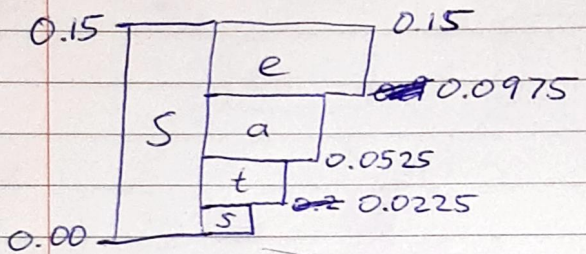
Maxwell Maia CT102 2021
21236277 Assignment 2

Q1 order by smallest frequency.

s	t	a	e
0.15	0.20	0.30	0.35

Intervals: [0.00; 0.15) [0.15, 0.35) [0.35, 0.65) [0.65, 1.00)

Encode: "set"



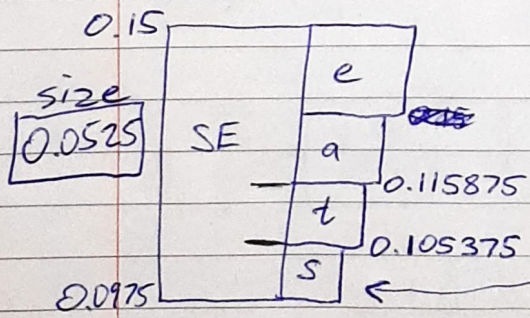
① size of $s_2 = 0.15 \times 0.15$
$= 0.0225$
So $[0, 0+0.0225)$

② size of $t_2 = 0.2 \times 0.15$
$= 0.03$

③ size of $a_2 = 0.3 \times 0.15$
$= 0.045$

size of block (interval) = (current letter) × (parent block)

Therefore interval "SE" is [0.0975, 0.15)



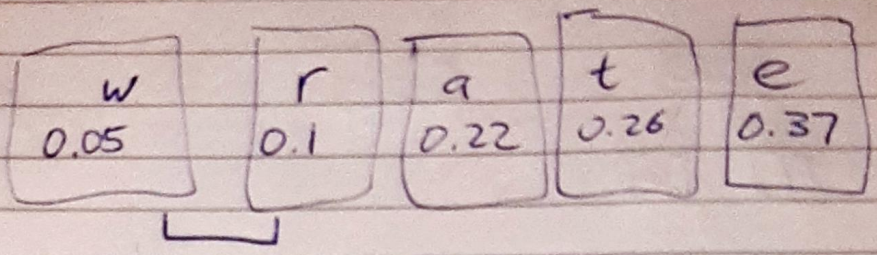
→ Find new intervals
$s = 0.007875$

$t = 0.0525 \times 0.2$
$= 0.0105$

~~$a_2 = 0.0525 \times 0.3$~~

"set" is represented by
[0.105375, 0.115875)

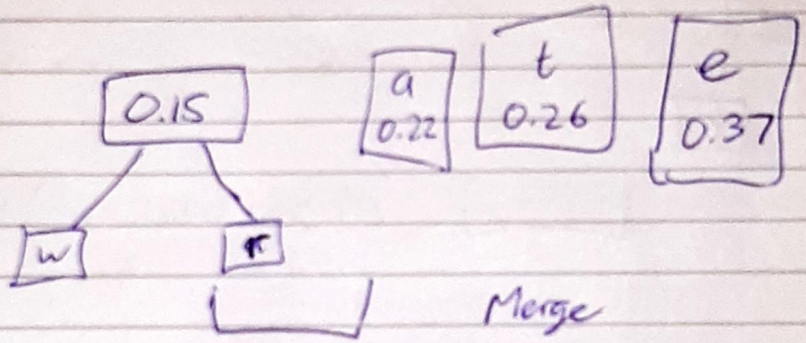
Q2



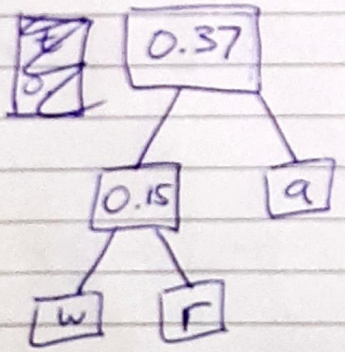
① Create nodes and put in correct order 1. freq 2. node size

