Epoch & ~~pass~~ [last one 4th epoch]

NOTES.



Normalized Cuts :-



	A	B	S	C	D	e	t	f	g
A	20	0	0	0	0	0	0	0	0
B	0	50	0	0	0	0	0	0	0
S	0	0	50						
C	0			40					
D	0				50				
e	0					60			
t	0						80		
f	0							65	
g	0								45

NOTES

	A	B	S	C	D	e	t	f	g
A	0	10	0	10	0	0	0	0	0
B	10	0	30	0	10	0	0	0	0
S	0	30	0	0	0	20	0	0	0
C	10	0	0	0	10	0	20	0	0
D	0	10	0	10	0	20	0	10	0
e	0	0	20	0	20	0	0	0	20
t	0	0	0	20	0	0	0	30	0
f	0	0	0	0	10	0	30	0	25
g	0	0	0	0	0	20	0	25	0

A

	A	B	S	C	D	e	t	f	g
A	20	-10	0	-10	0	0	0	0	0
B	-10	50	-30	0	-10	0	0	0	0
S	0	-30	50	0	0	-20	0	0	0
C	-10	0	60	40	-10	0	-20	0	0
D	0	-10	0	-10	50	-20	0	-10	0
e	0	0	-20	0	-20	60	0	0	-20
t	0	0	0	-20	0	0	50	-30	0
f	0	0	0	0	-10	0	-30	65	-25
g	0	0	0	0	0	-20	0	-25	45

$$L = \Delta - A$$

NOTES.

$C_1 = \{A, B, S\}$

$\{A, C, D, e, t, f, g\}$

$C_1 = [0, 1, 1, 0, 0, 0, 0, 0, 0]^T$

$C_2 [1, 0, 0, 1, 1, 1, 1, 1, 1]^T$

$C_1^T \Delta C_1$

$= [0, 50, 50, 0, 0, 0, 0, 0, 0] C_1$

$= 100 = Vol(A)$

$C_2^T \Delta C_2 = 330 = Vol(B)$

$CUT(A, B) = 40$

$NCUT(A, B) = 40\left(\frac{1}{100} + \frac{1}{330}\right)$

$= \frac{86}{165} \simeq 0.521$

$C_2 \Rightarrow \{A, B, S, D, e\}$

$\{C, t, f, g\}$

$CUT(A, B) = 50$

$C_1 = [0, 1, 1, 1, 0, 1, 1, 0, 0, 0]$

$C_2 = [0, 0, 0, 1, 0, 0, 1, 1, 1]$

$C_1^T \Delta C_1 = 230$

$C_2^T \Delta C_2 = 200$

$NCUT = 50\left(\frac{1}{230} + \frac{1}{200}\right)$

$= 0.47$

C_1 is better a graph cut but worse as a NCUT

C_2 is better a NCUT but worse as a graph cut

(3) a) ~~A = Distance Matrix~~

~~Similarity Matrix = K(x,y) = Δ~~

A = Similarity

Δ = Distance Matrix

(4) 1. Conditional Entropy $H(T|C_i) = -\sum_{j=1}^{s} \left(\frac{n_{ij}}{n_i}\right) \log\left(\frac{n_{ij}}{n_i}\right)$

2. Purity$_i = \frac{1}{n_i} \max_k n_{ij}$, Purity $= \sum_{i=1} \frac{n_i}{n}$ Purity$_i$

3. Pairwise measure (Jaccard and Rand index)

4. Max matching when # of clusters = 3, $= \operatorname{argmax}\left\{\frac{W(M)}{n}\right\} w(M) = \sum_{e \in M} w(e)$

5. F - Measure $Prec_i =$ Purity$_i$

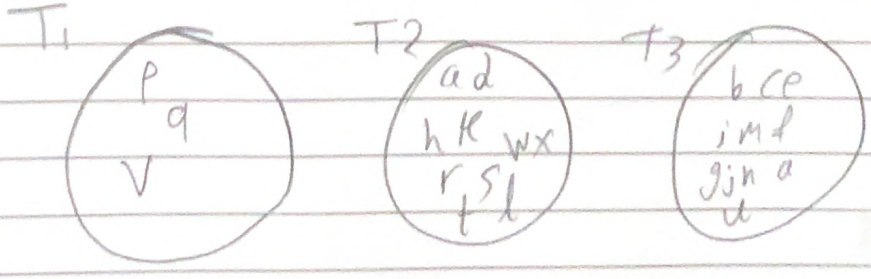
$$rec_i = \frac{n_{ji}}{|T_{ji}|}$$

$$F_i = \frac{2 * Prec_i \times rec_i}{Prec_i + rec_i}$$

$$F = \frac{1}{r} \sum_{i=1}^{r} F_i$$

$$\text{Jaccard Index} = \frac{TP}{TP+FN+FP}$$

$$\text{Define Rand Index} = \frac{TP+TN}{TP+TN+FN+FP}$$



ii) $BetaCV = \frac{W_{in}}{N_{in}} / \frac{W_{out}}{N_{out}}$

$$NC = \sum_{i=1}^{K} \frac{1}{\frac{W(C_{ai}, C_{bi})}{W(C_{ai}, C_{ai})} + 1}$$

#