What Makes a Web Page Occurring First in a Google Search Result?

Idea of the PageRank Method

Sergey Brin & Lawrence Page: The anatomy of a large-scale hypertextual web search engine, 1998.

The model tries to model the bahavior of a "random" web surfer.

- The surfer starts at a randomly chosen web page.
- Then she clicks at random on one of the links on that page.
- Afterwards the process is repeated (ad infinitum).
- If the page arrived at does not have any outgoing link, the process starts at the very beginning, that is, with a randomly chosen web page.

Disclaimer: The true algorithm used by Google is kept secret and it changes constantly.

Discrete Probability

We have a finite set Ω={1,2,...,n} of elementary events (or states). Each event occurs with a given probility p(k), where 0≤p(k)≤1. This induces

```
> ps := proc(p, A)
    local s, i; s := 0;
    for i from 1 to nops(A) do
      s := s + p(A[i]);
    end do;
    return s;
  end proc:
```

This is the probability of rolling the number x with a fair dice.

> $p := x \rightarrow \frac{1}{6}$

$$p := x \rightarrow \frac{1}{6} \tag{1.2.1}$$

> $ps(p, \{2, 3\})$

$$\frac{1}{3} \tag{1.2.2}$$

It is required that ps(Ω)=1.

> $ps(p, \{1, 2, 3, 4, 5, 6\})$

$$1 \tag{1.2.3}$$

The probability of two independent events is multiplied.

> $p(2) \cdot p(3)$

$$\frac{1}{36} \tag{1.2.4}$$

Stochastic Matrices

An non-negative real-valued *mxn*-matrix is *stochastic* if each column sums to 1. Each column defines a discrete probability space on the set *{1,2,...,m}*.

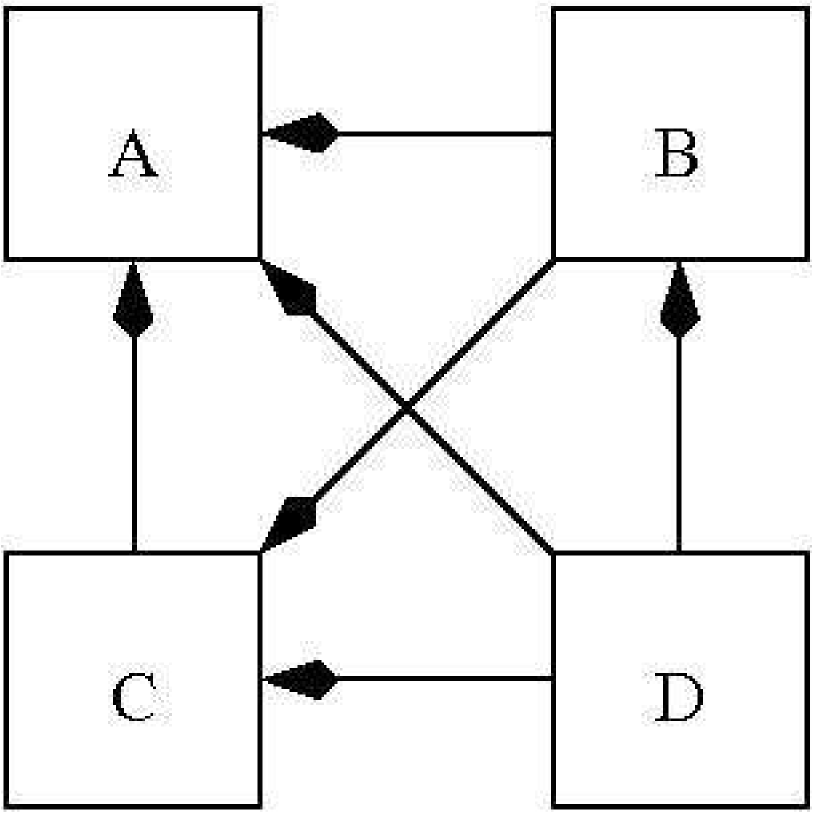
> $with(LinearAlgebra)$:

Here is an example with m=n=4.

