Machines and Algorithms
http://www.knovell.org/mna

Research Article

Learning Scholarly Network of Influential Authors Impact of author's position on H-Index

Masood Ahmed[1,*] and Khalid Iqbal[1]

[1]COMSATS University Islamabad (Attock Campus), Attock, 43600, Pakistan
[*]Corresponding Author: Masood Ahmed. Email: me.masood.ahmed@hotmail.com
Received: 7 December 2022; Revised: 30 December 2022; Accepted: 28 February 2023; Published: 6 March 2023
AID: 002-01-000020

Abstract: There have previously been several algorithms developed for academic networks to determine the author's productivity and influence. The primary goal of these algorithms is to compute bibliometric characteristics like as publication and citation counts. There are some that are similar to the h-index, I-index, and G-Index. However, all of these are primarily concerned with citation and publication counts, and they have certain drawbacks as well. All of these algorithms are primarily used to determine the productivity and influence of authors. They most important factor which these algorithms lack is identify the contribution of an author in a research paper. Our purpose is to create author's network using the available dataset (DBLP dataset) and then build position-based algorithm which take care of author position in research paper to find out the author actual productivity in related field. Base line of this algorithm will be h-index. To calculate the points author, gain due to his/her position in paper we will consider only that paper of an author which fall inside his h-index range.

Keywords: scholarly networks; H-index; author's position; g-index; social network metrics; DBLP; R & AR indices;

1. Introduction

We have different kind of networks in computer world; some of these include social networks, scholarly networks and computer network etc. For each network there are few basic requirement-like nodes, connectivity media etc. Research articles are regarded nodes in a scholarly network of writers. Each study article contains references to other studies that are related to the one under consideration. When a paper is quoted in another publication, its relevance score rises [1], affecting the author's worth as well as the journal or conference that published the work [2]. Different mathematical principles are used to examine these writers' work, both qualitatively and quantitatively [3].

When discussing how to assess the impact of a research project, we have two key components to consider: number and quality. Quantity refers to the number of research papers published by an author, whereas quality refers to the influence of a research article in its subject. Quantity is easier to quantify than quality since quality involves many other factors [4].

We already have few rankings and indexing algorithms to analyze the behavior of a scholarly network. The main purpose of these algorithms is to analyze an author work and find how important their work is and what a researcher has achieved in terms of impact of their work in research domain or productivity. Comparison is made among the researchers working in same field. This comparison related algorithm is being used by different organizations to allocate scholarships and funds to academic institutions on the basis of contribution to their respective research field and their performance in this field. In following we will discuss few of the ranking and indexing algorithms already developed to assess the research performance of an author.

Hirsch's H-index [5] is regarded as the first academic networks method that is used to measure the relevance of a scientist's work. The H-index is thought to be the best technique for determining the relevance of a scientist's research output in a certain field. Hirsch works shows that it is possible to not only identify the significance of a research work but it can also tell magnitude of discoveries and efforts made by authors.

New research work is being done continuously and as time pass by all new work is added in the scholarly network world. This new research work refers to research papers of other scientist who have worked or currently working in the same field or in same area of problem c current author is working on. Logically when the work of a research scientist is referenced by other researchers in their work then the significance of referenced work should be amplified. But when we are using the Hirsh's [5] method Specifically, the H-Index does not take into account new references made to the work in other researcher's papers after a publication has already been used to quantify the h-index, implying that the h-index does not rise.

To address the h-index issue, Egghe [6] created a more efficient approach that takes into consideration freshly added references to a publication that is already in the h-index. The approach developed by Egghe [6] is known as g-index. In this technique we have an edge "g-index" over h-index in a way to show maximum useful links of a research author in more visible and clearer way.

Once g-index is done we have another issue which is consider the evaluation of work. Work done by an old researcher and a fresh researcher cannot be combined. To address this issue, the "h-index" was devised to be viewed as h-index per year value. That is, the h-index should be divided by the number of years spent by scholars in their field of study. This is known as the m-index. [7].

