A Literature Survey on Page Rank Algorithm

[1]Ritu Sachdeva (Sharma) [2]Dr. Sachin Gupta
[1]Research Scholar, Head, [2]Computer Science Deptt.
MVNU, Palwal
Email id: [1]ritusach08@gmail.com, [2]sachin.gupta@mvn.edu.in

Abstract: There is a mass of web data available in the form of web pages on the World Wide Web (WWW). So whenever a user makes a query, a lot of search results having different web links corresponding to user's query are generated. Out of generated results, only some are relevant while the rest are irrelevant. The relevancy of a web page is calculated by search engines using page ranking algorithms. These algorithms are either based on web structure mining or web content mining to calculate the relevancy of a web page. In the beginning, Google introduced Page Rank Algorithm which was set as standard algorithm because no other algorithm meant for ranking web pages was in existence. Later a number of factors like weights or number of visits of a web link was incorporated in standard page rank algorithm by different authors. This paper includes a detailed literature survey of different variations of page rank algorithm. This paper also covers literature survey on Search Engine Optimization as well as Recommendation system as these approaches also play an important role in setting the accurate Page Rank of a Web Page.This paper also covers numerical & graphical analysis of Page Rank & its variant Page Rank based on visits of Links(VOL)

Keywords: inlinks, outlinks, search engine, web mining, World Wide Web (WWW), PageRank, Weighted page rank, VOL

1 INTRODUCTION

World Wide Web is a vast resource of information like text, image, audio, video and metadata. It is growing by multi fold day by day. With the expansion of information on the WWW as well as extreme user dependence on web, there is difficulty in managing web information and satisfying the users with respect to their queries. First user make the query, then information retrieval techniques filter the web pages according to relevancy and on behalf of it, an index is created. This indexing is related to the rank of web page. Lower the index value, higher will be the rank of the web page.

1. Data Mining over Web

1.1 Web Mining

Data Mining is defined as the procedure of extracting relevant information from huge sets of data via various processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation. Web mining is an immense area with dozens of developments and technological enhancements.

1.2. Web Mining Categories

According to literature, there are three categories of web mining: Web Content Mining (WCM), Web Structure Mining (WSM) and Web Usage Mining (WUM)

WCM includes the web page information. In it, the actual content pages whether semi structured hypertext or multimedia information are used for searching purposes.

WSM uses the central part linkage that flows through the entire web. The linkage of web content is called hyperlink. This hyperlinked structure is used for ranking the retrieved web pages on the basis of query generated by the user.

WUM returns the dynamic results with respect to users' navigation. This methodology uses the server logs (the logs that are created during user navigation via searching). WUM is also called as Web Log Mining because it extracts knowledge from usage logs.

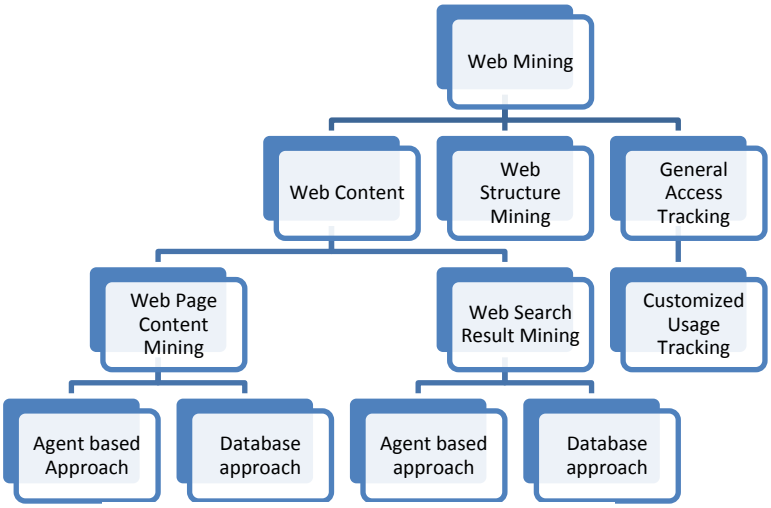


Fig. Taxonomy of Web Mining

2. Page Rank Algorithm (By Google)

This is the original PageRank algorithm. It was postulated by Lawrence Page and Sergey Brin[8] . The formula is:

$$PR(A) = (1-d) + d(\frac{PR(Ti)}{C(Ti)}.+\cdots\ldots\ldots\ldots..+\frac{PR(Tn)}{C(Tn)}) \cdot \quad \text{.....(1)}$$

where

- $PR(A)$stands for PageRank of page A
- $PR(Ti)$gives PageRank of pages Ti which link to page A
- $C(Ti)$counts the number of outbound links on page Ti
- d is a damping factor having valueset between 0 and 1

