

# How Google works

Christiane ROUSSEAU

*Département de mathématiques et de statistique and CRM  
Université de Montréal*

August 5, 2010

## 1 Introduction

From its very beginning, Google became “the” search engine. This comes from the supremacy of its ranking algorithm: the *PageRank algorithm*. Indeed, with the enormous quantity of pages on the World-Wide-Web, many searches end up with thousands or millions of results. If these are not properly ordered, then the search may not be of any help, since no one can explore millions of entries.

*How does the PageRank algorithm work?*

We will explain this. But before, let us make a search on Google. On June 4 2010, you get 16,300,000 results for *Klein project*, even though the project is just beginning. On that precise date, the first entry is

<http://www.mathunion.org/icmi/other-activities/klein-project/introduction/>  
rather than

<http://www.kleinproject.org/>

The first url is the url of a *page*, which is located on the website of the International Mathematical Union: <http://www.mathunion.org>. Because the International Mathematical Union is an important body, its official website comes first when you make the search “International Mathematical Union”. Moreover, it communicates some of its importance to all of its pages, one of which is

<http://www.mathunion.org/icmi/other-activities/klein-project/introduction/>  
In a few months or years from now, we could expect that the page

<http://www.kleinproject.org/>  
will appear first in a search for *Klein project*.

To explain the algorithm, we model the web as an oriented graph. The vertices are the *pages*, and the oriented edges are the *links* between pages. As we just explained, each page corresponds to a different url. Hence, a website may contain many pages. The model makes no difference between the individual pages of a website and its front page. But, most likely, the algorithm will rank better the front page of an important website.

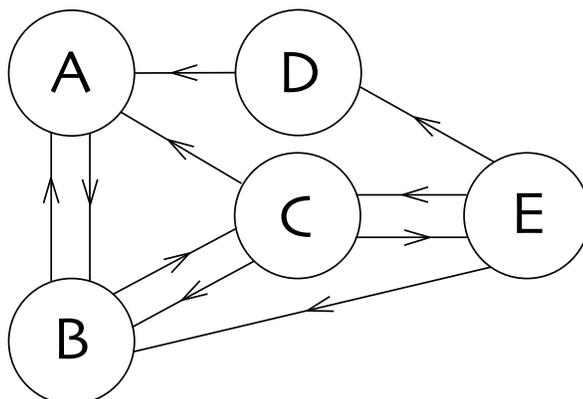


Figure 1: A simple web.

## 2 A simple example

Let us look at the simple web of Figure 1 with five pages named  $A$ ,  $B$ ,  $C$ ,  $D$ , and  $E$ . This web has few links. If we are on page  $A$ , then there is only one link to page  $B$ , while, if we are on page  $C$ , we find three links, and we can choose to move to either page  $A$ , or  $B$ , or  $E$ . Notice that there is at least one link from each page.

We play a game, which is simply a *random walk* on the oriented graph. Starting from a page, at each step we choose at random a link from the page where we are, and we follow it. For instance, in our example, if we start on page  $B$ , then we can go to  $A$  or to  $C$  with probability  $1/2$  for each case while, if we start on  $D$ , then we necessarily go to  $A$  with probability 1. We iterate the game.

*Where will we be after  $n$  steps?*

To automatize the process, we summarize the web in the following matrix  $P$ , where each column represents the departing page, and each row, the page where we arrive.

$$P = \begin{matrix} & \begin{matrix} A & B & C & D & E \end{matrix} \\ \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{3} & 1 & 0 \\ 1 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & 0 & 0 \end{pmatrix} & \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} \end{matrix}$$

Let us remark that the sum of entries of any column of  $P$  is equal to 1 and that all entries are greater or equal to zero. Matrices having these two properties are very special: each such matrix is the matrix of a *Markov chain process*, also called *Markov transition matrix*. It always has 1 as an eigenvalue and there exists an eigenvector with eigenvalue 1, all the components of which are less than or equal to 1 and greater than or equal to 0, with sum equal to 1. But, before recalling the definitions of eigenvalue and eigenvector, let us explore the advantage of the matrix representation of the web graph.

