CS – 540 – Fall 2019 – UW NETID: sshekhar4

HW 3

Problem 1: THIS IS THE WORKING. DENODROGRAM AT THE BOTTOM.

Part (a) – Hierarchical Agglomerative Clustering

The initial table is:-

	TYO	DEL	SHA	BJ	MUM	OSA	DHK	CCU	SEL
TYO	0	5847	1767	2099	6740	403	4894	5138	1156
DEL	5847	0	4243	3777	1163	5476	1425	1304	4686
SHA	1767	4243	0	1070	5039	1364	3161	3403	871
BJ	2099	3777	1070	0	4756	1777	3026	3266	956
MUM	6740	1163	5039	4756	0	6355	1895	1664	5608
OSA	403	5476	1364	1777	6355	0	4500	4744	821
DHK	4894	1425	3161	3026	1895	4500	0	244	3793
CCU	5138	1304	3403	3266	1664	4744	244	0	4038
SEL	1156	4686	871	956	5608	821	3793	4038	0

The closest pair has a distance of 244 between CCU & DHK

We merge CCU & DHK to get a new node (CCU-DHK)

The updated distance table is:-

	TYO	DEL	SHA	BJ	MUM	OSA	CCU-DHK	SEL
TYO	0	5847	1767	2099	6740	403	4894	1156
DEL	5847	0	4243	3777	1163	5476	1304	4686
SHA	1767	4243	0	1070	5039	1364	3161	871
BJ	2099	3777	1070	0	4756	1777	3026	956
MUM	6740	1163	5039	4756	0	6355	1664	5608
OSA	403	5476	1364	1777	6355	0	4500	821
CCU - DHK	4894	1304	3161	3026	1664	4500	0	3793
SEL	1156	4686	871	956	5608	821	3793	0

The closest pair has a distance of 403 between TYO & OSA.

We merge TYO & OSA to get a new node (TYO - OSA).

	DEL	SHA	BJ	MUM	TYO-OSA	CCU-DHK	SEL
DEL	0	4243	3777	1163	5476	1304	4686
SHA	4243	0	1070	5039	1364	3161	871
BJ	3777	1070	0	4756	1777	3026	956
MUM	1163	5039	4756	0	6355	1664	5608
TYO – OSA	5476	1364	1777	6355	0	4500	821
CCU-DHK	1304	3161	3026	1664	4500	0	3793
SEL	4686	871	956	5608	821	3793	0

The closest pair has a distance of 821 between TYO-OSA & SEL.

We merge TYO-OSA & SEL to get a new node (TYO-OSA-SEL)

	DEL	SHA	BJ	MUM	TYO-OSA-SEL	CCU-DHK
DEL	0	4243	3777	1163	4686	1304
SHA	4243	0	1070	5039	871	3161
BJ	3777	1070	0	4756	956	3026
MUM	1163	5039	4756	0	5608	1664
TYO-OSA-SEL	4686	871	956	5608	0	3793
CCU-DHK	1304	3161	3026	1664	3793	0

The closest pair has a distance of 871 between SHA and TYO-OSA-SEL.

We merge TYO-OSA-SEL & SHA to get a new node (TYO-OSA-SEL-SHA).

	DEL	BJ	MUM	TYO-OSA-SEL-SHA	CCU-DHK
DEL	0	3777	1163	4243	1304
BJ	3777	0	4756	956	3026
MUM	1163	4756	0	5039	1664
TYO-OSA-SEL-SHA	4243	956	5039	0	3161
CCU-DHK	1304	3026	1664	3161	0

The closest pair has a distance of 956 between BJ and TYO-OSA-SEL-SHA.

We merge BJ and TYO-OSA-SEL-SHA to get a new node (TYO-OSA-SEL-SHA-BJ)

	DEL	MUM	TYO-OSA-SEL-SHA-BJ	CCU-DHK
DEL	0	1163	3777	1304
MUM	1163	0	4756	1664
TYO-OSA-SEL-SHA-BJ	3777	4756	0	3026
CCU-DHK	1304	1664	3026	0

The closest pair has a distance of 1163 between DEL & MUM.

We merge DEL and MUM to get a new node (DEL-MUM).

	DEL-MUM	TYO-OSA-SEL-SHA-BJ	CCU-DHK
DEL-MUM	0	3777	1304
TYO-OSA-SEL-SHA-BJ	3777	0	3026
CCU-DHK	1304	3026	0

The closest pair has a distance of 1304 between DEL-MUM and CCU-DHK.

