Development G-Index and H-Index : Dgh-Index

Ahmed AbdulHassan AL-Fatlawi
Computer Science and Information Technology, University of Sumer

Abstract

Of the most important indicators such as the h-index ,a method of measuring the productivity and impact of an academic's work, is often used as a component or metric in the ranking of higher education institutions and their staff then proposed the g index as a modification of the h index. So the g-index, have been Trying to develop and improve the disadvantage in h-index. and although the g-index can provide a more comprehensive measure of scientific contribution, but value g-index is integer where two authors or more than may be get the same g-index value although different number of citations and papers, making it difficult to differentiate performance between authors so in paper we suggest improvement index to resolve this problem is called dgh-index that gives new features to g-index and h-index that give us real number not an integer .

INTRODUCTION

The twentieth century may be described as the century of the development of metric science. Among the different metrics Scientrometric is the most interesting subject area in the field of library and information science, which can be applied to any discipline irrespective of their period of evolution. It involves quantitative studies of scientific activities. It is also one such useful metrics/technique which helps to solve the problems, challenges posed by so called information explosion. Over the years, several new terms have appeared in library and information science. They were known as

Librametrics (1940's) , Bibliometrics (1960's) Informetrics and Webometrics /Cybermetrics as the study of the quantitative aspects of the construction and use of information resources , structure and technologies on the www drawing on bibliometrics and informetric approaches[14,24].

The Scientometric the science of measuring the "quality" of science. It is a subject has seen tremendous growth in recent times. The availability of information through internet has made it much more interesting topic and eased the data collection process. Today large data can be analyzed easily with assistance web based technologies, there is need to use scientometrics for measuring the research output of an organization/institutions or individuals. Many of the universities know measure the research output of its scientists through scientometric tools such has h-index, g-index and citations for research publications. It comprises all the metrics studies related to science indicators , citation analyses , research evaluation ,etc. and the scientometrics research uses various online database, indices and tools in order to establish relationships between authors or their work.

Although a huge debate exists about which is the best methodology for the assessment of research performance of individual scientists, the use of different quantitative bibliometric indicators to support expert judgment is widely accepted as a good approach to improve objectivity and fairness in the evaluative process [15,17,24]. Specifically, the combined use of multiple quantitative indicators, instead of relying in just a single indicator, is strongly recommended by different authors [25,9]. However, several simplified indexes to characterize the significance of the scientific output of researchers have been suggested many of indexes and one of them the h-index is one of the most popular ones[12]. This index has been very well accepted by the scientific community [2] and has been calculated for scientists of different fields as well as the h-index is so popular that Scopus and Web of Science (WoS), less than two years after the introduction, have decided to present it as an indicator furthermore, "an entire issue of the journal Scientometrics was recently devoted to the h-index ,and the measure is now automatically calculated in the 'citation report' function of Web of Science" [18].. And by definition of the h-index, the papers on rank 1,…,h each have at least h citations, hence these h papers together have at least h2 citations and be the h-index better than other single-number indicators used to evaluate the scientific output of researchers from through (impact factor, total number of documents, total number of citations, citations per document rate and number of highly cited papers) but found some disadvantage in the evaluation of the output of scientists . In order to decrease some of the disadvantage of h-index, different modifications appearance to the h-index [4, 22].

One of the most interesting improvements is the "g-index", defined by Egghe [5]. It is defined as follows: A set of papers has a g-index g if g is the highest rank such that the top g papers have, together, at least g^2 citations. This also means that the top g+1 papers have less than $(g+1)^2$ papers. As well as g-index cannot be larger than the number of articles and academic has published.

What is G-index?