The PageRank algorithm is used to determine the rank of a web page individually. This algorithm is not meant to rank a web site.Moreover, the PageRank of a page say A, is recursively defined by the PageRanks of those pages which link to page A. The PageRank of pages Ti which link to page A does not influence the PageRank of page Aconsistently. In PageRank algorithm, the PageRank of a page T is always weighted by the number of outbound links C(T) on page T. It means, more outbound links a page T has, the less will page A benefit from a link to it on page T. The weighted PageRank of pages Ti is then added up. But an additional inbound link for page A will always increase page A's PageRank. In the end, the sum of the weighted PageRanks of all pages Ti is multiplied with a damping factor d which can be set between 0 and 1.Thus,

the extend of PageRank benefit for a page by another page linking to it is reduced.

They deem PageRank as a genre of user behaviour, where a user clicks on links rf

And omlyirrespective of content. This user visits a web page with a certain probability given by the number of links on that page. Thus, one page's PageRank is not completely passed on to a page it links to, but is divided by the number of links on the page. So, the probability accessing one page is the sum of probabilities following links to this page. Now, this probability is diminished by the damping factor d. Sometimes, user jumps to some other page instead of clicking straight to the links associated to a page. This probability is calculated by the damping factor d (also called as degree of probability having value between 0 and 1). Regardless of inbound links, the probability for the random surfer jumping to a page is always (1-d), so a page has always a minimum PageRank.

Later, Lawrence Page and Sergey Brin gave an updated version of Page Rank. In this algorithm, the PageRank of page A is given as

$$PR(A) = (1-d)/N + d(\frac{PR(Ti)}{C(Ti)}.+\cdots\ldots\ldots\ldots..+\frac{PR(Tn)}{C(Tn)}) \quad \text{.....(2)}$$

where N is the total number of all pages on the web. Thisversion is basically equivalent the original one. Thisversionis the actual

probability for a surfer reaching that page after clicking on many links. The sum of all page ranks of all pages will be one by calculating the probability distribution of all web pages.

But, these versions of the algorithm do not differ fundamentally from each other. A PageRank which has been calculated by using the second version of the algorithm has to be multiplied by the total number of web pages to get the according PageRank that would have been calculated by using the first version.

3. Dangling Nodes

A node is called a dangling node if it does not contain any out-going link, i.e., if the out-degree is zero. The hypothetical web graph taken in this paper is having a dangling node i.e. Node D.

2 Literature Survey

To effectively and efficiently discover valuable information from web social networks. Ranking web pages is very difficult due to the following reasons:

- Pages grow in a fashion of terabyte or even petabyte each day which needs a large amount of time to process these pages
- Link relations among these pages are very intricate which can affect the accuracy of page scoring

A number of Page Rank Algorithms are practiced.

The Standard Page Rank has following drawbacks:

- The page rank algorithm works in an offline fashion, so the scores of pages depend on the previously given pages. Actually, the web will accumulate a large volume of new pages that contain high quality information during a short period of time, these newly published pages with high authority will be frequently cited by other pages or web sites, which will increase their page rank values.
- The traditional page rank algorithms are biased to compute the scores among the type of URLs ending with ".com", since such pages are often portals or large sites which can easily capture more links or citations than some professional ones. However, these pages which provide special services for certain users are sometimes authoritative and have lots of in-links from other pages.
- Page rank cannot identify whether the hyperlinks of distinct pages are content-correlated (i.e. the themes are similar among pages) or not, which will cause the phenomena of "theme-draft".

S.No	Author	Refer-ence No.	Proposed Method	Striking Features	Method Description
1.	Dung B. Le	[7]	TS-Local Rank	Refines the relevance of web pages to user query by re-calculating the rank based on the inter-connectivity of web pages in search results.	This method is evaluated by constructed a meta-search engine named TS-Meta Search.
2.	Shaojie Qiaot	[8]	SimRank	Traditional page rank algorithm is improved by assigning a probability of browsing a page to be initial page rank value of each paper	It is based on similarity measure named to score web pages. This measurecomputes the similarity between pages &use it to partition a web database into several web social networks.
3.	Yang	[9]	New Page Rank		Accelerates the scoring of web pages

