On the mathematical background of Google PageRank algorithm

Alberto Peretti
Alberto Roveda
Department of Economics, University of Verona
Verona, Italy
e-mail: alberto.peretti@univr.it
e-mail: alberto.roveda@univr.it

December, 2014

Abstract

The PageRank algorithm, the kernel of the method used by Google Search to give us the answer of a search we are asking in the web, contains a lot of mathematics. Maybe one could say that graphs and Markov chains theories are in the background, while the crucial steps are in a linear algebra context, as the eigenvalues of a matrix are involved. In this working paper we deal with all the mathematics we need to explain how the PageRank method works.

Keywords. Graph theory, Linear mappings, Eigenvalues, Markov chains.

AMS Classification. 15A18, 05C, 60J10

JEL Classification. C65

Contents

Introduction

PageRank is an algorithm used by Google Search to rank websites in their search engine results. It was developed at Stanford University by Larry Page and Sergey Brin, the founders of Google, in 1996 as part of a research project about a new kind of search engine. PageRank was named after Larry Page. The idea under PageRank, contained in the Page and Brin's original paper, is to measure the importance of website pages depending on how many and how much "important" other pages are linked to them. According to Google: *PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites.*

PageRank is the first algorithm that was used by the company, and it is the best-known. It has been modified during the years and combined with other methods.

More in detail, PageRank is a link analysis algorithm and it assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of "measuring" its relative importance within the set. The algorithm may be applied to any collection of entities with reciprocal quotations and references. The numerical weight that it assigns to any given element is referred to as the PageRank of that element.

A PageRank results from a mathematical algorithm based on the so called webgraph, created by all World Wide Web pages as nodes and hyperlinks as edges. The rank value indicates an importance of a particular page. A hyperlink to a page counts as a vote of support. The PageRank of a page is defined recursively and depends on the number and PageRank metric of all pages that link to it. A page that is linked to by many pages with high PageRank receives a high rank itself.

A lot of research has been conducted on the PageRank method, also to identify falsely influenced PageRank rankings. Following these studies Google looked for improvements with the goal to find an effective means of ignoring links from documents with falsely influenced PageRank.

In this working paper we want to present a summary of all the mathematical background that has some relevance inside the PageRank algorithm. The mathematics that is involved is quite a lot actually, essentially covering three big mathematical areas: the graph theory, some linear algebra topics, mainly eigenvalues matters, and some simple issues on Markov chains.

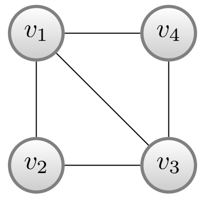
The first three sections are devoted to these three topics. In the fourth section we present the power method, an iterative method to compute the dominant eigenvalue and eigenvector, that is a fundamental part of the PageRank algorithm, and the original algorithm itself.

1 Some elements from graph theory

Definition 1 *A graph is a couple* $G = (V, E)$, *where* $V = \{v_1, v_2, \dots, v_n\}$ *is a finite set of elements (the nodes of the graph) and* $E = \{e_1, e_2, \dots, e_m\} \subseteq V \times V$ *is a subset of couples of nodes (the edges of the graph).*

If E *is a subset of ordered couples of nodes, the graph* G *is an oriented graph (or direct graph).*

In the picture below on the left we have a possible representation of a graph G with $V = \{v_1, v_2, v_3, v_4\}$ and $E = \{(v_1, v_2), (v_1, v_3), (v_1, v_4), (v_2, v_3), (v_3, v_4)\}$. On the right there is a representation of a direct graph G with $V = \{v_1, v_2, v_3, v_4\}$ and $E = \{(v_1, v_2), (v_1, v_3), (v_2, v_1), (v_2, v_3), (v_3, v_4), (v_4, v_1)\}$.



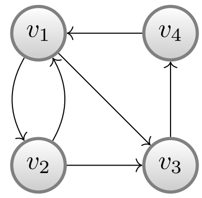


Figure 1: Graphs

As we will be concentrated in the following on direct graphs mainly, let's now recall some classical definitions for this kind of graphs.

Definition 2 *If* $G = (V, E)$ *is a direct graph, a direct path in* G *is a sequence of nodes* $(v_0, v_1, \dots, v_k)$ *such that* $(v_i, v_{i+1}) \in E$ *for each* $i = 0, 1, \dots, k-1$. *We say that* k *is the length of the path.*

If none of the nodes, with the possible exception of v_0 *and* v_k, *appears in the sequence twice, the path is simple. If* $v_0 = v_k$ *the path is closed. A path that is simple and closed is called a cycle.*

For the definition of the concept of connectivity in a direct graph it may be important to consider a path regardless of direction of the edges.

Definition 3 *In the same context of the previous Definition, a non-direct path in* G *is a sequence of nodes* $(v_0, v_1, \dots, v_k)$ *such that either* (v_i, v_{i+1}) *or* (v_{i+1}, v_i) *belongs to* E *for each* $i = 0, 1, \dots, k-1$.

Definition 4 *A direct graph is connected if for every pair of nodes* (v_i, v_j) *in* V *there exists a non direct path that joins them. A direct graph is strongly connected if for every pair of nodes* (v_i, v_j) *in* V *there exists a direct path that joins them.*

2 The algebraic background

In the PageRank algorithm important aspects are related to the dominant eigenvalue of the Google matrix and its associated eigenvector.

We are going to recall some topics in linear algebra that are important in the sequel.