Egghe proposed the g index to improve on the h-index and to be a very useful complement to the h-index for assessment outputs the scientists by giving more weight to highly -cited articles [5] . The aim is to avoid a

disadvantage of the h index   that "once a paper belongs to the top h papers, its subsequent citations no longer 'count' " [10] . And also this give for g-index a better measurement of authors' scientific contribution and the g-index is defined as follows:
 A set of articles ranked in decreasing order of the number of citations that they received, the g-index is the (unique) largest number such that the top g articles received (together) at least g2 citations. From this definition it is already clear that g ≥ h. And we can show it through the mathematical theory of the g-index based on Lotka's law (The general Lotkaian theory of the g-index is presented) and we show that

$$g = \left(\frac{\alpha - 1}{\alpha - 2} \right)^{\frac{\alpha - 1}{\alpha}} T^{\frac{1}{\alpha}}$$

hence by Glänzel or   Egghe and Rousseau [8,5,21], since one showed there that h = $T^{\frac{1}{\alpha}}$ we have:

$$g = \left(\frac{\alpha - 1}{\alpha - 2} \right)^{\frac{\alpha - 1}{\alpha}} h$$

Here α is the Lotka exponent and T denotes the total number of sources (in the citation application).

ADVANTAGE AND DISADVANTAGE FOR G-INDEX AND H-INDEX:
The Advantages of h-index [16]
1. The h-index work as a tool for quantifying the scientific productivity and the scientific impact of an individual researcher: "A scientist has index h if h of his or her n articles have at least h citations each, whereas the other n-h articles have at most h citations each.
2. h- index is a simple single number incorporating publication as well as citation data (hence comprising quantitative as well as qualitative or visibility aspects)
3. It is a mathematically simple index and easy to understand
4. It performs better than other single-number indicators used to evaluate the scientific output of researchers, such as (impact factor, total number of documents, total number of citations, citations per document rate and number of highly cited papers).
5. The h-index can be easily obtained by anyone with access to the Thomson ISI Web of Science .
6. The h-index is also robust in the sense that it is insensitive to an accidental set of lowly cited papers and also to one or several outstandingly highly cited papers.

Disadvantage of h-index[16]:
1. H-index depend on length of the scientific career therefore, puts newcomers at a disadvantage the h-index should provide a more realistic assessment of the academic achievement of academics that have started publishing at least 10 years ago.
2. Its inability to differentiate clearly between active and inactive scientists .
3. Ignores   h-index the number of citations to each individual article over and above what is needed to achieve a certain h-index.
4. The value of h-index limited by the total number of document
5. H-index should provide a more realistic assessment of the academic achievement of academics that have started publishing at least 10 years ago. I would argue that for more senior academics, assessing the impact of their own publications is preferable to assessing the journal impact factor of the journals they publish in. The latter is only a measure of how of the average article in the journal is cited.
6. It puts newcomers at a disadvantage since both publication output and citation rates will be relatively low. To solve this problem Hirsch proposes the m quotient which is computed by dividing the h index by the scientific age of the author.
7. It does not account for the number of authors in a paper. To address this problem Batista suggest to adjust the original h index by dividing it by the mean number of researchers in the h publications that determine the h index[4]. The new index is named individual h index.

Advantage of g-index:
1. G-index from more traditional bibliometric indicators in their ability to discriminate among different types of scientists or researchers in more sensitive   with a selective publication strategy (scientists with intermediate productivity but a high impact).
2. The g index gives more weight to highly cited papers. The aim is to avoid a disadvantage of the h index

and thus highlighting the impact of authors.

3. Egghe modified the index by replacing the idea of calculating the number of citations received by each article with calculating the total accumulated citations of the top g articles[5].

4. Egghe pointed out that the g-index value will always be higher than the h-index value and lower than the total publication number[5]. It compensates a shortcoming of h-index, which is insensitive to authors with few and lowly-cited (or non-cited) papers. The higher values of g-index make it easier to differentiate the performance of authors.

5. The value of the g-index is not limited by the total number of documents. This necessary to rank by decreasing order of citations all the documents of the unit.

Disadvantage of g-index:

1. The g-index is limited in discriminatory power because it is defined to be integers. This means that several researchers may receive the same g-index although they have significantly different numbers of citations.

2. The values g-index are integers and many authors may get the same g-index value, making it difficult to differentiate performance.