4.	Haveliwala	[10]	Topic sensitive page rank	Can handle the situation that some pages may not be considered to important in other fields although they get a high score in some field. This method can also help to avoid the problem of theme draft	Due to dated web pages, the efficiency of page scoring will fall down quickly can handle the problem of theme-draft. This approach is based on 16 basic topic vectors from open directory[14] on behalf of which similarity between the query from users and the given topic vectors is calculated.
5	Zheng Chen	[11]	Modified page rank		Construct implicit links by mining Users' access patterns and then apply re-rank web pages for small web search.
6	May Thu Aung	[12]	Refined page rank	Solves the sparsity of web links &diversity problem of user's access behavior.	An implicit link structure is constructed by using Frequent sequence Miner (FS-Miner) in user web log. Then web pages are re-ranked
7.	Fayyaz Ali	[13]	Ratio based Weighted Page Rank	It calculates ranking of web pages in terms of convergence. Convergence is the number of steps taken to get the ranks stabilized , so the change in the coming iterations is minimal This method performs better than page rank	Ratio Rank approach divides the page rank of a node among its referenced nodes so that every node receives its own share from the referee node according to its weightage.
8.	Bouchra Frikh	[14]	SALSA	SALSA (Stochastic Approach for Link Structure Analysis). It combines best features of Page Rank and HITS algorithms	The main strengths of SALSA are less susceptible to topic drift problem as well as spammers.
9.	Dr. Divya Gupta	[15]	User preference based page ranking	User preference based Page Rank algorithm regulates search quality &makes user search navigation experience in the results of a Search Engine. Ranking algorithm based on web structure mining are less relevant to user query because they don't consider user trends on current topics. Ranking algorithm based on web content mining totally ignores importance of web page.	It uses agents to determine pages context relevancy and considers user behavior while ranking web pages.

10	Rekha Singhal .	[26]	WIL PR (Weight age In-Link page rank)	This gives higher relevancy in web search results than original standard page rank	Irrespective of dividing the weight of an in-linked web-page, this method distributes it to all the out linked pages on the basis of their popularity and finally a new score of every individual webpage is calculated and web pages are ranked accordingly.

*Irrespective of different page rank algorithms, there is another approach via Search Engine Optimization. Due to tremendous growth of web search engine, users are facing a lot of problems in retrieving the most appropriate information out of it called OverKill problem. The main cause of this problem is non-optimization of web pages. Web Optimization started way back in 1997 and it is called Search Engine Optimization (SEO). It makes a website friendly and easy to navigate for users as well as search engine robots and to increase their traffic in an organic way.

S.No	Author	Refere-nceNo.	Research Description
1	Baeza-Yates	[16]	Study indicates the different types of usage of Search Engine Transaction Logs for development of different aspects of search engines.
2	Malaga	[17]	Observed that the search engines constantly change their ranking algorithms. This leads to situations in which a site might rank well one day and completely disappear from the search engine the next" i.e. a website's ranking is variable in time. This variation is linked to the updates of a search engine algorithm and to increase in the number of websites. To manage this volatile situation by site owners, some techniques or method could help SEO Practitioners to validate their SEO actions.
3	Utrich Guntzer	[18]	Proposed an algorithm to combine multi-feature results for efficient top ranked results. In this algorithm, the author mainly concerns with weighted queries and multimedia such as retrieval of images
4.	Wen	[19]	Proposed a technique to make clusters of similar kinds of queries to propose URLs with respect to often asked queries of a search engine. A much more work is done in this area.
5.	Shipra Kataria	[3]	Proposed a technique for investigation of transaction logs obtained by Search Engines to optimize rank of web pages and then resulting into the topic/subject relevant and user suitable documents at the top of the result pages of Search Engines
6	Sylvain Sagot	[20]	An Engineering model is used that analyzes a range of different approaches (from literature meta-model), their key differences and relationship among them. Also design of SEO integrates different models with the aim of improving and optimizing SEO. It is an adaptive process which should be assisted by intelligent models and tools integrated in a holistic approach
7	P. Ravi Kumar	[21]	Put a light on website optimization, its challenges as well as website optimization classes. The author provides a detail report on WSO " Onsite" and "Off-site" ranking factors, efficient methodologies and techniques for on-site and off-site ranking factors. There is also a list of black hat techniques that can potentially reduce the ranking of a website and even can remove the entry from the index of a search engine.
8	Y. Nawas Ahamad Khan	[22]	Introduced a system for optimizing, clustering and summarizing web search results using Intelligent Agents called OCS system. This system first analysis duplicates in content and link levels and then optimized contents are clustered using top frequent clustering approach by identifying an optimal threshold. Then clustered contents are summarized using extraction process

