

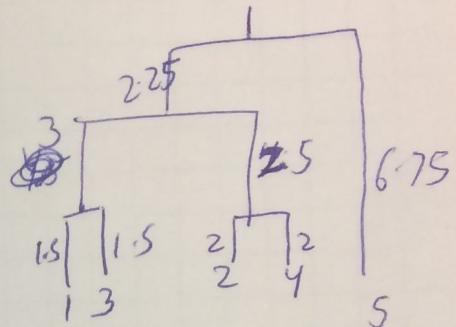
CS 576
 SPARSH AGARWAL
 9075905142

HW 3

AD

@ merging 1, 3

	1,3	2	4	5
1,3	4.5	4.5	5.5	
2		4	8	
4			8	
5				



merging 284.

	1,3	2,4	5
1,3	4.5	5.5	
2,4			8
5			

merging 1,3 8 2,4

	((1,3),(2,4))	5	6.75
((1,3),(2,4))			
5			

⑥

$$r_1 = (4+3+4+6)/3 = \cancel{17}/3$$

$$r_2 = (4+5+4+8)/3 = \cancel{27}$$

$$r_3 = (3+5+5+5)/3 = 6$$

$$r_4 = (4+4+5+8)/3 = 7$$

$$r_5 = (6+8+5+8)/3 = 9$$

	1	2	3	4	5
1				4	6
2				5	8
3					5
4					8
5					

$$D_{12} = 4 - \frac{17}{3} - 7 = -8.7$$

$$D_{23} = 5 - 7 - 6 = -8$$

$$D_{34} = 5 - 7 - 6 = -8$$

$$D_{45} = 8 - 7 - 9 = -8$$

$$D_{51} = 6 - \frac{17}{3} - 9 = -8.7$$

$$D_{13} = 3 - \frac{17}{3} - 6 = -8.7$$

$$D_{14} = 4 - \frac{17}{3} - 7 = -8.7$$

$$D_{24} = 4 - 7 - 7 = -10$$

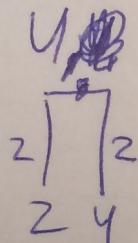
$$D_{25} = 8 - 7 - 9 = -8$$

$$D_{35} = 5 - 6 - 9 = -10$$

\therefore min are D_{35} or D_{24} so choosing D_{24} (U_1 , being node ~~odd~~)

$$d_{2U_1} = \cancel{\frac{4}{2}} + \left(\frac{7-7}{2} \right) = 2$$

$$d_{4U_1} = \frac{4}{2} \left(\frac{7-7}{2} \right) = 2$$



$$r_{U_1} = (2+3+6)/2 = 5.5$$

$$r_1 = (2+3+6)/2 = 5.5$$

$$r_3 = (3+3+5)/2 = 5.5$$

$$r_5 = (6+6+5)/2 = 8.5$$

	U_1	1	3	5
		2	3	6
			3	6
				5
		3		
			5	

$$D_{U_11} = 2 - 5.5 - 5.5 = -9$$

$$D_{U_13} = 3 - 5.5 - 5.5 = -8$$

$$D_{U_15} = 6 - 5.5 - 8.5 = -8$$

$$D_{13} = 3 - 5.5 - 5.5 = -8$$

$$D_{15} = 6 - 5.5 - 8.5 = -8$$

$$D_{35} = 5 - 5.5 - 8.5 = -9$$

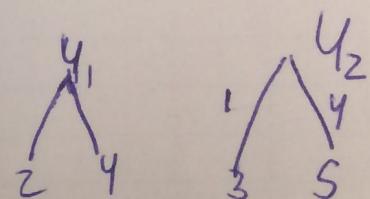
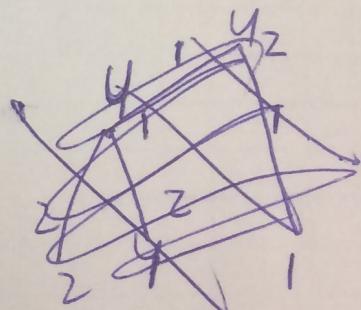
$\therefore D_{U_1}$ & D_{35} are min \therefore choosing $U_3, 5$ (U_2 being node)