3. g-index is not an appropriate indicator when evaluating a small group of authors.

Calculation to g-index and h-index

In order to obtain the g-index of a scientist or other unit of analysis, it is necessary to rank by decreasing order of citations all the documents of the unit. By definition g-index which states that the rank r is the g-index: r = g of this system if r is the highest value such that $g(r) \geq r^2$.

Also the g-index is defined as the highest rank such that the cumulative sum of the number of citations received is larger than or equal to the square of this rank .So the position where the square of the rank position is equal to the accumulated number of citations corresponds to the g-index .also can be calculated h-index from the position where the rank is equal to the citation. In example will know how calculate the g-index and h-index for any researcher at mode simple and from through using the program is called pop (perish or publish) where giving information about number papers publication and total number of citations and the rank for each author but we calculated h-index and g-index for one author.

	Authors	Paper	h-index	g-index
1	Seifedine karee			
2	Ali Kalakh			
3	Osama Zen			
4	Hader Harmanani			
5	Abbas Hajj			
6	Ali Hamie			

Table 1.1: Before calculation the g-index and h-index for the authors

In the table below select author Ali Kalackech:
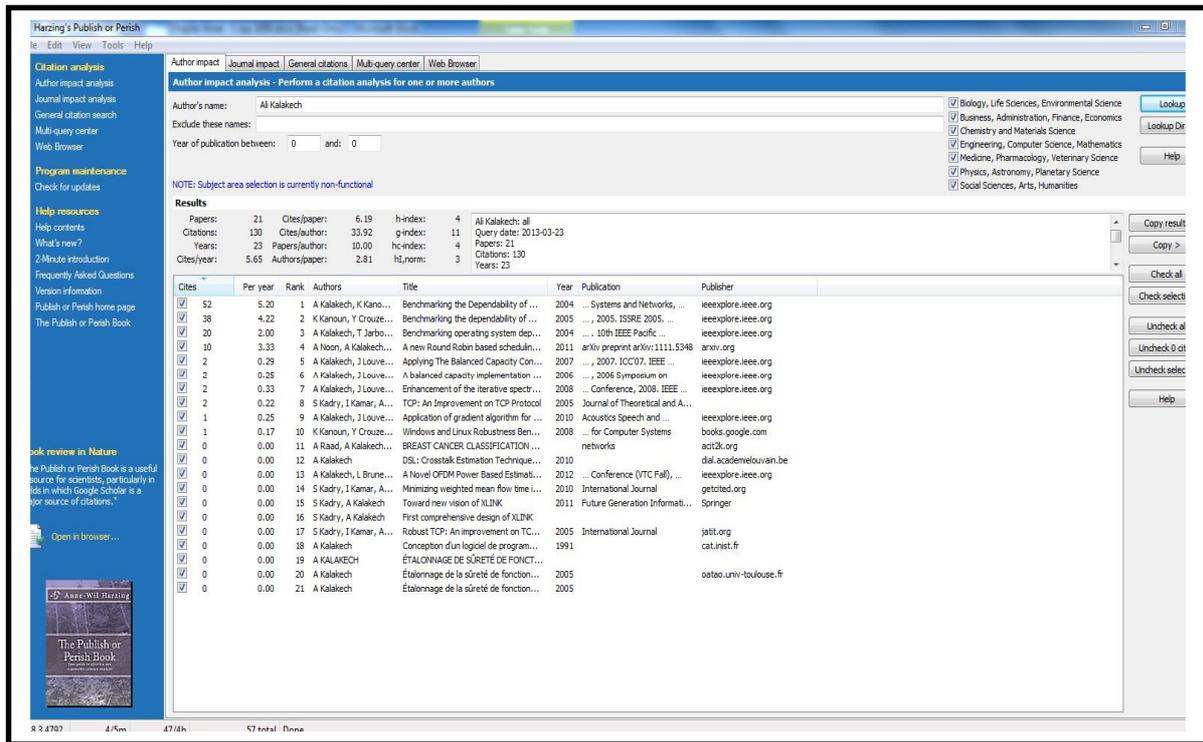
Figure 1.1 : Software Pop using to find g-index and h-index