Then we have the R-index [8]. The purpose of this index is same as what we have in index but this index is considered to be more appropriate than "g-index". Jon M. Kleinberg [9] proposed the Hypertext Induced Topic Selection (HITS) technique. It is consider being an ancestor of "PageRank" [10,11] algorithm. HITS use the link structure of a hypertext document as its input data to get the information related to the text of the paper. Same way the PageRank [10,11] algorithm ranks the page by getting the information from incoming links of document. To expedite the review and typesetting process, authors must follow the Microsoft Word template provided for preparing their manuscripts. This template must be strictly adhered to when formatting the manuscript for submission.

2. Literature Review

For quantifying an author's effect and productivity, a number of ranking and indexing algorithms have been created, some of which are detailed below.

2.1. H-index

At first, we have certain bibliometrics such as citation counts and paper counts. These indicators are routinely used to assess an author's success. There were several serious issues with these measures, such as the fact that the number of research publications represents the amount of effort and has nothing to do with the significance of the research activity. The amount and quality of work cannot be represented by a single numerical figure. This has a huge impact on how researchers work since it forces them to produce fewer papers in order to avoid having a low citation count. Hirsch's H-index [5] is regarded as the first academic networks method that is used to measure the relevance of research effort. The H-index is thought to be the best method for judging the relevance of a scientist's research output in a certain field. Hirsch's study demonstrates that it is feasible to determine not only the value of a research endeavor, but also the scale of discoveries and efforts made by writers.

Hirsch [5] proposed organizing the papers in descending order based on their citation count. Hirsch [5] assigned each document a serial number R beginning with 1. The researcher's h-index is now the sequence number at which R must be larger or equal to citation counts, but R+1 must be less than citation counts.

2.2. The g-index

When utilizing the Hirsh's [5], it is crucial to remember that the h-index does not rise if new references are added to a publication after it has been included in the quantified h index in other articles. Consequently, Egghe [6] developed a more useful method to account for the recently obtained references. The g index is the name of this novel method. The g-index is superior than the h-index in that it more clearly, conspicuously, and prominently reflects the maximal helpful output of a scientist.

The publications are sorted by citation count in decreasing order in which they have been recognized in other publications. The square rank of each publication is then computed in conjunction with the cumulative sum of citations to determine the g-index. The rank at which (rank)2 is less than the cumulative total of citations in a way that makes (rank+1)2 bigger than the cumulative sum is the g-index value.

Egghe's g-index is a highly helpful variant of the pioneer index, which is the h-index, even though it is not particularly popular or has attracted academic attention for scientific output. It inherits the h-index's benefits. It is superior to the h-index in that it presents a scientist's maximal helpful output in a clearer, more conspicuous, and more apparent manner. The H-index is a straightforward and basic calculation [5].

2.3. M-Index

The H-Index is not a good tool to evaluate young scientists' research projects. Using m-index is the workaround for this h-index constraint [7]. M-quotient is another name for the m-index. The value of the h-index is divided by the researcher's years to get the m-index [5, 7].

While the m-quotient is a useful metric for evaluating the work of novice researchers based on the number of years spent conducting research, one limitation of the m-index is that a small fluctuation in the h-index value will result in a significant shift in the m-quotient value [7]. Furthermore, the author of [5] stated that the h-index should not be determined only on a scientist's initial publication, especially if the author contributed very little to the work.

2.4. R and AR indices:

Even if a publication receives twice or three times as many citations after it is listed in the h index, these citations have no bearing on the h-index. Thus, authors with the same h-index but differing citation counts cannot be distinguished by the h-index. Let's take an example where writers A and B both have two published papers and a two h-index. Paper PA2 only obtained two citations, but Author A's piece PA1 received twenty. Similarly, author B received two citations for manuscript PB2 and five for article PB1. Both writers have the same rank and influence according to the h-index, yet author A has a bigger impact than author B. To address this issue, JIN BiHui et al. develop the R-index [7]. The following formula for the R index is suggested by the authors.

$$R = \sqrt{\sum_{j=1}^{h} cit_j}$$

R-index may be used to identify various scientists that appear to be the same when looking at the h index measure but are actually not. The AR-index is the option to use when it's necessary to give newer work more relevance and older work less significance. It's computed by:

$$AR = \sqrt{\sum_{j=1}^{h} \frac{cit_j}{a_j}}$$

The formula for calculating age is (k-0.5), where k is the number of years. When it is necessary to give recent articles a higher priority than older studies, this index is utilized.