Connected graph   Strongly connected graph

Figure 2: Connectivity

2.1 The adjacency matrix

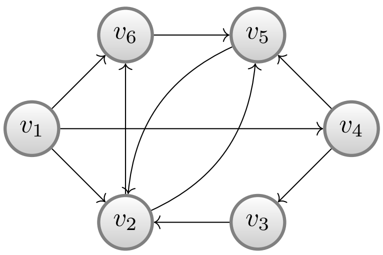
First of all, there are important ways to represent a graph. We may for instance put all the relevant information about the graph in a matrix.

Suppose $G = (V, E)$ is a direct graph.

Definition 5 *We call the adjacency matrix A of G the $n \times n$ square matrix, where $n = |V|$, such that*

$$a_{ij} = \begin{cases} 1 & \textit{if } (v_i, v_j) \in E \\ 0 & \textit{if } (v_i, v_j) \notin E. \end{cases}$$

Example. Here is an example of a direct graph and its adjacency matrix.



The graph

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

The adjacency matrix

Figure 3: Graphs and matrices

There can be many matrices associated to a graph. Often it is much easier to operate on a matrix that on a geometrical representation of the graph.

2.2 Permutation matrices

Definition 6 *An $n \times n$ matrix P whose rows are obtained by any permutation of the rows of the identity matrix is called a permutation matrix.*

Example. An example of permutation matrix is

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

The matrix P has been obtained with the permutation (3 1 2) of the rows of the identity matrix or equivalently with the permutation (2 3 1) of its columns. And the two permutations are the inverse of each other.[1]

Regarding the effect of the multiplication by a permutation matrix, let's consider a simple example:

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} b & c & a \\ e & f & d \\ h & i & g \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = \begin{pmatrix} g & h & i \\ a & b & c \\ d & e & f \end{pmatrix}.$$

[1] Think that if the first row goes to the k-th position, and then $p_{k1} = 1$, then the k-th column goes to the first position.

The effect of a right multiplication of a 3×3 matrix A by the permutation P is the permutation (2 3 1) of the columns of A. The effect of a left multiplication of A by P is the permutation (3 1 2) of the rows of A.

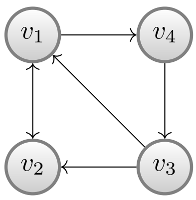
Remark. In general, since the rows of an $n \times n$ permutation matrix P are the fundamental vectors in $\mathbb{R}^n$, the result of PP^T is the identity matrix and then the inverse of P is its transpose. In other words P is an orthogonal matrix.

The reason is in what we have just seen: if P has a certain permutation of rows, then P^T has that permutation on its columns. In this way, looking at PP^T as the right permutation of P by P^T, we permute the columns of P with the permutation of its rows. But the two are inverse and then we obtain the identity matrix. In a similar way if we look at PP^T as the left permutation of P^T by P.

Definition 7 *Given any $n \times n$ square matrix A, the result $A' = PAP^T$, where P is a permutation matrix, is called a symmetric permutation of A.*

Remark. By considering the effect of right and left multiplication by a permutation matrix, the effect of PAP^T is a certain permutation of the rows followed by the same permutaion of the columns. We may remark also that from the associative property of the matrix multiplication the result is the same if we permute the columns first and the rows secondly or the rows first and the columns at a second time.

Example. To see the algebraic effects on the matrix and the corresponding effects on the graph, let's make an example on the adjacency matrix of a simple graph. Consider the direct graph G in Figure 4 together with its adjacency matrix A.



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

Figure 4: Example on permutations

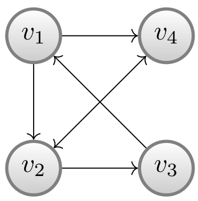
Let's consider also the permutation matrix

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

obtained by the permutation (3 1 4 2) of the rows of the identity 4×4 matrix. We have

$$PAP^T = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

The corresponding graph is the following



It is easy to see that the effect of the symmetric permutation on the adjacency matrix of a graph is just to rename the nodes of the graph. On the graph there has been the permutation (3 1 4 2) of the nodes, in the sense that: the new node 1 is the old node 3, the new node 2 is the old 1, the new node 3 is the old 4 and the new node 4 is the old 2.

It is worthwhile to remark that the only left multiplication of the adjacency matrix by a permutation P is not interesting and is not related to a permutation of the nodes in the graph. For example in our case

$$PA = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

In Figure 5 we can see the effects of the non symmetric permutation on the original graph. It is evident that in the result some important topological properties of the original graph are destroyed.

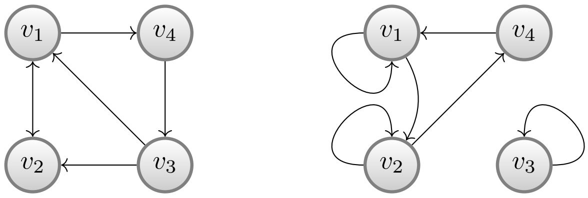


Figure 5: A non simmetric permutation on a graph

2.3 Eigenvalues and eigenvectors

Fundamental steps of the PageRank algorithm are related to eigenvalues. We recall that

Definition 8 *Given an $n \times n$ real or complex matrix A, a real or complex number λ and a real or complex non null vector v are respectively an eigenvalue and an associated eigenvector of A if*

$$Av = \lambda v.$$

In a similar way, a real or complex non null vector w is a left eigenvector associated to λ if

$$w^T A = \lambda w^T.$$

We recall that the eigenvalues of A are the roots of $P(\lambda) = \det(A - \lambda I)$, the characteristic polynomial of A. The multiplicity of λ as a root of the characteristic polynomial is the so called algebraic multiplicity $m_a(\lambda)$ of λ.