Table 1.2: The calculate g-index for the author(ali Kalakech)

Cites (TC)	Rink	∑ Cites (∑ TC)	Rink ²
52	1	52	1
38	2	90	4
20	3	110	9
10	4	120	16
2	5	122	25
2	6	124	36
2	7	126	49
2	8	128	64
1	9	129	81
1	10	130	100
0	11	130	121
0	12	130	144
0	13	130	169
0	14	130	196
0	15	130	225
0	16	130	256
0	17	130	289
0	18	130	324
0	19	130	361
0	20	130	400
0	21	130	441

In the table 1.2, used TC denote to for the total number of citations for each paper on rank and ∑ TC denote to for the cumulative number of citations to the papers on rank 1,...,r (for each r) in the example g-index equal 11. Indeed this is the highest rank such that the top 11 papers have at least(11)^2=121 citations (here 130>121); on rank 12 we have $(130 < (12)^2 =144)$ citations and h-index =4 that is the highest rank such that all papers on rank 1,…,h have at least 4 citations (and hence the papers on rank 5 or higher have not more than 4 citations).

After we calculated g-index of author Ali Kalakech as an example and will fill the table 1.3 (for all authors) such as same way calculate g-index.and h-index.

Table 1.3:after calculation the g-index and h-index

	Auther	paper	h-index	g-index
1	Seifedine kadry	71	6	8
2	Ali Kalakech	20	4	11
3	Oussama Zein	26	5	5
4	Hader Harmanani	49	8	17
5	**Abbas Hajj**	21	2	7
6	Ali Hamie	45	9	17

The Development g–index and h-index

Current research suggests that these indexes (h-index and g-index) do not substitute each other but that they are complementary. Where the h index depends on both the number of a scientist's publications and the impact of the papers on the scientist's peers. This index has been very well accepted by the scientific community [2], but it may not discriminate sufficiently well between researchers. So proposed g-index in order to overcome the disadvantage of h-index and The g-index presents two important improvements as compared to h-index: first, the weighting of the citations received by the documents is considered in the g-index calculation; and secondly, the g-index for a given scientist is not limited by his/her total number of publications. According to these features, g-index might be more adequate than h-index for assessing selective scientists, who are less likely to obtain high values and although the g-index can provide a more comprehensive measure of scientific contribution, we believe that the values g-index are integers and many authors may get the same g-index value, making it difficult to differentiate performance so we are in this chapter we will deal with this problem by using a new index suggest to resolve this problem is called dg-index that gives new features to g-index and that give us real number not an integer number as well as that depends on the number of citation at each scientist this will be a step to make a different between 3 scientists have the same g-index and different number of papers.

Mathematical for the Dgh-index:
First we will give a mathematically exact definition of the g-index, The Web of Science (WoS) allows an automatic arrangement of the publication list in decreasing order according to the number of citations C(r), where r is the rank attributed to the paper.
Calculating the sum S(r) of the number of citations up to rank r,

$$S(r) = \sum_{r'=1}^{r} C(r') \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

allows us to determine the g-index analogously from

$$g^2 \le s(g) \text{ while } s(g+1) < (g+1)^2 \ldots\ldots\ldots\ldots\ldots(2)$$
$$\text{Need of citation's (NoC)} = (g+1)^2 - s(g) \ldots\ldots\ldots\ldots\ldots(3)$$
$$\text{Need of Article (NoA)} = \begin{cases} g = \text{Numberof Article} & \text{NoA} = 1 \\ g \ne \text{Number of Article} & \text{NoA} = 0 \end{cases} \ldots(4)$$

The relation (2) formally expresses the above specification that g is the highest number of articles that together received g2 or more citations.

$$\text{Dg-index} = g + \frac{1}{NoC + NoA + 0.001}$$

The Fraction of this number orient us to know how many citation you need to be g +1 such as the dg-index will be like 8.64 that mean the 0.64 calculated by dividing 1 by the summation of the needed of citation to be g+1 and need of articles by calculated the dg-index we will understand more detail about the scientist.