			at query level and anchor text level. Google Engine is used as a case study to implement this OCS(Optimizing, clustering & Summarizing) system.
9	Jain	[23]	As the SEO can be done by two ways: On page and links to particular web page i.e. off-page to improve organic search ranking. explains that optimization process mostly initiates with on-page techniques. But the author focused on content and tags. Off-Page SEO measures relevant back linking to the website and to present the web page in front of web spiders to crawl successfully. The author describes the techniques as well as the value of Off-Page SEO. It is considered that improving webpage visibility by enhancing massive semantic content using on-page optimization.
10	Meenakshi Bansal	[4]	Emphasizes that website optimization is about making important modifications to almost concerned sections of the website. Though it is viewed individually, some of the change might seem like gradual incremental improvements. SEO is a process which requires considerable time. This work on on-page optimization include actual code merged with various languages, keyword placement and keyword density. As search engine employs combination of automated algorithms, manually edited directories and advertisements to generate results for users' queries.
As time has passes, there comes a recommendation system. There are basically two approaches to make a recommendation system i.e. content-based recommendation systems and calloborative filtering recommendation systems.			
11	Anamika Rajput	[24]	Suggested a recommender system for user rating and synonyms based ranking of the websites. In this system, when the user searches for a keyword, keyword and its synonyms are searched and based on the its search, the websites are displayed and user is asked to give rating. The user rating is used to upgrade or degrade the page rank in database. The Map Reduce Algorithm has been used to differentiate and reduce the data retrieved. The recommender system has created is more efficient and works at a faster speed. It takes input from the user to rank a website. Thus, the system created is a new approach based and fulfills the demand of the user searching.

3 RELATED WORK

Brin and Page [8] (Algorithm: Google Page Rank)

This algorithm is used to calculate rank of web pages using hyperlinked structure of the web. This algorithm is used by Google based on the results produced by keyword based search. It works on the principle that if a web page has significant links towards it, then the links of this page to other pages are also considered imperative. Thus, it depends on the backlinks to calculate the rank of web pages. The page rank is calculated by the formula given in equation 1.

$$PR(u) = c\sum_{V\in B(U)}\frac{PR(V)}{Nv} \quad (3)$$

Where U depicts a Web Page
PR(u)&PR(v)= Page Ranks Of Web Pages U And V Respectively
B(U)= A Set Of Web Pages Pointing To U

Nv= Total Numbers Of Outlinks Of Web Page

V And C = Factor Used For Normalization

Original page rank algorithm was modified considering that all users don't follow direct links on web data. Thus, the modified formula for calculating page rank is given in equation 2.

$$PR(u) = (1-d) + d\sum_{v\in B(u)}\frac{\boldsymbol{PR(v)}}{\boldsymbol{Nv}} \quad (4)$$

Where D = Dampening Factor representingthe Probability of user Using Direct Links having value set between 0 & 1.

Wenpu Xing And Ali Ghorbani [30] (Algorithm: Weighted Page Rank)

The authors gave this method by extending standard pagerank. It works on the theory that if a page is vital, it has many inlinks and outlinks. Unlike standard pagerank, it does not equally distribute the page rank of a page among its outgoing linked pages. The page rank of a web page is divided among its outgoing linked pages in proportional to the importance or popularity (its number of inlinks and outlinks).

$W^{in}(v,u)$, the popularity from the number of inlinks, is calculated based on the number of inlinks of page u and the number of inlinks of all reference pages of page v as given in equation 3.

$$W^{in}(v,u) = \frac{I_u}{\sum_{p \in R(v)} I_p} \quad (5)$$

Where I_u&I_p= Number Of Inlinks Of Page U &P Respectively

R(v)= A Set Of Web Pages Pointed By V.

$W^{out}(v,u)$= Number Of Outlinks of Page U and the number Of Outlinks Of All Reference Pages Of Page V As Given In Equation. 4.

$$W^{out}(v,u) = \frac{o_u}{\sum_{p \in R(v)} o_p} \quad (6)$$

Whereo_u&o_p= Number Of Outlinks Of Page U & P Respectively

R(v)= The Set Of Web Pages Pointed By V.

