

# Information Retrieval

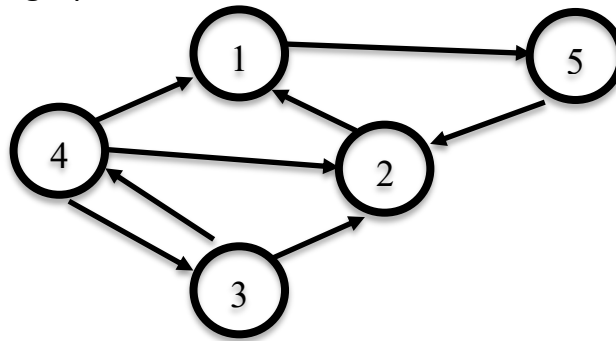
15 June 2020 – time 30 minutes

**Question #1.** Assume that you are given a set of strings  $S = \{\text{bar, bat, bes, bet, bit}\}$ .

- Build the data structure that efficiently searches for an arbitrary pattern  $P$  with 1 edit distance.
- Show how it is executed the 1-edit search for  $P = \text{“bas”}$

**Question #2.** Show how to synchronize via rsync the new file “bacaddabb” (on the server) using the old file “acabbbdabac” (on the client) and blocks of size 3 chars.

**Question #3.** Given the graph:



- Compute one step of PageRank by assuming a uniform starting distribution, and setting  $\alpha=0.5$
- Compute one step of **Personalised** PageRank with respect to node 3 and assuming a uniform starting distribution, and setting  $\alpha=0.5$

**Question #4.** Let us assume that you are given a set of records  $R_1, R_2, \dots, R_n$  each one formed by  $F$  fields. Sketch the design of an algorithm that finds the most similar records in terms of the number of fields which contain the same values.