$$\cancel{d_{3U_2}} = \cancel{\frac{2}{2}} + \cancel{\left(\frac{5.5 - 5.5}{2} \right)} = \cancel{0}$$

$$\cancel{d_{1U_2}} = \cancel{\frac{3}{2}} + \cancel{\left(\frac{5.5 - 5.5}{2} \right)} = \cancel{0}$$

$$d_{3U_2} = \frac{5}{2} + \left(\frac{5.5 - 8.5}{2} \right) = 1$$

$$d_{5U_2} = \frac{5}{2} + \left(\frac{8.5 - 5.5}{2} \right) = 4$$



$$r_{41} = (2+2)/1 = 4$$

$$r_1 = (2+2)/1 = 4$$

$$r_{42} = (2+2)/1 = 4$$

	<u>u_1</u>	<u>1</u>	<u>u_2</u>
u_1		2	2
	1		
u_2			2

$$D_{411} = 2 - 4 - 4 = -6$$

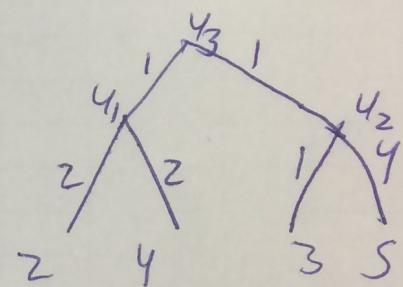
$$D_{u_1 u_2} = 2 - 4 - 4 = -6$$

$$D_{u_2 1} = 2 - 4 - 4 = -6$$

all are same so choosing u_1, u_2 (u_3 being ~~rooted~~ node)

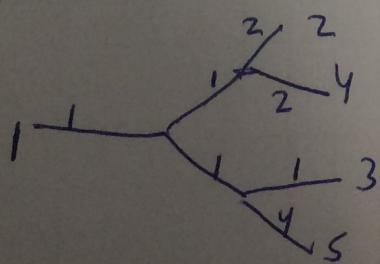
$$d_{u_1 u_3} = \frac{2}{2} + \left(\frac{4-4}{2} \right) = 1$$

$$d_{u_2 u_3} = \frac{2}{2} + \left(\frac{4-4}{2} \right) = 1$$



	<u>u_3</u>	<u>1</u>
u_3		1
	1	

∴ The first ~~rooted~~ tree is

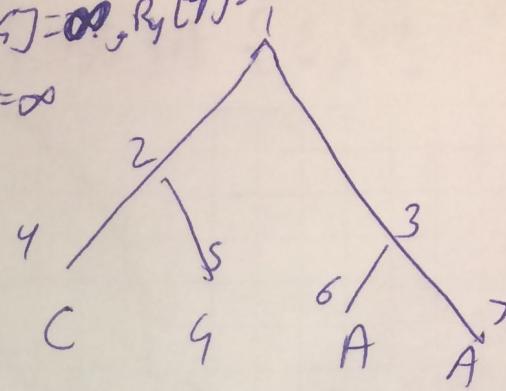


(C) neighbour joining algorithm is better for this data because the data does not have ~~all~~ ultrametric property for all node distances, since this is assumption for UPGMA, it will produce wrong tree. Thus neighbour joining algorithm is better

A2) @ $R_4[A] = \infty, R_4[C] = 0, R_4[G] = \infty, R_4[T] = \infty$
 $R_5[A] = \infty, R_5[C] = \infty, R_5[G] = 0, R_5[T] = \infty$
 $R_6[A] = 0, R_7[A] = 0$

$$R_3[A] = \min_A(0 + 0) + \min_A(0 + 0)$$

$$R_3[A] = 0$$



\therefore A will come at node 3, 6, 7
 Now for node 2, either C or G will come.

for node 2, change from A to G is less costly
 than change from A to C
 \therefore we will need G in node 2.