The Page Rank Using Weighted Pagerank Algorithm Is Calculated By The Formula As Given In Equation

$$PR(u) = (1-d) + d\sum_{v \in B(u)} \frac{PR(v)}{Nv} W^{in}(v,u) W^{out}(v,u) \quad .....(7)$$

Gyanendra Kumar Et. Al. [32] (Algorithm : Page Rank With Visits Of Links (Vol))

This methodology includes the browsing behavior of the user. The prior algorithms were either based on WSM or WCM. But it incluses page ranking based on visits of links (vol). It modifies the basic page ranking algorithm by considering the number of visits of inbound links of web pages. It assists to prioritize the web pages on the basis of user's browsing behavior. Also, the rank values are assigned in proportional to the number of visits of links in this algorithm. The more rank value is assigned to the link which is most visited by user. The page ranking based on visits of links (vol) can be calculated by the formula given in equation 6.

$$PR(u) = (1-d) + d\sum_{v \in B(u)} \frac{PR(v)L_u}{TL(v)} .....(8)$$

WherePR(u)&PR(v)= Page Ranks Of Web Pages U & V Respectively

D= Dampening Factor

B(U) =A Set Of Web Pages Pointing To U

Lu = Number Of Visits Of Links Pointing From V To U

TL(V) = Total Number Of Visits Of All Links From V

Neelam Tyagi & Simple Sharma [2] (Algorithm: Weighted Page Rank Algorithm Based On Number Of Visits Of Links Of Web Page)

The authors incorporate weighted pagerank algorithm and the number of visits of links (vol). This algorithm consigns more rank to the outgoing links having high vol. It is based on the inlink popularity ignoring the outlink popularity. In this algorithm, number of visits of inbound links of web pages are taken into consideration in addition the weights of page. The rank of web page using this algorithm can be calculated as given in equation 7.

$$WPR(u) = (1-d) + d\sum_{v \in B(u)} \frac{L_u WPR_{VOL}(v) W^{in}(v,u)}{TL(v)} (9)$$

WhereWPR_{Vol}(U, V)= Page Rank Of Web Page U& V Respectively

D = Dampening Factor

B(U) =A Set Of Web Pages Pointing To U

Lu = Number Of Visits Of Links Pointing From V To U

Tl(V)= Total Number Of Visits Of All Links From V

$W^{in}(v,u)$ = Number Of Inlinks Of U.

Sonal Tuteja [1] (Algorithm: Enhancement In Weighted Page Rank Using Visits Of Link (Vol))

The Author Incorporated $W^{in}{}_{VOL}$ (v, u)I.E. The Weight Of Link(V,U) And Calculated Based On The Number Of Visits Of Inlinks Of Page U.

$W^{out}{}_{VOL}$ (v, u) The Popularity From The Number Of Visits Of Outlinks Are Used To Calculate The Value Of Page Rank.

$W^{in}{}_{VOL}$ (v, u)Is The Weight Of Link(V, U) Which Is Calculated Based On The Number Of Visits Of Inlinks Of Page U And The Number Of Visits Of Inlinks Of All Reference Pages Of Page V As Given In Equation 8.

$$W^{in}{}_{VOL}(v,u) = \frac{I_{u(VOL)}}{\sum_{p \in R(v)} I_{p(VOL)}} \quad (10)$$

Where$I_{u(VOL)}$&$I_{p(VOL)}$= Incoming Visits Of Links Of Page U And P Respectively

R(V) =A Set Of Reference Pages Of Page V.

$W^{in}{}_{VOL}$ (v, u)= Weight Of Link(V, U) and is calculated Based On The Number Of Visits Of Outlinks Of Page U And The Number Of Visits Of Outlinks Of All Reference Pages Of Page V As Given In Equation 9.

$$W^{out}{}_{VOL}(v,u) = \frac{O_{u(VOL)}}{\sum_{p \in R(v)} O_{p(VOL)}} \quad (11)$$

Where $O_{u(VOL)}$ And $O_{p(VOL)}$= Outgoing Visits Of Links Of Page U And V Respectively

R(V) =A Set Of Reference Pages Of Page V.

Now These Values Are Used To Calculate Page Rank Using Equation (10)

$$EWPR(u) = (1-d) + d\sum_{v\in B(u)} WPR_{VOL}(v) W^{in}{}_{VOL}(v,u) W^{out}{}_{VOL}(v,u)$$

.....(12)

Where D = Dampening Factor

B(U) = A Set Of Pages That Point To U

WPRvol (U) And WPRvol(V) = Rank Scores Of Page U And V Respectively

$W^{in}{}_{VOL}$ (v, u)=Number Of Visits Of Inlinks

$W^{out}{}_{VOL}$ (v, u)=Number Of Visits Of Outlinks

I. Comparison chart of various Ranking Algorithms

Algorithm	PR	PR with VOL	WPR	WPRV	EWPRV
Mining Technique	Web Structure Mining	Web Structure Mining, Web Usage Mining	Web Structure Mining	Web Structure Mining, Web Usage Mining	Web Structure Mining, Web Usage Mining
I/t Parameters	Backlinks	Backlinks, VOL	Backlinks, VOL	Backlinks, VOL	Backlinks, Forelinks,VOL
Importance	More	More	More	More	More
Relevancy	Less	Less	More	More	More
Year of Introduce	1996	2011	2004	2012	2013
Methodology	Calculate PR based on number of Backlinks	Assign more rank value to outgoing links which is most visited by users	Assign more PR value to popular page	Assign more rank value to outgoing links which is most visited by users	Assign more rank value to ingoing links which is most visited by users