2.5. HITS:

Jon M. Kleinberg [9] introduces the concept of Hypertext Induced Topic Selection (HITS). The PageRank algorithm may be traced back to its predecessor. It exploits the link structure of hypertext documents. The underlying assumption is that the link structure of a document can provide insights into its content.

The HITS algorithm, often referred to as the Hubs and Authorities algorithm, utilizes the concept of a hub. A hub is a page that lacks substantial informative material but is rich in links to other pages. It operates in a manner similar to a directory. The pages represented by the hub are referred to as authority pages due to their provision of authoritative information on a certain issue.

The process starts by identifying a collection of sites referred to as the root set, which are pertinent to a given query. Subsequently, the hub and authority values are computed for this set of pages. In order to measure the extent of authority, the collective magnitude of the pages that reference this specific authority is aggregated. When determining the size of the hub, the authorities' magnitudes are merged.

It is crucial to acknowledge that this process relies on the ideas of in-degree and out-degree, which can lead to problems like mutual reinforcing and subject drift [9]. Both of these challenges may be resolved by assigning weights to the algorithm.

2.6. The Page Rank

Larry Page [10] and Sergey Brin [11] introduced the concept of PageRank as a means to measure the relative significance of websites on the internet. The algorithm takes into account the number of links sent towards a page, however not all inbound connections are given equal weight. The incoming connections have a role in defining the quantitative importance depending on their relative worth. This approach operates by first assigning equal weight to all pages that have outbound links to a page A, whose rank is being sought. When a page has links to both page A and other external websites, these linkages are taken into account and the initial weight assigned to page A is evenly distributed across all of these links. Only a small portion of the weight of this page is currently utilized in determining the weight of the page being evaluated.

Incoming links that possess a high rank have a greater impact or influence. Web pages inside a web environment are equipped with forward connections known as outgoing links and backward connections referred to as incoming links. Estimating the forward links of an internet page is rather simple, however locating the back-links is exceedingly challenging. The stability and immutability of links on a webpage are not obligatory. If there exists a set of pages Q(k), where each page y has a hyperlink to page k, then the formula for PageRank [12] proposed by authors is as follows:

$$PR(k) = (1 - d) + d \sum_{y=Q(k)} \frac{PR(y)}{N(y)}$$

The symbol "d" in the above equation represents the damping factor. One drawback is that at starting, all outgoing and incoming connections are given identical weight, resulting in the rank being distributed evenly among them. However, in actuality, not all links hold the same level of relevance since various linkages have varying degrees of significance. Another limitation is that it is designed for a generic internet user, but not all users fit this profile. For instance, it may not consistently yield satisfactory outcomes for a researcher. PageRank assigns greater rankings to older pages, which might be considered a disadvantage of the algorithm.

2.7. Weighted PageRank:

W. Xing et al [12] describe a weighted version of the PageRank algorithm that adds weights to pages that use both forward and backward links rather than treating all web pages equally at first. The number of forward and backward linkages provides a decent concept for assigning weights. Web sites that are more popular are given a higher weight than web pages that are less popular [13].

Both forward and backward links help figure out how popular a web page is, and a page's rank is based on how popular it is. One problem with weighted PageRank is that, like the original PageRank, it only looks at forward and backward links. It doesn't look at other things, like how semantically sound a web page is. This is how the weighted PageRank [12] is shown mathematically:

$$PR(u) = (1 + D) + d \sum_{v \in B(u)} PR(v)\, W^{in}_{(v,u)} W^{out}_{(v,u)}$$

The above equation is based on two important metrics: and. These represent the back-links to u and the forward-links to and from u, respectively. The equation then figures out the score for the page u. Here's how to figure out both and:

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

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

In the last expression, Iu and Ou show how many backlinks and forward links there are for u, and Ip and Op show how many backlinks and forward links there are for p. The author of [13] used weighted page rank to figure out how well-known experts were. The following calculation was used to figure out weighted PageRank:

$$PR_{w(pi)} = (1 - d) * \frac{w(pi)}{\sum_{k=1}^{N} w(pk)} + d \sum_{pj \in M(pi)} \frac{PR\_w(pj)}{L(pj)}$$

3. PROPOSED METHODOLOGY

This study is divided into two sections. In the first, we will use an existing index, h-index, to identify the top-ranked author in our scholarly dataset, DBLP. Once these findings are obtained, we will apply our newly developed method to determine the author's h-index rank based on his/her position in various papers.