$$g < Dg < g+1$$

the Dg-index is improvement g-index to solve problem in same g-index for more one scientist but different number of citation's.

we found new index and by using Dg- index by depended on h-index is called Dhg - index, the Dhg - index is same the idea hg index is introduced by Alonso [1], and defined as the geometric mean of the h- and g-indices of a researcher:

$$hg = \sqrt{Dg * h}$$

we see such as in table (3.1) and Must be the result of the Dg less than g +1, the Dg-index flowchart is represented as a step of plan.
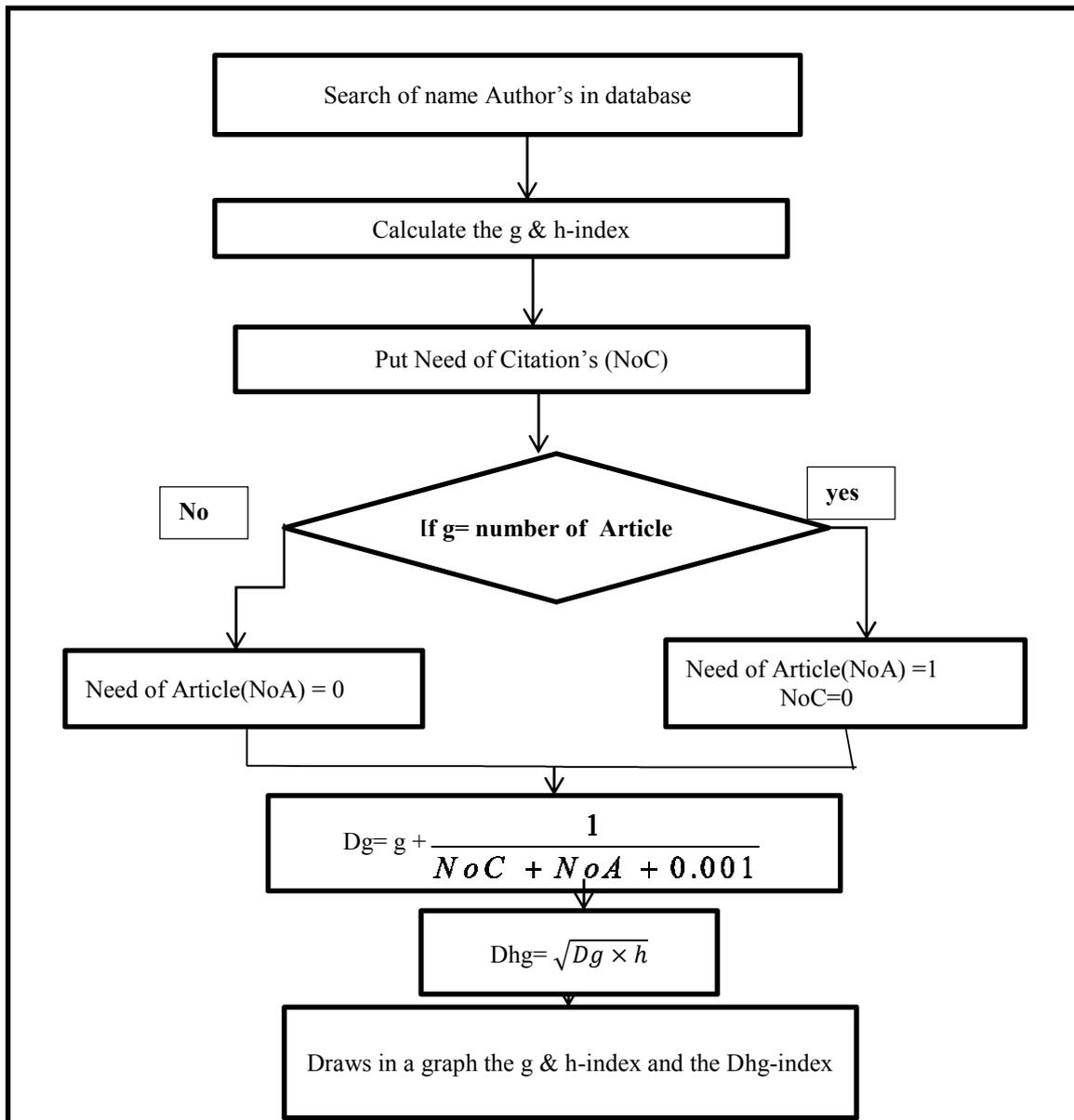
Figure 1.2: Flowchart for Dgh-index

