Silesian J. Pure Appl. Math.
vol. 6, is. 1 (2016), 137–153

Katarzyna ADRIANOWICZ, Iwona NOWAK

Institute of Mathematics, Silesian University of Technology, Gliwice, Poland

MATHEMATICAL METHODS IN ALGORITHM FOR WEBSITES POSITIONING

Abstract. Nowadays it is more and more common to treat the Internet as one of the first sources of information. Given key words, different types of web search engines generate a list of websites ranked by priority (theoretically corresponding to the query). The page position on the list depends on many factors. The method presented herein is a version of a *PageRank* algorithm introduced by Google to designate one of them. The *PageRank* algorithm ranks a webpage, depending on the number and quality of links leading to it and thus determines its position on the list. In its simplest version, the method can operate using just the basic operations on matrices. This paper presents also the more advanced version based on probabilistic approach.

1. Introduction

It is now much easier and faster than ever before to acquire information. Increasingly, the Internet is regarded as the first (unfortunately sometimes the only) source of information. We usually use different types of search engines for searching the Web. After typing chosen keywords in the search box, we get in response a list of websites ranked by priority (theoretically corresponding to the query).

The site position on the list depends on many factors. Most search engines determine the position of the page on the basis of over 200 parameters associated

2010 Mathematics Subject Classification: 15A18, 60J20.
Keywords: PageRank algorithm, linear algebra application, eigenvectors, Markov chain.
Corresponding author: I. Nowak (Iwona.Nowak@polsl.pl).
Received: 27.07.2016.

not only with the content placed on it but also taking into account its popularity on the Internet [1, 2, 4].

The method presented in this article is a simplified version of *PageRank* algorithm, introduced by Google to designate a single factor used to determine the page rank on the basis of its interaction with other network resources.

The *PageRank* algorithm assigns each web site its rank, depending on the number and quality of links leading to that page.

The idea of the algorithm in terms of linear algebra is presented at the beginning.

2. Linear algebra point of view

The approach presented here does not require advanced mathematical knowledge as it is based almost exclusively on the basic operations on matrices.

Example 1[1]

Suppose that the network consists of just four websites referencing each other in the manner shown in the Figure 1.

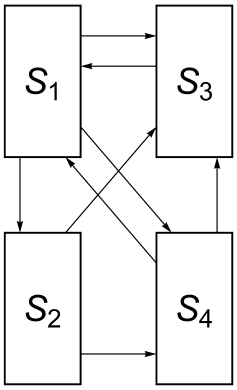

Fig. 1. A simply internet network

The arrows between pages indicate the connections (links) between the websites. Let's denote rank (as the "validity") of the website S_j by r_j.

[1]The example comes from [3].

In the simplest case the rank of the page may be dependent on the number of links directed to it. It will be true under the assumption that the quality of the page is directly proportional to the number of sites referring to it[2]. In the considered network, it means that $r_3 = 3$, $r_1 = r_4 = 2$ and $r_2 = 1$.

Such order is not quite correct. Intuition tells us that a page with single connection from a popular website should be more important than a page with several incoming links from rarely visited www.

It seems much more reasonable to take into account not only the number of links directing to the website but also the rank of pages on which those links were placed.

The above remark will be taken into account if the page rank is determined according to the formula:

$$r_k = \sum_{j \in \mathcal{L}_k} \frac{r_j}{n_j}, \tag{1}$$

where $\mathcal{L}_k$ is a set of pages which are pointed to S_k and n_j is the total number[3] of outgoing links from the page Sj. According to the previous agreement r_j indicates the rank of the page S_j.

In the network presented on Figure 1, the site S_1 is linked from the pages S_3 and S_4. Thus the formula (1) leads to the following dependence:

$$r_1 = \frac{r_3}{1} + \frac{r_4}{2}.$$

Similarly:

$$r_2 = \frac{r_1}{3},$$
$$r_3 = \frac{r_1}{3} + \frac{r_2}{2} + \frac{r_4}{2},$$
$$r_4 = \frac{r_1}{3} + \frac{r_2}{2}.$$

All equations built in such a way form a system of linear equations, which can be written in a matrix form as:

$$\mathbf{r} = \mathbf{P}\mathbf{r}, \tag{2}$$

[2]It is similar to the number of citations.

