CS 576

SPARSH AGARWAL

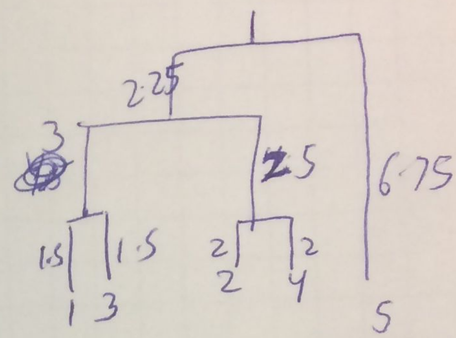
9075905142

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 2 & 4.

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

merging 1, 3 & 2, 4

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

6

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

$r_1 = (4+3+4+6)/3 = 17/3$

$r_2 = (4+5+4+8)/3 = 7$

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

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

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

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

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

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

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

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

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

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

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

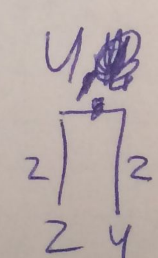
$D_{25} = 8 - 7 - 9 = -8$

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

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

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

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



$r_{u_1} = (2+3+6)/2 = 5.5$

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

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

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

	u_1	1	3	5
u_1		2	3	6
1			3	6
3				5
5				

$D_{u_1 1} = 2 - 5.5 - 5.5 = -9$

$D_{u_1 3} = 3 - 5.5 - 5.5 = -8$

$D_{u_1 5} = 6 - 5.5 - 8.5 = -8$

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

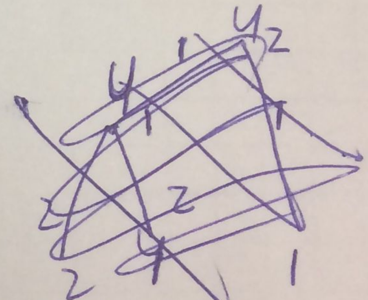
$D_{15} = 6 - 5.5 - 8.5 = -8$

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

$\therefore D_{u_1 1}$ & D_{35} are min $\therefore$ choosing $D_{3,5}$ (u_2 being node)

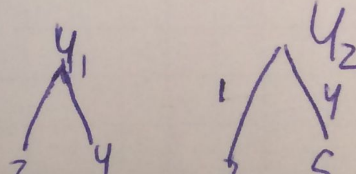
$\cancel{d_{3u_2} = \frac{2}{2} + \left(\frac{5.5 - 5.5}{2}\right) = 1}$

$\cancel{d_{1u_2} = \frac{2}{2} + \left(\frac{8.5 - 5.5}{2}\right) = 1}$



$d_{3u_2} = \frac{5}{2} + \left(\frac{5.5 - 8.5}{2}\right) = 1$

$d_{5u_2} = \frac{5}{2} + \left(\frac{8.5 - 5.5}{2}\right) = 4$



$r_{u_1} = (2+2)/1 = 4$

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

$r_{u_2} = (2+2)/1 = 4$

	u_1	1	u_2
u_1		2	2
1			2
u_2			

$D_{u_1 1} = 2-4-4 = -6$

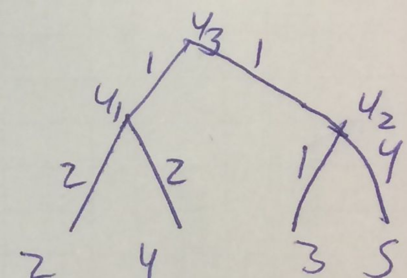
$D_{u_1 u_2} = 2-4-4 = -6$

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

all are same so choosing $u_1 u_2$ (u_3 being node)

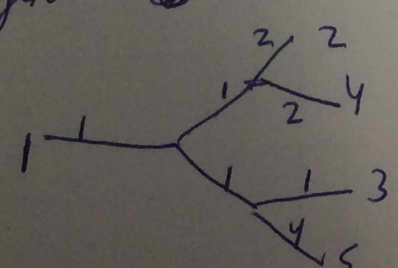
$d_{u_1 u_3} = \frac{2}{2} + \left(\frac{4-4}{2}\right) = 1$

