Google's secret and Linear Algebra

Pablo Fernández Gallardo (Madrid, Spain)

1. Introduction

Some months ago newspapers all around the world considered Google's plan to go public. The significance of this piece of news was not only related to the volume of the transaction, the biggest since the dot-com "irrational exuberance" in the 90's, but also to the particular characteristics of the firm. A few decades ago there was a complete revolution in technology and communications (and also a cultural and sociological revolution), namely the generalization of use and access to the Internet. Google's appearance has represented a revolution comparable to the former as it became a tool that brought some order into this universe of information, a universe that was not manageable before.

The design of a web search engine is a problem of *mathematical engineering*. Notice the adjective. First, a deep knowledge of the context is needed in order to translate it into models, into mathematics. But after this process of abstraction and after the relevant conclusions have been drawn, it is essential to carry out a thorough, detailed and efficient design of the computational aspects inherent in this problem.

2. The Google engine

The Google search engine was designed in 1998 by Sergei Brin and Lawrence Page, two computer science doctorate students at Stanford – two young men, now in their thirties, who have become multimillionaires. The odd name of the firm is a variation of the term *googol*, the name that somebody[1] invented to refer to the overwhelming number 10^{100}. This is one of those numbers that mathematicians are comfortable with but are perhaps bigger than the number of particles in the whole universe.

The scale of the question we are concerned with is also immense. In 1997, when Brin and Page were to start working on Google's design, there were about 100 million web pages. Altavista, the most popular search engine in those days, attended to 20 million daily queries. Today, these figures have been multiplied; Google receives some hundred million daily queries and indexes several billion web pages.

Therefore, the design of a search engine must efficiently solve some computational aspects, namely the way to store that enormous amount of information, how it is updated, how to manage the queries, the way to search the databases, etc. But, although interesting, we are not going to treat these questions here. The point of interest can be formulated in a simple manner. Let us suppose that, after a certain query, we have determined that, say, one hundred web pages enclose information that might, in some sense, be relevant to the user. Now, *in which order should they be displayed?* The objective, as explicitly posed[2] by Brin and Page (see [6]), is that in the majority of attempts, at least one of, say, the first ten displayed pages contains useful information for the user.

We now ask the reader (quite possibly a *google-maniac* himself) to decide, from his own experience, whether Google fulfils this objective or not. We are sure the common response will be affirmative … and even amazingly affirmative! It seems to be magic[3] but it is just mathematics, mathematics requiring no more than the tools of a first year graduate course, as we will soon see.

To tackle our task, we need an **ordering criterion**. Notice that if we label each web page with symbols $P_1,\dots,P_n$, all we want is to assign each P_j a number x_j, its **significance**. These numbers might range, for example, between 0 and 1. Once the complete list of web pages, along with their significances, is at our disposal, we can use this ordering each time we answer a query; the selected pages will be displayed in the order as prescribed by the list.[4]

3. The model

Let us suppose that we have collected all the information about the web: sites, contents, links between pages, etc. The set of web pages, labelled $P_1 \dots P_n$, and the links between them can be modelled with a (directed) **graph** G. Each web page P_j is a vertex of the graph and there will be an edge between vertices P_i and P_j whenever there is a link from page P_i to page P_j. It is a gigantic, overwhelming graph, whose real structure deserves some consideration (see Section 8).

When dealing with graphs, we like to use drawings in the paper, in which vertices are points of the plane, while edges are merely arrows joining these points. But, for our purposes, it is helpful to consider an alternative description, with matrices. Let us build an $n \times n$ matrix $\mathbf{M}$ with zero-one entries, whose rows and columns are labelled with symbols $P_1,\dots,P_n$. The matrix entry m_{ij} will be 1 whenever there is a link from page P_j to page P_i and 0 otherwise.

The matrix $\mathbf{M}$ is, except for a transposition, the *adjacency matrix* of the graph. Notice that it is not necessarily symmetric because we are dealing with a directed graph. Observe also that the sum of the entries for P_j's column is the number of P_j's outgoing links, while we get the number of ingoing links by summing rows.

We will assume that the significance of a certain page P_j "is related to" the pages linking to it. This sounds reasonable; if there are a lot of pages pointing to P_j, its information must have been considered as "advisable" by a lot of web-makers.

The above term "related to" is still rather vague. A **first attempt** to define it, in perhaps a naïve manner, amounts to supposing that the significance x_j of each P_j *is proportional to the number of links to* P_j. Let us note that, whenever we have the matrix $\mathbf{M}$ at our disposal, the computation of each x_j is quite simple; it is just the sum of the entries of each row P_j.