II. Proposed Algorithm- Numerical Analysis Of Various Page Rank Algorithms

To demonstrate the working of Page Rank, consider a Hypothetical web structure as shown below:

1. Page Rank (Brin & Page)

Using equation 2, the ranks for pages A, B, C are calculated as follows:

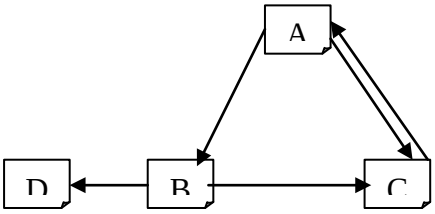
$PR(A) = (1-d) + d(PR(C)/2)$ (1)
$PR(B) = (1-d) + d((PR(A)/1))$ (2)
$PR(C) = (1-d) + d((PR(A)/1) + ((PR(B)/1))$(3)
$PR(D) = (1-d) + d(PR(B)/1)$ (4)

Having value d=0.25, 0.5, 0.85, the page ranks of pages A, B and C become:

Dampening Factor	PR(A)	PR(B)	PR(C)	PR(D)
0.25	0.9	0.975	1.22	0.99
0.5	0.8	0.9	1.35	0.95
0.85	0.85	0.829	1.53	0.357

From the results, it is concluded that
PR(C)> PR(D)> PR(B)> PR(A)

2. Page Rank with Visits of Links (VOL) (Gyanendra Kumar)



Using equation 6, the ranks for pages A, B, C are calculated as follows:

PR_{VOL}(A)=(1-d)+d(PR_{VOL} $(C)/2$. $L_A/TL(C)$)(1)

PR_{VOL}(B)=(1-d)+d(PR_{VOL} $(A)/1$. $L_B/TL(A)$)(2)

PR_{VOL}(C)=(1-d)+d($PR_{VOL}(A)/1 \cdot L_C/TL(A)$)+ $PR_{VOL}(B)/2 \cdot L_C/TL(B)$) (3)

PR_{VOL}(D)=(1-d)+d(PR_{VOL} $(B)/1 \cdot L_D/TL(B)$)(4)

The intermediate values can be calculated as:

$L_A/TL(C) = 2/3$
Similarly other values after calculation are:
$L_B/TL(A) = 1/3$
$L_C/TL(A)$ =2/3
$L_C/TL(B) = 1/3$
$L_D/TL(B) = 1/3$

Figure showing a web graph having three web pages i.e. A, B, C, D

Having value d=0.25,0.5, 0.85 the page ranks of pages A, B and C become:

Dampening Factor	PR(A)	PR(B)	PR(C)	PR(D)
0.25	0.83	0.82	1.23	0.818
0.5	0.635	0.606	0.808	0.6
0.85	0.2478	0.22	0.3449	0.1123

From the results, it is concluded that

PR(C)> PR(A)> PR(B)> PR(D)

VI Analysis of Page Rank value of different Page Ranking algorithm on different iteration

Table 1: Values of Page A on different Page Rank Algorithms and different d values

D	Page Rank	Page Rank with VOL
0.25	0.9	0.83
0.5	0.8	0.635
0.85	0.85	0.2478

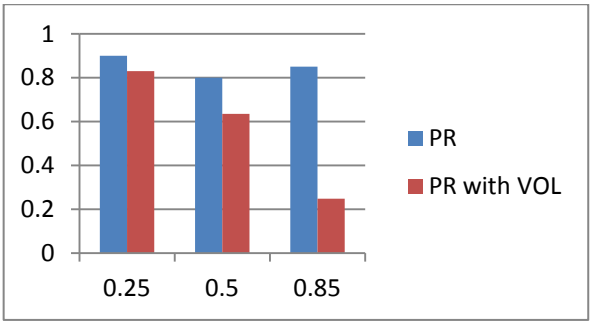


Table 2: Values of Page B on different Page Rank Algorithms and different d values

