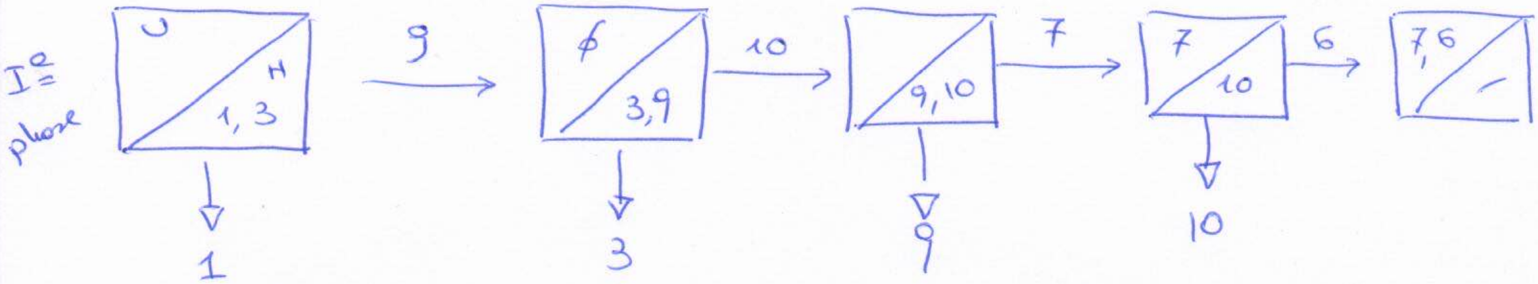
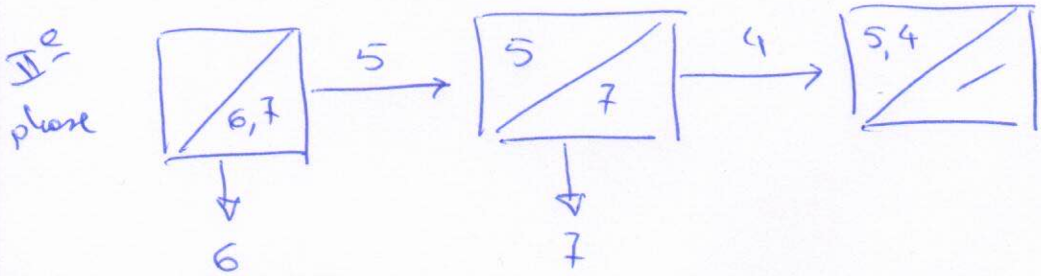
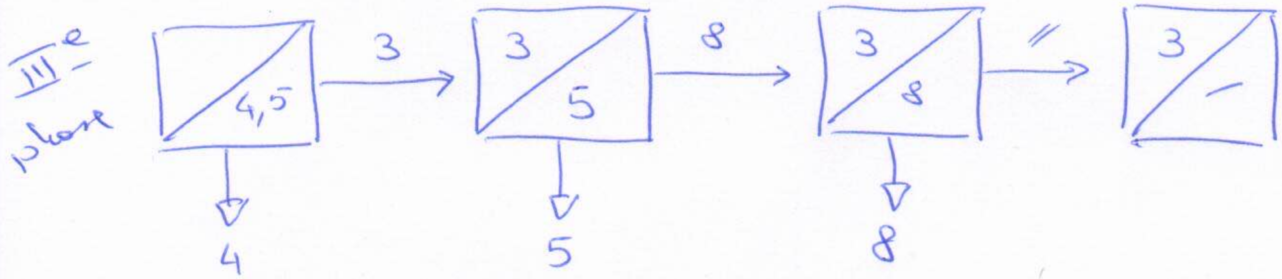
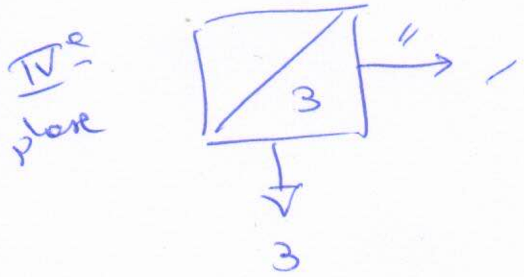
MidTerm 11. 11. 2020

Q #1. We have $M = 2$ and $S = (1, 3, 9, 10, 7, 6, 5, 4, 3, 8)$









we created 4 runs

1, 3, 9, 10

6, 7

4, 5, 8

3

- A sequence of length 10 generating 5 runs is a sequence of decreasing numbers.