As we all know by the standard of research paper format, an author is positioned in the paper based on his/her contribution in the research work, person who contributed most in the research work is places in first position and person who contributed least in the research work is places at the last.

The second part of this research is mainly focused on this criterion, the algorithm which we are going to build will take in consideration the position of an author. We will apply three different point gain methods to find out the author's total point gain on the basis of author position in the research papers under the consideration and then we will take the average of it to find out the actual author point gain in research papers.

Mathematically it is given below:

$$PA = \frac{\sum_{i=1}^{3} PA_i}{3} \tag{1}$$

In the above equation, (1) 1-3 represent the total number of point gain methods we will use to determine the influential authors' points in research papers, represent an author's position point gain calculated using one measure, and PA is the sum of all point gain calculated using the proposed point gain measure.

The idea of point gain rank will be applied on results of these ranking and indexing algorithms to determine that how much percentage of authors are below from the current author. In order to find out the research consistency of an author during his academic career, standard deviation will be applied.

Below are the three proposed point gain algorithms we will use the find out the author points in his/her research paper.

3.1. Authors points gain based on total authors in paper

In first author's point gain method we will assign the point to each author based on his position in total number of authors. In this system, each author receives his or her fair portion of the overall number of citations in a publication. Following is the proposed algorithm to calculate point gain of an author:

i. Select an author from the list
ii. Calculate the h-index of each article this scholar generated. Select the paper which lie within the H-index of author
iii. Iterate through each paper of author we selected in last step. Calculate the points gain by the author based on following two values
   a. Number of citations of current paper
   b. Author position in the paper
   c. Author's point will be calculated using the following formula.
      i. $P_{(g)} = C_{a\,*}\,\frac{100-(10*\sum_{1}^{3} Pa)}{100}$
      ii. Here $P_{(g)}$ represents the point gain of an author in current paper, C_a represents the citation count of a paper, $\sum_{i}^{n+1} A_i$ represent the total author in a paper and P_a represent the position of author under
      consideration in current paper
iv. Sum all the points gain in last step.
v. Calculate the Authors total point using the following formula.

$$A_{(pg)} = \frac{\sum_{i}^{n} P(g)}{H_i}$$

Here $A_{(pg)}$ represents the author total point gain in all paper lies inside h-index, $\sum_{i}^{n} P_{(g)}$ represents the sum of point gain in all paper.

Example: Let's suppose we have a paper with citation count of 100, and the top author will receive 100 points, the second will receive 80 points, the third will receive 60 points, the fourth will receive 40 points, and the last will receive 20 points.

3.2. Author points gain based on threshold assign to position

In second author's point gain method we will assign the point to each author based on his position in total number of authors. In this methodology each author gets his/her due share with respect to total number of citations a paper has. Following is the proposed algorithm to calculate point gain of an author:

i. Select an author from the list
ii. Calculate the h-index of each article this scholar generated. Select the paper which lie within the H-index of author
iii. Iterate through each paper of author we selected in last step. Calculate the points gain by the author based on following two values
   a. Number of citations of current paper

b. Author position in the paper
c. Threshold value assign on position change
   i.e., in our case we assign 5 and after position # 19 all author will get equal share
d. Author's point will be calculated using the following formula.
e.
   i. $P_{(g)} = C_{a\,*}\ \frac{100-(5*\sum_1^{19} Pa)}{100}$
   ii. Here $P_{(g)}$ represents the point gain of an author in current paper, C_a represents the citation count of a paper, $\sum_1^{19} Pa$ represent the position of author under consideration in current paper

iv. Sum all the points gain in last step.
v. Calculate the Authors total point using the following formula.

$$A_{(pg)} = \frac{\sum_i^n P(g)}{H}$$

Here $A_{(pg)}$ represents the author total point gain in all paper lies inside h-index, $\sum_i^n P_{(g)}$ represents the sum of point gain in all paper

Example: As discussed in last example we have a paper with citation count of 100, and there are 5 authors involve in this paper then first author will get 100 points, second will get 95 points, third will get 90 points, fourth will get 85 points and last one will get 80 points. Incase total authors are more than 19 then all author from 19 onwards will get only 5 points.