This model does not adequately grasp a situation deserving attention, i.e., when a certain page is cited from a few *very relevant* pages, e.g., from www.microsoft.com and www.

amazon.com. The previous algorithm would assign it a low significance and this is not what we want. So we need to enhance our model in such a way that a strong significance is assigned both to highly cited pages and to those that, although not cited so many times, have links from very "significant pages".

Following this line of argument, the **second attempt** assumes that the significance x_j of each page P_j *is proportional to the sum of the significances of the pages linking to* P_j. This slight variation completely alters the features of the problem.

Suppose, for instance, that page P_1 is cited on pages P_2, P_{25} and P_{256}, that P_2 is only cited on pages P_1 and P_{256}, etc., and that there are links to page P_n from P_1, P_2, P_3, P_{25} and P_{n-1}. Following the previous assignment, x_1 should be proportional to 3, x_2 to 2, etc., while x_n should be proportional to 5. But now, our assignment $x_1,\dots,x_n$ must verify that

$$\begin{aligned} x_1 &= K\,(x_2 + x_{25} + x_{256})\,, \\ x_2 &= K\,(x_1 + x_{256})\,, \\ &\vdots \\ x_n &= K\,(x_1 + x_2 + x_3 + x_{25} + x_{n-1})\,, \end{aligned}$$

where K is a certain proportionality constant. In this way, we face an enormous system of linear equations, whose solutions are all the admissible assignments $x_1 \dots x_n$. Below these lines we write the system of equations in a better way, using matrices.

$$\begin{array}{ccccccccccccccccc} & & P_1 & P_2 & & & & P_{25} & & & P_{256} & & & P_{n-1} & \\ & & \downarrow & \downarrow & & & & \downarrow & & & \downarrow & & & \downarrow & \end{array}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = K \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Let us call the significance vector $\mathbf{x}$. The $n \times n$ matrix of the system is exactly the matrix $\mathbf{M}$ associated with the graph. So we can state that the significance assignment is a solution of

$$\mathbf{M}\mathbf{x} = \lambda \mathbf{x}\,.$$

We have already used the symbol λ for the constant of proportionality. This is so because, as anyone who has been exposed to a linear algebra first course will recognize, the question has become a problem of **eigenvalues** and **eigenvectors**; the significance vector $\mathbf{x}$ is no more than an eigenvector of the matrix $\mathbf{M}$. You might recall that this matrix contains all the information about the web structure, i.e., the vertices and adjacency relations.

Perhaps this is not enough to arouse the reader's enthusiasm yet. Alright, an eigenvector. But which one? There are so many. And also, how could we compute it? The matrix is inconceivably huge. Remember, it is built up of a thousand million rows (or columns). Patience, please. For the time being, it sounds reasonable to demand the entries of our vector (the significance of the web pages) to be non-negative (or, at least, with the same sign). This will be written as $\mathbf{x} \geq 0$. We ask the reader to excuse this abuse of notation. But also, since we intend the method to be useful, we need this hypothetical non-negative vector to be **unique**. If there were more than one, which of them should be chosen?

4. The random surfer

Google's approach to the question follows a slightly different point of view. At the present stage, a page P_j distributes a "1" to every page where there is an outgoing link. This means that pages with many outgoing links have a great influence, which surely is not reasonable. It is more fair to assign each page P_j a "total weight" 1, which is equally distributed among the outgoing links. So we should consider a new matrix instead of $\mathbf{M}$ (the matrix of the graph, with entries 0 and 1). Let N_j be the number of P_j's outgoing links (that is, the sum of the entries in the column labelled P_j in $\mathbf{M}$). The new matrix $\mathbf{M}'$ is built from the original $\mathbf{M}$ by replacing each entry m_{ij} by $m'_{ij} = m_{ij}/N_j$. The entries of $\mathbf{M}'$ will be non-negative numbers (between 0 and 1) and the sum of the entries for *each column* will be 1. And now we are interested in the non-negative vector of the corresponding[5] problem $\mathbf{M}'\mathbf{x} = \lambda\mathbf{x}$. The matrix $\mathbf{M}'$ is called a **stochastic (or Markovian) matrix**.

This new point of view leads us to a nice interpretation. Let us imagine a user surfing the web. At some moment he will reach some page, say P_1. But, probably bored with the contents of P_1, he will jump to another page, following P_1's outgoing links (suppose there are N_1 possibilities). But, to which one? Our brave navigator is a *random* surfer – and needless to say, also blond and suntanned. So, in order to decide his destination, he is going to use chance, and in the most simple possible way: with a regular (and virtual, we presume) die, which has the same number of faces as the number of outgoing links from P_1. In technical terms, the choice of destination follows a (discrete) *uniform* probability distribution in $[1, N_1]$. Say, for instance, that there are three edges leaving P_1 to vertices P_2, P_6 and P_8. Our navigator draws his destination, assigning a probability of 1/3 to each vertex.