[3]Note that always $n_j > 0$ because there exists at least one link from S_j to k.

where $\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix}$, while $\mathbf{P} = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}$.

The equation (2) implies that the vector $\mathbf{r}$ is the eigenvector of matrix $\mathbf{P}$ corresponding to the eigenvalue $\lambda = 1$.

So, every eigenvector ($\mathbf{r} \neq \mathbf{0}$), which satisfies the equation:

$$(\mathbf{P} - \mathbf{I})\mathbf{r} = \mathbf{0}. \tag{3}$$

is the solution of equation (2).

The construction of matrix $\mathbf{P}$ causes that it is always a left stochastic matrix i.e. a non-negative matrix in which sum of elements in each column equals 1. It is easy to show that every left stochastic matrix has an eigenvalue equal 1. Thus there is a non-trivial solution of (3) what implies that there are infinitely many solutions (dependent on parameter t):

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} 2t \\ \frac{2}{3}t \\ \frac{3}{2}t \\ t \end{bmatrix}.$$

It is convenient to treat the eigenvector components, which sum up to 1 ($t = \frac{6}{31}$), as a standardized measure of the "importance" of the relevant pages. For the matrix $\mathbf{P}$ in the example considered, the corresponding eigenvector components are:

$$r_1 \approx 0.387, \quad r_2 \approx 0.129, \quad r_3 \approx 0.29 \quad \text{and} \quad r_4 \approx 0.194, \tag{4}$$

what means that in the network shown in Figure 1, the webpage order is S_1, S_3, S_4, S_2.

3. Probabilistic point of view

Another possible way of PageRank determination is the probabilistic approach. In this case, a page rank will reflect the probability that a (slightly confused) surfer, following hyperlinked websites randomly, will eventually land on the testing page.

The more incoming links to the page, the higher its rank. The rank will also be bigger if the page gets links from important or frequently visited websites.

Let a set of $S_1, S_2, \ldots S_N$ create a simple internet network. Imagine that the surfer begins to search it from a random page and moves around the network using the so-called "random walk": in each successive step, he chooses randomly the link from his current position and follows it to the next page. It is assumed that surfer continues his random decision process indefinitely.

To create a formal mathematical model we introduce a sequence of random variables $X_n,\ n = 1, 2, \ldots$ such that:

X_0 – means a website at which the surfer starts his walk,

$X_k,\ k = 1, 2, \ldots$ – means a website that the surfer landed on after the k-th step of his random walk.

$$X_n \in \{S_1, S_2, \ldots S_N\} \text{ for all } n = 0, 1, 2, \ldots$$

PageRank of S_i is defined as the limit of probabilities that the random variable X_n will achieve the value S_i:

$$r_i = \lim_{n\to\infty} P(X_n = S_i) \tag{5}$$

and means the probability, that an infinite random walk will "end" on page S_i.

Note that the sequence of random variables defined by the random walk of the surfer creates a Markov chain. The value of a random variable X_{n+1}, i.e. the page that the surfer would reach after $(n+1)$ steps, depends entirely on the page where the surfer is in his n-th step (i.e. the value of X_n). It means that it does not depend on what has happened previously. Thus that the probability of transition from one page onto another in a single step depends only on these sites and is not dependent on the surfing history or the step number. It has no memory of the past.

If the p_{ij} denote the probability that the surfer will go from S_j to the S_i, the transitive matrix:

$$\mathbf{P} = [p_{ij}] = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ \vdots & & & \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix} \tag{6}$$

has only nonnegative entries and the sum of elements in each column equals 1. So it is a left stochastic matrix, matrix of Markov chain process called "Markov transition matrix".