In the sequel we shall be interested essentially on the real case, but it is important to keep in mind that for a polynomial, even with real coefficients, we can guarantee the existence just of complex roots. The fundamental theorem of algebra states that every $n \geq 1$ degree single-variable polynomial with complex coefficients has at least one complex root. The theorem may also be stated as follows: every n degree single-variable polynomial with complex coefficients has, counted with multiplicity, exactly n roots. The equivalence of the two statements can be proved through the use of successive polynomial division.

This includes polynomials with real coefficients. There are important properties of the matrix A that bring to the existence of real eigenvalues (at least one or all). A fundamental result is that symmetric matrices have all the eigenvalues that are real.

Our interest is on real eigenvalues $\lambda \in \mathbb{R}$ and associated real eigenvectors $v \in \mathbb{R}^n$.

We shall indicate with:

- V_λ the eigenspace of the eigenvalue λ, that is the subspace of $\mathbb{R}^n$ spanned by the eigenvectors associated to λ.
- $\mathrm{Sp}(A)$ the spectrum of the matrix A, that is the set of the eigenvalues of A.
- $\rho(A)$ the spectral radius of A, that is $\rho(A) = \max_{1 \le i \le n} |\lambda_i|$.

Sometimes we call *spectral circle* the circle in the complex plane with center at the origin and radius equal to the spectral radius. Obviously all the eigenvalues belong to the spectral circle.

Remark. If λ is an eigenvalue of A, its associated eigenspace V_λ is the set of solutions of the linear homogeneous system $(A - I)v = 0$. The dimension of V_λ as a subspace of $\mathbb{R}^n$ is the so called geometric multiplicity $m_g(\lambda)$ of λ. For every eigenvalue λ we have that $m_g(\lambda) \le m_a(\lambda)$.

Definition 9 *If A is an $n \times n$ symmetric matrix and x is a non null vector in $\mathbb{R}^n$, the Rayleigh quotient $R(A, x)$ is defined as*

$$R(A, x) = \frac{x^T A x}{x^T x}.$$

It can be shown that

$$\min_{x \in \mathbb{R}^n} \frac{x^T A x}{x^T x} = \lambda_{\min}$$

that is the smallest eigenvalue of A. The minimum is attained at $x = v_{\min}$, where $v_{\min}$ is an eigenvector associated to $\lambda_{\min}$. Similarly

$$\max_{x \in \mathbb{R}^n} \frac{x^T A x}{x^T x} = \lambda_{\max},$$

the largest eigenvalue of A. The maximum is attained at $x = v_{\max}$, where $v_{\max}$ is an eigenvector associated to $\lambda_{\max}$.

The Rayleigh quotient is used in eigenvalue algorithms to obtain an eigenvalue approximation from an eigenvector approximation. We are going to see this shortly in the power method algorithm.

2.4 The dominant eigenvalue

Fundamental results hold regarding the leading (or dominant) eigenvalue, that is the eigenvalue with maximum module (absolute value). The most famous among these results is probably the *Perron-Frobenius* theorem, that has an important role in the PageRang algorithm.

Let us first give some definitions about matrices.

Definition 10 *A matrix A is positive if all its elements are positive, that is $a_{ij} > 0$ per every i, j.*

A matrix A is non-negative if all its elements are non-negative, that is $a_{ij} \ge 0$ per every i, j.

The Perron-Frobenius theorem essentially says that a real square matrix with positive elements has a unique largest real eigenvalue. Moreover, an associated eigenvector can be chosen to have positive components. A similar statement holds for certain classes of non-negative matrices.

The first results are due to Perron, regarding positive matrices. A generalization was obtained later by Frobenius and deals with non-negative matrices.

The theorem has important applications to probability theory (ergodicity of Markov chains), to economic theory (Leontiev's input-output model) and many other fields.

Theorem 1 (Perron-Frobenius for positive matrices) *Let A be an $n \times n$ positive matrix. Then the following statements hold.*

(i) There is an eigenvalue r (the so called Perron-Frobenius eigenvalue), that is real and positive, and for any other eigenvalue λ we have $|\lambda| < r$. Then $r = \rho(A)$.

(ii) r is a simple eigenvalue, that is a simple root of the characteristic polynomial. In other words its algebraic multiplicity is one ($m_a(r) = 1$). As a consequence the eigenspace V_r associated with r is one-dimensional.

(iii) There exists a positive eigenvector v associated with r. Respectively, there exists a positive left eigenvector w.

(iv) There are no other positive eigenvectors of A, except (positive) multiples of v (respectively, left eigenvectors except w).

(v) $\lim_{k \to +\infty} \left(\frac{A}{r}\right)^k = vw^T$, where the right and left eigenvectors are normalized, so that $w^T v = 1$. Moreover, the matrix vw^T is the projection onto the eigenspace V_r, the so called Perron projection.

(vi) The Perron-Frobenius eigenvalue r satisfies the inequalities

$$\min_i \sum_j a_{ij} \le r \le \max_i \sum_j a_{ij}.$$

Example. Let's consider the positive matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

With the help of a scientific software[2] we get:

```
-->A=[1,2,3;4,5,6;7,8,9]
 A  =
    1.    2.    3.
    4.    5.    6.
    7.    8.    9.

-->spec(A)
 ans  =
    16.116844
  - 1.116844
  - 1.304D-15

-->[R,diagevals] = spec(A)
 diagevals  =
    16.116844    0            0
    0          - 1.116844     0
    0            0          - 1.304D-15

 R  =
  - 0.2319707  - 0.7858302    0.4082483
  - 0.5253221  - 0.0867513  - 0.8164966
  - 0.8186735    0.6123276    0.4082483
```

At the first prompt the command `A=[...]` just defines the matrix A. At the second prompt the `spec(A)` command gives the eigenvalues of A, namely

$$\lambda_1 = 16.116844 \quad , \quad \lambda_2 = -1.116844 \quad , \quad \lambda_3 = 1.304 \cdot 10^{-15}.$$ [3]

[2] We have been using Scilab for the numerical computations.