Our model is no longer *deterministic* but *probabilistic*. We do not know where he will be a moment of time later but we do know *what his chances are* of being in each admissible destination. And it is a *dynamic* model as well because the same argument may be applied to the second movement, to the third one, etc. In our example, if the first movement is from P_1 to P_2 and there are four edges leaving P_2, then he is to draw again, now with probability 1/4 for each possible destination. Our surfer is following what is known as a **random walk** in the graph.

And what about the matrix $\mathbf{M}'$? Let us say that the surfer is on page (vertex) P_k at the beginning, that is in probabilistic terms, he is on page P_k with a probability of 100 %. We represent this initial condition with the vector $(0,\dots,1,\dots,0)$, the 1 being in position k. Recall that the surfer draws among the N_k destinations, assigning probability $1/N_k$ to each of them. But when we multiply the matrix $\mathbf{M}'$ by this initial vector, we get $(m'_{1k}, m'_{2k}, \dots, m'_{nk})$, a vector with entries summing to 1: the m'_{jk} are either 0 or $1/N_k$ and there are exactly N_k non-zero entries. Notice that the vector we get exactly describes the probability of being, one moment later, on each page of the web, assuming he began at P_k. More than that, in order to know the probabilities of being on each page of the web after *two* moments of time, it is enough to repeat the process. That is, to multiply $(\mathbf{M}')^2$ by the initial vector. And the same for the third movement, the fourth, etc.

Following the usual terminology, we consider a certain number of **states**, in our case just being the vertices of the graph G. The matrix $\mathbf{M}'$ is (appropriately) called the **transition matrix** of the system; each entry m'_{ij} describes the probability of going from state (vertex) P_j to state (vertex) P_i. And the entries of the successive powers of the matrix give us transition probabilities between vertices as time goes by. The well versed reader may have already deduced the relation with the previous ideas: the stationary state of this Markov chain turns out to be precisely the non-negative vector of the problem $\mathbf{M}'\mathbf{x} = \lambda\mathbf{x}$.

It might happen that some pages have no outgoing links at all (with only zeros in the corresponding columns). This would not give a stochastic matrix. We will discuss Google's solution to this problem in Section 8.

5. Qualifying for the playoffs

We will illustrate the ordering algorithm[6] with the following question. Let us imagine a sports competition in which teams are divided in groups or conferences[7]. Each team plays the same number of games but not *the same number of games against each other*; it is customary they play more games against the teams from their own conference. So we may ask the following question. Once the regular season is finished, which teams should classify for the playoffs? The standard system computes the number of wins to determine the final positions but it is reasonable (see [10]) to wonder whether this is a "fair" system or not. After all, it might happen that a certain team could have achieved many wins just because it was included in a very "weak" conference. What should be worthier: the number of wins or their "quality"? And we again face Google's dichotomy!

Say, for example, that there are six teams, $E_1, \dots, E_6$, divided into two conferences. Each team plays 21 games in all: 6 against each team from its own conference, 3 against the others. These are the results of the competition:

	E_1	E_2	E_3	E_4	E_5	E_6	
E_1	–	3/21	0/21	0/21	1/21	2/21	→ 6/21
E_2	3/21	–	2/21	2/21	2/21	1/21	→ 10/21
E_3	6/21	4/21	–	2/21	1/21	1/21	→ 14/21
E_4	3/21	1/21	1/21	–	2/21	2/21	→ 9/21
E_5	2/21	1/21	2/21	4/21	–	2/21	→ 11/21
E_6	1/21	2/21	2/21	4/21	4/21	–	→ 13/21

To the right of the table, we have written the number of wins of each team. This count suggests the following ordering: $E_3 \to E_6 \to E_5 \to E_2 \to E_4 \to E_1$. But notice, for instance, that the leader team E_3 has collected a lot of victories against E_1, the worst one.

Let us now assign *significances* $\mathbf{x} = (x_1, \dots, x_6)$ to the teams with the mentioned criterion: x_j is proportional to the number of wins of E_j, *weighted* with the significance of the other teams. If $\mathbf{A}$ is the above table, this leads, once more, to $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. And again, we want to find a non-negative eigenvector of $\mathbf{A}$ (a unique one, if possible).

Even in such a simple example as this one, we need to use a computer. So we ask some mathematical software to perform the calculations. We find that the moduli of the six eigenvalues are 0.012, 0.475, 0.161, 0.126, 0.139 and 0.161. So $\lambda = 0.475$ is the biggest (in modulus) eigenvalue, its associated eigenvector being

$$\mathbf{x} = (0.509, 0.746, 0.928, 0.690, 0.840, 1).$$

And this is the **only** eigenvector that has real non-negative entries! The components of the vector suggest the following ordering: $E_6 \to E_3 \to E_5 \to E_2 \to E_4 \to E_1$. And now E_6 is the best team!