We consider a random variable  $X_n$  with values in the set of pages  $\{A, B, C, D, E\}$ , which contains  $N$  pages (here  $N = 5$ ).  $X_n$  represents the page where we are after  $n$  steps of the random walk. Hence, if we call  $p_{ij}$  the entry of the matrix  $P$  located on the  $i$ -th row and  $j$ -th column, then  $p_{ij}$  is the conditional probability that we are on the  $i$ -th page at step  $n + 1$ , given that we were on the  $j$ -th page at step  $n$ :

$$p_{ij} = \text{Prob}(X_{n+1} = i \mid X_n = j).$$

Note that this probability is independent of  $n$ ! We say that a Markov chain process has *no memory of the past*.

It is not difficult to figure that the probabilities after two steps can be summarized in the matrix  $P^2$ .

Let us prove it (you can skip the proof if you prefer). The theorem of total probabilities yields

$$\text{Prob}(X_{n+2} = i \mid X_n = j) = \sum_{k=1}^N \text{Prob}(X_{n+2} = i \text{ and } X_{n+1} = k \mid X_n = j).$$

By definition of conditional probability

$$\text{Prob}(X_{n+2} = i \mid X_n = j) = \sum_{k=1}^N \frac{\text{Prob}(X_{n+2} = i \text{ and } X_{n+1} = k \text{ and } X_n = j)}{\text{Prob}(X_n = j)}.$$

We use a familiar trick: multiply and divide by the same quantity:

$$\begin{aligned} \text{Prob}(X_{n+2} = i \mid X_n = j) &= \sum_{k=1}^N \frac{\text{Prob}(X_{n+2} = i \text{ and } X_{n+1} = k \text{ and } X_n = j)}{\text{Prob}(X_{n+1} = k \text{ and } X_n = j)} \frac{\text{Prob}(X_{n+1} = k \text{ and } X_n = j)}{\text{Prob}(X_n = j)}. \end{aligned}$$

The first quotient is equal to

$$\text{Prob}(X_{n+2} = i \mid X_{n+1} = k \text{ and } X_n = j) = \text{Prob}(X_{n+2} = i \mid X_{n+1} = k),$$

since a *Markov chain process has no memory past the previous step*. Hence,

$$\begin{aligned} \text{Prob}(X_{n+2} = i \mid X_n = j) &= \sum_{k=1}^N \text{Prob}(X_{n+2} = i \mid X_{n+1} = k) \text{Prob}(X_{n+1} = k \mid X_n = j) \\ &= \sum_{k=1}^N p_{ik} p_{kj} \\ &= (P^2)_{ij}. \end{aligned}$$

□

In our example

$$P^2 = \begin{pmatrix} A & B & C & D & E \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{11}{18} \\ 0 & \frac{2}{3} & \frac{4}{9} & 1 & \frac{1}{9} \\ \frac{1}{2} & 0 & \frac{5}{18} & 0 & \frac{1}{6} \\ 0 & 0 & \frac{1}{9} & 0 & 0 \\ 0 & \frac{1}{6} & 0 & 0 & \frac{1}{9} \end{pmatrix} \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix}$$

Iterating this idea, it is clear that the entry  $(P^m)_{ij}$ , of the matrix  $P^m$  describes the probability  $\text{Prob}(X_{n+m} = i \mid X_n = j)$ . For instance,

$$P^{32} = \begin{pmatrix} A & B & C & D & E \\ 0.293 & 0.293 & 0.293 & 0.293 & 0.293 \\ 0.390 & 0.390 & 0.390 & 0.390 & 0.390 \\ 0.220 & 0.220 & 0.220 & 0.220 & 0.220 \\ 0.024 & 0.024 & 0.024 & 0.024 & 0.024 \\ 0.073 & 0.073 & 0.073 & 0.073 & 0.073 \end{pmatrix} \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix}$$

All columns of  $P^{32}$  are identical if we choose precision to 3 decimals, and the same as the columns of  $P^n$  when  $n > 32$ . If we choose a higher precision, we also find stabilization, but for  $n$  larger than 32. Hence, after  $n$  steps, where  $n$  is sufficiently large, the probability of being on a page is independent from where we started!