[3] λ_3 is actually zero, as the matrix is singular.

At the third prompt again the `spec(A)` command, in a more complete form, gives the diagonal matrix similar to A (the eigenvalues are in the main diagonal) and a second matrix where the columns are normalized eigenvectors, respectively associated to the corresponding eigenvalues in the first matrix.

Clearly λ_1 is the Perron-Frobenius eigenvalue r; for the other eigenvalues the inequalities $|\lambda_2| < r$ and $|\lambda_3| < r$ hold.

The first column v_1 of the last matrix is a generator for the eigenspace V_{λ_1}. We can see that the eigenvector v_1 is negative and this confirms the thesis of Perron-Frobenius theorem, that is satisfied of course by $-v_1$. Note also that the other two normalized eigenvectors v_2 and v_3 both have positive and negative components at the same time and then they do not have positive multiples.

We may get the left eigenvectors with the same commands above, applied to the transpose matrix of A (A' in Scilab). Here is Scilab:

```
-->[R,diagevals] = spec(A')
 diagevals  =
    16.116844    0            0
    0          - 1.116844     0
    0            0          - 5.701D-16

 R  =
  - 0.4645473  - 0.8829060    0.4082483
  - 0.5707955  - 0.2395204  - 0.8164966
  - 0.6770438    0.4038651    0.4082483
```

The eigenvalues are of course the same (the last is zero). The opposite of first column of the last matrix is the positive left eigenvector of the Perron-Frobenius eigenvalue. To check point 5 of the thesis we may compute some powers of A^k/r^k for increasing values of k. Some attempts show that already with $k = 5$ we are near to converge. Here is Scilab:

```
-->A^5/16.116844^5
 ans  =
    0.1120296    0.1376534    0.1632772
    0.2537054    0.3117313    0.3697573
    0.3953811    0.4858093    0.5762374
```

We may define the vectors v and w as the positive right and left eigenvectors, compute the product vw^T and we get[4]

```
-->v = [0.2319707,0.5253221,0.8186735]
 v  =
    0.2319707    0.5253221    0.8186735

-->norm(v)
 ans  =
    1.

-->w = [0.4645473,0.5707955,0.6770438]
 w  =
    0.4645473    0.5707955    0.6770438

-->norm(w)
 ans  =
    1.

-->v'*w
 ans  =
    0.1077614    0.1324078    0.1570543
    0.2440370    0.2998515    0.3556661
    0.3803126    0.4672951    0.5542778
```

[4]In Scilab you see actually $v^T w$ because the vectors are defined as row vectors.

It's interesting that the two matrices are not exactly the same. The elements are different more or less for a 5 percent. The reason is the condition on the normalization. The two eigenvectors must be normalized in order to ensure $w^T v = 1$. As they are norm one vectors, we may just divide one of the two by $w^T v$. In Scilab we have

```
-->v*w'
 ans  =
    0.9618907

-->v = v/(v*w')
 v  =
    0.2411612    0.5461349    0.8511087

-->v*w'
 ans  =
    1.

-->v'*w
 ans  =
    0.1120308    0.1376537    0.1632767
    0.2537055    0.3117314    0.3697573
    0.3953802    0.4858090    0.5762379
```

Now the two matrices are exactly the same.

To conclude the example, it's immediate to verify point 6 of the thesis, as

$$\min_i \sum_j a_{ij} = 6 \quad \text{and} \quad \max_i \sum_j a_{ij} = 24.$$

Frobenius extended Perron's results to the case of non-negative matrices, with both important similarities and differences. It is easy to obtain for this kind of matrices the existence of a non-negative eigenvalue r, with a corresponding non-negative eigenvector, such that $r \geq |\lambda|$ for every other eigenvalue λ. The problem is to guarantee the uniqueness. In fact there may be eigenvalues with the same absolute value. Consider for example the matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

that is non-negative and has -1 and 1 as eigenvalues.

Moreover non-negative matrices may have the maximal eigenvalue that is not strictly positive and is not a simple root of the characteristic polynomial. Take for example the matrix

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

whose only eigenvalue is 0 with algebraic multiplicity 2.

The main properties of the maximum eigenvalue seem to fail with no hope.

Frobenius idea was to consider a restriction on the class of non-negative matrices, in order to bring the generalization to its maximum extent. Anyway, as we shall see, some properties of the positive matrices Perron eigenvalue are lost.

The matrices for which Frobenius was able to extend Perron results are the so called *irreducible* matrices. They may be defined by means of one of a few equivalent properties, some purely algebraic, some other related to graph theory. We report some of these properties.

Let A be an $n \times n$ square matrix. A is irreducible if any one of the following equivalent properties 1–2–3–4 holds.

The first is a geometric property related to the linear transformation on $\mathbb{R}^n$ associated to the matrix A. Let's make a remark first.

We call a *coordinate subspace* any linear subspace of $\mathbb{R}^n$ spanned by a proper non-trivial subset of the $\mathbb{R}^n$ canonical basis $\{e^1, e^2, \ldots, e^n\}$.

Property 1 *A does not have invariant coordinate subspaces.*

More explicitly, for every coordinate subspace its image through the linear transformation associated to the matrix A is not the subspace itself or a subset of it.

