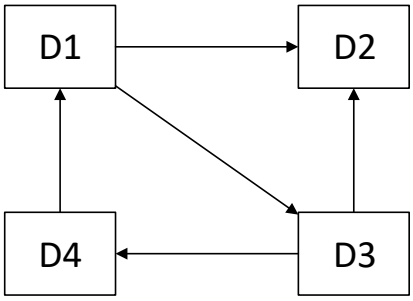
LAB 8: Exercises

1. **HITS:** Given is the network shown in the image below. Find hubs and authorities vectors for this network. Complete the matrix L and L^T for this network and calculate matrix LL^T. Use online eigenvector calculator to find vectors h and a.



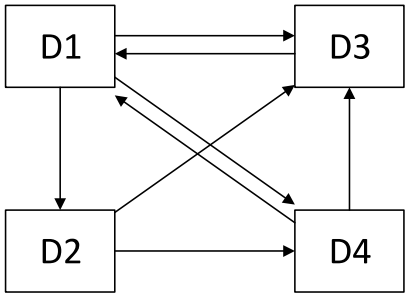
$$L = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad L^T = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$LL^T = \begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad L^TL = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

h = [1 , 0 , 1 , 0], a = [0 , 2 , 1 , 1]
h_{norm} = [0.5 , 0 , 0.5, 0], a_{norm} = [0 , 0.5 , 0.25 , 0.25]

The best hub is page: D1 and D3, the best authority is page: D2…….

2. **PageRank:** Given is the network shown in the picture below. Find stochastic matrix M, write and solve the equation system for finding PageRank values for this network (use basic PageRank model – without a damping factor).



$$M = \begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$

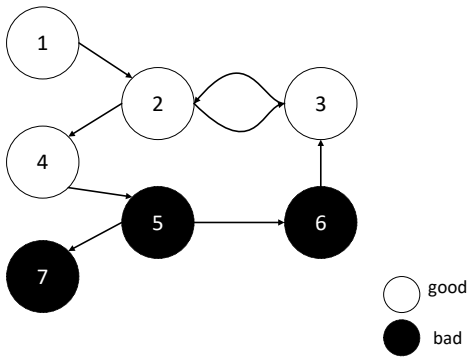
Equation system:

PR1 = PR3 + 1/2*PR4
PR2 = 1/3*PR1
PR3 = 1/3*PR1 + 1/2*PR2 + 1/2*PR4
PR4 = 1/3*PR1 + 1/2*PR2
PR1 + PR2 + PR3 + PR4 = 1

PR2 = 1/3PR1
PR4 = 1/3PR1 + 1/2PR2 = 1/3PR1 + 1/6PR1 = 1/2PR1
PR3 = 1/3PR1 + 1/2PR2 + 1/2PR4 = 1/3PR1 + 1/6PR1 + 1/4PR1 = 3/4PR1
PR1 + 1/3PR1 + 1/2PR1 + 3/4PR1 = 1
31/12PR1 = 1

PR1 = 12/31
PR2 = 4/31
PR3 = 9/31
PR4 = 6/31

3. **TrustRank:** Find initial TrustRank vector d (seed = {2, 4, 5} and write equations for finding TrustRank for pages 2, 3, and 5, q = 0.15.



M =

0	0	0	0	0	0	0
1	0	1	0	0	0	0
0	0.5	0	0	0	1	0
0	0.5	0	0	0	0	0
0	0	0	1	0	0	0
0	0	0	0	0.5	0	0
0	0	0	0	0.5	0	0

d = [….0….., ….1…., ….0…..,….1…..,….0….,….0….,….0…..]

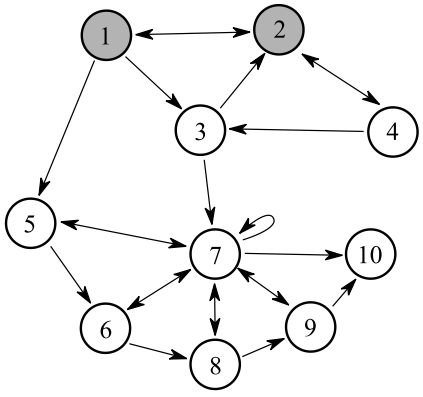
TR(2) = 0.15*1 + 0.85*(TR(1) + TR(3))

TR(3) = 0.15*0 + 0.85*(0.5*TR(2) + TR(6))

TR(5) = 0.15*0 + 0.85*(TR(4))

4. **Programming Assignment (deadline +1 week)**

Given is the following web structure:



Download the pr_tr.py python script from the lab directory. The above structure is kept in L matrix (matrix of indices). Complete the TODOs:

TODO 1. Compute stochastic matrix M (function getM).

TODO 2. Compute pagerank vector and return the results (sorted pairs -> [page id : **pagerank**]). Which pages have the greatest pagerank? Why?

TODO 3. Which pages do you think belong to the link farm? Compute trustrank vector. Pages 1 and 2 are marked as "good". Analyze the results. What has changed?

TODO 4. Repeat TODO3 but remove connections 1->5 and 3->7. Analyze the computed trustrank vector.