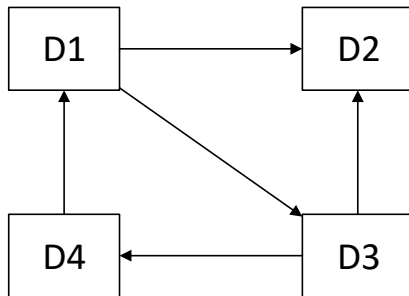


## LAB 6: Exercises

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  $a$ .



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

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

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

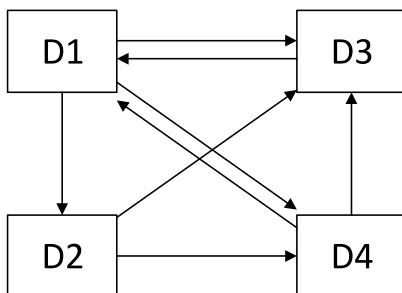
$$L^TL = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

$$h = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix}, \quad a = \begin{bmatrix} 0 & 2 & 1 & 1 \end{bmatrix}$$

$$h_{\text{norm}} = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix}, \quad a_{\text{norm}} = \begin{bmatrix} 0 & 1 & 1/2 & 1/2 \end{bmatrix}$$

The best hub is page: 1 or 3, the best authority is page: 2

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).

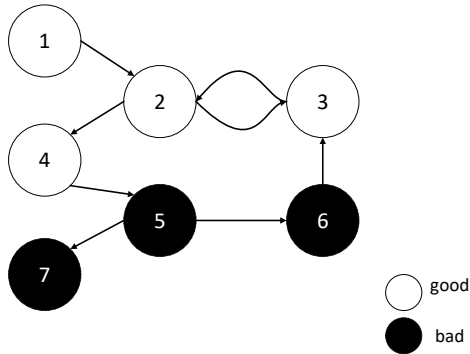


$$M = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1/3 & 0 & 0 & 1 \\ 1/3 & 1/2 & 0 & 0 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$$

**Equation system:**

$$\begin{aligned} v_1 &= v_3 \\ v_2 &= 1/3 \cdot v_1 + v_4 \\ v_3 &= 1/3 \cdot v_1 + 1/2 \cdot v_2 \\ v_4 &= 1/3 \cdot v_1 + 1/2 \cdot v_2 \end{aligned}$$

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



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

$$d = [ \dots, 1/4, \dots, 1/4, \dots, 1/4, \dots, 1/4, \dots, 0, \dots, 0, \dots, 0, \dots ]$$

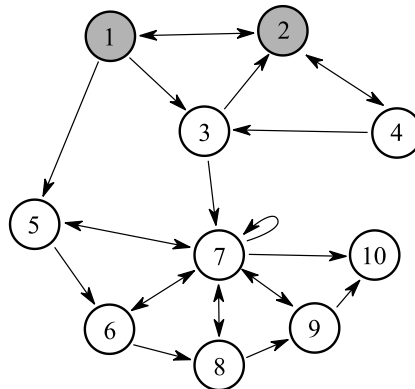
$$TR(2) = 1/4$$

$$TR(3) = 1/4$$

$$TR(5) = 0$$

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  $\rightarrow$  [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 $\rightarrow$ 5 and 3 $\rightarrow$ 7. Analyze the computed trustrank vector.