planetmath.org

Math for the people, by the people.

Google matrix

Canonical name	GoogleMatrix
Date of creation	2013-03-22 17:01:05
Last modified on	2013-03-22 17:01:05
Owner	Mathprof (13753)
Last modified by	Mathprof (13753)
Numerical id	6
Author	Mathprof (13753)
Entry type	Definition
Classification	msc 15A51

Google's PageRank algorithm uses a particular stochastic matrix called the Google matrix. The purpose of the PageRank algorithm is to compute a stationary vector of the Google matrix. The stationary vector is then used to provide a ranking of the pages on the internet.

A directed graph D is constructed whose vertices correspond to web pages and a directed arc from vertex i to vertex j exists if and only if page i has a link out to page j. Then a stochastic matrix $A = (a_{ij})$ is constructed from D: for each i, j set

$$a_{ij} = 1/d(i)$$

if the outdegree of vertex $i, d(i)$ is positive and there is an arc from i to j in D. Set

$$a_{ij} = 0$$

if $d(i) > 0$ but there is no arc from i to j in D.

Set

$$a_{ij} = 1/n$$

if $d(i) = 0$, where n is the order of the matrix.

Having defined A choose a positive row vector v^T such that $v^T \mathbf{1} = 1$ where $\mathbf{1}$ is a vector of all ones. Finally, choose a constant $c \in (0, 1)$. The *Google matrix* G is

$$G = cA + (1 - c)\mathbf{1}v^T.$$

Clearly, G is stochastic. For the actual matrix that Google uses c is about .85.