**Information Retrieval**

**5 June 2015**

**Ex 1 [ranks 5]** Let us given a graph G of directed edges {(1,2), (2,1), (1,4), (3,1)}. Simulate the execution of **two steps** of PageRank, starting with nodes with PR=1 (unnormalized), and assuming a teleportation step which favors node 3.

**Ex 2 [points 4+5]** Given the sequence S = (1, 2, 1, 1, 4, 7, 7, 4, 1, 3, 9), show:

* the PForDelta encoding with b=2 and base=0;
* the Elias-Fano encoding (*warning*: remind that EF encodes monotonic sequences).

**Ex 3 [points 4+4]**

* Describe the implementation of the Rank data structure which adds o(m) bits to a binary array B of length m, and answers any Rank-query in constant time.
* Then, show the data structure on B = [0011 0101 1110 1010 0000 1111 1010 0000] by setting z=4 and Z=8.

**Ex 4 [points 4]** Describe how the cosine similarity between a LSI-projected document and a LSI-projected query is computed in the reduced space of k dimensions.

**Ex 5 [points 4]** Given the following matrix of pair-wise similarities between 5 items

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| I1 | 1,00 |  |  |  |  |
| I2 | 0,80 | 1,00 |  |  |  |
| I3 | 0,20 | 0,70 | 1,00 |  |  |
| I4 | 0,70 | 0,60 | 0,40 | 1,00 |  |
| I5 | 0,50 | 0,50 | 0,30 | 0,80 | 1,00 |

Show the next cluster formed by the agglomerative clustering algorithm based on AVG similarity-function, given that we have already formed the following clusters {(I1,I2), I3, (I4,I5)}