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



	u_3	1
u_3		1
1		

∴ The final rooted tree is

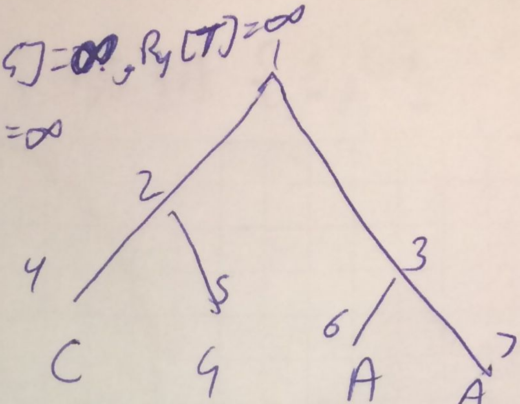


(c) neighbour joining algorithm is better for this data because the data doesnot have ~~all~~ ultrometric property for all node distances, since this is assumption for UPGMA, it will produce wrong tree. Thus neighborn joining algorithm is better

A2) (a) $R_4[A] = \infty$, $R_4[C] = 0$, $R_4[G] = \infty$, $R_4[T] = \infty$

$R_5[A] = \infty$, $R_5[C] = \infty$, $R_5[G] = 0$, $R_5[T] = \infty$

$R_6[A] = 0$, $R_7[A] = 0$



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

$R_3[A] = 0$

$\therefore$ A will come at node 3, 6, 7

Now for node 2, either C or G will come.

for node 1, change from A to G is less costly than change from A to C

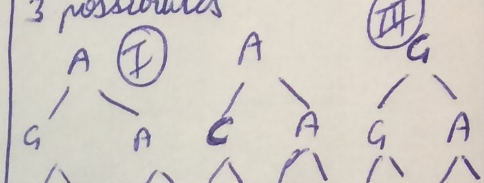
$\therefore$ we will need G in node 2.

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

$R_3[A] = 0$

$R_1[A] = 0 + 1 = 1$

3 possibilities



For middle one we will have larger cost so I or III are possible options

$\therefore$ min cost of tree is $1 + 2 + 0 + 0 + 0 + 0 + 0$
$R_1 \; R_2 \; R_3 \; R_4 \; R_5 \; R_6 \; R_7$

$\therefore$ Minimum cost is 3.