Let us summarize. In this particular matrix with non-negative entries (that might be regarded as a small-scale version of the Internet matrix) we are in the best possible situation; there is a unique non-negative eigenvector, the one we need to solve the ordering question we posed. Did this happen by chance? Or was it just a trick, an artfully chosen matrix to persuade the unwary reader that things work as they should? The reader, far from being unwary, is now urgently demanding a categorical response. And he knows that it is time to welcome a new actor to this performance.

6. Mathematics enters the stage

Let us distil the common essence of all the questions we have been dealing with. Doing so, we discover that the only feature shared by all our matrices (being stochastic or not) is that all their entries are **non-negative**. Not a lot of information, it seems. They are neither symmetric matrices nor positive definite nor... Nevertheless, as shown by Perron[8] at the beginning of the 20th century, it is enough to obtain interesting results:

Theorem (Perron, 1907). *Let* $\mathbf{A}$ *be a square matrix with positive entries,* $\mathbf{A} > 0$. *Then*
(a) *there exists a (simple) eigenvalue* $\lambda > 0$ *such that* $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, *where the corresponding eigenvector is* $\mathbf{v} > 0$*;*
(b) λ *is bigger (in modulus) than the other eigenvalues;*
(c) *any other positive eigenvector of* $\mathbf{A}$ *is a multiple of* $\mathbf{v}$.

Perron's result points to the direction we are interested in but it is not enough because the matrices we deal with might contain zeros. So we need something else. The following act of this performance was written several years later by Frobenius[9] where he deals with the general case of non-negative matrices. Frobenius observed that if we only have that $\mathbf{A} \geq 0$ then, although there is still a *dominant* (of maximum modulus) eigenvalue $\lambda > 0$ associated to an eigenvector $\mathbf{v} \geq 0$, there might be other eigenvalues of the same "size". Here is his theorem:

Theorem (Frobenius, 1908–1912). *Let* $\mathbf{A}$ *be a square matrix with non-negative entries,* $\mathbf{A} \geq 0$. *If* $\mathbf{A}$ *is* ***irreducible***,[10] *then*
(a) *there exists a (simple) eigenvalue* $\lambda > 0$ *such that* $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, *where the corresponding eigenvector is* $\mathbf{v} > 0$. *In addition,* $\lambda \geq |\mu|$ *for any other eigenvalue* μ *of* $\mathbf{A}$.
(b) *Any eigenvector* ≥ 0 *is a multiple of* $\mathbf{v}$.
(c) *If there are k eigenvalues of maximum modulus, then they are the solutions of the equation* $x^k - \lambda^k = 0$.

Notice firstly that Frobenius' theorem is indeed a generalization of Perron's result, because if $\mathbf{A} > 0$, then $\mathbf{A}$ is ≥ 0 and irreducible. Secondly, if $\mathbf{A}$ is irreducible then the question is completely solved: there exists a unique non-negative eigenvector associated to the positive eigenvalue of maximum modulus (a very useful feature, as we will see in a moment).

These results, to which we will refer from now on as the Perron–Frobenius Theorem, are widely used in other contexts (see Section 9). Some people even talk about "Perron–Frobenius Theory", this theorem being one of its central results.

The proof is quite complicated and here we will just sketch an argument (in the 3×3 case) with some of the fundamental ideas. Let us start with a non-negative vector $\mathbf{x}\geq 0$. As $\mathbf{A}\geq 0$, the vector $\mathbf{Ax}$ is also non-negative. In geometric terms, the matrix $\mathbf{A}$ maps the positive octant into itself. Let us consider now the mapping $\boldsymbol{\alpha}$ given by $\boldsymbol{\alpha}(\mathbf{x})=\mathbf{Ax}/\|\mathbf{Ax}\|$. Notice that $\boldsymbol{\alpha}(\mathbf{x})$ is always a unit length vector. The function $\boldsymbol{\alpha}$ maps the set $\{\mathbf{x}\in\mathbb{R}^3 : \mathbf{x}\geq 0, \|\mathbf{x}\|=1\}$ into itself. Now, applying the Brouwer Fixed Point Theorem[11], there exists a certain $\tilde{\mathbf{x}}$ such that $\boldsymbol{\alpha}(\tilde{\mathbf{x}})=\tilde{\mathbf{x}}$. Therefore,

$$\boldsymbol{\alpha}(\tilde{\mathbf{x}})=\frac{\mathbf{A}\tilde{\mathbf{x}}}{\|\mathbf{A}\tilde{\mathbf{x}}\|}=\tilde{\mathbf{x}} \quad\Longrightarrow\quad \mathbf{A}\tilde{\mathbf{x}}=\|\mathbf{A}\tilde{\mathbf{x}}\|\,\tilde{\mathbf{x}}.$$

Summing up, $\tilde{\mathbf{x}}$ is an eigenvector of $\mathbf{A}$ with non-negative entries associated to an eigenvalue >0. For all other details, such as proving that this eigenvector is (essentially) unique and the other parts of the theorem, we refer the reader to [1], [4], [13] and [14].