It can be assumed[4] that each page $S_1, S_2, \ldots S_N$ in the internet network can be selected by the surfer as the first page of his random walk with the same probability equaling to $\frac{1}{N}$. Thus, the random variable X_0 has discrete uniform distribution. That is:

$$p_i(0) = \frac{1}{N} \quad \text{for all } i = 1, \ldots, N,$$

which can be written in vector notation as:

$$\mathbf{p}(0) = \begin{bmatrix} \frac{1}{N} \\ \vdots \\ \frac{1}{N} \end{bmatrix}.$$

Therefore the vector $\mathbf{p}(0)$ represents X_0 distribution.

Let us denote distribution of X_1 by the vector $\mathbf{p}(1) = \begin{bmatrix} p_1(1) \\ p_2(1) \\ \vdots \\ p_N(1) \end{bmatrix}$.

The values of its components could be calculated from the formula for the total probability:

$$p_i(1) = P(X_1 = S_i) = \sum_{k=1}^{N} P(X_1 = S_i | X_0 = S_k) P(X_0 = S_k) = \\ = \sum_{k=1}^{N} p_{ik} p_k(0) \qquad \text{for all } i = 1, 2, \ldots, N,$$

where p_{ik} means the probability of transition to S_i from S_k. So, these are entries of the transitive matrix $\mathbf{P} = [p_{ij}]$.

In matrix notation this relationship has the following brief form:

$$\mathbf{p}(1) = \mathbf{P} \cdot \mathbf{p}(0).$$

[4]Actually (what is shown in the example), the distribution of X_0 can be arbitrary and it has no effect on the final result.

The distribution of X_2, given by the vector $\mathbf{p}(2)$, has the following components:

$$p_i(2) = P(X_2 = S_i) = \sum_{k=1}^{N} P(X_2 = S_i | X_1 = S_k) P(X_1 = S_k) = \\ = \sum_{k=1}^{N} p_{ik} p_k(1) \qquad \text{for all } i = 1, 2, \ldots, N,$$

What in matrix notation produces:

$$\mathbf{p}(2) = \mathbf{P} \cdot \mathbf{p}(1) = \mathbf{P}(\mathbf{P} \cdot \mathbf{p}(0)) = \mathbf{P}^2 \mathbf{p}(0).$$

Similarly, it could be shown that for any $n = 1, 2, \ldots$ the distribution vector for variable X_n is:

$$\mathbf{p}(n) = \mathbf{P} \cdot \mathbf{p}(n-1) = \mathbf{P}^n \cdot \mathbf{p}(0). \tag{7}$$

Determination of all web pages ranking on the basis of (5) means finding the PageRank vector $\mathbf{r}$ whose components r_i for $i = 1, 2, \ldots, N$ are:

$$r_i = \lim_{n \to \infty} P(X_n = S_i),$$

that is, according to procedure introduced above

$$r_i = \lim_{n \to \infty} p_i(n),$$

or in vector notation: $\mathbf{r} = \lim_{n \to \infty} \mathbf{p}(n)$.

If n tends to infinity the formula (7) becomes:

$$\mathbf{r} = \lim_{n \to \infty} \mathbf{p}(n) = \lim_{n \to \infty} \mathbf{P} \cdot \mathbf{p}(n-1) = \mathbf{P} \cdot \lim_{n \to \infty} \mathbf{p}(n-1),$$

and so

$$\mathbf{r} = \mathbf{P}\mathbf{r}. \tag{8}$$

Note that this equality is analogous to the relationship (2), which appeared in a model based on an algebraic approach.

Let's examine the model of "random walk" for a very simple network, analysed previously in Example 1.

