Maths Project Report

# Topic: Pagerank Algorithm

### Done By Abijith and Rohith.G

INDEX

What is Pagerank Algorithm-------------------------------1-2

Damping Factor-----------------------------------------------3

Advantages and Disadvantages of using Pagerank algorithm--------------------------------------------------------4

Implementation of Pagerank Algorithm using Matlab------------------------------------------------------------------------5-6

Result---------------------------------------------------------------7

Applications-------------------------------------------------------8

Bibliography-------------------------------------------------------9

## What is Pagerank Algorithm

PageRank (PR) is an algorithm used by Google Search to rank web pages in their search engine results. One of the reasons why Google is such an effective search engine is the Pagerank algorithm developed by Google’s founders Larry Page and Sergey Brin.

Imagine Surfing the Web from page to page by randomly choosing an outgoing link from one page to another. This may lead to surfing ending up on pages with no more outgoing links (Sink) or cycles around cliques of interconnected pages (Wikipedia). This random walk is known as the Markov process or the Markov chain. The limiting probability that a random surfer visits any particular page in a set of Web pages is known as that page’s Pagerank (PR). A page will have a high pagerank if and only if other pages with high rank link to it.

## Algorithm simplified

Assume a Web with pages A, B, C and D. Consider the following equation below:



This means that A is web page with pages B, C, and D have outbound links coming towards A.

D

C

B

A



C

B

A

D



General Case algorithm:



# Damping Factor (d)

The probability, at any step, that the person will continue is a damping factor d. It is generally assumed that the damping factor will be set around d=0.85. The damping factor is included in the algorithm to avoid sinks and cliques with interconnected pages. The probability 1-d shows that an arbitrary page is randomly chosen and 1-d/N, where N is the total number of web pages, shows the probability that a web page is randomly chosen.





# Modification of Algorithm using Eigenvector concept.

The PageRank values are the entries of the dominant right eigenvector of the modified adjacency matrix rescaled so that each column adds up to one. This makes PageRank a particularly elegant metric: the eigenvector is



{\displaystyle \mathbf {R} ={\begin{bmatrix}PR(p\_{1})\\PR(p\_{2})\\\vdots \\PR(p\_{N})\end{bmatrix}}}

where **R** is the solution of the equation



{\displaystyle \mathbf {R} ={\begin{bmatrix}{(1-d)/N}\\{(1-d)/N}\\\vdots \\{(1-d)/N}\end{bmatrix}}+d{\begin{bmatrix}\ell (p\_{1},p\_{1})&\ell (p\_{1},p\_{2})&\cdots &\ell (p\_{1},p\_{N})\\\ell (p\_{2},p\_{1})&\ddots &&\vdots \\\vdots &&\ell (p\_{i},p\_{j})&\\\ell (p\_{N},p\_{1})&\cdots &&\ell (p\_{N},p\_{N})\end{bmatrix}}\mathbf {R} }

where the adjacency function {\displaystyle \ell (p\_{i},p\_{j})}

 is the ratio between number of links outbound from page j to page i to the total number of outbound links of page j. The adjacency function is 0 if page j{\displaystyle p\_{j}} does not link to page i{\displaystyle p\_{i}}ipagsa and normalized such that, for each j

{\displaystyle \sum \_{i=1}^{N}\ell (p\_{i},p\_{j})=1}

i.e. the elements of each column sum up to 1, so the matrix is a stochastic matrix. Thus this is a variant of the eigenvector centrality measure used commonly in network analysis.Because of the large eigengap of the modified adjacency matrix above, the values of the PageRank eigenvector can be approximated to within a high degree of accuracy within only a few iterations.

# Advantages and Disadvantages of using Pagerank Algortihm

## Advantages:

* Less Query time cost: The query-time cost of incorporating the precomputed Pagerank importance score for a page is low as compared to other algorithms.
* Less susceptibility to localized links: Furthermore, as Pagerank is generated using the entire Web graph, rather than a small subset, it is less susceptible to localized link spam.
* More efficient: In contrast, Pagerank computes a single measure of quality for a page at crawl time. This measure is then combined with a traditional information retrieval score at query time.
* Feasibility: It performs computations at crawl time than query time. Hence, it is more feasible.

## Disadvantages:

* Raw sinks: When in network pages get in infinite link cycles.
* Spider traps: This is another problem in this algorithm. A group pages is a spider trap if there are no links from within the group to the outside the group.
* Dead ends.

# Implementation of Pagerank using Matlab

% Parameter M adjacency matrix where M\_i,j represents the link from 'j' to 'i', such that for all 'j'

% sum(i, M\_i,j) = 1

% Parameter d damping factor

% Parameter v\_quadratic\_error quadratic error for v

% Return v, a vector of ranks such that v\_i is the i-th rank from [0, 1]

function [v] = rank2(M, d, v\_quadratic\_error)

N = size(M, 2); % N is equal to either dimension of M and the number of documents

v = rand(N, 1);

v = v ./ norm(v, 1); % This is now L1, not L2

last\_v = ones(N, 1) \* inf;

M\_hat = (d .\* M) + (((1 - d) / N) .\* ones(N, N));

while(norm(v - last\_v, 2) > v\_quadratic\_error)

last\_v = v;

v = M\_hat \* v;

% removed the L2 norm of the iterated PR

end

end %function

































# Result:

M=[1 0.5 1 1/3;0 0 0 1/3;0 0 0 1/3;0 0.5 0 0]

M =

1.0000 0.5000 1.0000 0.3333

0 0 0 0.3333

0 0 0 0.3333

0 0.5000 0 0

>> rank2(M,0.85,0.0001)

ans =

0.8298

0.0547

0.0547

0.0608

# Applications

* For the analysis of protein networks in biology PageRank is also a useful tool.
* In neuroscience, the PageRank of a neuron in a neural network has been found to correlate with its relative firing rate.
* Twitter, Facebook, Instagram, etc.
* E-shopping (Amazon, Flipkart, etc.)

# Bibliography

Links:

[www.wikipedia.org/wiki/PageRank](http://www.wikipedia.org/wiki/PageRank)

Books:

Experiments with Matlab, Cleve Moler