Make a giant $n \times n$ matrix A where $n =$ the number of existing webpages (url's)

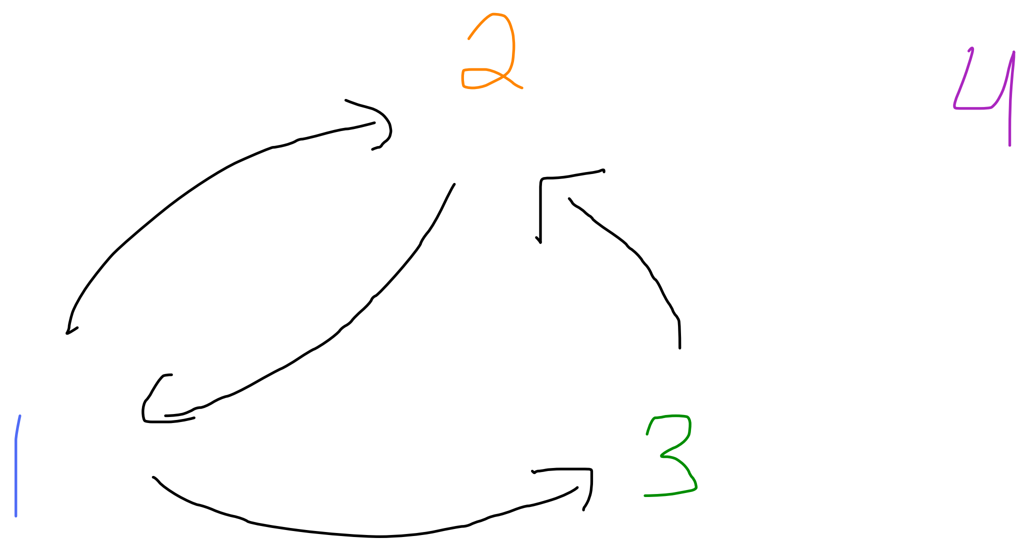
Matrix A

$$A_{i,j} = \begin{cases} 1 & \text{if page } j \text{ links to page } i \\ 0 & \text{if page } j \text{ doesn't link to page } i \end{cases}$$

Note: A is (in general) not symmetric!

Simple Model.

4 webpages, arrow indicates link



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

(4×4)

Page / Brin normalize

A by making all columns add up to one in the following manner:

Let $$B_{i,j} = \frac{A_{i,j}}{\sum_{i=1}^{n} A_{i,j}}$$

Problem What if $\sum_{i=1}^{n} A_{i,j} = 0$?

Can't divide – this means page j doesn't link to anything.

When the webpage doesn't link to anything, the "random surfer" can't click on any new links.

They must pick a new URL to continue the search.

How to deal with this?

Page & Brin fix the problem by changing all the zeroes in that column to ones, creating a matrix A' and recalculating B with A'.

Back to example

Replace A with

just like A

$$A' = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Zeroes changed to ones

Then

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

They then used d to divide the Page Rank of page j evenly among the remaining pages.

(they use $d = .85$)

Finally, they defined a matrix C by

$$C_{i,j} = d B_{i,j} + \frac{1-d}{n}.$$

(n = # of webpages)

Observe that the columns of C still sum to one

Sum of columns

$$\sum_{i=1}^{n} C_{i,j} = \sum_{i=1}^{n}\left(d(B_{i,j}) + \frac{1-d}{n}\right)$$

$$= d\left(\sum_{i=1}^{n} B_{i,j}\right) + 1-d$$

$$= d \cdot 1 + 1 - d = \boxed{1}$$

In matrix terms,

$$C = dB + \left(\frac{1-d}{n}\right)E$$

where E is the $n \times n$ matrix

$$E_{i,j} = 1 \quad \text{for all } 1 \leq i,j \leq n$$

For our example

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

$$F = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$C = .85B + \underbrace{\frac{.15}{4}}_{= .0375} F$$

$$= \begin{bmatrix} .0375 & .8875 & .6375 & 1/4 \\ .4625 & .0375 & .8875 & 1/4 \\ .4625 & .0375 & .0375 & 1/4 \\ .0375 & .0375 & .0375 & 1/4 \end{bmatrix}$$

PageRank: C has $\lambda = 1$ as an eigenvalue and a unique eigenvector of length one with all positive entries

If we denote this vector by

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}, \quad \text{then}$$

the PageRank of webpage i is equal to v_i.