Example 1a

Consider the Internet network of websites connected by links as shown in Figure 1. At the beginning, the surfer randomly selects the home (starting) page. It may be one of the four sites S_1, S_2, S_3 or S_4. Therefore, the probability that the web surfer will begin his walk from site S_j is as follows:

$$p_j(0) = 0.25 \quad \text{for all } j = 1, 2, 3, 4.$$

We assume that the choice of every link from the S_j site to the next page has the same probability. Therefore, the probability p_{ij} denoting that the surfer goes from S_j to S_i depends only on the number of outgoing from S_j links and is:

$$p_{ij} = \frac{1}{\text{the number of outgoing links from } S_j}.$$

For example, in the considered network, $p_{21} = p_{31} = p_{41} = \frac{1}{3}$ because the website S_1 has 3 outcoming links to the S_2 or S_3 or S_4.

It is easy to build all probabilities: p_{ij} for all $i, j = 1, 2, 3, 4$ and use them to build a transition matrix[5] $\mathbf{P}$:

$$\mathbf{P} = [p_{ij}] = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}. \tag{9}$$

All likelihoods determined for the network are additionally shown in Figure 2.

The probability $p_i(1)$, that after the first step the web surfer will be on the website S_i can be calculated using the formula for total probability. For this purpose, we propose to use the tree diagram (the graphical form of the total probability formula) presented in Figure 3, with branches correspond to the respective probabilities.

[5]Matrix $\mathbf{P}$ is the same as in the formula (2).

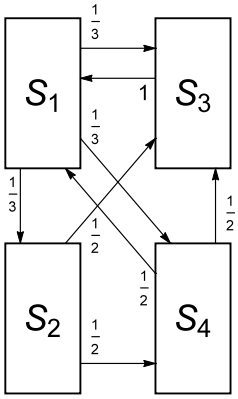

Fig. 2. The transition probabilities

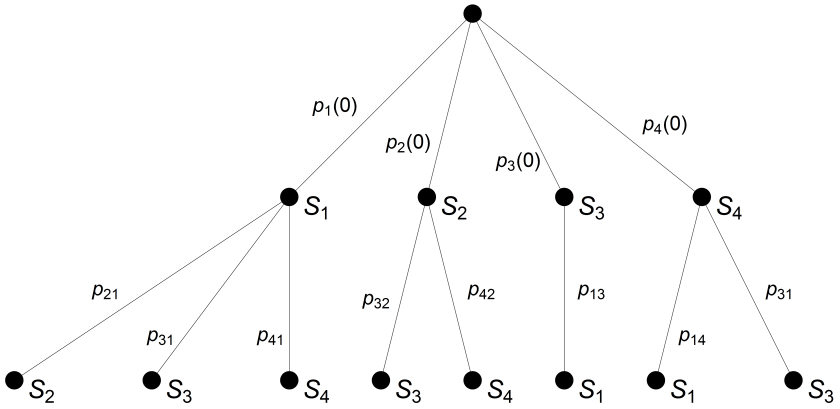

Fig. 3. Total probability calculation tree

The first level of branches (from the top) illustrates the selection of a home page, while the second one illustrates the transition in every single step from one website onto another[6]. Thus,

$$p_i(1) = \sum_{k=1}^{4} p_{ik} p_k(0) \quad \text{for all } i = 1, 2, 3, 4$$

or in matrix notation

$$\mathbf{p}(1) = \mathbf{P} \cdot \mathbf{p}(0).$$

[6]In order to calculate an appropriate probability, the numbers of stacked along one branch are multiplied, and the individual branches needed for the calculation are added.