3.3. Author points gain based on top five positions

In third author's point gain method we will assign the point to each author based on his position in total number of authors. In this technique points are given only to top five authors after that author receive no points. Following is the proposed algorithm to calculate point gain of an author:

i. Select an author from the list
ii. Calculate the h-index of each article this scholar generated. Select the paper which lie within the H-index of author
iii. Iterate through each paper of author we selected in last step. Calculate the points gain by the author based on following two values
   i. Number of citations of current paper
   ii. Author position in the paper
   iii. Author's point will be calculated using the following formula.
      i. $P_{(g)} = C_{a\,*}\ \frac{\sum_i^{n+1} A_i}{(\sum_i^{n+1} A_i - Pa)}$
      ii. Here $P_{(g)}$ represents the point gain of an author in current paper, represents the citation count of a paper, represent the total author in a paper and represent the position of author under consideration in current paper
   iv. Sum all the points gain in last step.
   v. Calculate the Authors total point using the following formula.

$$A_{(pg)} = \frac{\sum_i^n P(g)}{H_i}$$

Here $A_{(pg)}$ represents the author total point gain in all paper lies inside h-index, $\sum_{i}^{n} P(g)$ represents the sum of point gain in all paper

Example: Considering the same example we discussed in last two sections for a paper with citation count of 100, and there are more than 3 authors involve in this paper then first author will get 100 points , second will get 90 point , third will get 80 points, author listed after the 5th position will get no points from that paper.

4. EXPERIMENT AND ASSESSMENT

In this part we are going to discuss the setup we used for our data collection and for experiments on the data.

4.1. Data Collection

Data for the experiment was collected from Aminer portal, following data is gather from this portal. Below is the detail of data collected from Aminer portal.

Table 1: Complete data set

Parameter	Representation	No. of records
Paper	P	2,092,356
Citations	C	8,024,869
Author	A	1,712,433
Author Collaboration	AC	4,258,615

Once this data is collected, we created a collaboration network in our database. H-Index was calculated on whole dataset. After the calculation of H-index we created another network in which we import the top 100 author on H-index to perform experiment on less data to make it in presentable form below is the detail of dataset we use for all proposed algorithms.

Table 2: Top 100 Authors data set

Parameter	Representation	No. of records
Paper	P	18,767
Citations	C	562,391
Author	A	100
Author Collaboration	AC	63, 313

4.2. Experimental setup

By using the Microsoft visual studio 2013 data was imported in Microsoft SQL server 2008 database. Once the data is imported, we executed our proposed algorithm using the SQL server query editor. For implementation the Dell system was used having processor of 2.40 GHz core i5 with 8GB RAM. We executed all three proposed algorithms and then store their result in excel sheet to generate graphs.

5. EXPERIMENTAL RESULTS

After applying all three proposed algorithms on the dataset and then calculating their average it shows that author position plays a very important role when calculating the author ranking. Results shows that a high ranked author in H-index list does not mean his contribution are also more than the other authors. A low ranked author over takes the high ranked just because of his position i.e., contribution in some influential papers.

In first step we calculated the H-index so that we know the list of influential authors we have on the basis of H-Index. Once we have calculated the list of influential authors, we created a separate network to top 100 authors.

Below table shows the top 25 influential authors from that list calculated on the bases of H-Index using the DBLP data set.

Table 3: Top 25 Authors using H-index

Author	H-Index
Hector GarciaMolina	60
Scott Shenker	56
Jiawei Han	53
David E. Culler	51
Anil K. Jain	50
Chris Faloutsos	50
Jeffrey D. Ullman	49
Ian Foster	46
P S Yu	46
Christos H. Papadimitriou	45
D Estrin	45
Hari Balakrishnan	45
Jennifer Widom	45
Jon M. Kleinberg	45
W Bruce Croft	44
Ben Shneiderman	43
Rakesh Agrawal	43
T. Anderson	43
M. Naor	42
Michael Stonebraker	42
Mihir Bellare	42
Moshe Y. Vardi	42
Pat Hanrahan	42
Tom Henzinger	42
David A. Patterson	41

Once we have the list of influential authors using the h-index our next step was to execute our proposed algorithms on top 100 influential authors to find out the impact of author positioning on his/her ranking. Following are the tables which shows the result we obtain by applying the proposed algorithm one by one