> $L := \left\langle \left\langle \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right\rangle \middle| \left\langle \frac{1}{2}, 0, \frac{1}{2}, 0 \right\rangle \middle| \langle 1, 0, 0, 0 \rangle \middle| \left\langle \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0 \right\rangle \right\rangle$

$$L := \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & 1 & \frac{1}{3} \\ \frac{1}{4} & 0 & 0 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{3} \\ \frac{1}{4} & 0 & 0 & 0 \end{bmatrix} \quad (1.3.1)$$

Surfing the WWW (with 4 Web Pages)



Our surfer begins with the following distribution.

> $Start := \left\langle \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right\rangle$

$$Start := \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix} \quad (1.4.1)$$

We now see that L(i,j) is the number of links from page j to page i devided by the number of links on page j.

A page without a link is modelled by the column (1/n,1/n,...,1/n).

The state after one click:

> $L.Start$

$$\begin{bmatrix} \frac{25}{48} \\ \frac{7}{48} \\ \frac{13}{48} \\ \frac{1}{16} \end{bmatrix} \quad (1.4.2)$$

For instance, the first entry, corresponding to web page A gives the probability that after one click we arrive at page A. In the previous step we had been at A, B, C, or D, where each page makes a contribution. This sums up as follows:

> $\frac{1}{16} + \frac{1}{8} + \frac{1}{4} + \frac{1}{12}$

$$\frac{25}{48} \quad (1.4.3)$$

... after two clicks:

> $L^2.Start$

$$\begin{bmatrix} \frac{95}{192} \\ \frac{29}{192} \\ \frac{43}{192} \\ \frac{25}{192} \end{bmatrix} \quad (1.4.4)$$

> $evalf(L^2.Start)$

$$\begin{bmatrix} 0.4947916667 \\ 0.1510416667 \\ 0.2239583333 \\ 0.1302083333 \end{bmatrix} \quad (1.4.5)$$

... and so on:

> $seq(evalf(L^k.Start), k = 8..11)$

$$\begin{bmatrix} 0.4799513169 \\ 0.1600239601 \\ 0.2399326372 \\ 0.1200920858 \end{bmatrix}, \begin{bmatrix} 0.4799631417 \\ 0.1600185245 \\ 0.2400305045 \\ 0.1199878292 \end{bmatrix}, \begin{bmatrix} 0.4800264953 \\ 0.1599867285 \\ 0.2399959908 \\ 0.1199907854 \end{bmatrix}, \begin{bmatrix} 0.4799929073 \\ 0.1600035523 \\ 0.2399969166 \\ 0.1200066238 \end{bmatrix} \quad (1.4.6)$$

Eigenvalues and Eigenvectors

An eigenvalue of a matrix A is a scalar λ such that there is some nonzero vector x with Ax = λx. The vector x is called eigenvector of A with respect o λ.

> $M := Matrix\left(1..5, 1..5, \left\langle 1, 3, -3, 0, \frac{3}{7} \right\rangle, shape = diagonal\right)$

$$M := \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{7} \end{bmatrix} \quad (1.5.1)$$

> $Eigenvalues(M)$

$$\begin{bmatrix} 0 \\ 1 \\ 3 \\ -3 \\ \frac{3}{7} \end{bmatrix} \quad (1.5.2)$$

> $(\lambda, V) := Eigenvectors(M)$

$$\lambda, V := \begin{bmatrix} 1 \\ 0 \\ 3 \\ \frac{3}{7} \\ -3 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (1.5.3)$$

> $\lambda[1].Column(V, 1), M.Column(V, 1)$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (1.5.4)$$

> $(\lambda, V) := Eigenvectors(L)$

$$\lambda, V := \begin{bmatrix} -\frac{1}{4} + \frac{1}{12} \mathrm{I}\sqrt{15} \\ -\frac{1}{4} - \frac{1}{12} \mathrm{I}\sqrt{15} \\ -\frac{1}{4} \\ 1 \end{bmatrix}, \left[\left[\right. \right. \quad (1.5.5)$$

$$-\frac{68}{3}\,\frac{-\frac{3}{4}-\frac{3}{4}\,\mathrm{I}\sqrt{15}+22\left(-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)^2}{\left(\frac{5}{4}+\frac{7}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{5}{2}+\frac{23}{6}\,\mathrm{I}\sqrt{15}\right)\left(-2+\frac{1}{3}\,\mathrm{I}\sqrt{15}\right)},$$