	Page Rank	Page Rank with VOL
0.25	0.975	0.82
0.5	0.9	0.606
0.85	0.829	0.22

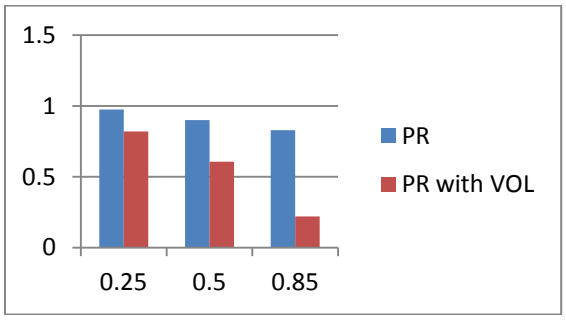


Table 3: Values of Page C on different Page Rank Algorithms and different d values

	Page Rank	Page Rank with VOL
0.25	1.22	1.23
0.5	1.35	0.808
0.85	1.53	0.3449

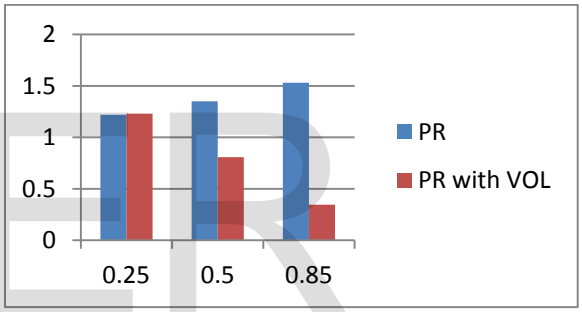
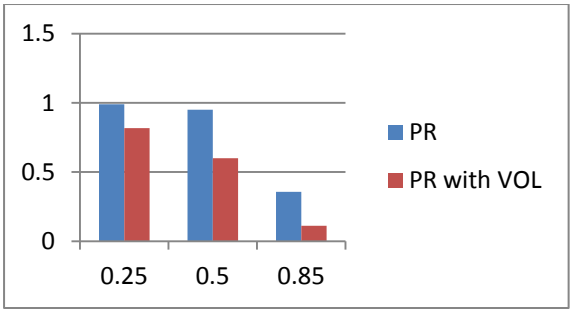


Table 4: Values of Page D on different Page Rank Algorithms and different d values

	Page Rank	Page Rank with VOL
0.25	0.99	0.818
0.5	0.95	0.6
0.85	0.357	0.1123



4 Conclusion

In this paper, literature survey regarding the ranking score of web pages is done. This survey

includes different variations of Page Rank Algorithm as well as its various aspects like Search Engine Optimization and/or Recommendation Systemsthat affect the ranking of a web rank and contribute to find more accurate results in ranking score of a web page. Also, a numerical analysis has been done by taking a hypothetical graph and two Page rank algorithms are implemented. It is too be emphasized that three different values of dampening factor i.e. 0.25, 0.5 & 0.85 have been taken and graphically, a node wise comparison is done. In all nodes, page rank is maximum at dampening factor 0.25 and minimum at 0.85.

III. FUTURE SCOPE

A practical analysis of more different variants of Page Rank Algorithm will be done using a hypothetical Web Graph. Generally, d=0.5 is considered. A deep survey or study will be done with respect to different values of dampening factor and its affect on page rank score. Also,Search Engine Optimization have been formalized as well as own Recommendation system can be created to give the more accurate results of the user query.