5.1. Author points gain based on total authors in paper

Table 4: Top 25 author using total author algorithms

Author	H-Index
Rakesh Agrawal	305.06
Leslie Lamport	180.55
E. M. Clarke	176.60
A Shamir	165.05

Anil K. Jain	160.87
Jiawei Han	154.01
Ian Foster	143.43
James N. Gray	136.80
Oded Goldreich	130.98
Jeffrey D. llman	129.92
Mihir Bellare	124.22
Philip A. Bernstein	122.07
Rajeev Alur	119.73
R. E. Schapire	119.53
Jon M.Kleinberg	119.29
Robert Morris	118.98
Dan Boneh	116.03
R Motwani	114.26
Serge Abiteboul	109.23
David R. Karger	109.13
Ronald Fagin	107.70
Andrew R.	107.28
S Osher	106.20
P Raghavan	102.75
I. Stoica	101.78

This list shows that the author which were part of top 25 influential author list using h-index are no more there which means that author position does have impact on the ranking of author and also it helps in calculating how much contribution an author made so far regardless of h-index. To verify our proposed research methodology, we apply our two other proposed algorithm to see how much impact they make in calculation of points gain. Table 5 & 6 the result we calculated using the "Author points gain based on threshold assign to position" and "Author points gain based on top five positions" respectively.

Table 5: Top 25 author using threshold algorithm

Author	**H-Index**
Rakesh Agrawal	316.24
Jeffrey D. Ullman	223.87
A Shamir	211.79
Leslie Lamport	196.03
Robert Morris	194.85
Jitendra Malik	183.03
Anil K. Jain	182.51
E. M. Clarke	180.65
Jiawei Han	178.34
Hari Balakrishnan	177.64
R Motwani	167.94
David E. Culler	163.89
Michael I. Jordan	159.21
R. E. Schapire	158.42

P Raghavan	156.83
M.Frans Kaashoe	155.83
Ian Foster	155.33
David R. Karger	148.85
S Osher	143.85
James N. Gray	143.57
Andrew Zisserman	142.35
Jon M. Kleinberg	142.28
Philip A. Bernstei	140.43
Oded Goldreich	139.96
I. Stoica	138.65

Table 6: Top 25 author using top five author algorithms

Author	**H-Index**
Rakesh Agrawal	295.08
Jeffrey D. Ullman	192.46
A Shamir	191.44
Leslie Lamport	181.89
E. M. Clarke	167.62
Anil K. Jain	167.02
Jiawei Han	162.08
Robert Morris	159.43
Jitendra Malik	157.89
Hari Balakrishnan	143.97
R Motwani	143.65
R. E. Schapire	141.52
Ian Foster	141.28
David E. Culler	135.51
P Raghavan	133.47
Michael I. Jordan	132.85
James N. Gray	129.73
Oded Goldreich	129.16
Philip A. Bernstei	128.80
Jon M. Kleinberg	127.97
S Osher	126.87
David R. Karger	124.94
Andrew Zisserma	122.21
M. Frans Kaashoek	121.88
Andrew R. McCallum	121.76

As we can see there is not a lot of difference in second and third algorithm but result in first algorithm are a bit different from these twos, to overcome this issue we decided to take the mean of results we achieved through these algorithm

Table 7: Top most author comparison

H-Index Top Most Author		**Position base Top Author**	
Hector Garcia-Molina		Rakesh Agrawal	
Position	Paper count	Position	Paper count
1	77	1	117
2	211	2	25
3	70	3	20
4	22	4	5
5	8	5	1
6	5	6	0
7	1	7	0
8	2	8	0
9	2	9	0
Total paper	398	**Total paper**	168

In Table 8 we listed all 25 authors and with their h1-index, points gain through three proposed algorithm and their final contribution points.