7. And what about the computational aspects?

The captious reader will be raising a serious objection: Perron–Frobenius' theorem guarantees the existence of the needed eigenvector for our ordering problem but *says nothing* about how to compute it. Notice that the proof we sketched is not a constructive one. Thus, we still should not rule out the possibility that these results are not so satisfactory. Recall that Google's matrix is overwhelming. The calculation of our eigenvector could be a cumbersome task!

Let us suppose we are in an ideal situation, i.e., in those conditions[12] that guarantee the existence of a positive eigenvalue λ_1 *strictly bigger* (in modulus) than the other eigenvalues. Let $\mathbf{v}_1$ be its (positive) eigenvector. We could, of course, compute *all* the eigenvalues and keep the one of interest but even using efficient methods, the task would be excessive. However, the structure of the problem helps us again and make the computation easy. It all comes from the fact that the eigenvector is associated to the dominant eigenvalue.

Suppose, to simplify the argument, that $\mathbf{A}$ is diagonalizable. We have a basis of $\mathbb{R}^n$ with the eigenvectors $\{\mathbf{v}_1,\dots,\mathbf{v}_n\}$, the corresponding eigenvalues being decreasing size ordered: $|\lambda_1|>|\lambda_2|\geq|\lambda_3|\geq\cdots\geq|\lambda_n|$. We start, say, with a certain $\mathbf{v}_0\geq 0$ that may be written as $\mathbf{v}_0=c_1\mathbf{v}_1+c_2\mathbf{v}_2+\cdots+c_n\mathbf{v}_n$, where the numbers $c_1,\dots,c_n$ are the $\mathbf{v}_0$ coordinates in our basis. Now we multiply vector $\mathbf{v}_0$ by matrix $\mathbf{A}$ to obtain $\mathbf{Av}_0=c_1\lambda_1\mathbf{v}_1+c_2\lambda_2\mathbf{v}_2+\cdots+c_n\lambda_n\mathbf{v}_n$ because the vectors $\mathbf{v}_1,\dots,\mathbf{v}_n$ are eigenvectors of $\mathbf{A}$. Let us repeat the operation, say k times: $\mathbf{A}^k\mathbf{v}_0=c_1\lambda_1^k\mathbf{v}_1+c_2\lambda_2^k\mathbf{v}_2+\cdots+c_n\lambda_n^k\mathbf{v}_n$. Let us suppose that $c_1\neq 0$. Then,

$$\frac{1}{\lambda_1^k}\mathbf{A}^k\mathbf{v}_0=c_1\mathbf{v}_1+c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\mathbf{v}_2+\cdots+c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\mathbf{v}_n \xrightarrow{k\to\infty} c_1\mathbf{v}_1$$

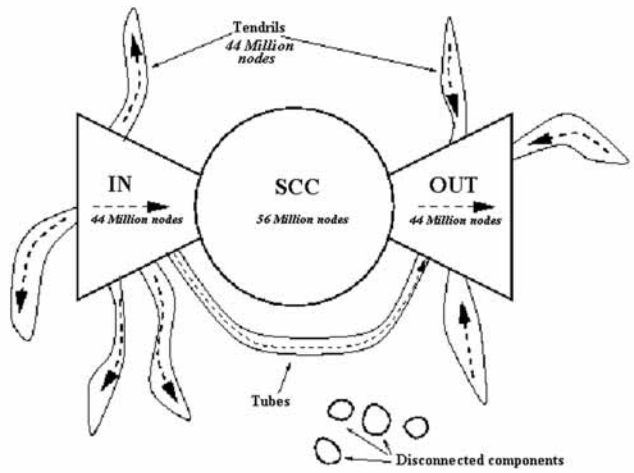
since $|\lambda_j/\lambda_1|<1$ for each $j=2,\dots,n$ (recall that λ_1 was the dominant eigenvalue).

Therefore, when repeatedly multiplying the initial vector by the matrix $\mathbf{A}$, we determine, more precisely each time, the *direction* of interest, namely the one given by $\mathbf{v}_1$. This numerical method is known as the **power method** and its rate of convergence depends on the *ratio* between the first and the second eigenvalue (see in [8] an estimate for Google's matrix).

