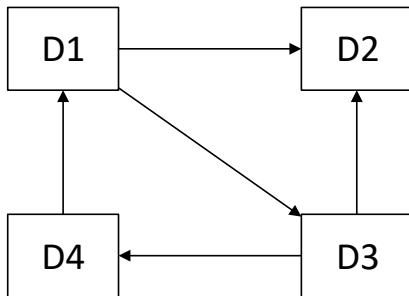


## LAB 6: Exercises

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  $a$ .



$$L = \begin{array}{c|cccc} & 0 & 1 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 \end{array}$$

$$L^T = \begin{array}{c|cccc} & 0 & 0 & 0 & 1 \\ \hline 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 \end{array}$$

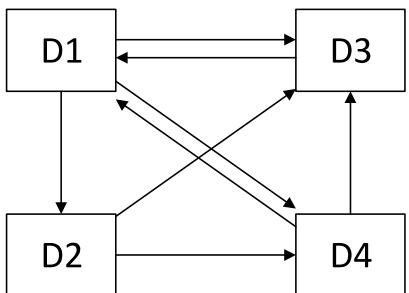
$$LL^T = \begin{array}{c|cccc} & 2 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 2 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \end{array}$$

$$L^T L = \begin{array}{c|cccc} & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 2 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ \hline 0 & 1 & 0 & 1 & 0 \end{array}$$

$$h = [1, 0, 1, 0], \quad a = [0, 2, 1, 1] \\ h_{\text{norm}} = [1, 0, 1, 0], \quad a_{\text{norm}} = [0, 1, 1/2, 1/2]$$

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).

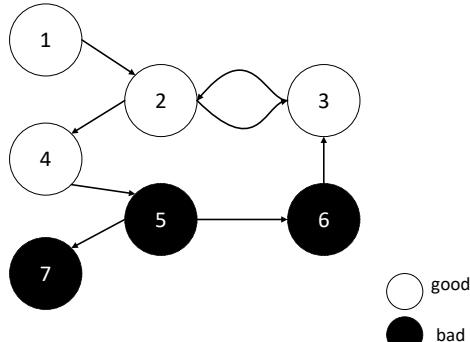


$$M = \begin{array}{c|cccc} & 0 & 0 & 1 & 0 \\ \hline 1/3 & 0 & 0 & 0 & 1 \\ 1/3 & 1/2 & 0 & 0 & 0 \\ \hline 1/3 & 1/2 & 0 & 0 & 0 \end{array}$$

**Equation system:**

$$\begin{aligned} v1 &= v3 \\ v2 &= 1/3*v1 + v4 \\ v3 &= 1/3*v1 + 1/2*v2 \\ v4 &= 1/3*v1 + 1/2*v2 \end{aligned}$$

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



$$M = \begin{array}{c|ccccccc} \hline & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

$$d = [ \dots, \frac{1}{4}, \dots, \frac{1}{4}, \dots, \frac{1}{4}, \dots, 0, \dots, 0, \dots, 0, \dots ]$$

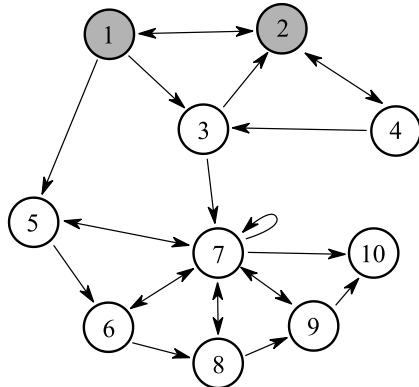
$$\text{TR}(2) = \frac{1}{4}$$

$$\text{TR}(3) = \frac{1}{4}$$

$$\text{TR}(5) = 0$$

#### 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->5 and 3->7. Analyze the computed trustrank vector.