In the next step, the probabilities of transition to the side S_i from S_j are the same as those presented in the matrix $\mathbf{P}$. Selection of a new landing page depends only on the number of outbound links from the current page and does not depend on earlier choices of the surfer. Therefore, according to (7), the vector of probability distributions in the n-th step has following form:

$$\mathbf{p}(n) = \mathbf{P}^n \cdot \mathbf{p}(0).$$

According to the above formula the probability of landing on a particular site depends on the power of the matrix $\mathbf{P}$ and the starting vector $\mathbf{p}(0)$. Let's consider selected powers of transition matrix $\mathbf{P}$:[7]

$$\mathbf{P}^2 = \begin{bmatrix} 0.5 & 0.75 & 0 & 0.5 \\ 0 & 0 & 0.333333 & 0.166667 \\ 0.333333 & 0.25 & 0.33333 & 0.166667 \\ 0.166667 & 0 & 0.33333 & 0.166667 \end{bmatrix},$$

$$\mathbf{P}^{10} = \begin{bmatrix} 0.387153 & 0.389757 & 0.384838 & 0.3886 \\ 0.128279 & 0.127315 & 0.130787 & 0.129051 \\ 0.290895 & 0.290509 & 0.289931 & 0.28941 \\ 0.193673 & 0.192419 & 0.19444 & 0.192708 \end{bmatrix},$$

$$\mathbf{P}^{25} = \begin{bmatrix} 0.387097 & 0.387097 & 0.387097 & 0.387097 \\ 0.129032 & 0.129032 & 0.129032 & 0.129032 \\ 0.290323 & 0.290323 & 0.290323 & 0.290323 \\ 0.193548 & 0.193548 & 0.193548 & 0.193548 \end{bmatrix}.$$

It can be observed that with the increase of n, the matrix $\mathbf{P}$ stabilizes. Not only do we observe the values of matrix elements changing less and less between successive powers, but also the elements in the columns are getting more and more similar to each other. In the matrix P^{25} the columns are identical with accuracy to the sixth decimal places.

From the considerations above the following conclusions can be extended:

1) The probability of reaching the page S_i after appropriate number of steps (in this example, after at least 25) does not depend on the place where the surfer began to walk.

[7]Calculations were performed using *Mathematica*.

2) If we find n for which the matrix $\mathbf{P}^n$ has the same (with the assumed accuracy) columns, the approximate PageRank vector $\mathbf{r}$, specifying the rank of sites, is determined by the (any) column of this matrix.

In this example, it can be assumed that:

$$\mathbf{r} \approx \begin{bmatrix} 0.387097 \\ 0.129032 \\ 0.290323 \\ 0.193548 \end{bmatrix},$$

what gives the following order of the websites in the network: S_1, S_3, S_4, S_2[8].

3) Finally, because at some point $\mathbf{P}^n \approx \mathbf{P}^{n+1}$, $\mathbf{p}(n) \approx \mathbf{p}(n+1) \approx \mathbf{r}$, the relationship (8) is obtained again. The example above can be considered as an illustration of the classic power method for the determination of the dominant eigenvector $\mathbf{r}$ matrix $\mathbf{P}$.

The conclusions presented above are valid and the method of page positioning works well, provided that process presented converges, i.e. that there exist $\lim_{n\to\infty} \mathbf{p}(n)$ or $\lim_{n\to\infty} \mathbf{P}^n$ (both conditions are equivalent). If it is not the case, the modification of the transitive matrix are necessary, so that the process used for modified matrix is convergent.

4. Modifications

It can be shown that if the matrix $\mathbf{P}$ is stochastic and it meets certain subtle conditions (detailed description of the necessary conditions can be found in [6]), there is one vector of PageRank $\mathbf{r}$ and the web pages rank is unequivocally determined. When there is no proper convergence and determination of the rank pages by these algorithms is not possible, relevant modifications need to be introduced. This is illustrated by the examples below.

Example 2

Consider a network with so-called "dangling nodes", i.e. sites with no ourgoing links. For example, in the network presented in Figure 4, page S_3 does not have any outgoing links.

[8]Note that the order of pages received here as well as the value of ranks are very similar to those obtained by Method 1 and shown in (4).