Our problem is finally solved, at least if we are in the best possible conditions (a non-negative irreducible matrix). The answer does exist, it is unique and we have an efficient method to compute it at our disposal (according to Google's web page, a few hours are needed). But …

8. Are we in an ideal situation?

To make things work properly, we need the matrix $\mathbf{M}$ associated to the web-graph G to be irreducible. In other words, we need G to be a strongly connected graph[13]. As the reader might suspect, this is not the case. Research developed in 1999 (see [7]) came to the conclusion that, among the 203 million pages under study, 90 % of them laid in a gigantic (weakly connected) component, this in turn having a quite complex internal structure, as can be seen in the following picture, taken from [7].



This is a quite peculiar structure, which resembles a biological organism, a kind of colossal amoeba. Along with the central part (SCC, Strongly Connected Component), we find two more pieces[14]: the IN part is made up of web pages having links to those of SCC and the OUT part is formed by pages pointed from the pages of SCC. Furthermore, there are sort of tendrils (sometimes turning into tubes) comprising the pages not pointing to SCC's pages nor accessible from them. Notice that the configuration of the web is something dynamic and that it is evolving with time. And it is not clear whether this structure has been essentially preserved or not[15]. We refer here to [3].

What Google does in this situation is a standard trick: try to get the best possible situation in a reasonable way[16]. For instance, *adding* a whole series of transition probabilities to all the vertices. That is, considering the following matrix,

$$\mathbf{M}''=c\mathbf{M}'+(1-c)\begin{pmatrix}p_1\\ \vdots\\ p_n\end{pmatrix}(1,\dots,1),$$

where $p_1, \ldots, p_n$ is a certain probability distribution ($p_j \geq 0$, $\sum_j p_j = 1$) and c is a parameter between 0 and 1 (for Google, about 0.85).

As an example, we could choose a uniform distribution, $p_j = 1/n$, for each $j = 1, \ldots, n$ (and the matrix would have positive entries). But there are other reasonable choices and this degree of freedom gives us the possibility of making "personalized" searches. In terms of the random surfer, we are giving him the option (with probability $1-c$) to get "bored" of following the links and to jump to any web page (obeying a certain probability distribution)[17].

9. Non-negative matrices in other contexts

The results on non-negative matrices that we have seen above have a wide range of applications. The following two observations (see [13]) may explain their ubiquity:

- In most "real" systems (from physics, economy, biology, technology, etc.) the measured interactions are positive, or at least non-negative. And matrices with non-negative entries are the appropriate way to encode these measurements.
- Many models involve linear iterative processes: starting from an initial state $\mathbf{x}_0$, the generic one is of the form $\mathbf{x}_k = \mathbf{A}^k \mathbf{x}_0$. The convergence of the method depends upon the size of $\mathbf{A}$'s eigenvalues and upon the ratios between their sizes, particularly between the biggest and all the others. And here is where Perron–Frobenius' theorem has something to say, as long as the matrix $\mathbf{A}$ is non-negative.

The probabilistic model of Markov chains is widely used in quite diverse contexts. Google's method is a nice example, but it is also used as a model for population migrations, transmission of diseases, rating migrations in finance, etc. But, as mentioned before, Perron–Frobenius' Theory also plays a central role in many other contexts (we refer the reader again to [13]). Let us mention just a pair:

Biological models: a well known population model, in some sense a generalization of the one developed by Fibonacci, is encoded with the so called *Leslie matrices*. Their entries are non-negative numbers, related to the transition fractions between age classes and survival rates. If λ_1 is the dominant eigenvalue then the system behaviour (extinction, endless growth or oscillating behaviour) depends upon the precise value of λ_1 ($\lambda_1 > 1$, $\lambda_1 = 1$ or $\lambda_1 < 1$ being the three cases of interest).

Economic models: in 1973, Leontief was awarded the Nobel Prize for the development of his *input-output model*. A certain country's economy is divided into sectors and the basic hypothesis is that the j^{th} sector's input of the i^{th} sector's output is proportional to the j^{th} sector's output. In these conditions, the existence of the solution for the system depends upon the value of the dominant eigenvalue of the matrix that encodes the features of the problem.

Finally, there are several extensions of Perron–Frobenius' Theory that the reader might find interesting:

Cones in $\mathbb{R}^n$: the key point of Perron–Frobenius' theorem is that any $n \times n$ matrix with non-negative entries preserve the "positive octant". There is a general version dealing with (proper convex) cones[18] (see [1, 4]).

Banach spaces: those readers versed in functional analysis and spectral theory will be aware of the generalization to Banach spaces known as the Krein–Rutman theorem (see [12] and [5]). And those engaged in partial differential equations will enjoy proving, using Krein–Rutman Theorem, that the first eigenfunction of the Laplacian in the Dirichlet problem (in an open, connected and bounded set $\Omega \subset \mathbb{R}^n$) is *positive* (see the details in the appendix to Chapter 8 of [9]).

