Indian Journal of Science and Technology, *Vol 8(34), DOI: 10.17485/ijst/2015/v8i34/86120, December 2015*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Weighted Page Rank Algorithm based on In-Out Weight of Webpages

B. Jaganathan* and Kalyani Desikan

Division of Mathematics,
School of Advanced Sciences (SAS), VIT-Chennai, Chennai – 600048,
Tamil Nadu, India; jaganathan.b@vit.ac.in, kalyanidesikan@vit.ac.in

Abstract

In its classical formulation, the well known page rank algorithm ranks web pages only based on in-links between web pages. We propose a new in-out weight based page rank algorithm. In this paper, we have introduced a new weight matrix based on both the in-links and out-links between web pages to compute the page ranks. We have illustrated the working of our algorithm using a web graph. We notice that the page rank values of the web pages computed using the original page rank algorithm and our proposed algorithm are comparable. Moreover, our algorithm is found to be efficient with respect to the time taken to compute the page rank values.

Keywords: Algorithm, In-Weight, Out-Weight, Page Rank, Web Pages

1. Introduction

World Wide Web (WWW) comprises of billions of web pages that hold a huge amount of information. Search engines help users to retrieve relevant information from this large collection of information. However, user's interest for high quality information search services is not fully satisfied by the current search engines.

This poses challenges for information retrieval and to navigate within the search results various ranking methods are applied. Page rank algorithms are well known for ordering web pages. Ranking methods have become an important tool to sort and fetch the relevant web pages based on the user's interest.

The structure of this paper is as follows: Section 1 presents the need for ranking algorithms, Section 2 deals with the different types of ranking algorithms. In Section 3, the relationship between Web pages and Web graph is presented. In Section 4, the adjacency matrix based page rank algorithm is presented. In Section 5, weighted page rank algorithm is presented. In Section 6 we present our proposed in-out weight based page ranking method and its illustration. In Section 7, the comparison between the page rank algorithms is given. In Section 8, the conclusion and the possible future work are presented.

2. Different Types of Ranking Algorithms

Sergey Brin and Larry Page proposed the page rank algorithm[1,2] at Stanford University. A new approach known as weighted page rank algorithm[3] was put forth by Wenpu Xing et al. This algorithm is an extension of the original page rank algorithm. Recently, we developed two page rank algorithms: Category-based page rank method[4] and penalty-based page rank algorithm[5].

In this paper, we propose an efficient page ranking algorithm based on in-out weights of web pages for finding more relevant web pages to users' query.

**Author for correspondence*

3. Web Page and Graph Relationship

Before proceeding further, we must understand the relation between web pages and a web graph. The World Wide Web (WWW) is generally represented as a directed web graph. The vertices are the web pages and directed edges represent the hyperlinks between web pages (out-link/in-link) [5, 6]. A degenerate edge of a graph which joins a vertex to itself is called a loop. A web graph with no loops is called a simple directed graph.

An example of a simple directed graph representing 5 web pages connected by hyperlinks is given in Figure 1.

In graph theory, an adjacency matrix[7] for a directed graph is a matrix for representing adjacent vertices (or nodes) of the graph. For any directed graph G with n vertices, the adjacency matrix is a $n \times n$ matrix with matrix elements being 1 for vertices (or nodes) which point to other vertices and 0 otherwise. This can be mathematically represented as:

$$a_{i,j} = \begin{cases} 1 & \text{if } i \neq j \text{ and } v_i \text{ is pointing/links to } v_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

And it is denoted by A (G). In simple graphs with no loops the adjacency matrix consists of only zeros and ones with diagonal entries being zero.

Adjacency Matrix for Figure 1 is as follows:

$$A(G) = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (2)$$

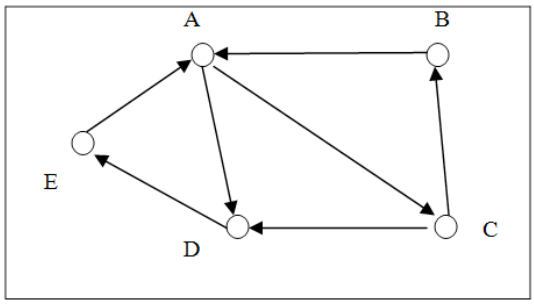

Figure 1. Web graph.

4. Page Rank Algorithm

The page rank algorithm[1,2] was originally proposed by Larry Page and Sergey Brin. A brief explanation of adjacency matrix based page rank algorithm used in the Google search engine is given below.