Examples. The matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

is not irreducible, as in $\mathbb{R}^2$ the coordinate subspace spanned by e^1 is invariant. In fact

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ 0 \end{pmatrix} = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}.$$

Conversely the matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is irreducible. As the only coordinate subspaces in $\mathbb{R}^2$ are either the span of e^1 or the span of e^2, it is easy to be convinced that none of the two is invariant.

To make things clearer on how the invariance must be intended the further following example is worthwhile. With the matrix

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

the images of the coordinate subspaces are

$$A \begin{pmatrix} x_1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ x_1 \\ x_1 \end{pmatrix} \quad , \quad A \begin{pmatrix} 0 \\ x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad , \quad A \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and

$$A \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ x_1 \\ x_1 \end{pmatrix} \quad , \quad A \begin{pmatrix} x_1 \\ 0 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ x_1 \\ x_1 \end{pmatrix} \quad , \quad A \begin{pmatrix} 0 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The matrix is not irreducible as the are three cases in which the image is a subspace of the coordinate subspace.

Property 2 *A cannot be transformed into block upper triangular form by a permutation matrix P, that is*

$$PAP^{-1} \quad \textit{cannot be of the form} \quad \begin{pmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{pmatrix}$$

where B_{11} and B_{22} are square matrices of order greater or equal to 1.

Examples. The first matrix of the previous example is clearly already block upper triangular.

The second matrix of the previous example cannot be put in block upper triangular form because the only non-trivial permutation matrix in $\mathbb{R}^2$ is the matrix itself and $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

In $\mathbb{R}^3$ the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

is not irreducible. In fact, taking the permutation matrix

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

we have

$$PAP^T = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

and the result is block upper triangular.

Properties 1 and 2 are characterizations of an irreducible general matrix. For non-negative matrices other properties can be proved. The following is a typical algebraic property, that may have though a geometric interpretation on a corresponding graph. We shall see in the next sections what is the geometric version of the power of a matrix.

Property 3 *For every pair of indices i and j there exists a natural number m such that $\left(A^m\right)_{ij}$ is strictly positive.*

Remark. Going back to a previous example, the matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

is not irreducible as it is easy to see that the $(2,1)$ element of every power of the matrix is zero.

Conversely, the matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is irreducible as

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

and then if $i \neq j$ we have $\left(A^m\right)_{ij} > 0$ with $m = 1$ and if $i = j$ we have $\left(A^m\right)_{ij} > 0$ with $m = 2$.

We remark that property 3 does not mean of course that there is a power of A that is a positive matrix.

Property 4 *The $n \times n$ non-negative matrix A can be associated to a direct graph G, with n nodes, having an edge from node i to node j if and only if $a_{ij} > 0$. The matrix is irreducible if and only if the graph G is strongly connected.*

Example. Let's consider again the example, already seen before, in Figure 6.

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

with associated graph

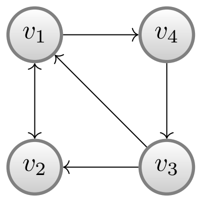


Figure 6: Adjacency and connectedness

The graph is strongly connected and A is irreducible. We may use Property 1: it is not difficult to see that for each coordinate subspace we have in $\mathbb{R}^4$, none of them is invariant.

If we consider instead the example in Figure 7 it easy to see that the graph is not strongly connected as for example there is no way to go from v_2 to v_1. The adjacency matrix is not irreducible, as for example

$$A \begin{pmatrix} x_1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

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

with associated graph

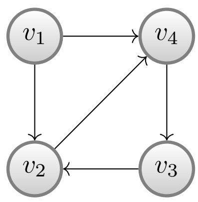


Figure 7: Adjacency and connectedness

Sometimes it is easier to decide if a matrix is irreducible by giving a look at the corresponding graph, but if the graph is very large and complicated it is reasonable to think it may be easier to study the algebraic aspects of the matrix.

Before we go into the details of the thorem in the Frobenius form, the concept of *period* of a matrix is important.

Let A be a non-negative square matrix.

Definition 11 *Given an index i, the* period of index i *is the greatest common divisor of all natural numbers m such that $\left(A^m\right)_{ii} > 0$.*

Example. Let's take the permutation matrix we've used before

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

By observing that

$$A^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = A^T \quad \text{and} \quad A^3 = A \cdot A^2 = A \cdot A^T = I_{3\times 3},$$

we may conclude that for each index $i = 1, 2, 3$ the period of index i is 3.

When A is irreducible the period of every index is the same and is called the *period of the matrix A*. In the previous example the matrix is not irreducible, but the indexes are the same however. The matrix

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

is irreducible (its graph is strongly connected) and we have

$$A^2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad , \quad A^3 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad , \quad A^4 = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \quad , \quad A^5 = \begin{pmatrix} 2 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

It means that the period is 1.

Definition 12 *The matrix A is* aperiodic *if its period is one.*

There can be of course an interpretation of the period of a matrix in terms of the properties of the corresponding graph. An interesting characterization exists involving irreducible, aperiodic non-negative matrices. We need another definition first.

Definition 13 *A matrix A is* primitive *if it is non-negative and its m-th power is positive for some natural number m.*[5]

The characterization is the following:

Theorem 2 *A matrix A is primitive if and only if A is non-negative, irreducible and aperiodic.*

The last example shows a primitive matrix.

Here is the Perron-Frobenius theorem for non-negative irreducible matrices.[6]

[5]Sometimes primitive matrices are defined as non-negative matrices having only one eigenvalue $r = \rho(A)$ in the spectral circle. It means of course that it is real and simple. In this way our definition can be converted in a theorem: a non-negative matrix A is primitive if and only if there exists $m > 0$ such that A^m is positive.