10. Coda

The design of a web search engine is a formidable technological challenge. But in the end, we discover that the key point is mathematics: a wise application of theorems and a detailed analysis of the algorithm convergence. A new confirmation of the *unreasonable effectiveness* of mathematics in the natural sciences, as Eugene Wigner used to say – as in so many other fields, we might add. We hope that these pages will encourage the readers to explore for themselves the many problems we have briefly sketched here – and hopefully, they have been a source of good entertainment. And a very fond farewell to Perron–Frobenius' theorem, which plays such a distinguished role in so many questions. Let us bid farewell with a humorous (but regretfully untranslatable[19]) *coplilla manriqueña*:

Un hermoso resultado
que además se nos revela
indiscreto;
y un tanto desvergonzado,
porque de Google desvela
su secreto.

11. To know more

The following book is an excellent and very recent reference: *Google's page rank and beyond: the science of search engine rankings*. (Amy N. Langville and Carl D. Meyer. Princeton University Press, 2006).

Other references cited throughout the note:

[1] Bapat, R. B. and Raghavan, T. E. S.: *Nonnegative matrices and applications*. Cambridge University Press, 1997.

[2] Barabási, A.-L.: *The physics of the web*. Physics World (july 2001). Available at www.nd.edu/~alb.

[3] Barabási, A.-L.: *Linked, the new science of networks. How everything is connected to everything else and what it means*. Plume Books, 2003.

[4] Berman, A. and Plemmons, R. J.: *Nonnegative matrices in the Mathematical Sciences*. Academic Press, 1979.

[5] Brézis, H.: *Análisis funcional. Teoría y aplicaciones*. Alianza Editorial, Madrid, 1984.

[6] Brin, S. and Page, L.: *The anatomy of a large-scale hypertextual web search engine*. www-db.stanford.edu/~sergey/

[7] Broder, A. et al.: *Graph structure in the web*. www9.org/w9cdrom/160/160.html.

[8] Haweliwala, T. S. and Kamvar, S. D.: *The second eigenvalue of the Google matrix*. www.stanford.edu/taherh/papers/secondeigenvalue.pdf.

[9] Dautray, R. and Lions, J.-L.: *Mathematical analysis and numerical methods for science and technology. Volume 3: Spectral Theory and Applications*. Springer-Verlag, Berlin, 1990.

[10] Keener, J. P.: *The Perron–Frobenius Theorem and the ranking of football teams*. SIAM Review **35** (1993), no. 1, 80–93.

[11] Kendall, M. G.: *Further contributions to the theory of paired comparisons*. Biometrics **11** (1955), 43–62.

[12] Krein, M. G. and Rutman, M. A.: *Linear operators leaving invariant a cone in a Banach space*. Uspehi Matem. Nauk (N.S.) **3** (1948), no. 1, 3–95 [Russian]. Translated to English in Amer. Math. Soc. Transl. **10** (1962), 199–325.

[13] MacLauer, C. R.: *The many proofs and applications of Perron's Theorem*. SIAM Review **42** (2000), no. 3, 487–498.

[14] Minc, H.: *Nonnegative matrices*. John Wiley & Sons, New York, 1988.

[15] Wei, T. H.: *The algebraic foundations of ranking theory*. Cambridge Univ. Press, London, 1952.

[16] Wilf, H. S.: *Searching the web with eigenvectors*.