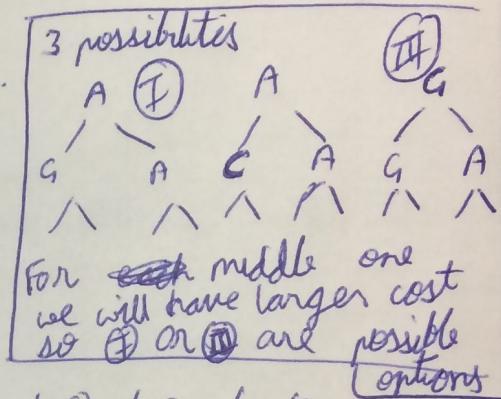
$$\therefore R_2[G] = 0 + 2 = 2$$

$$R_3[A] = 0$$

~~$R_3[A] = 0 + 1 = 1$~~

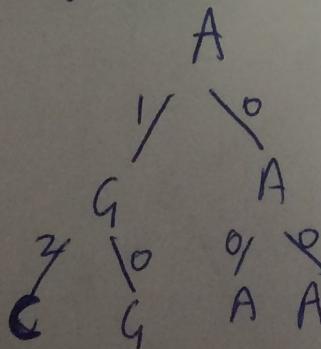
\therefore min cost of tree is $1 + 2 + 0 + 0 + 0 + 0 + 0$

$$R_1 \quad R_2 \quad R_3 \quad R_4 \quad R_5 \quad R_6 \quad R_7$$



\therefore Minimum cost is 3.

(b) The assignments will be



If A is main root node,
 & G at ②
 Then cost of $R_3[A] = 0$
 $R_2[G] = 0 + 2 = 2$
 $R_1[A] = 0 + 1 = 1$

If G is main node & G is at ②
 then cost of $R_3[A] = 0$
 $R_2[G] = 0 + 2 = 2$
 $R_1[G] = 0 + 1 = 1$

\therefore The (I) & (II) are 2 possible explanations

AB) ① Step 1

$$\{A\} \cup \{G\} = \{A, G\}$$

$$\{A, C\} \cap \{A\} = \{A\}$$

$$\{A\} \cup \{C\} = \{A, C\}$$

$$\{G\} = \{A, G\} \cap \{C, G\}$$

$$\{C, G\} = \{C\} \cup \{G\}$$

Step 2

from $\{A\}$ we cannot
keep $\{A\}$ as only $\{G\}$ is present

$\{G\}$ from $\{G\}$ we choose $\{G\}$

from $\{A\}$ we keep
 $\{A\}$

from $\{A, C\}$
we choose
 $\{A\}$.