$$-\frac{68}{3}\,\frac{-\frac{3}{4}+\frac{3}{4}\,\mathrm{I}\sqrt{15}+22\left(-\frac{1}{4}-\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)^2}{\left(\frac{5}{4}-\frac{7}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{5}{2}-\frac{23}{6}\,\mathrm{I}\sqrt{15}\right)\left(-2-\frac{1}{3}\,\mathrm{I}\sqrt{15}\right)},\ -1,\ 4\Bigg],$$

$$\Bigg[\frac{4}{3}\,\frac{-\frac{19}{4}+\frac{103}{12}\,\mathrm{I}\sqrt{15}+66\left(-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)^2}{\left(\frac{5}{4}+\frac{7}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{5}{2}+\frac{23}{6}\,\mathrm{I}\sqrt{15}\right)},$$

$$\frac{4}{3}\,\frac{-\frac{19}{4}-\frac{103}{12}\,\mathrm{I}\sqrt{15}+66\left(-\frac{1}{4}-\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)^2}{\left(\frac{5}{4}-\frac{7}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{5}{2}-\frac{23}{6}\,\mathrm{I}\sqrt{15}\right)},\ -\frac{1}{3},\ \frac{4}{3}\Bigg],$$

$$\Bigg[-\frac{2}{3}\,\frac{-\frac{5}{4}+\frac{3}{4}\,\mathrm{I}\sqrt{15}}{\frac{5}{4}+\frac{7}{12}\,\mathrm{I}\sqrt{15}},\ -\frac{2}{3}\,\frac{-\frac{5}{4}-\frac{3}{4}\,\mathrm{I}\sqrt{15}}{\frac{5}{4}-\frac{7}{12}\,\mathrm{I}\sqrt{15}},\ \frac{1}{3},\ 2\Bigg],$$

$$\Big[1,\ 1,\ 1,\ 1\Big]\Big]\Big]$$

> $\lambda[1].Column(V,1),\ L.Column(V,1)$

$$\Big[\Big[\qquad \textbf{(1.5.6)}$$

$$-\frac{68}{3}\left(\left(-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{3}{4}-\frac{3}{4}\,\mathrm{I}\sqrt{15}+22\left(-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)^2\right)\right)\Big/\left(\left(\frac{5}{4}+\frac{7}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{5}{2}+\frac{23}{6}\,\mathrm{I}\sqrt{15}\right)\left(-2+\frac{1}{3}\,\mathrm{I}\sqrt{15}\right)\right)\Big],$$

$$\Big[\frac{4}{3}\left(\left(-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{19}{4}+\frac{103}{12}\,\mathrm{I}\sqrt{15}+66\left(-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)^2\right)\right)\Big/\left(\left(\frac{5}{4}+\frac{7}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{5}{2}+\frac{23}{6}\,\mathrm{I}\sqrt{15}\right)\right)\Big],$$

$$\left[-\frac{2}{3}\,\frac{\left(-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{5}{4}+\frac{3}{4}\,\mathrm{I}\sqrt{15}\right)}{\frac{5}{4}+\frac{7}{12}\,\mathrm{I}\sqrt{15}}\right],$$

$$\left[-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right]\Bigg]\Bigg],\Bigg[\Bigg[$$

$$-\frac{17}{3}\,\frac{-\frac{3}{4}-\frac{3}{4}\,\mathrm{I}\sqrt{15}+22\left(-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)^2}{\left(\frac{5}{4}+\frac{7}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{5}{2}+\frac{23}{6}\,\mathrm{I}\sqrt{15}\right)\left(-2+\frac{1}{3}\,\mathrm{I}\sqrt{15}\right)}$$

$$+\frac{2}{3}\,\frac{-\frac{19}{4}+\frac{103}{12}\,\mathrm{I}\sqrt{15}+66\left(-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)^2}{\left(\frac{5}{4}+\frac{7}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{5}{2}+\frac{23}{6}\,\mathrm{I}\sqrt{15}\right)}$$

$$\left.-\frac{2}{3}\,\frac{-\frac{5}{4}+\frac{3}{4}\,\mathrm{I}\sqrt{15}}{\frac{5}{4}+\frac{7}{12}\,\mathrm{I}\sqrt{15}}+\frac{1}{3}\right],$$