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

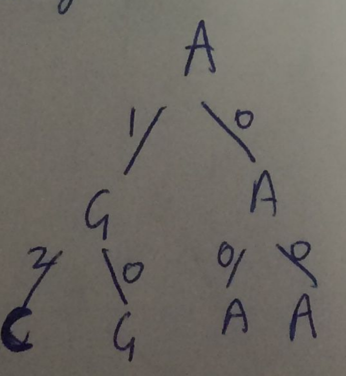
then cost of $R_3[A] = 0$

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

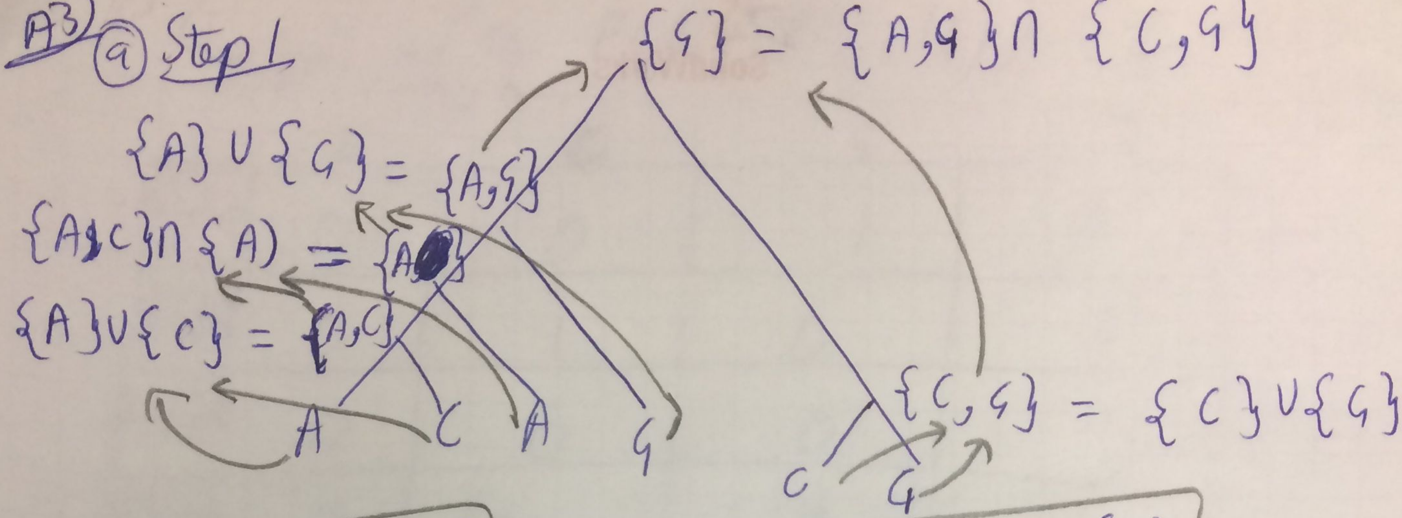
$R_1[G] = 0 + 1 = 1$

$\therefore$ The I & II are 2 possible explainations

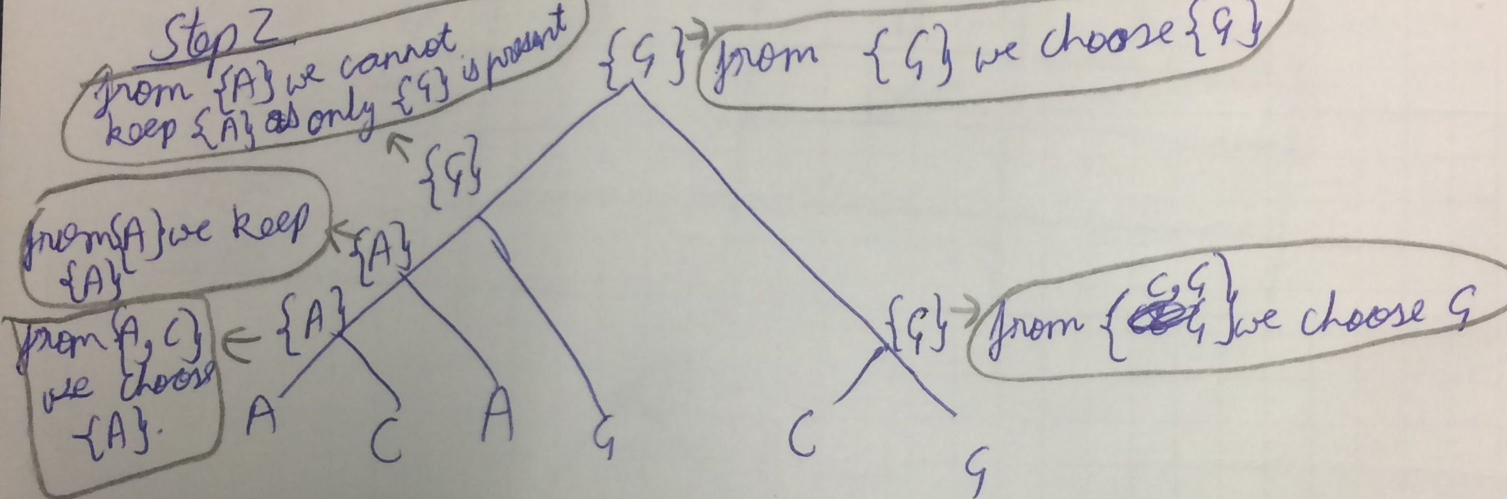
(b) The assignments will be



A3) (a) Step 1



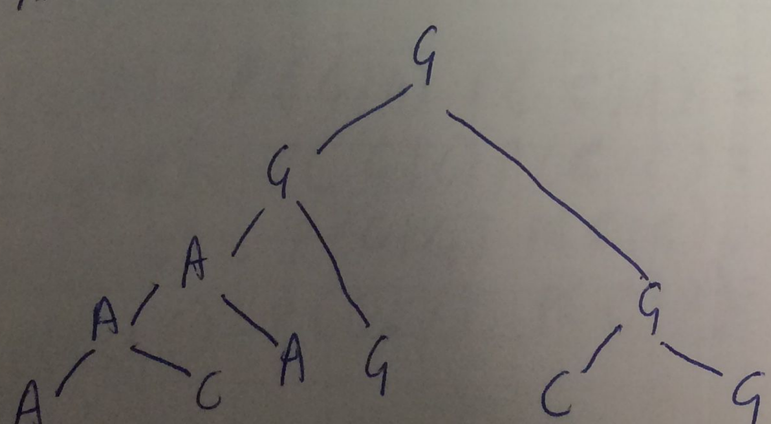
Step 2



∴ Minimum number of changes found = 3

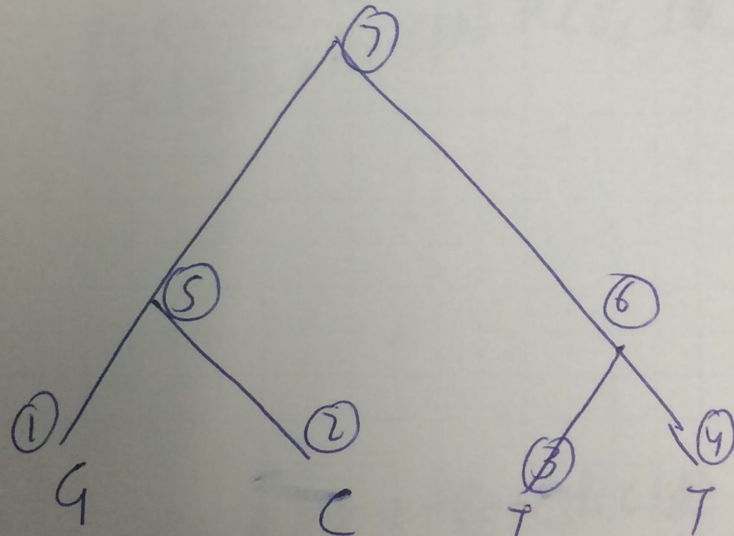
$$\begin{pmatrix} G \to C \\ G \to A \\ A \to C \end{pmatrix}$$

(b) Ancestral bases are



A4) $P(L_k|a) = \left(\sum_b P(b|a)P(L_i|b)\right)\left(\sum_c P(c|a)P(L_j|c)\right)$

	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
$P(L_5|x)$				
$P(L_6|x)$				
$P(L_7|x)$				



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

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

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

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

$P(L_6|A) = P(T|A)\,P(T|A) = 0.0025$

$P(L_6|C) = (P(T|C))^2 = 0.01$

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

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