Simulation

In this simulation, we suppose that we have database contain information for author's (number of papers and number of citations)can consider is worked Web of science (WoS) and we want to calculate g-index and Dg-index. We design interface of program contain one choice, the choice to calculate dgh-index by author's and show the result in graph.

Figure 1.3: Interface of programming to calculate the Dgh-index
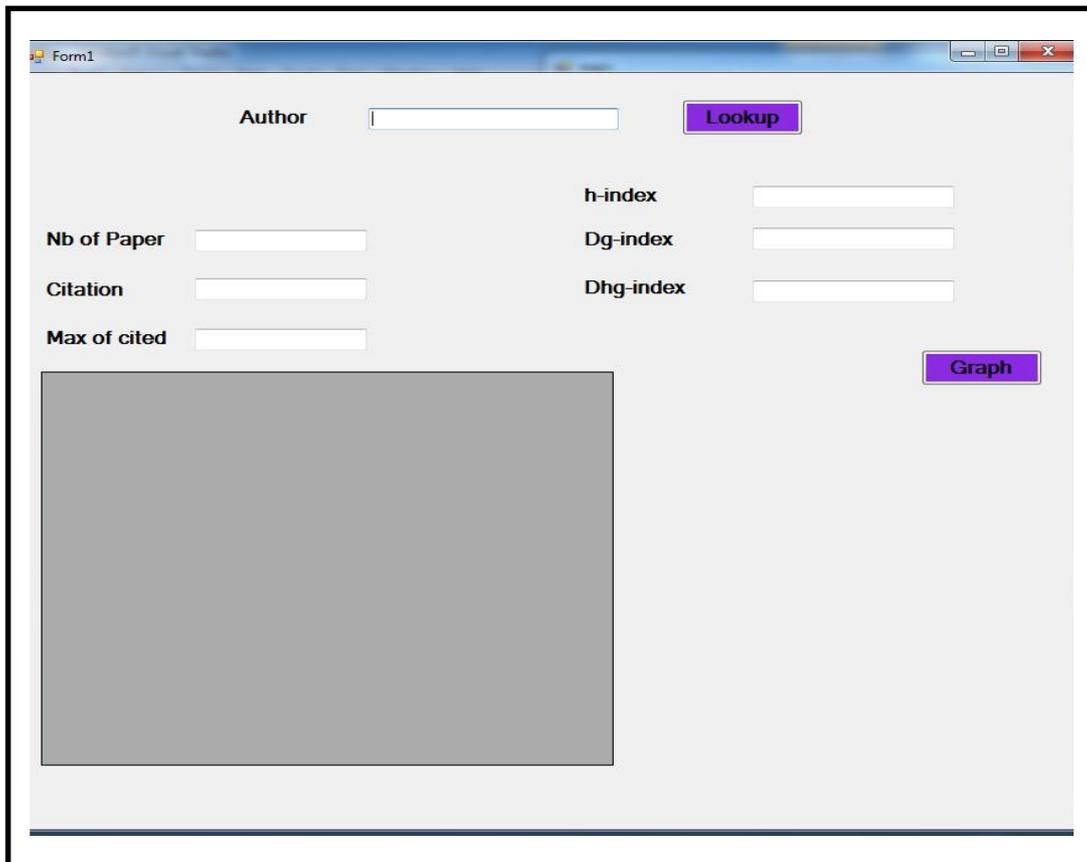
Figure 1.4: Interface of programming to calculate the Dg-index for one author