$$\left[-\frac{17}{3}\,\frac{-\frac{3}{4}-\frac{3}{4}\,\mathrm{I}\sqrt{15}+22\left(-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)^2}{\left(\frac{5}{4}+\frac{7}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{5}{2}+\frac{23}{6}\,\mathrm{I}\sqrt{15}\right)\left(-2+\frac{1}{3}\,\mathrm{I}\sqrt{15}\right)}\right.$$

$$\left.+\frac{1}{3}\right],$$

$$\left[-\frac{17}{3}\,\frac{-\frac{3}{4}-\frac{3}{4}\,\mathrm{I}\sqrt{15}+22\left(-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)^2}{\left(\frac{5}{4}+\frac{7}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{5}{2}+\frac{23}{6}\,\mathrm{I}\sqrt{15}\right)\left(-2+\frac{1}{3}\,\mathrm{I}\sqrt{15}\right)}\right.$$

$$\left.+\frac{2}{3}\,\frac{-\frac{19}{4}+\frac{103}{12}\,\mathrm{I}\sqrt{15}+66\left(-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)^2}{\left(\frac{5}{4}+\frac{7}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{5}{2}+\frac{23}{6}\,\mathrm{I}\sqrt{15}\right)}+\frac{1}{3}\right],$$

$$\left[-\frac{17}{3}\,\frac{-\frac{3}{4}-\frac{3}{4}\,\mathrm{I}\sqrt{15}+22\left(-\frac{1}{4}+\frac{1}{12}\,\mathrm{I}\sqrt{15}\right)^2}{\left(\frac{5}{4}+\frac{7}{12}\,\mathrm{I}\sqrt{15}\right)\left(-\frac{5}{2}+\frac{23}{6}\,\mathrm{I}\sqrt{15}\right)\left(-2+\frac{1}{3}\,\mathrm{I}\sqrt{15}\right)}\right]\Bigg]$$

The largest real eigenvalue of L is 1.

> *evalf*(λ)

(1.5.7)

$$\begin{bmatrix} -0.2500000000 + 0.3227486122\,I \\ -0.2500000000 - 0.3227486122\,I \\ -.2500000000 \\ 1. \end{bmatrix} \quad \textbf{(1.5.7)}$$

> $evalf(V)$

$$[[-1.000000000 + 1.290994449\,I, -1.000000000 - 1.290994449\,I, -1., 4.], [0.5000000000 - 0.6454972241\,I, 0.5000000000 + 0.6454972241\,I, -.3333333333, 1.333333333], [-0.5000000001 - 0.6454972246\,I, -0.5000000001 + 0.6454972246\,I, 0.3333333333, 2.], [1., 1., 1., 1.]] \quad \textbf{(1.5.8)}$$

This is an eigenvector to the eigenvalue 1. All its entries are non-negative.

> $x := Column(V, 4)$

$$x := \begin{bmatrix} 4 \\ \frac{4}{3} \\ 2 \\ 1 \end{bmatrix} \quad \textbf{(1.5.9)}$$

> $nx := x[1] + x[2] + x[3] + x[4]$

$$nx := \frac{25}{3} \quad \textbf{(1.5.10)}$$

Scaling does not change the property of being an eigenvector.

> $x := \frac{1}{nx} \cdot x :$

> $x, L.x$

$$\begin{bmatrix} \frac{12}{25} \\ \frac{4}{25} \\ \frac{6}{25} \\ \frac{3}{25} \end{bmatrix}, \begin{bmatrix} \frac{12}{25} \\ \frac{4}{25} \\ \frac{6}{25} \\ \frac{3}{25} \end{bmatrix} \quad \textbf{(1.5.11)}$$

Because of the re-scaling x is a probability distribution, the *limit distribution* of the Markov chain defined by the transition matrix L.

> $evalf(L^{1000}.Start), evalf(x)$

$$\begin{bmatrix} 0.4800000000 \\ 0.1600000000 \\ 0.2400000000 \\ 0.1200000000 \end{bmatrix}, \begin{bmatrix} 0.4800000000 \\ 0.1600000000 \\ 0.2400000000 \\ 0.1200000000 \end{bmatrix} \quad \textbf{(1.5.12)}$$

>