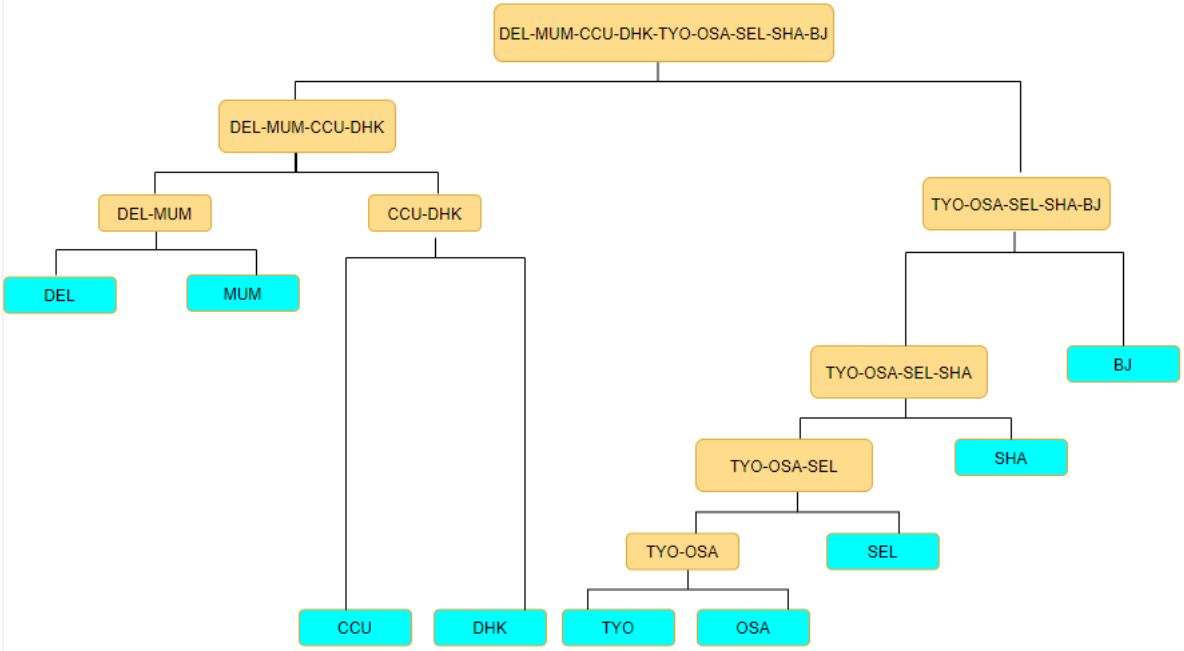
We merge DEL-MUM & CCU-DHK to get a new node (DEL-MUM-CCU-DHK)