Consider the web as a directed graph G = {V,E}, where V is the vertex (node) set, that is, the set of all web pages and E is the set of all directed edges in G, that is, hyperlinks of the web graph. Let n be the total number of pages in the web graph. Adjacency matrix A (G) for a web graph G can be calculated using equation 1.

The initial page rank for each of the n web pages is PR_0 = (PR_0 (1), PR_0 (2),… PR_0 (n)) and their value is set as 1.

The formula for computing the m^{th} iteration of the page rank is given by:

$$PR_m = (1-d) + d * A(G) * PR_{m-1} \quad (3)$$

Where d is the damping factor[8,9]. The value of the damping factor ranges between 0 and 1 and it is usually taken as 0.85. At each iteration, the page rank value of each web page is calculated using equation 3. To determine the final page rank of a web page, iterations are carried out until they converge.

5. Weighted Page Rank Algorithm of Web Pages

The weighted page rank algorithm[3] was proposed by Wenpu Xing and Ghorbani Ali. A brief explanation of weighted page rank algorithm is given below.

In the weighted page rank algorithm, more important (popular) web pages are assigned larger page rank values. The popularity of a web page depends on the number of its in links and out links and each web page gets a proportional page rank value. The popularity of each page can be obtained using the in and out weights, $W^{in}_{(v,u)}$ and $W^{out}_{(v,u)}$, given by equations 4 and 5, respectively.

$$W^{in}_{(v,u)} = \frac{I_u}{\sum_{p \in r(v)} I_p} \quad (4)$$

$$W^{out}_{(v,u)} = \frac{O_u}{\sum_{p \in r(v)} O_p} \quad (5)$$

Here r (v) is the set of all Web pages that have in links from node v (reference page list of page v). These weights

depend on the number of in links and out links of page u and the sum of the number of in links and out links of all reference pages of page v, respectively.

The initial page rank for each of the n Web pages is given by $PR_0 = (PR_0\ (1), PR_0\ (2),\ldots\ PR_0\ (n))$ and their value is set as 1.

The formula for computing the weighted page rank of Web page v is given by:

$$PR(v) = (1-d) + d^* \sum_{v \in B(u)} PR(v)^* W^{in}_{(v,u)} {}^* W^{out}_{(v,u)'} \qquad (6)$$

Where B (u) is the set of all web pages that point to u and d denotes the damping factor[8,9] .

6. Proposed Weighted Page Rank Algorithm based on In-Out Weight of Web Pages

We propose a new weighted page rank algorithm. In this algorithm we first calculate the in-weight and out-weight of the web pages using equations 7 and 8.

$$W^{in}_{(v,u)} = \frac{I_u}{\sum_{p \in R(v)} I_p} \qquad (7)$$

$$W^{out}_{(v,u)} = \frac{O_u}{\sum_{p \in R(v)} O_p} \qquad (8)$$

Where I_u and I_p in equation 7 denote the number of in links of page u and page p respectively. Also, O_u and O_p in equation 8 are the number of out links of page u and page p respectively. R (v) is the set of all web pages that point to v (reference page list of page v).

We calculate the weight matrix W (G) using equation 9.

$$W_{(v,u)}(G) = W^{in}_{(v,u)} + W^{out}_{(v,u)} \qquad (9)$$

And its denoted as W (G), where u, v represent the row and column respectively of the weight matrix W (G).

We make use of the weight matrix given in 9 to compute the m^{th} iteration of the page rank as given below:

$$PR_m = (1-d) + d * W(G) * A(G) * PR_{m-1}. \qquad (10)$$

Where d is the damping factor[8,9] that always lies between 0 and 1 and it is usually set as 0.85. The initial page rank for each of the n web pages is $PR_0 = (PR_0\ (1), PR_0\ (2),\ldots\ PR_0\ (n))$ and their value is originally set as 1.

The page rank value of each page, at each iteration, is calculated using equation 10. To determine the final page rank of a web page iterations are carried out until they converge.

We now explain the working of the original and our proposed in-out weighted page rank algorithms by considering the hyperlinked web graph shown in Figure 1, consisting of five pages A, B, C, D and E.