[6]There are actually some other possible statements associated to the Perron-Frobenius theorem that we do not mention.

Theorem 3 (Perron-Frobenius for non-negative irreducible matrices) *Let A be a non-negative irreducible $n \times n$ matrix with period p and spectral radius $\rho(A) = r$. Then the following statements hold.*

(i) r is positive and it is an eigenvalue of the matrix A, called the Perron-Frobenius eigenvalue.

(ii) r is simple. Both right and left eigenspaces associated with r are one-dimensional.

(iii) A has a left eigenvector v and a right eigenvector w associated with r, whose components are positive for both and the only eigenvectors with all positive components are the ones associated with r.

(iv) The matrix A has exactly p (the period) complex eigenvalues with module r. Each of them is a simple root of the characteristic polynomial and is the product of r with a p-th complex root of the unity.

(v) $\lim\limits_{k\to+\infty} \left(\frac{A}{r}\right)^k = vw^T$, where the right and left eigenvectors are normalized, so that $w^T v = 1$. Moreover, the matrix vw^T is the projection onto the eigenspace V_r, the so called Perron projection.

(vi) The Perron-Frobenius eigenvalue r satisfies the inequalities

$$\min_i \sum_j a_{ij} \le r \le \max_i \sum_j a_{ij}.$$

To conclude this section we mention another result regarding primitive matrices.

Theorem 4 *If A in non-negative and irreducible, with $r = \rho(A)$, A is primitive if and only if*

$$\lim_{k\to+\infty} \left(\frac{A}{r}\right)^k \quad \textit{exists.}$$

In this case

$$\lim_{k\to+\infty} \left(\frac{A}{r}\right)^k = \frac{pq^T}{q^T p},$$

where p and q^T are respectively a right and a left Perron-Frobenius vectors.

3 Some elements from Markov chains theory

For the use of the Markov chain model in the PageRank Algorithm a very simple approach to Markov chains is sufficient. So we will skip the general definition that involves the probabilistic concept of a general random process. We may just think of n possible states of a system, for which the times $t = 1, 2, 3, \ldots$ are relevant. At each time the system can be in one (and only one) of its states.

To give just a bit of formalization, we may think of the random variables X_t, $t = 1, 2, 3, \ldots$, where each one may take its values in the set of states $E = \{1, 2, \ldots, n\}$. We say that at time t the system is in its k-th state if the random variable X_t assumes the value k. Of course this will happen with probability $\mathrm{Prob}(X_t = k)$.

Our system is a Markov chain if the following property holds: for each $t = 1, 2, 3, \ldots$, for each pair of states i, j and for every sequence of states $\nu_0, \nu_1, \ldots, \nu_{t-1} \in E$

$$\mathrm{Prob}\big(X_{t+1} = j \bigm| X_0 = \nu_0\,,\, X_1 = \nu_1\,,\, \ldots\,,\, X_{t-1} = \nu_{t-1}\,,\, X_t = i\big) = \mathrm{Prob}\big(X_{t+1} = j \bigm| X_t = i\big).$$

The meaning is: the probability the system goes from state i to state j at time t is independent on the states the system was in before time t.

For example, taking $t = 1$,

$$\text{Prob}\big(X_2 = j \mid X_0 = \nu_0\,,\, X_1 = i\big) = \text{Prob}\big(X_2 = j \mid X_t = i\big),$$

that is the transition at time $t = 1$ does not depend on the initial state ν_0, but depends only on the state i at that time. The conditional probability

$$\text{Prob}\big(X_{t+1} = j \mid X_t = i\big) = p_{ij}(t)$$

is called the probability of transition at time t from state i to state j. Of course this probability may or may not depend on t. The Markov chain is said to be *homogeneous* or *time stationary* if the probability of transition does not depend on t, namely

$$p_{ij}(t) = p_{ij} \quad \text{for each } t.$$

It is clear that we can associate both a matrix (independent on t) and a graph to a stationary Markov chain. The matrix $P = (p_{ij})$ is called the *transition matrix* of the chain and it has the following properties:

1. $p_{ij} \geq 0$ for each i, j;

2. $\sum_{j=1}^{n} p_{ij} = 1$ for each i.

A matrix with properties 1 and 2 is called a *stochastic matrix.*

Remark. Property 2 may be written in matrix form as $Pu = u$, where $u = (1, \ldots, 1)^T$. This means that any stochastic matrix has the eigenvalue 1 and u is a corresponding eigenvector. It can be proved that stochastic matrices have spectral radius 1.

It is quite natural to associate the transition matrix P of a Markov chain to a direct graph. In fact P is a sort of adjacency matrix, with an extra information: it tells not just which transitions are possible but also what is the probability of any single transition.

Given the square transition matrix $P = (p_{ij})$ of order n, the associated graph has n nodes $\{v_1, \ldots, v_n\}$ and has the edge (v_i, v_j) if and only if $p_{ij} > 0$. It is what is called a weighted graph, a graph in which each edge is a positively weighted edge, weights being probabilities in this case. Of course a state j is reachable from a state i if and only if there exists a direct path from v_i to v_j in the graph.

Definition 14 *A Markov chain is irreducible if its transition matrix P is irreducible.*

Remark. Among the various characterizations of irreducible matrices we have seen before it means that the graph is strongly connected, namely for any two nodes v_i and v_j there is a direct path that goes from v_i to v_j.

Definition 15 *A Markov chain is aperiodic if it is irreducible and its transition matrix P is primitive.*

Let's call *probability vector* any non-negative row vector $p^T = (p_1, \ldots, p_n)$ such that $\sum_{i=1}^{n} p_i = 1$.

