Midterm 16. 11. 2020

Q1

14 → 3, 10, 11, 13, 14, 17, 19, 21, 25

15 → 5, 10, 11, 13, 14

16 → 3, 10, 11, 13, 21, 25, 30

We encode 15's list and 16's list w.r.t 14's list because this list allows to copy the most for both.

node	outd	ref	Copy list	extra nodes
14	9	0	–	~~ 14's list
15	5	1	011110000	5
16	7	2	111100011	30

From copy-lists to copy-blocks, shown only for 15 and 16

node	outd	ref	# blocks	Copy blocks	extra nodes
15	5	1	3	0, 0, 3, (3)	5
16	7	2	3	1, 3, 2, (1)	30

there could be dropped and then derived from outd(14).

In fact

3 (+1) = outd(14) - (0+1) - (3+1)

= 9 - 1 - 4 = 4 ✓

Q2.

A = 01100
B = 00000
C = 11111
D = 00010
E = 10010

We apply the clustering approach based on LSH, by setting $k = 2$, $L = 2$, and projections $I_1 = \{1,2\}$ and $I_2 = \{3,4\}$

Given the first projection I_1:

A = 01
B = 00
C = 11
D = 00
E = 10

linking ⟹ B — D

After the second projection I_2

A = 10
B = 00
C = 11
D = 01
E = 01

linking ⟹ D - E

We form only the cluster $\{B, D, E\}$ while A and C remain singletons

Q3.

$b → bud, budy
ud → bud, budy
bu → bud, budy
dy → budy, Turdy
$T → Turdy, Tus
Tu → Turdy, Tus
uz → Tuzdy
zd → Tuzdy
us → Tus

Q = $Tusdy
↓
$T, Tu, us, sd, dy
(sd: x)

budy ⇒ 1
Turdy ⇒ 3
Tus ⇒ 3

$L - k \cdot e = 5 - 2 \cdot 1 = 3$

Result: {Turdy, Tus}

Q4.

F_old = "cane␣ratto␣matt" (blocks: h_1, h_2, h_3, h_4, h_3) B = 3

↓ $\{h_1, h_2, h_3, h_4\}$

F_new = "cane␣gatto␣pane" (blocks: h_1, h_3)

↓

h_1, e, ␣, g, h_3, o, ␣, p, a, n, e

If, instead, we apply zsync.

F_new = "cane␣gatto␣pane" (blocks: h_1, h_2, h_3, h_4, h_5)

↓ h_1, h_2, h_3, h_4, h_5

F_old = "cane␣ratto␣matt" (blocks: h_1 / h_5, h_3, h_3)

↓ 10101

server compresses : canattane | e␣go␣p via LZ77

<1,1,␣> <∅,∅,g> <∅,∅,o> <3,1,p>

Q5

bababc

<0,0,b> <0,0,a> <2,3,c>