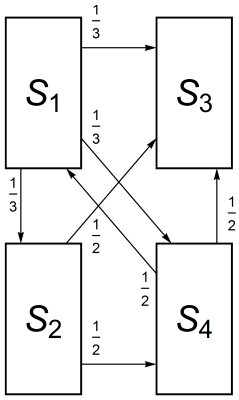

Fig. 4. Example of network with dandling nod (S_3)

In such a situation, the column consisting of all zeros appears in the transitive matrix $\mathbf{P}$ (for the considered network, it is the 3^{ed} column):

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

Obviously, this matrix is not a stochastic one and in its subsequent powers $\mathbf{P}^n$ zero column will always appear.

In this case the solution is quite simply, the matrix needs to be modified by just replacing with the zero column with the one consisting of all entries equal $\frac{1}{N}$ (here $\frac{1}{4}$). These values reflect the probability of the random (uniform) selection of the site in the considered network. The new transitive matrix:

$$\mathbf{P}' = \begin{bmatrix} 0 & 0 & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{4} & 0 \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{4} & 0 \end{bmatrix}$$

is now a left stochastic matrix and its subsequent powers fairly quickly lead to a matrix consisting of the same columns (with reasonable accuracy). In the matrix

$\mathbf{P'}^{15}$ the columns are identical with accuracy to the fifth decimal places:

$$\mathbf{P'}^{15} = \begin{bmatrix} 0.216495 & 0.216497 & 0.216495 & 0.216494 \\ 0.164947 & 0.164948 & 0.164948 & 0.164945 \\ 0.371135 & 0.371134 & 0.371134 & 0.371133 \\ 0.247423 & 0.247421 & 0.247423 & 0.247422 \end{bmatrix}.$$

It means we can assum the vector of PageRank $\mathbf{r} \approx \begin{bmatrix} 0.216495 \\ 0.164947 \\ 0.371135 \\ 0.247423 \end{bmatrix}$ and on this basis determine the order of pages according to their rank: S_3, S_4, S_1, S_2.

In some cases, a stochastic matrix could have more than one eigenvector associated with eigenvalue 1. This situation also causes that presented method not to be convergent. The example of the network in which such a problem occurs will be discussed below.

Example 3

Consider the network presented in Figure 5.

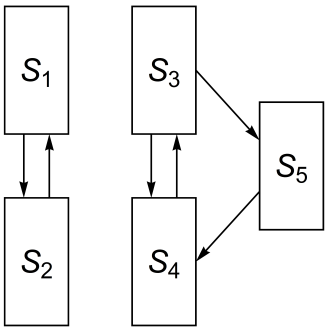

Fig. 5. A simple Internet network – illustration of Example 3

In this case, the transitive matrix has the following form

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 1 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

The construction of the first two rows and columns causes that for even powers $\mathbf{P}^n$ the first two rows and columns are always the same as in $\mathbf{P}^2$:

$$\mathbf{P}^2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 1 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0 \end{bmatrix},$$

while for an odd powers, the first two rows and columns always look like in $\mathbf{P}$.

So there is no chance of convergence, although $\mathbf{P}$ is a left stochastic matrix. In this simple example, it is easy to calculate that one of the eigenvalues of matrix $\mathbf{P}$ is $\lambda = 1$ but there are two lineary independent eigenvectors associated with:

$$\mathbf{r}^{(1)} = \begin{bmatrix} 0 \\ 0 \\ \frac{2}{5} \\ \frac{2}{5} \\ \frac{1}{5} \end{bmatrix} \qquad \text{oraz} \qquad \mathbf{r}^{(2)} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

That is why, it is not clear (not unequivocal), which vector should be chosen in order to determine the ranking of sites.

The reason of the situation shown in Example 3 is that the network in Figure 5 in fact consists of two unconnected subnets[9]. Of course, such a situation in reality occurs often.

In this case, the modification (described in details in [3]) of the presented method needs to be used.

[9]If the network W consists of r subnets $W_1, \ldots, W_r$, the number of linearly independent eigenvectors associated with an eigenvalue $\lambda = 1$ is $\geqslant r$.