The in and out weights for pages A, B, C, D and E are calculated using equations 7 and 8 respectively. We now illustrate, for a few web pages, how the in and out weights are computed. For example:

$$W^{in}_{AC} = \frac{I_C}{I_B + I_E} = \frac{1}{1+1} = \frac{1}{2}$$

$$W^{in}_{AD} = \frac{I_D}{I_B + I_E} = \frac{2}{1+1} = \frac{1}{1} = 1$$

$$W^{out}_{AC} = \frac{O_C}{O_B + O_E} = \frac{1}{1+1} = \frac{1}{2}$$

$$W^{out}_{AC} = \frac{O_D}{O_B + O_E} = \frac{1}{1+1} = \frac{1}{2}$$

In the same way we can compute the in and out weights for the remaining pages and we get the following matrices:

$$W^{in} = \begin{bmatrix} 0 & 0 & 1/2 & 1 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/3 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$W^{out} = \begin{bmatrix} 0 & 0 & 1/2 & 1/2 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/4 \\ 2 & 0 & 0 & 0 & 0 \end{bmatrix}$$

From the above two matrices we can form the weight matrix W (G) using equation 9 as follows:

$$W(G) = \begin{bmatrix} 0 & 0 & 1 & 3/2 & 0 \\ 3 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 7/12 \\ 3 & 0 & 0 & 0 & 0 \end{bmatrix}$$

After computing the weight matrix W (G), we make use of equation 10 to compute the page ranks. We have computed the page ranks for three different values of the damping factor viz., d = 0.85, d = 0.7 and d = 0.5. The values to which the page ranks converge in these three cases are given in Table 1.

Table 1. Page rank computations using our proposed page rank algorithm for various damping factor (d) values for web graph in Figure 1

Proposed Page Rank Algorithm			
Web page	Page rank (d = 0.85)	Page rank (d = 0.7)	Page rank (d = 0.5)
A	0.5343	0.8149	1.0043
B	0.2783	0.4580	0.6371
C	0.5314	0.7885	0.9446
D	0.4368	0.6603	0.8172
E	0.2783	0.4580	0.6371

7. Comparison between Page Rank Algorithms

We have calculated the page rank for the web pages for the web graph (Figure 1.) using the adjacency matrix based original page rank algorithm[1,2], weighted page rank algorithm method and our proposed in-out weighted matrix based page rank algorithm.

The Tables 2, 3 and 4 shows the page ranks computed for the web pages using the adjacency matrix, weighted page rank algorithm method and in-out weighted matrix based page rank algorithm for the three different values of the damping factor. The web pages (vertices) are arranged in the tables in increasing order of page rank value.

The page rank algorithm and our proposed in-out weighted based page rank algorithm method give the same rank to the web pages. Calculating page rank using our proposed page rank method takes lesser time compared to the original page rank algorithm. Time taken to

Table 2. Page rank computations using original page rank algorithm, weighted page rank algorithm and our proposed page rank algorithm for *d=0.85* in the web graph (shown in Figure 1)

Original Page Rank Algorithm		Weighted Page Rank Algorithm		Proposed Page Rank Algorithm	
Web page	Page rank	Web page	Page rank	Web page	Page rank
A	1.4773	A	0.7008	A	0.5343
C	1.1559	E	0.4580	C	0.5314
D	0.8112	D	0.3624	D	0.4368
B	0.7778	C	0.2824	B	0.2783
E	0.7778	B	0.1900	E	0.2783

Table 3. Page rank computations using original page rank algorithm, weighted page rank algorithm and our proposed page rank algorithm taking *d = 0.7* for the web graph (Figure 1.)

Original Page Rank Algorithm		Weighted Page Rank Algorithm		Proposed Page Rank Algorithm	
Web page	Page rank	Web page	Page rank	Web page	Page rank
A	1.4068	A	1.0364	A	0.8149
C	1.1538	E	0.6982	C	0.7885
D	0.8547	D	0.5688	D	0.6603
B	0.7924	C	0.4612	B	0.4580
E	0.7924	B	0.3638	E	0.4580

Table 4. Page rank computations using original page rank algorithm, weighted page rank algorithm and our proposed page rank algorithm taking d=0.5 for the web graph (Figure 1.)

Original Page Rank Algorithm		Weighted Page Rank Algorithm		Proposed Page Rank Algorithm	
Web page	Page rank	Web page	Page rank	Web page	Page rank
A	1.2982	A	1.2115	A	1.0043
C	1.1404	E	0.8702	C	0.9446
D	0.9123	D	0.7404	D	0.8172
B	0.8246	C	0.6346	B	0.6371
E	0.8246	B	0.5529	E	0.6371

compute page rank values, with d = 0.85, using page rank algorithm is 15.00292 seconds and for our proposed page rank method it is 10.79334 seconds. The same behavior is observed for d = 0.70 and d = 0.5.

8. Conclusions and Future Work

From the previous section on in-out weight based page rank algorithm we notice that our algorithm is more efficient when compared to the original page rank algorithm with respect to time. It can also be seen that the ranking of the web pages using our algorithm agrees with that obtained by using the original page rank algorithm. While the ranking obtained using the weighted page rank algorithm does not agree with the original page rank algorithm.

