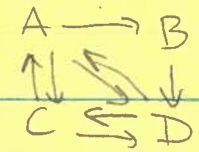
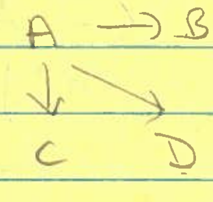
Lab 09



TRANSITION MATRIX: First column of matrix:

$j = A$: LINKS IN A :



#LINKS $= L(A) = 3$

$$\Rightarrow \quad t_{iA} = \begin{bmatrix} 0 \\ 1/L(A) \\ 1/L(A) \\ 1/L(A) \end{bmatrix} = \begin{bmatrix} 0 \\ 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

If you were to click on a random link in page A, you would arrive at page B 1/3 of the time, C 1/3 of the time, and D 1/3 of the time.

The entries of the column are the probabilities that you arrive @ each of the pages when you click on a randomly chosen link in A. Since there are only four possible outcomes (A, B, C, D), the probabilities must add to 1 (provided page A contains at least one link)

Second column of matrix:

$j = B$: LINKS IN B: $B \to D$; $L(B) = 1$

$$\Rightarrow \quad t_{iB} = \begin{array}{c} \\ \\ \\ \\ \end{array} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1/1 \end{bmatrix} \begin{array}{c} A \\ B \\ C \\ D \end{array}$$

Clicking on a link in B takes you to D, all of the time.

Third column of matrix:

$j = C$ LINKS IN C:

$$\begin{tikzcd} A & \\ C \arrow[u] \arrow[r] & D \end{tikzcd}$$

$L(C) = 2$

$$t_{iC} = \begin{bmatrix} 1/2 \\ 0 \\ 0 \\ 1/2 \end{bmatrix}$$

Fourth column of matrix

$j = D$ LINKS IN D:

$$\begin{tikzcd} A & \\ C & D \arrow[l] \arrow[ul] \end{tikzcd}$$

$L(D) = 2$

$$t_{iD} = \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \\ 0 \end{bmatrix}$$

PAGERANK VECTOR:

$$p_0 = \begin{bmatrix} p_A \\ p_B \\ p_C \\ p_D \end{bmatrix}$$ = probabilities of choosing each page at random.

eg. $p_A = p_B = p_C = p_D = 1/4$

$\Rightarrow$ each page is equally likely to be chosen by a surfer as his "starting" page

MARKOV CHAIN

$$p_1 = T p_0$$

$$= \begin{bmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{4}\left(\frac{1}{2}+\frac{1}{2}\right) \\ \frac{1}{4}\left(1/3\right) \\ 1/4\left(\frac{1}{3}+\frac{1}{2}\right) \\ 1/4\left(\frac{1}{3}+1+\frac{1}{2}\right) \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 \\ 1/3 \\ 5/6 \\ 11/6 \end{bmatrix}$$

Sum the elements of p_1:

$$\frac{1}{4}\left[1+\frac{1}{3}+\frac{5}{6}+\frac{11}{6}\right] = \frac{1}{4}\,\frac{6+2+5+11}{6} = \frac{1}{4}\,\frac{24}{6}$$

$$= 1$$

In fact,

$$\|p\|_1 = 1$$

where

$$\|x\|_1 = \ell_1\text{-norm. (taxicab metric)}$$
$$= \sum_{i=1}^{N} |x_i|$$

In general:

$$\|p_k\|_1 = 1.$$

$$p_A^{(1)} = \sum_j t_{Aj}\, p_j^{(0)}$$
$$= t_{AA}\, p_A^{(0)} + t_{AB}\, p_B^{(0)} + t_{AC}\, p_C^{(0)} + t_{AD}\, p_D^{(0)}$$

= probability that surfer arrives at page A after one random click in a random page [if you haven't taken a probability course, then just take my word on this one!]

Thus argmax $\{p_A^{(1)}, p_B^{(1)}, p_C^{(1)}, p_D^{(1)}\}$ is the page the random surfer is most likely to arrive at by chance after one click, and is therefore assumed to be the most relevant.