In order to determine the validity of websites, the matrix of weighted averages is applied instead of matrix $\mathbf{P}$:

$$\mathbf{M} = (1 - m) \cdot \mathbf{P} + m \cdot \mathbf{S} \tag{10}$$

where $0 \leqslant m \leqslant 1$,[10] and $\mathbf{S}$ is a matrix of size $n \times n$, whose all elements are equal to $\frac{1}{n}$.

$\mathbf{S}$ is left stochastic matrix, therefore it has an eigenvalue 1 and, as can be easily checked, it has one eigenvector associated with it (which sums up to 1). Note that the matrix $\mathbf{M}$ is also a left stochastic one.

Paper [3] presents the proof that for $m \in (0, 1\rangle$ matrix $\mathbf{M}$ has always only one eigenvector associated with $\lambda = 1$.

Let's apply this modification to Example 3.

According to formula (10) and assuming $m = 0.15$:

$$\mathbf{M} = 0.85 \cdot \mathbf{A} + 0.15 \cdot \mathbf{S} = \begin{bmatrix} 0.03 & 0.88 & 0.03 & 0.03 & 0.03 \\ 0.88 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.88 & 0.03 \\ 0.03 & 0.03 & 0.455 & 0.03 & 0.88 \\ 0.03 & 0.03 & 0.455 & 0.03 & 0.03 \end{bmatrix}.$$

Now the eigenvector of matrix $\mathbf{M}$, related to the value $\lambda = 1$ and summing up to 1, needs to be determined. In this simple case, it can be done by solving the matrix equation:

$$\mathbf{M} \cdot \mathbf{r} = \mathbf{r}.$$

The search solution is:

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.232674 \\ 0.23844 \\ 0.128886 \end{bmatrix},$$

therefore, the order of pages in the network shown in Figure 5 is: S_4, S_3, S_1 and S_2, S_5.

[10] According to [5], Google was originally using $m = 0.15$.

5. Conclusions

The paper presents PageRank method, which was developed and introduced by Google firm as one of the tools for positioning the websites.

Two approaches, algebraic and probabilistic were discussed. Both of them are use to assess the level in which the website match the user's inquiry. The main goal of the work was to demonstrate how mathematical methods can be applied to build the so-called transitive matrix that models the surfer's behaviour in a single step of his Internet walk. This matrix is often a stochastic one. Subsequent decisions taken by the surfer create a sequence of states building a Markov chain. If this sequence is convergent, the sequence limit enables unequivocal determination of website ranking.

Due to the structure of the Internet network, the created sequence is often not convergent. Depending on the reasons behind non-convergence, a relatively modified transition matrix is applied for positioning.

The work also discussed the modifications required in case of dangling nodes and the networks consisting of disjoint sub-networks, with their operations illustrated with simple examples.

Acknowledgment

The author would like to thank to the anonymous reviewer for his valuable comments and suggestions.

References

1. Andersson F.K., Silvestrov S.D.: *The Mathematics of Internet Search Engines*. Acta Appl. Math. **104**, no. 2 (2008), 211–242.
2. Austin D.: *How Google Finds Your Needle in the Web's Haystack*. Online: http://www.ams.org/samplings/feature-column/fcarc-pagerank, 2016.
3. Bryan K., Leise T.: *The $25,000,000,000 eigenvector: The linear algebra behind Google*, SIAM Review **48**, no. 3 (2006), 569–581.
4. Haveliwala T., Jeh G., Kamvar S.: *An Analytical Comparison of Approaches to Personalizing PageRank*. Technical Report, Stanford University 2003.
5. Langville A., Meyer C.: *Deeper Inside PageRank*, Internet Math. **1**, no. 3 (2005), 335–380.

6. Langville A., Meyer C.: *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton Univ. Press, Princeton 2016.
7. Tanase R., Radu R.: *Lecture #3: PageRank algorithm – the mathematics of Google Search*. Online: http://www.math.cornell.edu/m̃ec/Winter2009/RalucaRemus/Lecture3/lecture3.html, 2016.