Google's PageRank Algorithm

General Process

1. Crawl the WWW
2. Index
3. **Global Ranking of Webpages (PageRank)** (This is what we are looking at)
4. Query → Find subset of pages (subset of WWW)

Pages → Nodes
Connections → directed arcs

Create a directed graph

In-Link = Link to a page
Out-Link = Link from a page

An Example

Let x_i be 'importance' of page i

4 pages

Let importance = number of in-links of page i

$x_1 = 2$
$x_2 = 1$
$x_3 = 3$
$x_4 = 2$

However, I can artificially boost my ranking by making a load of dummy pages and making them link to my page

If there is a link from Page i to Page j, the value of this link = x_i/N_i, where N_i = number of out-links from page i, and x_i = importance of page i

$x_1 = x_3 + x_4/2$
$x_2 = x_1/3$
$x_3 = x_1 + x_2/2 + x_4/2$
$x_4 = x_1/3 + x_2/2$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

Solution = Eigen vectors for eigen value = 1

E_1 = span{$\begin{bmatrix} 12 \\ 4 \\ 9 \\ 6 \end{bmatrix}$}

$P_1 > P_3 > P_4 > P_2$

Problems?

1. What if a page has no in-links?
2. What if a page has no out-links?
3. What if $\dim(E_1) > 1$?
4. A is huge. Is it computationally feasible to compute the eigen space E_1?

Dealing with Problem 3

All columns of A add up to 1

A = column-stochastic

1. $a_{ij} \geq 0 \ \forall i, j$
2. $\sum_{i=1}^{n} a_{ij} = 1 \ \forall j$

Row-stochastic = Similar

$\det(A) = \det(A^T)$

A and A^T have the same eigen values

If A is column-stochastic, then 1 is an eigenvalue of A

Dealing with Problems 1 and 2

'Damping'

Like Taxation

M = (1 - d)A + dS

where S = $\begin{bmatrix} 1/n & 1/n & \dots & 1/n \\ 1/n & 1/n & \dots & 1/n \\ \dots & \dots & \dots & 1/n \\ 1/n & 1/n & 1/n & 1/n \end{bmatrix}$ = nxn matrix

and where $0 \leq d \leq 1$

Technique = Smoothing of Probability

M = still column-stochastic

Every entry is > 0

An Example

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

$d = 1/2$

M = (1 - d)A + dS = 0.5A + 0.5S = 0.5$\begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}$ + 0.5

$\begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$

Back to General case...

$M = (1 - d)A + dS$

M = column-stochastic

Every entry is > 0

Solving Problem 4

Take two arbitrary eigenspaces $E_{\lambda_1}(M)$ and $E_{\lambda_2}(M)$

Where $\lambda_2 > \lambda_1 > 0$

Dominant Eigen space, $\forall\lambda\ |\lambda| < |\lambda_d|$

Power Convergence Theorem: $\lim_{k\to\infty} M^k\vec{x_0} \in E_{\lambda_d}$

To show: $E_1(M)$ is the dominant Eigenspace

If λ is an eigenvalue of M, then $|\lambda| \leq 1$

Assume there exists \lambda > 1 is the dominant eigenvalue of M

Let $\vec{x} \in E_\lambda(M^T)$

$M^T\vec{x} = \lambda\vec{x}$

Let x_{max} be the maximum entry in $\vec{x}$

$x_{max} < \lambda x_{max}$ is the maximum value in $\lambda\vec{x}$

$$M^T\vec{x} = \begin{bmatrix} \sum_{j=1}^n m_{j1}x_j \\ \sum_{j=1}^n m_{j2}x_j \\ \cdots \\ \sum_{j=1}^n m_{jn}x_j \end{bmatrix}$$

$\sum_{j=1}^n m_{ji}x_j < \sum_{j=1}^n m_{ji}x_{max} = x_{max}\sum_{j=1}^n m_{ji} = x_{max}$

So max value of $M^T\vec{x} \leq x_{max}$

Contradiction!

Summary

1. $M = (1-d)A + dS$
   1. In original paper, $d = 0.15$
2. $M^k\vec{x_0}$ converges to $\vec{x^*} \in E_1(M)$
   1. k = 45?
   2. $\vec{x^*}$ = PageRank values