Moreover, let us consider the vector

$$\pi^t = (0.293, 0.390, 0.220, 0.024, 0.073)$$

( $\pi$  is a vertical vector, and its transpose  $\pi^t$  is a horizontal vector). It is easily checked that  $P\pi = \pi$ . If we consider the  $i$ -th coordinate of the vector  $\pi^t$  as the probability of being on page  $i$  at a given time  $n$ , and hence  $\pi^t$  as the probability distribution of pages at time  $n$ , then it is also the probability distribution at time  $n + 1$ . For this reason, the vector  $\pi$  is called the *stationary distribution*. This stationary distribution allows to order the pages. In our example, we order the pages as  $B, A, C, E, D$ , and we declare  $B$  the most important page.

### 3 The general case

The general case can be treated exactly as our example. We represent the web as an oriented graph in which the  $N$  vertices represent the  $N$  pages of the web, and the oriented edges represent the links between pages. We summarize the graph in an  $N \times N$  matrix,  $P$ , with the  $j$ -column representing the  $j$ -th departing page and the  $i$ -row, the  $i$ th arrival page. In our example we found a vector  $\pi$  satisfying  $P\pi = \pi$ . This vector is an eigenvector of the eigenvalue 1. Let us recall the definition of eigenvalue and eigenvector.

**Definition 3.1** *Let  $P$  be an  $N \times N$  matrix. A number  $\lambda \in \mathbb{C}$  is an eigenvalue of  $P$  if there exists a nonzero vector  $X \in \mathbb{C}^N$  such that  $PX = \lambda X$ . Any such vector  $X$  is called an eigenvector of  $P$ .*

We also recall the method to find the eigenvalues and eigenvectors.

**Proposition 3.2** *Let  $P$  be an  $N \times N$  matrix. The eigenvalues of  $P$  are the roots of the characteristic polynomial  $\det(\lambda I - P) = 0$ , where  $I$  is the  $N \times N$  identity matrix. The eigenvectors of an eigenvalue  $\lambda$  are the nonzero solutions of the linear homogeneous system  $(\lambda I - P)X = 0$ .*

The following deep theorem of Frobenius guarantees that for the matrix associated to a web graph, we will always find a stationary solution.

**Theorem 3.3** *(Frobenius theorem) We consider an  $N \times N$  Markov transition matrix  $P = (p_{ij})$  (i.e.  $p_{ij} \in [0, 1]$  for all  $i, j$ , and the sum of entries on each column is equal to 1, namely  $\sum_{i=1}^N p_{ij} = 1$ ). Then*

- (i)  $\lambda = 1$  is one eigenvalue of  $P$ .
- (ii) Any eigenvalue  $\lambda$  of  $P$  satisfies  $|\lambda| \leq 1$ .
- (iii) There exists an eigenvector  $\pi$  of the eigenvalue 1, all the coordinates of which are greater or equal than zero. Without loss of generality we can suppose that the sum of its coordinates be 1.

It is now time to admire the power of this theorem. For this purpose, we will make the simplifying hypothesis that the matrix  $P$  has a basis of eigenvectors  $\mathcal{B} = \{v_1, \dots, v_n\}$  and we suppose that  $v_1$  is the vector  $\pi$  of Frobenius theorem. For each  $v_i$ , there exists  $\lambda_i$  such that  $Pv_i = \lambda_i v_i$ . Let us take any nonzero vector  $X$  with  $X^t = (x_1, \dots, x_N)$ , where  $x_i \in [0, 1]$  and  $\sum_{i=1}^N x_i = 1$ . We decompose  $X$  in the basis  $\mathcal{B}$ :

$$X = \sum_{i=1}^N a_i v_i.$$

A technical proof that we skip allows to prove that  $a_1 = 1$ . Now, let us calculate  $PX$ :

$$PX = \sum_{i=1}^N a_i P v_i = \sum_{i=1}^n a_i \lambda_i v_i,$$

since  $v_i$  is an eigenvector of the eigenvalue  $\lambda_i$ . And if we iterate we get

$$P^n X = \sum_{i=1}^n a_i \lambda_i^n v_i.$$

If all  $\lambda_i$  for  $i > 1$  satisfy  $|\lambda| < 1$ , then

$$\lim_{n \rightarrow \infty} P^n X = a_1 v_1 = \pi.$$

This is exactly what happens in our example!

Notice however that the theorem does not guarantee that any matrix  $P$  satisfying the hypotheses of the theorem will have this property. Let us describe the possible pathologies and the remedy.

