Lecture 4

[MATH OF PAGE RANK]

0

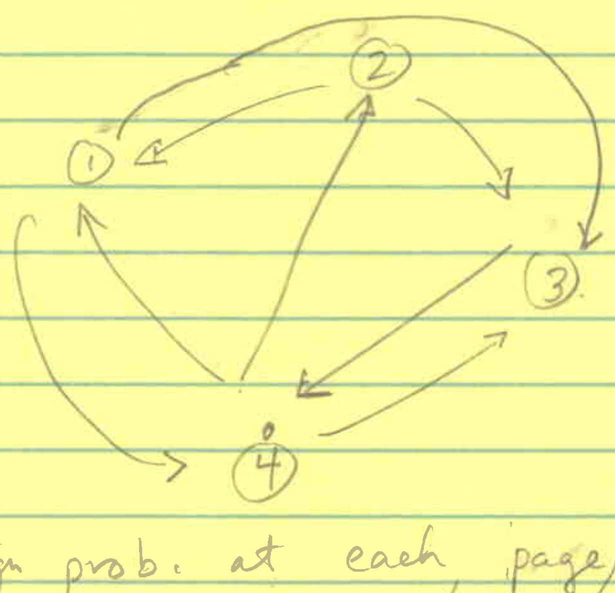
2/12

- How do we determine importance of a webpage? In the old days (BG) text based ranking was used
  BAD: a document containing the phrase "Uni High" 10^{10} times will be most important.

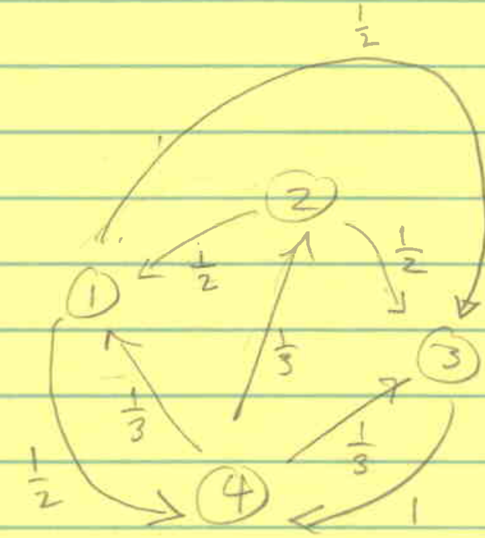
- Better idea: Look at how many pages point to a page.

- Better still: create a dynamical process and iterate it

—— // ——



Assign prob. at each page, simply proportional to number of links out:



Lecture 4

Build a transition matrix

$$A \left\{ \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & 1 & 0 \end{bmatrix} \right. \begin{pmatrix} \vec{V}(n) \end{pmatrix} = \begin{pmatrix} \vec{V}(n+1) \end{pmatrix}$$

"out" vector at ① " " " " "out" vector at ④

So, we want to know the "equilibrium" state $\lim_{n\to\infty} \begin{pmatrix} a_1(n) \\ \vdots \\ a_4(n) \end{pmatrix} = \lim_{k\to\infty} A^k \begin{pmatrix} a_1(0) \\ \vdots \\ a_n(0) \end{pmatrix}$

How do we find $A^{\text{real big}}$? Of course, by diagonalizing!

If we start w/ equal prob at all nodes, $\vec{V}(0) = \frac{1}{4}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, we want to know $\lim_{n\to\infty} A^n V(0) \cong \begin{pmatrix} .2 \\ .1 \\ .3 \\ .4 \end{pmatrix}$

But there are problems!

① The web is not a connected graph.

2. You don't always follow a link on your current page; you can jump to a new page, at random!!

How in the world can we account for a random jump?

Answer: (Page + Brin)
- we decide on a probability for a random jump (typically $p = .15$)
- make the random jump land on all nodes w/ equal liklihood.

$$G = (1-p)\,A + p\left(\frac{1}{n}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & & 1 \end{bmatrix}\right)$$

the Google matrix

i^{th} column gives links from node i

i^{th} col is all $\frac{1}{n}$, says there is a chance to jump to any other node in the web.

Note: G is a column stochastic matrix (for node w/ no out pointers, use $p = 1$, since they must jump).

Thm (Perron-Frobenius thm).

A positive, column stochastic matrix has 1 as an eigenvalue, and the correspondy eigenvector has all + entries

$\Rightarrow$ There is a unique eigen vect for eigenvalue 1, whose entries sum to 1

Thm If v^* is the eigenvector above, then

$V(0) = 1/n \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, $G\,V(0)$, $G^2(V(0))$, ... converges to V^*

Soln: To find the Google rank, just solve

$G\vec{v} = 1\vec{v}$, and scale so entries of $\vec{v}$ sum to 1. ///

The resulting vector is the page rank.

Warning in fact, it is easier to just get a rough approx $G^m \cdot V(0)$ by taking a small power of m.

The End.