$\{G\}$

$\{A\}$

$\{G\}$

$\{A\}$

$\{G\}$

$\{A\}$

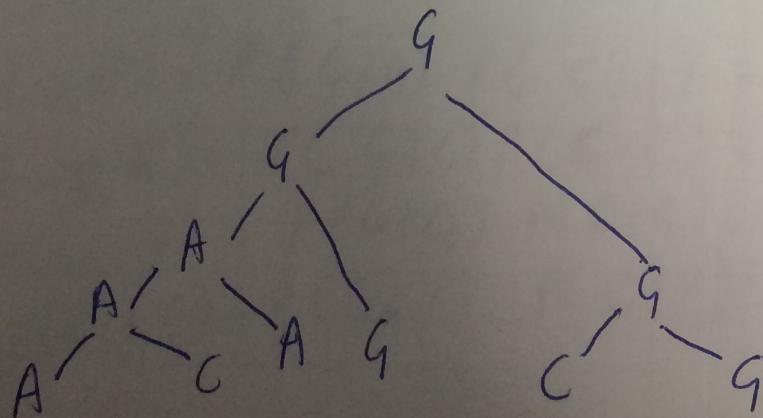
$\{G\}$

G

∴ Minimum number of changes found = 3

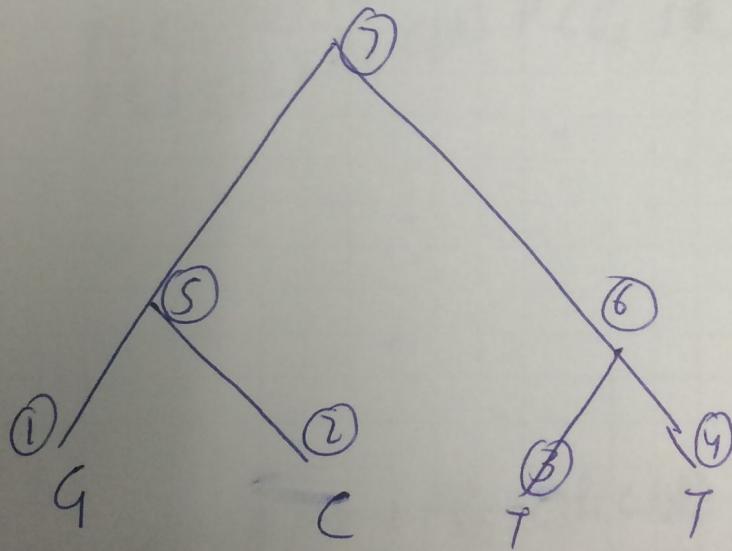
$$\begin{pmatrix} G \rightarrow C \\ G \rightarrow A \\ A \rightarrow C \end{pmatrix}$$

② Ancestral bases are



$$\text{Ans} P(L_k | a) = \left(\sum_b P(b|a) P(L_i | b) \right) \left(\sum_c P(c|a) \frac{P(L_j | c)}{P(L_j | a)} \right)$$

	A	C	G	T
P(L_1 x)	0	0	1	0
P(L_2 x)	0	1	0	0
P(L_3 x)	0	0	0	1
P(L_4 x)	0	0	0	1
P(L_5 x)	0			
P(L_6 x)				
P(L_7 x)				



$$P(L_5 | A) = P(G | A) P(C | A) = 0.1 \times 0.05 = 0.005$$

$$P(L_5 | C) = P(G | C) P(C | C) = 0.05 \times 0.8 = 0.04$$

$$P(L_5 | G) = P(G | G) P(C | G) = 0.8 \times 0.05 = 0.04$$

$$P(L_5 | T) = P(G | T) P(C | T) = 0.05 \times 0.1 = 0.005$$

$$P(T|A) = P(T|A) P(G|A) = 0.0025$$

$$P(T|C) = (P(T|C))^2 = 0.01$$

$$P(L_6|G) = (P(T|G))^2 = 0.0025$$

$$P(L_6|T) = (P(T|T))^2 = 0.64$$

$$\begin{aligned} P(L_7|A) &= P(L_5|A) P(A|A) + P(L_6|A) P(A|A) + \\ &\quad P(L_5|C) P(C|A) + P(L_6|C) P(C|A) + \\ &\quad P(L_5|G) P(G|A) + P(L_6|G) P(G|A) + \\ &\quad P(L_5|T) P(T|A) + P(L_6|T) P(T|A) + \\ &\quad P(L_5|A) P(A|A) + P(L_6|C) P(C|A) + \\ &\quad P(L_5|A) P(A|A) + P(L_6|G) P(G|A) + \\ &\quad P(L_5|A) P(A|A) + P(L_6|T) P(T|A) + \\ &\quad P(L_5|A) P(C|A) + P(L_6|A) P(C|A) + \\ &\quad P(L_5|C) \end{aligned}$$

⋮
⋮
⋮
⋮

$$= [P(L_5|A) P(A|A) + P(L_5|C) P(C|A) + P(L_5|G) P(G|A) + \\ P(L_5|T) P(T|A)] [P(L_6|A) P(A|A) + P(L_6|C) P(C|A) + \\ P(L_6|G) P(G|A) + P(L_6|T) P(T|A)]$$

$$= [(0.005 \times 0.8) + (0.04 \times 0.05) + (0.04 \times 0.1) + \\ (0.005 \times 0.05)] [(0.0025 \times 0.8) + (0.01 \times 0.05) \\ + (0.0025 \times 0.01) + (0.64 \times 0.05)]$$

$$P(L_7 | A) = \frac{[0.004 + 0.002 + 0.004 + 0.00025]}{[0.0005 + 0.0024 + 0.0005 + 0.00025 + 0.0032]}$$

$$= 0.01025 \times 0.06595$$

$$= 0.0000609875$$