② merge 2 smallest

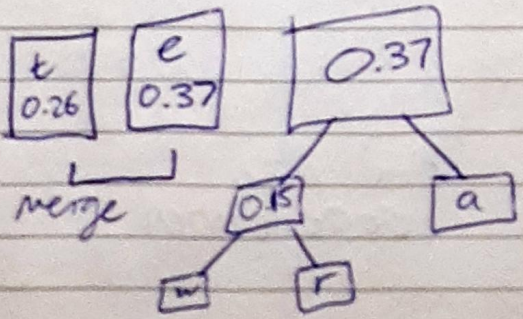
new queue

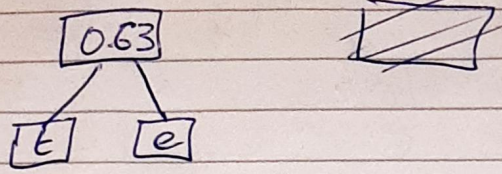


new queue

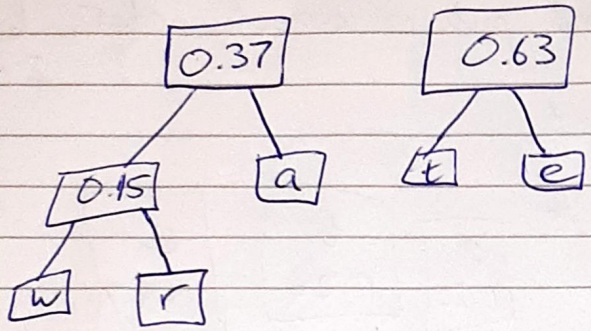


new queue



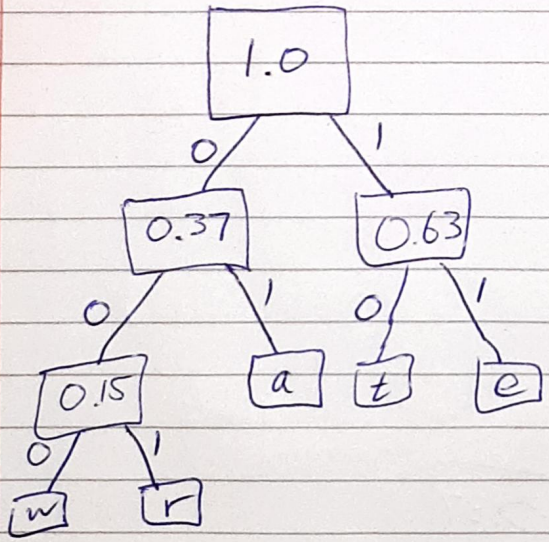


new queue



Merge

new



"water"

w: 000
a: 01
t: 10
e: 11
r: 001

Final code:
00001101 1001

only include items that are common

Q3. a)

$$w_{a,u} = \frac{\sum_{i=1}^{m} (r_{a,i} - \bar{r}_a) \times (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^{m} (r_{u,i} - \bar{r})^2} \times \sqrt{\sum_{i=1}^{m} (r_{a,i} - \bar{r}_a)^2}}$$

Hugo's average $= \frac{5+4+1}{3} = \frac{10}{3} = \bar{r}_a$

Luke's average $= \frac{2+5+3+4}{4} = \frac{7}{2} = \bar{r}_u$

$$corr(Hugo, Luke) = \frac{(5-\frac{10}{3}) \times (5-\frac{7}{2}) + (4-\frac{10}{3}) \times (4-\frac{7}{2})}{\sqrt{(5-\frac{10}{3})^2 + (4-\frac{10}{3})^2} \times \sqrt{(5-\frac{7}{2})^2 + (4-\frac{7}{2})^2}}$$

$corr(Hugo, Luke) =$ ~~1.0298830306~~ $\frac{\frac{17}{6}}{\frac{\sqrt{290}}{6}}$

$corr(Hugo, Luke) = 0.998$

b). A high positive correlation shows that 2 users like similar content. A prediction can be made for a user's rating of an item based off of the other user's rating.