**Possible pathologies.**

- The eigenvalue 1 may be a multiple root of the characteristic polynomial  $\det(\lambda I - A) = 0$ .
- The matrix  $P$  may have other eigenvalues  $\lambda$  than 1, with modulus equal to 1.

What do we do in this case?

We call a Markov transition matrix with no pathology a *regular Markov transition matrix*, namely

**Definition 3.4** *A Markov transition matrix is regular if*

- *The eigenvalue 1 is a simple root of the characteristic polynomial  $\det(\lambda I - A) = 0$ .*
- *All other eigenvalues  $\lambda$  of  $P$  than 1 have modulus less than 1.*

Remark that most matrices  $P$  are regular. So, if the Markov transition matrix of a web is not regular, then the strategy is to deform it slightly to a regular one.

**Remedy.** We consider the  $N \times N$  matrix  $Q = (q_{ij})$ , with  $q_{ij} = \frac{1}{N}$  for all  $i, j$ . We replace the matrix  $P$  of the web by the matrix

$$P_\beta = (1 - \beta)P + \beta Q, \tag{1}$$

for some small  $\beta \in [0, 1]$ . (The value  $\beta = 0.15$  has been used by Google.) Notice that the matrix  $P_\beta$  still has nonnegative entries and that the sum of the entries of each column is still equal to 1, so it is still a Markov transition matrix. The following theorem guarantees that there exists some small  $\beta$  for which we have fixed the pathologies.

**Theorem 3.5** *Given any Markov transition matrix  $P$ , there always exists a positive  $\beta$ , as small as we wish, such that the matrix  $P_\beta$  is regular. Let  $\pi$  be the eigenvector of 1 for the matrix  $P_\beta$ , normalized so that the sum of its coordinates be 1. For the matrix  $P_\beta$ , given any nonzero vector  $X$ , where  $X^t = (p_1, \dots, p_N)$  with  $p_i \in [0, 1]$  and  $\sum_{i=1}^N p_i = 1$ , then*

$$\lim_{n \rightarrow \infty} (P_\beta)^n X = \pi. \tag{2}$$

## 4 Link with the Banach fixed point theorem

Another (coming) vignette treats the Banach fixed point theorem. The theorem above can be seen as a particular application of it. You can skip this section if you have not read the other vignette. Indeed,

**Theorem 4.1** *Let  $P$  be a regular Markov transition matrix. We consider  $\mathcal{S} = \{X \mid X^t = (p_1, \dots, p_N), p_i \in [0, 1], \sum_{i=1}^N p_i = 1\}$ , with an adequate distance  $d(X, Y)$  between points (this distance depends on  $P$ ). On  $\mathcal{S}$ , we consider the linear operator  $L : \mathcal{S} \rightarrow \mathcal{S}$  defined by  $L(X) = PX$ . The operator  $L$  is a contraction on  $\mathcal{S}$ , namely there exists  $c \in [0, 1[$  such that for all  $X, Y \in \mathcal{S}$ ,*

$$d(L(X), L(Y)) \leq c d(X, Y).$$

*Then, there exists a unique vector  $\pi \in \mathcal{S}$  such that  $L(\pi) = \pi$ .*

*Moreover, given any  $X_0 \in \mathcal{S}$ , we can define the sequence  $\{X_n\}$  by induction, where  $X_{n+1} = L(X_n)$ . Then,  $\lim_{n \rightarrow \infty} X_n = \pi$ .*

**Definition of the distance  $d$ .** The definition of the distance  $d$  is a bit subtle and can be skipped. We include it for the sake of completeness for the reader who wants to push far the details. We limit ourselves to the case where the matrix  $P$  is diagonalizable. Let  $\mathcal{B} = \{v_1 = \pi, v_2, \dots, v_N\}$  be a basis of eigenvectors. Vectors  $X, Y \in \mathcal{S}$  can be written in the basis  $\mathcal{B}$ :

$$X = \sum_{i=1}^N a_i v_i, \quad Y = \sum_{i=1}^N b_i v_i,$$

where  $a_i, b_i \in \mathbb{C}$ . Then we define

$$d(X, Y) = \sum_{i=1}^n |a_i - b_i|.$$

With this distance,  $\mathcal{S}$  is a complete metric space, i.e. all Cauchy sequences converge.