The Page Rank is
the probability of
a "random surfer"
reaching your page.

Problems

How do we know $\lambda = 1$ is an eigenvalue, and how do we get that there is an eigenvector with all positive coordinates?

Perron - Frobenius Theorem

(early 1900's !)

Let P be any matrix with all $\boxed{\text{entries positive}}$.

Then the eigenvalue of largest absolute value of P is positive and admits a

unique eigenvector with all positive entries

Point 1: Just because all entries of P are positive doesn't mean all eigenvalues are. So the 1st part actually has content

Example

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

$$T - \lambda I = \begin{bmatrix} 1-\lambda & 2 \\ 1 & 1-\lambda \end{bmatrix}$$

$$\det(T - \lambda I) = \lambda^2 - 2\lambda - 1$$

roots $\lambda = \dfrac{2 \pm \sqrt{8}}{2} = 1 \pm \sqrt{2}$

$\lambda = 1 - \sqrt{2}$ negative

Observe

$$(1-\sqrt{2})^2 = 3-2\sqrt{2}$$

$$(1+\sqrt{2})^2 = 3+2\sqrt{2}$$

bigger,

so $|1+\sqrt{2}| > |1-\sqrt{2}|$

This fits with the theorem.

$$C = \begin{bmatrix} .0375 & .8875 & .0375 & 1/4 \\ .4625 & .0375 & .8875 & 1/4 \\ .4625 & .0375 & .0375 & 1/4 \\ .0375 & .0375 & .0375 & 1/4 \end{bmatrix}$$

We said this matrix has $\lambda = 1$ as an eigenvalue and a <u>unique</u> associated eigenvector with all positive entries and of length one.

Point 2: Why is the largest eigenvalue always 1 here?

λ an eigenvalue for A iff

λ an eigenvalue for A^t.

(obvious for 2x2 matrices)

This is true for all matrices

Use C^t instead of C.

The columns of C add up to one, so the rows of C^t add up to one

For our

$$C = \begin{bmatrix} .0375 & .8875 & .0375 & 1/4 \\ .4625 & .0375 & .8875 & 1/4 \\ .4625 & .0375 & .0375 & 1/4 \\ .0375 & .0375 & .0375 & 1/4 \end{bmatrix}$$

$$C^t = \begin{bmatrix} .0375 & .4625 & .4625 & .0375 \\ .8875 & .0375 & .0375 & .0375 \\ .0375 & .8875 & .0375 & .0375 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$$

$$\vec{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$ Can check

$$C^t \vec{v} = \vec{v}$$

For our

$$C = \begin{bmatrix} .0375 & .8875 & .0375 & 1/4 \\ .4625 & .0375 & .8875 & 1/4 \\ .4625 & .0375 & .0375 & 1/4 \\ .0375 & .0375 & .0375 & 1/4 \end{bmatrix}$$

$$C^t = \begin{bmatrix} .0375 & .4625 & .4625 & .0375 \\ .8875 & .0375 & .0375 & .0375 \\ .0375 & .8875 & .0375 & .0375 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$$

$$\vec{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$ Can check

$$C^t \vec{v} = \vec{v}$$

C^t can't have an eigenvalue bigger than one since

if $C^t v = \lambda v$, then $\qquad v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$

$$\max_{1 \le j \le n} \{ |\lambda v_j| \} = \max_{1 \le j \le n} \left\{ \left| \sum_{i=1}^{n} c_{i,j} v_i \right| \right\}$$

triangle inequality

$$\left(\le \max_{1 \le j \le n} \sum_{i=1}^{n} |c_{i,j} v_i| \right)$$

$$\le \max_{1 \le j \le n} |v_j|$$

$$= |\lambda| \max_{1 \le j \le n} |v_j| \le \max_{1 \le j \le n} |v_j|$$

So $\lambda \le 1$.

Point 3: How do you know $\vec{v}$ exists and is unique?

Moreover, how do you find it?

That's the hard part...