Definition 16 *A probability vector p is a* stationary probability vector *for a Markov chain with transition matrix P if*

$$p^T P = p^T.$$

Remark. It comes out here the importance for P to be irreducible. In fact from the Perron-Frobenius theorem in that case the existence and uniqueness of a stationary probability vector is guaranteed. This is a crucial aspect in the PageRank algorithm.

It is clear that the probability vectors are the right theoretical ways to describe the system as time goes on. We have just to formalize the way the probability vectors change in time.

Suppose we have an initial probability vector

$$p^T(0) = (p_1(0), \dots, p_n(0)).$$

This vector gives the probabilities the system is in the state $1, \dots, n$ at $t = 0$. In the same way

$$p^T(t) = (p_1(t), \dots, p_n(t))$$

will be the probability vector at time t: it gives the probabilities the system is in the state $1, \dots, n$ at time t.

Now, given the probabilities at time t, let's compute the probability that the system is in the state j at time $t+1$.

$$\begin{aligned}
p_j(t+i) = \text{Prob}(X_{t+1} = j) &= \text{Prob}\Big(\big(X_{t+1} = j\big) \cap \big(X_t = 1 \cup \ldots \cup X_t = n\big)\Big) \\
&= \text{Prob}\Big(\big(X_{t+1} = j \cap X_t = 1\big) \cup \ldots \cup \big(X_{t+1} = j \cap X_t = n\big)\Big) \\
&= \sum_{i=1}^{n} \text{Prob}\big(X_{t+1} = j \cap X_t = i\big) \\
&= \sum_{i=1}^{n} \text{Prob}\big(X_{t+1} = j \bigm| X_t = i\big) \cdot \text{Prob}\big(X_t = i\big) \\
&= \sum_{i=1}^{n} p_{ij} \cdot p_i(t).
\end{aligned}$$

It means that the j-th component of the row vector $p^T(t+1)$ is given by the inner product of the $p^T(t)$ vector itself by the j-th column of the transition matrix P. In matrix-vector form

$$p^T(t+1) = p^T(t)P.$$

It means that for any time $t = 1, 2, \ldots$ we may write

$$p^T(t) = p^T(0)P^t.$$

As we shall see in the next subsection this is the characteristic main step of the power method.

4 The PageRank algorithm

The main step of Google's PageRank algorithm is the iteration of the method we describe in the following subsection.

4.1 The power method

For the purpose of finding the dominant eigenvalue of a matrix together with a corresponding eigenvector a great variety of methods have been designed. This method is a fundamental step in the Google's PageRank algorithm.

Let's consider a square $n \times n$ matrix A having n linearly independent eigenvectors associated to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and suppose

$$|\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_n|.$$

This means we are assuming that the dominant eigenvalue λ_1 has algebraic multiplicity one. Let's call $v^1, v^2, \ldots, v^n$ the corresponding eigenvectors. This means

$$Av^i = \lambda_i v^i, \quad i = 1, 2, \ldots, n.$$

The power method, starting from any arbitrarily chosen vector $x^{(0)}$, builds up a sequence of vectors $\{x^{(k)}\}$ that converges to the eigenvector associated to the dominant eigenvalue.

Suppose $x^{(0)}$ is an arbitrary vector in $\mathbb{R}^n$. We may write $x^{(0)}$ as a linear combination of $v^1, v^2, \ldots, v^n$ that, because the hypothesis of linear independence, is a basis of $\mathbb{R}^n$.

$$x^{(0)} = \sum_{i=1}^{n} \alpha_i v^i \quad \text{and suppose } \alpha_1 \neq 0.^{7}$$

Starting from $x^{(0)}$ we may build up the sequence

$$x^{(1)} = Ax^{(0)} \quad , \quad x^{(2)} = Ax^{(1)} \quad , \quad \ldots \quad , \quad x^{(k)} = Ax^{(k-1)} \quad , \quad \ldots$$

The following result holds:

Theorem 5 *For the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ we have that*

$$\lim_{k\to\infty} \frac{x_j^{(k+1)}}{x_j^{(k)}} = \lambda_1 \quad \textit{and} \quad \lim_{k\to\infty} \frac{x^{(k)}}{x_j^{(k)}} = cv^1. \tag{1}$$

where j is an index for which $x_j^{(k)} \neq 0$, for every value of k.

Proof 1 *From the definition of the sequence $\{x^{(k)}\}$ we have*

$$x^{(k)} = Ax^{(k-1)} = A^2Ax^{(k-2)} = \ldots = A^k x^{(0)} = A^k \sum_{i=1}^{n} \alpha_i v^i = \sum_{i=1}^{n} \alpha_i A^k v^i = \sum_{i=1}^{n} \alpha_i \lambda_i^k v^i.$$

By collecting λ_1^k we get

$$x^{(k)} = \lambda_1^k \left(\alpha_1 v^1 + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^k v^i \right)$$

and then

$$x^{(k+1)} = \lambda_1^{k+1} \left(\alpha_1 v^1 + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^{k+1} v^i \right).$$

Then for the indices j for which $x_j^{(k)} \neq 0$ and $v_j^1 \neq 0$ we may write

$$\frac{x_j^{(k+1)}}{x_j^{(k)}} = \lambda_1 \frac{\alpha_1 v_j^1 + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^{k+1} v_j^i}{\alpha_1 v_j^1 + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^{k} v_j^i}. \tag{2}$$

As $\left|\frac{\lambda_i}{\lambda_1}\right| < 1$ for $2 \leq i \leq n$, we have

$$\lim_{k\to\infty} \frac{x_j^{(k+1)}}{x_j^{(k)}} = \lambda_1.$$

[7] The condition means that we don't have to start from a point in the subspace spanned by the eigenvectors $v^2, v^3, \ldots, v^n$. We need $x^{(0)}$ to have a component in the subspace spanned by v^1.

