Final Term IR del 14/12/21

Q1

T1 = "White Xmas" T2 = "Xmas Xmas happy"

T3 = "happy white" T4 = "red red"

		γ-code	idf n = 4
happy →	2, 3	010011	$\log_2 4/2 = 1$
red →	4	00100	$\log_2 4/1 = 2$
white →	1, 3	1011	$\log_2 4/2 = 1$
xmas →	1, 2	1010	$\log_2 4/2 = 1$

	T1	T2	T3	T4
happy	0	1 · 1	1 · 1	0
red	0	0	0	2 · 2
white	1 · 1	0	1 · 1	0
xmas	1 · 1	2 · 1	0	0

Q = "red xmas" ⟶ [0, 1 · 2, 0, 1 · 1]

$\tilde{\cos}(Q, T1) = 1$

$\tilde{\cos}(Q, T2) = 2$

$\tilde{\cos}(Q, T3) = 0$

$\tilde{\cos}(Q, T4) = 8$ ⟵ most similar

Q2

	UB
$t_1 \rightarrow$ 1, $\boxed{5}$, 6, 7, 8, 11	1
$t_4 \rightarrow$ 2, 3, $\boxed{5}$, 6, 7, 8, 11	2
$t_2 \rightarrow$ 4, 13, 15	0.2
$t_3 \rightarrow$ $\boxed{5}$, 6, 8, 9	0.6

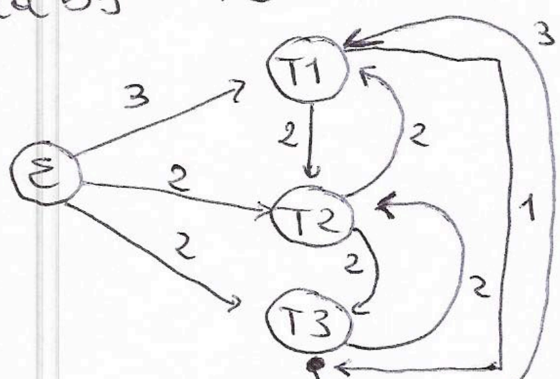
pivot

since we need to sum all UBs to be $> \theta = 3.7$

- The we move to 5 in all lists and find it only in t_1 and t_4 but not in t_2.
- After move, the re-check of the sum of UBs shows that it is $3.6 < \theta = 3.7 \rightarrow$ so 5 is not evaluated.

Q3 (part 1)

T1 = aabb T2 = aaaa T3 = bbbb



gzip (T1) = <0, a> <1, 1, b> <1, 1, EOF>

gzip (T2) = <0, a> <1, 3, EOF>

gzip (T3) = <0, b> <1, 3, EOF>

gzip (T1 | T2) = <1, 2, b> <1, 1, EOF>

gzip (T1 | T3) = <0, 0, a> <1, 1, b> <1, 1, EOF>

gzip (T2 | T1) = <4, 2, a> <1, 1, EOF>

gzip (T2 | T3) = gzip (T2)

gzip (T3 | T1) = <2, 4, EOF>

gzip (T3 | T2) = gzip (T3)

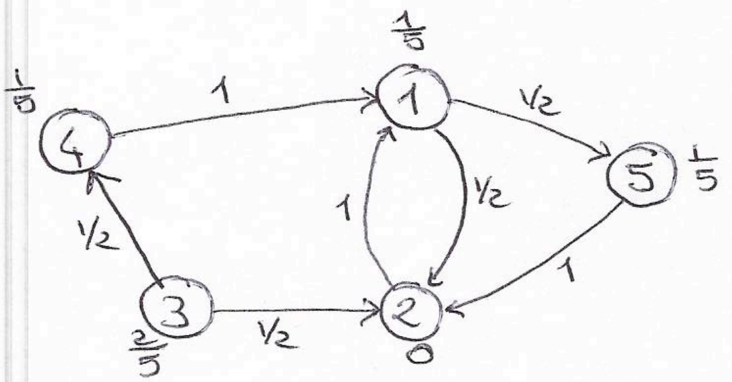
- A possible solution optimal of cost = **5** is

p.e.

$\varepsilon \xrightarrow{2} T2 \xrightarrow{2} T1 \xrightarrow{1} T3$

$\varepsilon \rightarrow aaaa \rightarrow aabb \rightarrow bbbb$

Q4



Personalized PR with $S = \{5\}$

$$pr(1) = \frac{1}{2}\left(\frac{1}{5}\cdot 1 + 0 \cdot 1\right) = \frac{1}{10}$$

$$pr(2) = \frac{1}{2}\left(\frac{1}{5}\cdot\frac{1}{2} + \frac{2}{5}\cdot\frac{1}{2} + \frac{1}{5}\cdot 1\right) = \frac{5}{20}$$

$$pr(3) = \frac{1}{2}(0) = 0$$

$$pr(4) = \frac{1}{2}\left(\frac{1}{2}\cdot\frac{2}{5}\right) = \frac{1}{10}$$

$$pr(5) = \frac{1}{2}\left(\frac{1}{5}\cdot\frac{1}{2}\right) + \frac{1}{2}\cdot 1 = \frac{1}{20} + \frac{1}{2}$$

The random walk over this graph will not converge to a single state independent of the starting probability because the graph is <u>not</u> irriducible, since it does not consists of a single SCC. Look e.g. at node 3 which is not part of any cycle.

Q3 (part 2)

$$f_{new} = \underbrace{AAB}_{h_1},\underbrace{AAB}_{h_1},\underbrace{AAA}_{h_2},\underbrace{BBB}_{h_3} \xrightarrow{h_1 h_1 h_2 h_3} \qquad f_{old} = \underbrace{\overbrace{AAA}^{h_1}}_{h_2},B$$

1110

string gzip-compressed

$$gzip(AABAABAAA \mid BBB) = \langle 4,1,B\rangle \langle 1,1,EOF\rangle$$