STATISTICAL STRUCTURED PREDICTION
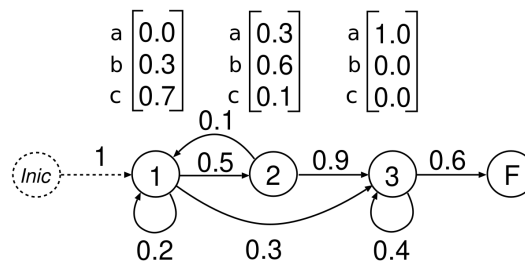
Question set (Part 2)

January 2024

Administrative issues

- This part represents 30% of the final score of PEE.

- The score of this part consists of 60% of theoretical questions and 40% of practical assignments.

- The score of each exercise will depend on the depth, quality and justification of the answer. both in the theoretical exercises and practical exercises.

- Please present the results in PDF documents and uplodad your document through a poliformat task.

- Deadline for delivering the theoretical questions and the report of practical assignment: **5 Feb 2024**

Theoretical questions

Question 1 (4 points)

Given the following HMM:

a) (1 point) Apply the *forward* algorithm to the string $cccaa$
b) (1 point) Apply the *backward* algorithm to the string $cccaa$.
c) (2 points) Define a string that contains each one of the symbols of the alfhabet at least once, of length larger than 4 with al least two paths. Then apply expression (3) and (4) of slide 18 to the previous HMM using as training sample string that you have defined and with the string in exercises a) and b). Plot the final HMM.

Question 2 (2 points)

Repeat exercice 2.c using the Viterbi algorithm

Practical questions

Question 3 (4 points)

Choose only one of the following exercices:

a) Perform a cross-validation experiment with `teds_bib.bib` bib file used in the practice sessions. Shuffle first the samples as `shuf teds_bib.bib`. Choose the adquate evaluation method and provide a short (less than 100 words) discussion on the results.

b) Look for a large public bib file in internet (more than 1000 bib entries) and replicate steps 1 to 6 in slide number 31. Provide a short (less than 100 words) discussion on the results.

c) Perfom last question on slide 77 and compare the results with the results reported in slide 76. Provide a short (less than 100 words) discussion on the results.