M. SOMBRA, 2020

VII. The PageRank algorithm and the eigenvalue problem

[BL] Bryan and Leise, The LA behind Google, SIAM 2006.

PageRank is the basic algorithm of the Google search machine (for Larry Page, founder of Google).

It has a huge influence on the development and structure of the internet: determines which kinds of information and services are accessed more often.

Three steps for ranking (web)pages:

(1) Locate all pages (with public access)

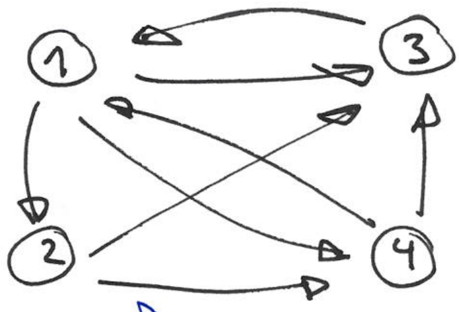
(2) Index data from (1) to be able to search for keywords & phrases.

(3) Rate the "importance" of these pages

We focuss on (3)

In an interconnected web of pages, how can we define and quantify "importance"

(A)



directed graph

"page 2 links to page 4"

The web is understood as a "democracy" where pages "vote" for the importance of other pages by linking to them.

We assume there are n pages, and for $k = 1, \dots, n$ we denote by

$$x_k \qquad \text{score of page } k$$

We also assume

- $x_k \geq 0 \quad \forall k$
- $\displaystyle\sum_{k=1}^{n} x_k = 1$

The simplest approach would consist in counting links. In the example it would give

$$x_1 = \frac{2}{8} \qquad x_2 = \frac{1}{8} \qquad x_3 = \frac{3}{8} \qquad x_4 = \frac{2}{8}$$

But this forgets an important aspect: links from "important" pages should count more:

For instance, pages 1 and 4 have the same number of backlinks, but 1 is linked by the important page 3, whereas 4 is linked by the less important page 1

Another aspect: a page should not increase its overall influence by increasing its number of links:
if page j has n_j links, each of them should contribute with the score $\boxed{x_j/n_j}$ to the page they link:
the overall influence of the page j is its score x_j.

The score vector $\underset{\substack{\| \\ (x_1,\dots,x_n)}}{x} \in \mathbb{R}^n_{\geq 0}$ satisfies the equations

$$x_k = \sum_{j \in L_k} \frac{x_j}{n_j} \qquad k = 1, \dots, n \qquad (*)$$

where $L_k \subset \{1, \dots, n\}$ is the set of pages linking to the page k.

The corresponding link matrix $A \in \mathbb{R}^{n \times n}$ is defined by

$$A_{ij} = \begin{cases} 1/n_j & \text{if page } j \text{ links page } i \\ 0 & \text{else} \end{cases}$$

Then the system of linear equations $(*)$ is equivalent to

$$\boxed{Ax = x}$$

The score vector is an eigenvector of A with eigenvalue 1.

In the example

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

The (normalized) eigenvector of A is

$$x = \left(\frac{12}{31}, \frac{4}{31}, \frac{9}{31}, \frac{6}{31}\right)^T = (0.387, 0.129, 0.290, 0.194)^T$$

The page 3 (linked by all the others) is less important than the page 1! the page 3 links only to page 1. Together with the link from page 2, gives the page 1 the highest score.

We assume that the web has <u>no dangling nodes</u>: (pages without outgoing links).

Then A is <u>column stochastic</u>: its entries are nonnegative and each column sums 1:

for $j = 1, \dots, n$

$$\boxed{\sum_i A_{ij} = 1}$$

This ensures that A has $\lambda = 1$ as one of its eigenvalues: indeed

$$A \quad \text{and} \quad A^T$$

have the same <u>characteristic polynomial</u>

$$\chi_{A^T} = \det(A^T - t\mathbb{1}_n) = \det(A - t\mathbb{1}_n)^T = \det(A - t\mathbb{1}_n) = \chi_A$$

The eigenvalues are the zeros of the characteristic polynomial, and so they coincide.

We have

$$A^T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

A^T is now stochastic!

Hence 1 is an eigenvalue of A^T, and so it is also an eigenvalue of A.

Set

$$V_1(A) = \{ x \in \mathbb{R}^n \mid Ax = x \}$$

eigenspace for the eigenvalue 1.

There are several desirable conditions for this idea to work properly: we would like to have

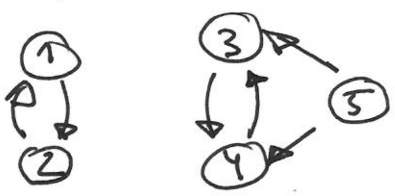
$\dim(V_1(A)) = 1$

so there is a unique eigenvector x with $\sum x_i = 1$

This ~~is~~ not always true

Example

B



gives

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

Then $x = \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ and $y = \begin{pmatrix} 0 \\ 0 \\ 1/2 \\ 1/2 \\ 0 \end{pmatrix}$

are eigenvectors, and $V_1(A) = \mathrm{Vect}(x, y)$ has dimension 2.
Which of the vectors in $V_1(A)$ should be used as a score?

More generally, this is the situation when the network is disconnected. if the web consists of

$$t \text{ subwebs}$$

then $\dim V_1(A) \geq t$ (prove this!)