REFERENCES

1. Sonal Tuteja ,"Enhancement in Weighted PageRank Algorithm using VOL", IOSR Journal of Computer Engineering (IOSR-JCE) ,e-=ISSN: 2278-0661, Volume 14, Issue 5(Sep – Oct, 2013), PP 135-141
2. Neelam Tyagi, Simple Sharma ,"Weighted Page Rank Algorithm based on Number of Visits of Links of Web Page", International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-2, Issue-3, July 2012
3. Shipra Kataria, Pooja Sapra, "A Novel Approach for Rank Optimization using Search Engine transaction logs" , 978-9-3805-4421-2, 2016, IEEE
4. Meenakshi Bansal, Deepak Sharma, ,"Improving webpage visibility in Search Engines by enhancing keyword Density using improved On-Page Optimization technique", IJCSIT, 5347-5352, ISSN-0975-9646, 2015
5. Page Ranking Based on Number of Visits of Links of Web Page By Gyanendra Kumar , Neelam Duhan , A. K Sharma, Department of Computer Engineering, YMCA University of Science & Technology, Faridabad, India, International Conference on Computer & Communication Technology (ICCCT)-2011
6. Page Ranking Algorithms: A Survey Neelam Duhan, A. K. Sharma, Komal Kumar Bhatia YMCA Institute of Engineering, Faridabad, India, 2009 IEEE International Advance Computing Conference (IACC 2009)
7. Dung B. Le, Sunita Prasad, "TS-LocalRank: A Topic Similarity Local Ranking Algorithm for Re-ranking Web Search Results", IEEE, International Conference on Advanced Technologies for Communications, 978-1-4244-5139-5/2009
8. Shavjie Qiaot, Yianrui Li, Hong Li Yan Zhu, "SimRank: A Page Rank Approach based on Similarity Measure", 978-1-4244-6793-8/2010, IEEE
9. Ling Zhang, Fanyan Ma, "Accelerated ranking : a new method to improve web structure Mining Quality, Journal of computer Research and Development, 41(1): 98-103,204
10. Taper H. Haveliwala, "Topic Sensitive Page Rank: a context-sensitive Ranking Algorithm for web search", IEEE Transactions on knowledge and Data Engineering, 15(4): 784-796, 2003
11. Yang Q. , Zhang H.H., Li T. Mining Web logs for prediction models in www caching and prefetching. In Proceeding of KDD'01, 473-478, August 2001
12. May Thu Aung, Khin Nive Ni Tun, "To construct Implicit, Link structure by using Frequent sequence Miner (FS- Miner), IEEE, 978-0-7695 /2009
13. Fayyaz Ali, Irfan Ullah, Shah Khusro, "An Empirical Investigation of Page Rank and its variants in Ranking Pages on the Web", 978-1-5090-5300-1/2016, IEEE
14. Bouchra Frikh, Brahim Quhbi, Amine Ameur, "A Comparative study of Link analysis Algorithms for Information Retrieval 2012 Next Generation Networks and Services NGNS, 2-4 December 2012 Portugal ISBN 978-9954-9091-2-6
15. Dr. Divya Gupta, Devika Singh, "User Preference Base Page Ranking Algorithm", ICCCCA 2016, ISBN: 978-1-5090-1666-2/2016, IEEE
16. R. Baeza-Yates, Web Usage Mining in Search Engines", Web Mining: Applications and Techniques, Anthony Scime, Editor, Idea Group, 2004
17. R.A. Malaga, "The Value of SEO: An Action Research Project at a new e-Cooerce site, "JECO, Volume 5m No 3, p. 68-82, 2007
18. Ulrich Gruntzer, wolf –Tilo Balke, Werner Kibling,"Optimizing Multi-Feature queries for Image Database", Proceedings of the 26th VLDB Conference, Cavio, Egypt, 2000
19. J. Wen Nie, H. Zhang, "Clustering User queries of a Search Engine, In Proceedings at 10th international world wide web conference, pp 162-168, W3C, 2001
20. Sylvain Sagot, Alain-Jerome Fougeres, Egon Ostrosi, Pauline Lacom,"Search Engine optimization: From Analysis based on an Engineering Meta-Model towards Integrative Approaches", International Conference on Information Society (1-Society 2014)
21. P. Ravi Kumar, Ashutosh K. Singh, "Efficient Mehodologies to optimize website for Link Structure based Search Engines", 978-1-4673-6126-2/2013, IEEE
22. Y. Nawaz Ahmed Khan, Bhaskaran Raman, Shanmugasundaram Hariharan, "OCS- A system for optimizing, Clustering and Summarizing web Search Results using Intelligent Agents", 978-1-4211-4/2009, IEEE
23. A. Jain (2013), "The Role Of Off-Page Search Engine ranking, International Journal of Advanced Research in Computer Science and software Engineering Vol 3, Issue 6, pp. 239-244
24. Anamika Rajput, Sushil K. Chsaturvedi, "User Rating and Synonyms Based Modified Ranking Technique for Recommender Systems", 978-1-5090-0076-0/2015, IEEE, DOI 10.1109/CUCN. 2015.196
25. Wenpu Xing, Ali Ghorbani, "Weighted Page Rank Algorithm", 0-7695-2096-0 / 2004, IEEE
26. Rekha Singhal, Saurabh Ranjan Ranjan Srivastava, "Enhancing the Page Ranking for Search Engine

Optimization Based on weightage of In-linked web Pages", ICRAIE-2016, IEEE, 978-1-5090-2807-8/2016

BIOGRAPHY

Ritu Sachdeva (Sharma)is a Research Scholar in the CSE Department, M.V.N. University. She received M.Tech (CS) in 2012 from Jamia Hamdard University, New Delhi, India. Her research area is Information Retrieval.

Dr. Sachin Gupta is Associate Professor and Head of Deptt. (Computer Science &Electronics), MVNU