Table 1.3: The same g-index for three authors

Author's	Article	citations	h-index	g-index
Ahmad Ghaddar	13	14	2	3
Hassan al-shalabi	4	12	2	3
Hussien chible	5	10	2	3
Ali hamdoun	8	31	3	5
Oussama zein	26	35	5	5
Walid fahs	9	35	1	5
Abbas khalaf	98	48	4	6
Ahmad faour	10	40	4	6
Imad issa	22	42	4	6
Abbas Hajj	21	56	2	7
Danielle Azar	16	62	5	7
Abu Salem, Fatima	23	67	4	7
Seifedine Kadry	76	102	6	8
Sanaa Sharafeddine	35	101	6	8
Hussein Hajj	24	67	6	8
Ali Kalakech	21	130	4	11
Azzam Mourad	34	140	8	11
Mcheick, Hamid	60	140	5	11
Ali hamie	45	345	9	17
Haidar harmanani	49	309	8	17
Keirouz , walid	25	298	9	17

We select three authors have the same g-index and h-index from table (1.3) as shown in the example [Dr. Abbas khalaf , Dr. Ahmad faour and Dr. Imad issa] .

In the application calculate the dgh-index by depending on (The software Pop which provides us with information on number of papers and number of citations etc.)

1.   Dr. . Abbas khalaf

Figure 1.5: The software Pop for author (Dr. . Abbas khalaf)

NoC=12, NoA = 0 , g=6, h=4 , Dgh =4.932

Figure 1.6:The interface of calculate Dgh-index for Dr. Abbas khalaf
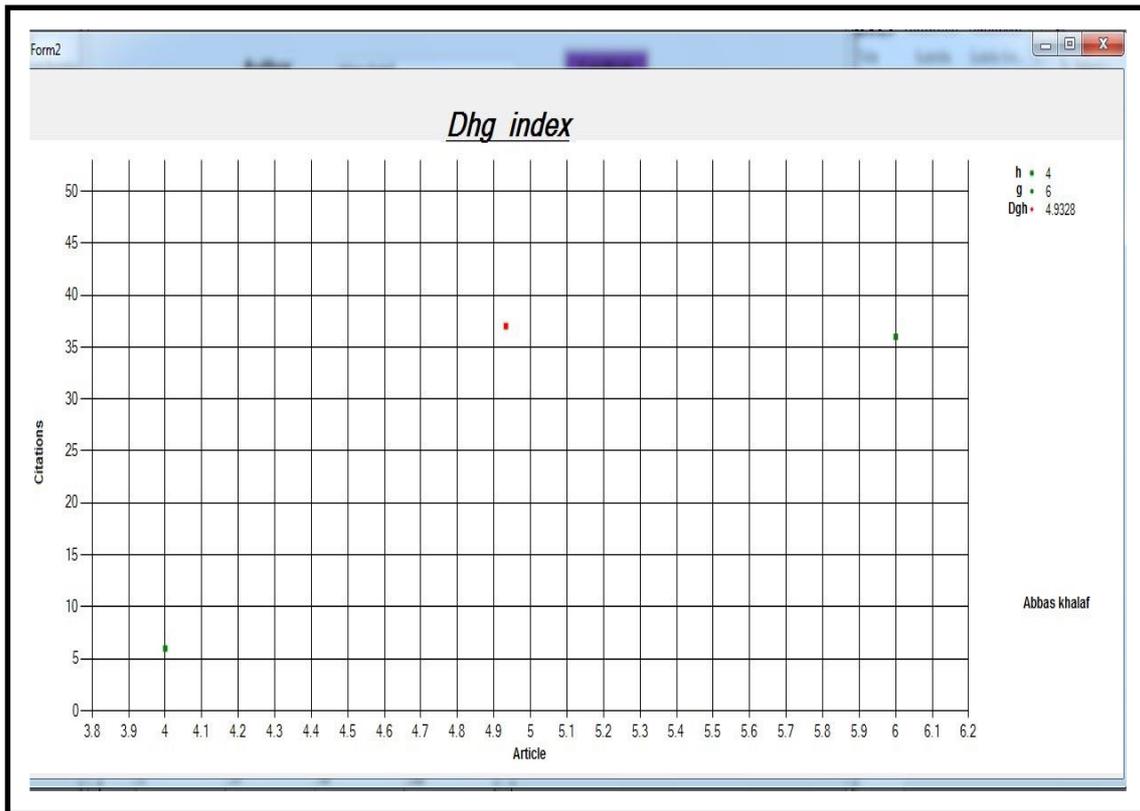
Figure 1.7 :The draw Dgh-index for Dr. . Abbas khalaf