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^{n} (r_{u,i} - \bar{r}) \times w_{a,u}}{\sum_{u=1}^{n} w_{a,u}}$$

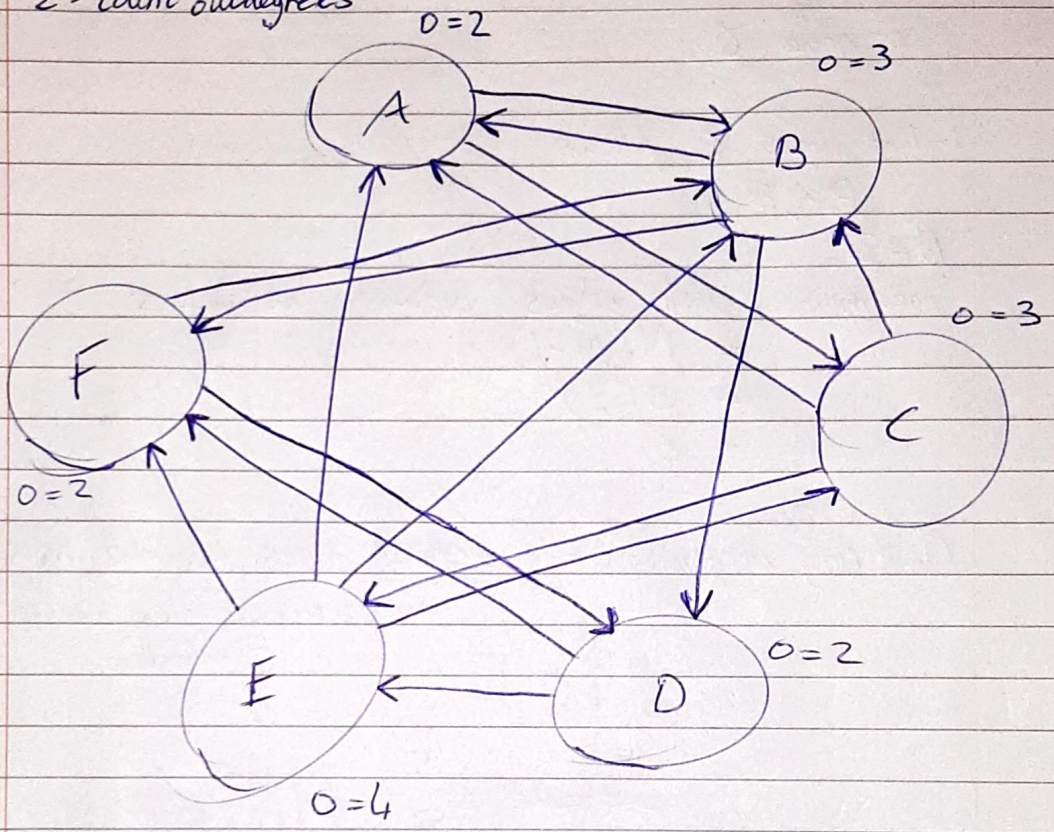
correlation

item rating minus average rating.

Q4.a).

1. draw network

2- count outdegrees



average outdegree = the sum of ~~n~~ node's outdegrees, divided by no. of nodes

$$\text{average outdegree} = \frac{2+3+3+2+4+2}{6} = \frac{8}{3}$$

$$\text{Edge density} = \frac{\text{actual number of edges}}{\text{no. of possible edges}} = \frac{16}{N(N-1)}$$

(no self ties)

$$= \frac{16}{30} = 0{,}53 = 53{,}33\%$$

b). Local clustering co-efficient of node C.

1. Find nodes C follows.

B, A, E

$N = 3$

maximum possible edges between neighbours

$= N(N-1)$

$= 3(2)$

$= 6$

Clustering coefficient = $\frac{\text{actual edges between friends}}{\text{possible edges between friends}}$

$= \frac{4}{6}$

$= \frac{2}{3} = 0.67$

c) Find all paths ~~bet~~ from node A to node F. choose the ~~a~~ one that passes the least edges.

Shortest path = 2

(from A to B, then B to F)

Q5. I am aware of what plagiarism is and include this here to confirm that this work is my own. ✓