$P(L_7|A) = P(L_5|A)P(A|A)\ P(L_6|A)P(A|A) +$
$P(L_5|C)\,P(C|A)\ P(L_6|C)P(C|A) +$
$P(L_5|G)\,P(G|A)\ P(L_6|G)P(G|A) +$
$P(L_5|T)\,P(T|A)\ P(L_6|T)\,P(T|A) +$
$P(L_5|A)P(A|A)\ P(L_6|C)P(C|A) +$
$P(L_5|A)P(A|A)\ P(L_6|G)P(G|A) +$
$P(L_5|A)P(A|A)P(L_6|T)\ P(T|A) +$
$P(L_5|A)\ P(A|A)\ P(L_6|A)\,P(A|A) +$
$P(L_5|C)\,P(C|A)$

⋮

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

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

$$P(L_7|A) = [0.004 + 0.002 + 0.004 + 0.00025]$$
$$[\cancel{0.0005}\ 0.002 + 0.0005 + 0.00025 + 0.0032]$$
$$= 0.01025 \times 0.00595$$
$$= 0.0000609875$$

$$P(L_7|C) = [P(L_5|A)P(A|C) + P(L_5|C)P(C|C) + P(L_5|G)P(G|C) + P(L_5|T)P(T|C)]$$
$$[P(L_6|A)P(A|C) + P(L_6|C)P(C|C) + P(L_6|G)P(G|C) + P(L_6|T)P(T|C)]$$

$$P(L_7|G) = [P(L_5|A)P(A|G) + P(L_5|C)P(C|G) + P(L_5|G)P(G|G) + P(L_5|T)P(T|G)]$$
$$[P(L_6|A)P(A|G) + P(L_6|C)P(C|G) + P(L_6|G)P(G|G) + P(L_6|T)P(T|G)]$$

$$P(L_7|T) = [P(L_5|A)P(A|T) + P(L_5|C)P(C|T) + P(L_5|G)P(G|T) + P(L_5|T)P(T|T)]$$
$$[P(L_6|A)P(A|T) + P(L_6|C)P(C|T) + P(L_6|G)P(G|T) + P(L_6|T)P(T|T)]$$

Final probability of given sequence $= [P(L_7|A) + P(L_7|G) + P(L_7|C) + P(L_7|T)]/4$