This theorem not only guarantees the existence of  $\pi$ , but gives a method to construct it, as the limit of the sequence  $\{X_n\}$ . We have seen an illustration of this convergence in our example. Indeed, the  $j$ -th column of  $P^n$  is the vector  $P^n e_j$ , where  $e_j^t = (0, \dots, 0, 1, 0, \dots, 0)$  is the  $j$ -th vector of the canonical basis. Of course, in our example, we could also have found directly the vector  $\pi$  by solving the system  $(I - P)X = 0$  with matrix

$$I - P = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{3} & -1 & 0 \\ -1 & 1 & -\frac{1}{3} & 0 & -\frac{1}{3} \\ 0 & -\frac{1}{2} & 1 & 0 & -\frac{1}{3} \\ 0 & 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & -\frac{1}{3} & 0 & 1 \end{pmatrix}.$$

We would have found that all solutions are of the form  $(4s, \frac{16}{3}s, 3s, \frac{1}{3}s, s)^t$  for  $s \in \mathbb{R}$ . The solution whose sum of coordinates is 1 is hence  $\pi$ , where

$$\pi^t = \left( \frac{12}{41}, \frac{16}{41}, \frac{9}{41}, \frac{1}{41}, \frac{3}{41} \right).$$

## 5 Practical calculation of the stationary distribution

We have identified the simple idea underlying the algorithm. However finding the stationary distribution  $\pi$ , i.e. an eigenvector of the eigenvalue 1 for the matrix  $P_\beta$  in (1), is not at all a trivial task when the matrix has billions of lines and columns: both the time of the computation and the memory space to achieve it represent real challenges. The usual method of Gauss elimination is useless in this case, both because of the size of the computation, and because it requires to divide by small coefficients. A more efficient algorithm makes use of the property (2) (see [2]). This makes the link with the coming vignette on the Banach fixed point theorem which will enlighten that the proof of Banach fixed point theorem provides an algorithm to construct the fixed point.

Indeed, we start with  $X_0$  such that

$$X_0^t = \left( \frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N} \right),$$

and we need to calculate  $\lim_{n \rightarrow \infty} (P_\beta)^n X_0$ . Usually,  $P^n X_0$  for some  $n$  between 50 and 100 gives a pretty good approximation of  $\pi$ . By induction, we calculate  $X_{n+1} = P_\beta X_n$ . Even such calculations are quite long. Indeed, due to its construction, the matrix  $P_\beta$  in (1) has no zero entries. On the other hand, most of the entries of the matrix  $P$  are zero. So, we must decompose the computation to take advantage of this fact, namely

$$X_{n+1} = (1 - \beta)PX_n + \beta QX_n.$$

Because of the special form of  $Q$ , it is easy to verify that if  $X$  is a vector the sum of entries of which is 1, then  $QX = X_0$ . Hence, it suffices to calculate the sequence

$$X_{n+1} = (1 - \beta)PX_n + \beta X_0.$$

## 6 Conclusion

We have presented the public part of Google's PageRank algorithm. Already you can experiment with simple webs and find the tricks to improve the ranking of your personal page by adding internal and external links in an optimal way. Some private, more sophisticated parts continue to be developed. Some of them consist in replacing the "neutral" matrix  $Q$  of (1) by matrices reflecting the *taste* of the surfer on the web. Others ensure that the ranking is not too sensitive to the manipulations of those who try to improve the ranking of their pages.

As general conclusion, what have we observed? A simple, clever idea has led to an immense breakthrough in the efficiency of search engines, and to the birth of a commercial empire. Even if the implementation is in itself a computational exploit, the initial idea required “elementary” mathematics, namely linear algebra and probability theory. These relatively standard mathematical tools, in particular the diagonalization of matrices, have given their full power when they have been used outside of their “normal” context. Also, we have highlighted the unifying ideas inside science, with Banach fixed point theorem having applications so remote from its origin.

## References

- [1] M. Eisermann, *Comment Google classe les pages web*, <http://images.math.cnrs.fr/Comment-Google-classe-les-pages.html>, 2009.
- [2] A. N. Langville and C. D. Meyer, A Survey of Eigenvector Methods for Web Information Retrieval, *SIAM Review*, Volume 47, Issue 1, (2005), pp. 135–161.
- [3] C. Rousseau and Y. Saint-Aubin, *Mathematics and technology*, SUMAT Series, Springer-Verlag, 2008 (A French version of the book exists, published in the same series.)