- The number of items that go to U is $\frac{1}{4} \cdot \tau$ (on average), according to the proof in the notes. Since $|U| = M$ at the end of a phase, we set $\frac{1}{4}\tau = M \Leftrightarrow \tau = 4M$ is the average length of a run, since we indicated with τ the #items read and this equals the #items going to M.

Q2 According to the proof of theo 8.7 of the notes, here adapted to work on a co-domain m which is not a power-two of the number of keys, we have that the Average # collisions is: $\binom{n}{2} \cdot \frac{1}{m} = \frac{n(n-1)}{2m} = \frac{10 \cdot 9}{2 \cdot 23} = \frac{45}{23}$

Q3. The optimal number of hash functions for the Bloom Filter of size m and n keys is $K = \frac{m}{n} \ln 2$.
Substituting this value in the formula describing the probability to have a cell equal to zero: $e^{-\frac{kn}{m}} = \frac{1}{2}$.

Q4. $s = 3,\ c = 1 \longrightarrow b = 2$ bits $\longrightarrow S = \{00, 01, 10\}$ $C = \{11\}$

the first 8 codewords are:

$$\underbrace{00,\ 01,\ 10}_{S},\ \underbrace{1100,\ 1101,\ 1110}_{C \cdot S},\ \underbrace{111100,\ 111101}_{C \cdot C \cdot S} \ldots$$

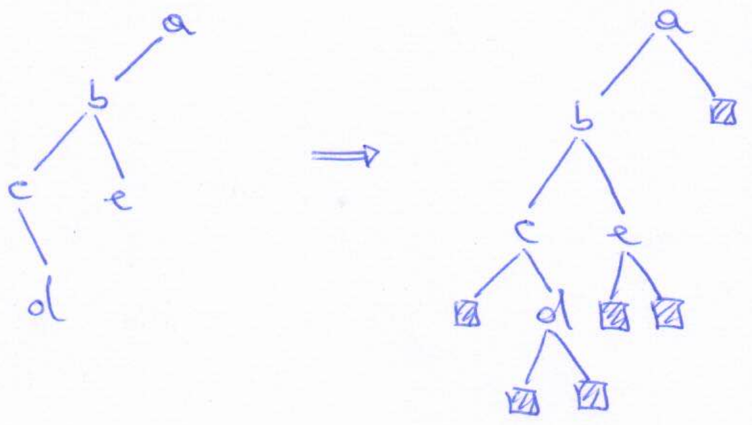
Q5. Access (8) given the Elias-Fano encoding

$$L = 01\,11\,10\,00\,10\,100\,11\,\overset{8^{th}}{\boxed{11}}\,00\,11\,00$$

$$H = 110\ 110\ 10\ 0\ 1010\ 10\ 110\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0$$

Since the original binary encoding is by 6 bits, and the number of 1s in H is 11, then EF encodes $n = 11$ integers.
So that $h = \lceil \log_2 11 \rceil = 4$ bits, and $l = 6 - 4 = 2$ bits.
So the 8^{th} pair of bits in L is $\boxed{11}$
For the high part we perform $Select_1(8) - 8 = 14 - 8 = 6 = (0110)_2$
Access $(8) = (011011)_2 = 16 + 8 + 2 + 1 = 27$

Q6.



encoding = 1 10 11 0 100 00

Q7.

$S = \{aa, ba, bb, bc, ca\}$

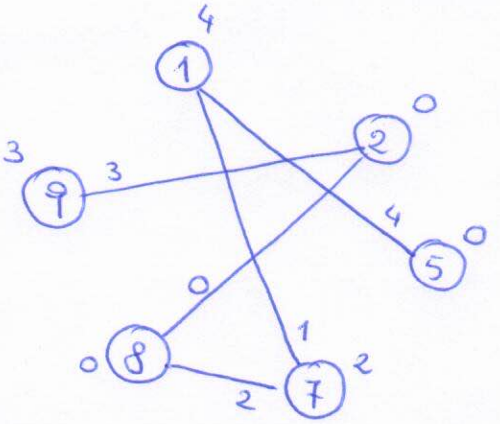
$h_1(c'c'') = 2 \cdot r' \cdot r'' \mod 11$

$h_2(c'c'') = 5 \cdot r' + r'' + 1 \mod 11$

	a	b	c
r	2	3	4

	h	h_1	h_2
aa	0	8	9
ba	1	1	7
bb	2	7	8
bc	3	2	9
ca	4	5	1

the graph is acyclic so the solution to MOPH does exist.



we draw only the nodes with incident edges. Remind that sum here is mod 5

g									
	4	0	0	0	0	0	2	0	3
	1	2	3	4	5	6	7	8	9

we set to zero all the other cells.