	DEL-MUM-CCU-DHK	TYO-OSA-SEL-SHA-BJ
DEL-MUM-CCU-DHK	0	3026
TYO-OSA-SEL-SHA-BJ	3026	0

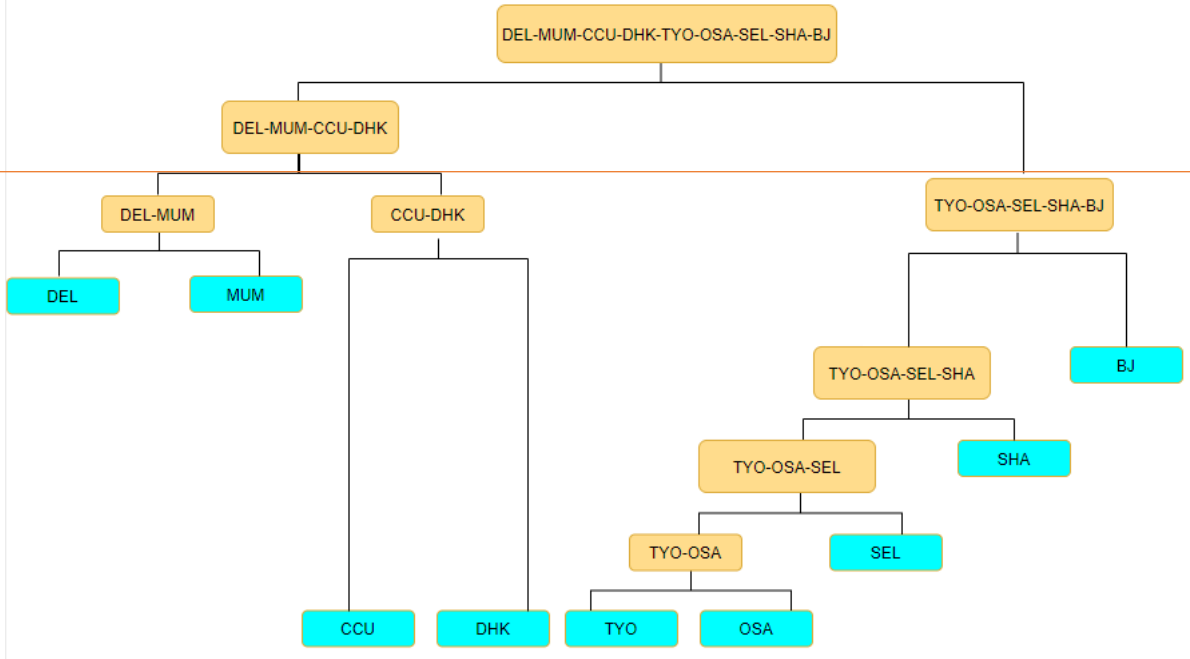
Finally only two clusters are left (DEL-MUM-CCU-DHK) and (TYO-OSA-SEL-SHA-BJ).

We can merge them to get a single node (DDU-MUM-CCU-DHK-TYO-OSA-SEL-SHA-BJ).

(i) The resulting dendrogram is:- ANSWER



(ii) If we want just three clusters: we can partition the dendrogram as:-



Thus the 3 clusters would be:- (DEL-MUM), (CCU-DHK) and (TYO-OSA-SEL-SHA-BJ). ANSWER

Part (b) – K- Means Clustering

Initial Cluster Centres – c1 (38.0, 127.0) & c2(27.0, 117.0)

(i)

Euclidean Distance Formula = sqrt((x2-x1)^2 + (y2-y1)^2)

	X	Y	Euclidean Distance from C1	Euclidean Distance from C2	**Decision**
TYO	35.7	139.7	12.9065	24.3100	**C1**
DEL	28.7	77.2	50.6609	39.8362	**C2**
SHA	31.2	121.5	8.7458	6.1554	**C2**
BJ	39.9	116.4	10.7689	12.9139	**C1**
MUM	19.1	72.9	57.3063	44.8020	**C2**
OSA	34.7	135.5	9.1181	20.0384	**C1**
DHK	23.8	90.4	39.2581	26.7917	**C2**
CCU	22.6	88.4	41.5586	28.9364	**C2**
SEL	37.6	127	0.4	14.5725	**C1**

ANSWER

In the clusters:-

C1:- TYO, BJ, OSA, SEL

C2:- DEL, SHA, MUM, DHK, CCU

(ii)

ANSWER

New cluster centre of C1:-

x-coordinate:- (35.7+39.9+34.7+37.6)/4 = 36.975

y-coordinate:- (139.7+116.4+135.5+127)/4 = 129.65

Thus, new cluster centre for C1 is (36.975, 129.65)

New cluster centre of C2:-

x-coordinate:- (28.7+31.2+19.1+23.8+22.6)/5 = 25.08

y-coordinate:- (77.2+121.5+72.9+90.4+88.4)/5 = 90.08

Thus, new cluster centre for C2 is (25.08, 90.08).

(iii)

Cluster Centres: C1**(36.975, 129.65)** & C2**(25.08, 90.08).**

	X	Y	Euclidean Distance from C1	Euclidean Distance from C2	**Decision**
TYO	35.7	139.7	10.1305	50.7437	**C1**
DEL	28.7	77.2	53.0987	13.3790	**C2**
SHA	31.2	121.5	9.9886	32.010	**C1**
BJ	39.9	116.4	13.5690	30.2055	**C1**
MUM	19.1	72.9	59.4985	18.1910	**C2**
OSA	34.7	135.5	6.2767	46.4275	**C1**
DHK	23.8	90.4	41.4022	1.3193	**C2**
CCU	22.6	88.4	43.6829	2.9954	**C2**
SEL	37.6	127	2.7227	39.9850	**C1**

ANSWER

The new clusters are:-

C1:- TYO, SHA, BJ, OSA, SEL

C2:- DEL, MUM, DHK, CCU