Let's consider now the sequence of vectors $\left\{\frac{x^{(k)}}{x_j^{(k)}}\right\}$*, taking again, for each value of* k*, an* $x_j^{(k)}$ *component that is non zero. Then*

$$\frac{x^{(k+1)}}{x_j^{(k)}} = \lambda_1 \frac{\alpha_1 v^1 + \sum_{i=2}^n \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^{k+1} v^i}{\alpha_1 v_j^1 + \sum_{i=2}^n \alpha_i \left(\frac{\lambda_i}{\lambda_1}\right)^k v_j^i}$$

and again, taking the limit for $k \to \infty$*, we get*

$$\lim_{k\to\infty} \frac{x^{(k+1)}}{x_j^{(k)}} = \frac{v^1}{v_j^1},$$

that is a "normalization" of the eigenvector v^1*.*

Remark. The eigenvalue can be obtained also as the limit of a different sequence, namely as the

$$\lim_{k\to\infty} \frac{x^{(k)^T} A x^{(k)}}{x^{(k)^T} x^{(k)}},$$

that is the limit of the Rayleigh quotients of the sequence $x^{(k)}$.

Remark. It is worthwhile to specify that in the practical implementations of the method some numerical problems are likely to arise. The method, if implemented as it has been presented, gives overflow/underflow problems. For this reason at each step it is convenient to normalize the vector $x^{(k)}$. The properties of the convergence are not modified and we prevent the norms becoming too large.

4.2 The PageRank algorithm

We give an essential description of the algorithm, while we leave to a later work a further analysis and some simulations.

Google is one of the most important search engines available on the web. Essentially the engine, in response to a question proposed by the user, replies with a classification in order of some importance of the web pages related to the query itself. As we all know, the answer is almost immediate.

The heart of the implementation of the mechanism is the algorithm called PageRank. It is based on the concept of "link popularity": a certain web page is important if, in addition to receive links from other important pages, has a limited number of links to other pages.

A formal representation of the concept is given by the following formula. P and Q are pages of the web and we indicate with $r(P)$ and $r(Q)$ the ranks of the pages, namely their importance.

$$r(P) = \sum_{Q\to P} \frac{r(Q)}{|Q|}, \tag{3}$$

where $|Q|$ is the number of external links of the page Q and the $Q \to P$ means that the summation is extended over the pages Q that have a link to the page P.

Remark. Of course the meaning is that for the ranking of P just the pages linked to P have relevance and the importance of these pages is reduced by the total number of links these pages have. The fewer external links a page has the better it is for the ranking of P.

We can give a matrix form to the equation. Let's $\{P_1, P_2, \ldots, P_n\}$ be the n pages on the web. We may define a transition matrix A, namely the matrix of the probabilities of a transition from a page to another page, in the following way:

$$a_{ij} = \text{Prob}(P_j \to P_i) = \begin{cases} \frac{1}{|P_j|} & \text{if } P_j \to P_i \\ 0 & \text{otherwise.} \end{cases}$$

Of course this setting assumes that if P_j is linked to P_i, the transition from P_j to P_i is just one of the $|P_j|$ possible transitions, each one with the same probability.

The reason why the dominant eigenvalue and its corresponding eigenvector are important in the PageRank algorithm is going to appear now. Suppose we collect the ranks of all the web pages in a vector r and suppose we want to compute it in an iterative way, starting from a previous evaluation of r itself. At the beginning the most reasonable setting, or maybe the only possible one, is to assume that all the pages have the same rank and, as probabilities are implicitly involved, $r = r^{(0)}$ is the "uniform distribution vector"

$$r^{(0)} = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)^T. \text{ [8]}$$

Let's rewrite equation (3) in terms of the n pages $P_1, P_2, \ldots, P_n$ of the web, and in particular for the i-th page P_i, after the first transition:

$$r_i^{(1)}(P) = \sum_{P_j \to P_i} \frac{r_j^{(0)}}{|P_j|} = \sum_{P_j \to P_i} \frac{1}{|P_j|} \cdot r_j^{(0)} = \sum_{j=1}^{n} a_{ij} \cdot r_j^{(0)} \quad i = 1, 2, \ldots, n.$$

In matrix/vector form the previous equation may be written as

$$r^{(1)} = Ar^{(0)}.$$

Clearly an iterative sequence is then defined, namely

$$r^{(k+1)} = Ar^{(k)} \quad k = 0, 1, \ldots$$

and this is the general iteration of the power method. We have seen that the limit of the sequence is an eigenvector associated to the dominant eigenvalue. The ranks of the web pages are the components of this limit vector, conveniently normalized.

Remark. A couple of immediate remarks. Firstly it appears evident the extreme heaviness of the computations. The number of the web pages exceeds a couple of billions and this is the length of the rank vector r and the order of the matrix A.

A more theoretical aspect is that the existence and the properties of the dominant eigenvalue depend on some properties of the matrix A, and for the moment we do not know if the transition matrix has those properties. We shall go into these details in a later work.

References

- L. R. Foulds, Graph Theory Applications, Sringer-Verlag, 1992
- G. H. Golub, C.F. Van Loan, Matrix Computations, The Johns Hopkins University Press, 1996
- J. R. Norris, Markov chains, Cambridge University Press, 1997
- PageRank, at `https://it.wikipedia.org/wiki/PageRank` and related pages

[8] Transposition is required in the next vector/matrix notations.