Table 8: Top 25 author complete table

Author	**HI**	**Proposed Methods A**	**Proposed Methods B**	**Proposed Methods C**	**Proposed Methods Points**
Rakesh Agrawal	43	305.06	316.24	295.08	305.46
A Shamir	34	165.05	211.79	191.44	189.43
Leslie Lamport	33	180.55	196.03	181.89	186.16
Jeffrey D. Ullman	49	129.92	223.87	192.46	182.08
E. M. Clarke	40	176.60	180.65	167.62	174.96
Anil K. Jain	50	160.87	182.51	167.02	170.13
Jiawei Han	53	154.01	178.34	162.08	164.81
Robert Morris	36	118.98	194.85	159.43	157.75
Jitendra Malik	35	99.21	183.03	157.89	146.71
Ian Foster	46	143.43	155.33	141.28	146.68
R Motwani	40	114.26	167.94	143.65	141.95
R. E. Schapire	35	119.53	158.42	141.52	139.82
Hari Balakrishnan	45	93.81	177.64	143.97	138.47
James N. Gray	33	136.80	143.57	129.73	136.70
Oded Goldreich	40	130.98	139.96	129.16	133.37
David E. Culler	51	100.46	163.89	135.51	133.29
P Raghavan	38	102.75	156.83	133.47	131.02
Philip A. Bernstein	34	122.07	140.43	128.80	130.43
Jon M. Kleinberg	45	119.29	142.28	127.97	129.85
David R. Karger	36	109.13	148.85	124.94	127.64
S Osher	35	106.20	143.85	126.87	125.64
Michael I. Jordan	35	83.53	159.21	132.85	125.20
M. Frans Kaashoek	37	88.31	155.83	121.88	122.01

6. CONCLUSION AND FUTURE WORK

A detailed analysis of effect of author position in research paper by selecting the research paper based on their h-index is covered in this paper. We have also compared the top most author of h-index and our position base ranking it clearly shows that author who top the position base ranking has a smaller number of papers as compare to h-index top author and also, he was first position author is his 2/3 of papers whereas h-index top rank author was first position author in his 1/5 of research papers see table 7. These algorithms can evolve over time, which means they can search a bigger area and change the goal threshold values on the fly. In this paper we only took the H-index as our base point to find out most effective papers in future we can use other scholarly indices to gather the affective paper and also, we can explore more on authors collaboration with each other using the author network we created.

References

[1] Mingers, John. "Measuring the research contribution of management academics using the Hirsch-index." *Journal of the Operational Research Society* 60, no. 9 (2009): 1143-1153.

[2] Yan, Erjia, and Ying Ding. "Discovering author impact: A PageRank perspective." *Information processing & management* 47, no. 1 (2011): 125-134.

[3] Katsaros, Dimitrios, Leonidas Akritidis, and Panayiotis Bozanis. "The f index: Quantifying the impact of coterminal citations on scientists' ranking." *Journal of the American Society for Information Science and Technology* 60, no. 5 (2009): 1051-1056.

[4] Kleinberg, Jon M. "Authoritative sources in a hyperlinked environment." *Journal of the ACM (JACM)* 46, no. 5 (1999): 604-632.

[5] Hirsch, Jorge E. "An index to quantify an individual's scientific research output." *Proceedings of the National academy of Sciences* 102, no. 46 (2005): 16569-16572.

[6] Egghe, Leo. "Theory and practise of the g-index." *Scientometrics* 69, no. 1 (2006): 131-152.

[7] Harzing, Anne-Wil, and Ron van der Wal. "A Google Scholar H-Index for journals: A better metric to measure journal impact in economics & business." In *Proceedings of the Academy of Management Annual Meeting*. Wiley, 2008.

[8] Jin, Bihui, LiMing Liang, Ronald Rousseau, and Leo Egghe. "The R-and AR-indices: Complementing the h-index." *Chinese science bulletin* 52, no. 6 (2007): 855-863.

[9] Kleinberg, Jon M. "Authoritative sources in a hyperlinked environment." *Journal of the ACM (JACM)* 46, no. 5 (1999): 604-632.

[10] Page, Larry, Sergey Brin, Rajeev Motwani, and Terry Winograd. *PageRank: Bringing order to the web*. Vol. 72. Stanford Digital Libraries Working Paper, 1997.

[11] Brin, Sergey, and Lawrence Page. "The anatomy of a large-scale hypertextual web search engine." *Computer networks and ISDN systems* 30, no. 1-7 (1998): 107-117.

[12] Xing, Wenpu, and Ali Ghorbani. "Weighted pagerank algorithm." In *Proceedings. Second Annual Conference on Communication Networks and Services Research, 2004.*, pp. 305-314. IEEE, 2004.

[13] Ding, Ying. "Applying weighted PageRank to author citation networks." *Journal of the American Society for Information Science and Technology* 62, no. 2 (2011): 236-245.