In our future work, based on this algorithm, we envisage to work with bigger web graphs. We also propose to introduce other weight based techniques for calculating the ranks of web pages.

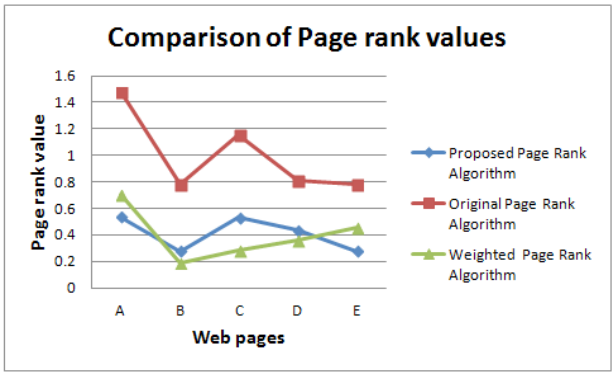

Figure 2. Shows the page rank values obtained using the three algorithms for d=0.85 for the web graph in Figure 1.

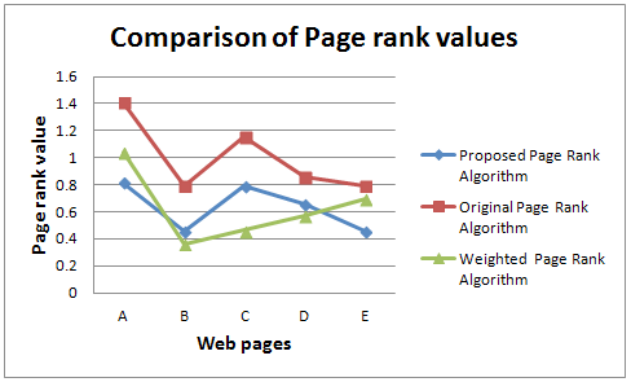

Figure 3. Shows the page rank values obtained using the three algorithms for d = 0.7 for the web graph in Figure 1.

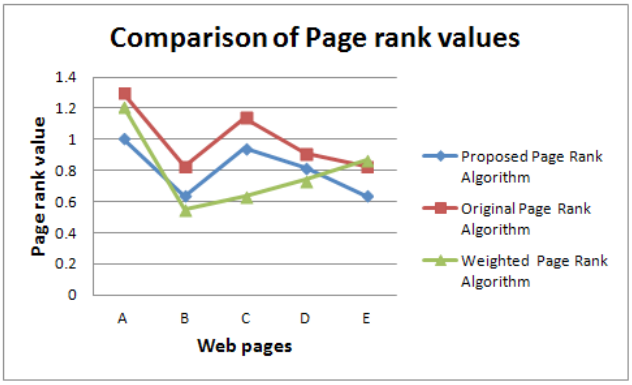

Figure 4. Shows the page rank values obtained using the three algorithms for d=0.5 for the web graph in Figure 1.

9. References

1. Page L, Brin S, Motwani R, Winograd T. The page rank citation ranking: Bringing order to the Web. Technical report. Stanford Digital Library Technologies Project; 1998 Jan. p. 1–17.

2. Page L, Brin S. The anatomy of a large scale hyper-textual web search engine. Computer Networks and ISDN Systems.1998 Apr; 30(1-7):107–17.
3. Xing W, Ghorbani A. Weighted page rank alogithm. Proceedings of the Second Annual Conference on Communication Networks and Services Research (CNSR'04); IEEE. 2004 May 19-21. p. 305–14.
4. Jaganathan B, Kalyani D. Category-based page rank algorithm. International Journal of Pure and Applied Mathematics. 2015 Aug; 101(5):811-820.
5. Jaganathan B, Kalyani D. Penalty-based page rank algorithm. ARPN Journal of Engineering and Applied Sciences. 2015 Mar; 10(5):2000–3.
6. Kleinberg JM. Authoritative sources in a hyperlinked environment. Journal of the ACM. 1999 Sep; 46(5):604–32.
7. Langville AN, Meyer CD. Google's page rank and beyond: The science of search engine rankings. Princeton, NJ: Princeton University Press; 2006.
8. Bressen M, Peserico E. Choose the damping, choose the ranking? Journal of Discrete Algorithms. 2010 Jun; 8(2):199–213.
9. Ali D, Faqir M, Hassan D, Hussain D. Ranking cricket teams. International Journal Information Processing & Management. 2015; 5:62–73.