Notes

1. The inventor of the name is said to be a nephew of the mathematician Edward Kasner. Kasner also defined the *googolplex*, its value being 10^{googol}. Wow!
2. They also intended the search engine to be "resistant" to any kind of manipulation, like commercially-oriented attempts to place certain pages at the top positions on the list. Curiously enough, nowadays a new "sport", *Google bombing*, has become very popular: to try to place a web page in the top positions, usually as only a recreational exercise. Some queries such as "miserable failure" have become classics.
3. Not to mention the incredible capacity of the search engine to "correct" the query terms and suggest the word one indeed had in mind. This leads us to envisage supernatural phenomena... well, let us give it up.
4. Although we will not go into the details, we should mention that there are a pair of elements used by Google, in combination with the general criterion we will explain here, when answering specific queries. On one hand, as is reasonable, Google does not give the same "score" to a term that is in the title of the page, in boldface, in a small font, etc. For combined searches, it will be quite different if, within the document, the terms appear "close" or "distant" to each other.
5. This is indeed a new model. Notice that, in general, matrices $\mathbf{M}$ and $\mathbf{M}'$ will not have the same spectral properties.
6. The ideas behind Google's procedure can be traced back to the algorithms developed by Kendall and Wei [11, 15] in the 1950's. At the same time that Brin and Page were developing their engine, Jon Kleinberg presented his HITS (Hypertext Induced Topic Search) algorithm, which followed a similar scheme. Kleinberg was awarded the Nevanlinna Prize at the recent ICM 2006.
7. The NBA competition is a good example although the dichotomy of "number of wins" *versus* their "quality" could also be applied to any competition.
8. The German mathematician Oskar Perron (1880–1975) was a conspicuous example of mathematical longevity and was interested in several fields such as analysis, differential equations, algebra, geometry and number theory, in which he published several text-books that eventually became classics.
9. Ferdinand Georg Frobenius (1849–1917) was one of the outstanding members of the Berlin School, along with distinguished mathematicians such as Kronecker, Kummer and Weierstrass. He is well known for his contributions to group theory. His works on non-negative matrices were done in the last stages of his live.
10. An $n \times n$ matrix $\mathbf{M}$ is *irreducible* if all the entries of the matrix $(\mathbf{I}+\mathbf{A})^{n-1}$, where $\mathbf{I}$ stands for the $n \times n$ identity matrix, are positive. If $\mathbf{A}$ is the adjacency matrix of a graph then the graph is *strongly* connected (see Section 8).
11. Notice that the part of the 2-sphere that is situated in the positive orthant is homeomorphic to a 2-disc.
12. A matrix $\mathbf{A}$ is said to be **primitive** if it has a dominant eigenvalue (bigger, in modulus, than the other eigenvalues). This happens, for instance, when, for a certain positive integer k, all the entries of the matrix $\mathbf{A}^k$ are positive.
13. Let us consider a directed graph G (a set of vertices and a set of directed edges). G is said to be strongly connected if, given any two vertices u and v, we are able to find a sequence of edges joining one to the other. The same conclusion, but "erasing" the directions of the edges, lead us to the concept of a weakly connected graph. Needless to say, a strongly connected graph is also a weakly connected graph but not necessarily the reverse.
14. Researchers put forward some explanations: The IN set might be made up of newly created pages with no time to get linked by the central kernel pages. OUT pages might be corporate web pages, including only internal links.
15. A lot of interesting questions come up about the structure of the web graph. For instance, the average number of links per page, the mean distance between two pages and the probability $P(k)$ of a randomly selected page to have exactly k (say ingoing) links. Should the graph be *random* (in the precise sense of Erdös and Rényi) then we would expect to have a binomial distribution (or a Poisson distribution in the limit). And we would predict that most pages would have a similar number of links. However, empirical studies suggest that the decay of the probability distribution is not exponential but follows a *power law*, $k^{-\beta}$, where β is a little bigger than 2 (see, for instance, [2]). This would imply, for example, that most pages have very few links, while a minority (even though very significant) have a lot. More than that, if we consider the web as an evolving system, to which new pages are added in succession, the outcome is that the trend gets reinforced: "the rich get richer". This is a usual conclusion in competitive systems (as in real life).
16. "Reasonable" means here that it works, as the corresponding ranking vector turns out to be remarkably good at assigning significances.
17. In fact, Google's procedure involves two steps: firstly, in order to make matrix $\mathbf{M}'$ stochastic, the entries of the zero columns are replaced by $1/n$. This "new" matrix $\mathbf{M}'$ is then transformed into $\mathbf{M}''$ as explained in the text. Notice that the original $\mathbf{M}'$ is a very sparse matrix, a very convenient feature for multiplication. In contrast, $\mathbf{M}''$ is a dense matrix. But, as the reader may check, all the vector-matrix multiplications in the power method are executed on the sparse matrix $\mathbf{M}'$.
18. A set $C \subset \mathbb{R}^n$ is said to be a *cone* if $a\mathbf{x} \in C$ for any $\mathbf{x} \in C$ and for any number $a \geq 0$. It will be a *convex cone* if $\lambda\mathbf{x} + \mu\mathbf{y} \in C$ for all $\mathbf{x}, \mathbf{y} \in C$ and $\lambda, \mu \geq 0$. A cone is *proper* if (a) $C \cap (-C) = \{0\}$, (b) $\text{int}(C) \neq \emptyset$; y (c) $\text{span}(C) = \mathbb{R}^n$.
19. More or less: "a beautiful result, which shows itself as indiscreet and shameless, because it reveals... Google's secret".

*Pablo Fernández Gallardo [pablo.fernandez@uam.es] got his PhD in 1997 from the Universidad Autónoma de Madrid and he is currently working as an assistant professor in the mathematics department there. His interests range from harmonic analysis to discrete mathematics and mathematical finance. An extended version of this note, both in English and Spanish, can be found at www.uam.es/pablo.fernandez. A previous Spanish version of this article appeared in Boletín de la Sociedad Española de Matemática Aplicada **30** (2004), 115–141; it was awarded the "V Premio SEMA a la Divulgación en Matemática Aplicada" of Sociedad Española de Matemática Aplicada in 2004.*