To avoid this phenomenon, we <u>modify</u> the <u>link matrix</u> by adding a multiple of the <u>uniform</u> matrix:

$$S = (1/n)_{ij} = \begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix}$$

$$M = (1-\alpha) A + \alpha S \quad \text{for } 0 \leq \alpha \leq 1$$

If $\alpha > 0$ then

$$\dim(V_1(M)) = 1$$

The value originally used by Google is

$$\alpha = 0.15$$

Examples:

For the graph in (A) the modified matrix M gives the scores

$$x_1 = 0.368 \qquad x_2 = 0.142 \qquad x_3 = 0.288 \qquad x_4 = 0.202$$

slightly different values giving same order

For the graph in (B) it gives

$$x_1 = 0.2 \qquad x_2 = 0.2 \qquad x_3 = 0.285 \qquad x_4 = 0.285 \qquad x_5 = 0.03$$

allows to compare pages in different subwebs.

The precise result is the following (from [E] L. Eldén, A note on the eigenvalues of the Google matrix, 2003

Let

$$\overset{1}{\lambda_1}, \lambda_2, \dots, \lambda_n$$

eigenvalues of A, ~~with the~~ repeated with their multiplicity (= dimension of the corresponding eigenspace).

Then

$$|\lambda_i| \leq 1 \qquad \forall i \qquad (*)$$

and the eigenvalues of M are

$$\overset{1}{\lambda_1}, (1-\alpha)\lambda_2, \dots, (1-\alpha)\lambda_n \qquad (**)$$

Proof (optional)

(*) for each i choose $x = (x_1, \dots, x_n) \neq 0$ s.t.

$$A x = \lambda_i x$$

Then

$$\| A x\|_1 \geq |\lambda_i| \, \|x\|_1 \qquad (1)$$

and

$$\| A x \|_1 = \sum_{i=1}^{n} \Big| \sum_{j=1}^{n} A_{ij} x_j \Big|$$

$$\leq \sum_{j=1}^{n} \underbrace{\Big(\sum_{i=1}^{n} A_{ij} \Big)}_{=1} |x_j| \leq \sum_j |x_j| = \|x\|_1 \qquad (2)$$

A is column stochastic

Then from (1) and (2) $|\lambda_i| \, \|x\|_1 \leq \|x\|_1$ and so

$$\boxed{|\lambda_i| \leq 1}$$

(**) Set $e = \begin{pmatrix} 1/\sqrt{n} \\ \vdots \\ 1/\sqrt{n} \end{pmatrix} \in \mathbb{R}^n$

It is a unit vector s.t. $S = e e^T$

Complete e to an orthogonal matrix

$$U = (e \ U_1) \qquad n \times n \text{ orthogonal}$$

Then

A column stochastic

$$U^T A U = \begin{pmatrix} e^T A \\ U_1^T A \end{pmatrix} (e \ \ U_1) = \begin{pmatrix} e^T \\ U_1^T A \end{pmatrix} (e \ \ U_1) = \begin{pmatrix} 1 & e^T U_1 \\ U_1^T A e & U_1^T A U_1 \end{pmatrix}$$

$$= \begin{array}{c|cc} & 1 & n-1 \\ \hline 1 & 1 & 0 \\ n-1 & w & T \end{array}$$

e orthogonal to U_1

$n \times n$

Since U^TAU is similar to A, it has the same eigenvalues. Hence T has the eigenvalues

$$\lambda_2, \dots, \lambda_n$$

We have that

$$U^T e = \begin{pmatrix} e^T \\ U_1^T \end{pmatrix} e = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^n$$

and so

$$U^T M U = (1-\alpha)\, U^T A U + \alpha\, U^T e e^T U$$

$$= (1-\alpha)\begin{pmatrix} 1 & 0 \\ w & T \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix}(1\ 0) = \begin{pmatrix} 1 & 0 \\ (1-\alpha)w & (1-\alpha)T \end{pmatrix}$$

Hence the eigenvalues of M are

$$1,\ (1-\alpha)\lambda_2,\ \dots,\ (1-\alpha)\lambda_n$$

proof end) $\boxtimes$

This implies that

$$\dim V_1(M) = 1$$

To compute the eigenvector $x \in \mathbb{R}^n$ we apply the <u>power method</u>:
start with a "typical" vector $x_0 \in \mathbb{R}^n$ and iterate

$$x_k \leftarrow \frac{M x_{k-1}}{\|M x_{k-1}\|_1} \qquad \text{for } k \geq 1$$

This iterative method converges towards x, the normalized eigenvector corresponding to the largest eigenvalue, $\lambda = 1$

The rate of convergence is <u>linear</u> and depends on the gap between the largest eigenvalue and the other ones:

$$\underbrace{-\log_b \|x - x_k\|_1}_{\approx \#\text{correct digits of } x_k} \geq k \log_b\left(\frac{1}{|\lambda_2|}\right) + \text{constant}$$

b base of the floating point system

In our situation, Eldén's theorem implies that

$$|\lambda_2| \leq 1-\alpha = 0.85$$

Taking $b = 10$ we have that

$$\log_{10}\left(\frac{1}{|\lambda_2|}\right) \geq 0.07$$

Hence

$$-\log_{10} \|x - x_k\|_1 \geq 0.07 \cdot k + \text{constant}$$

correct decimal digits of x_k

There are about $n = 4 \cdot 10^8$ active public webpages.
Each iteration of the power method consists in a matrix-vector multiplication.
In a practical implementation, it is computed as

$$Mv = (1-\alpha) Av + \alpha \begin{pmatrix} 1/n \\ \vdots \\ 1/n \end{pmatrix}$$

sparse

for $v \in \mathbb{R}^n_{\geq 0}$ with $\|v\|_1 = 1$ because

$$Sv = \begin{pmatrix} 1/n \\ \vdots \\ 1/n \end{pmatrix}$$

uniform column stochastic matrix