2.   Dr. Ahmad faour
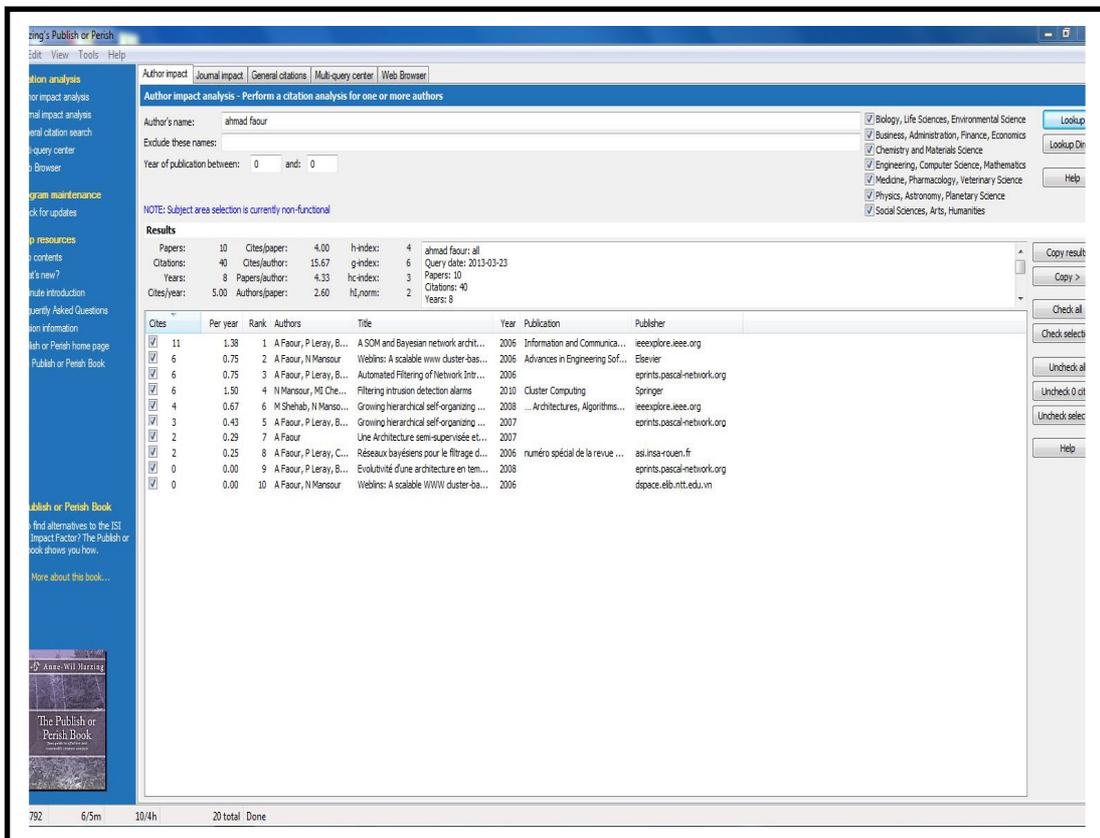
Figure 1.8: The software Pop for author (Dr. Ahmad faour)

NoC = 14 , NoA = 0 , g=6, h=4 , Dgh=4.930
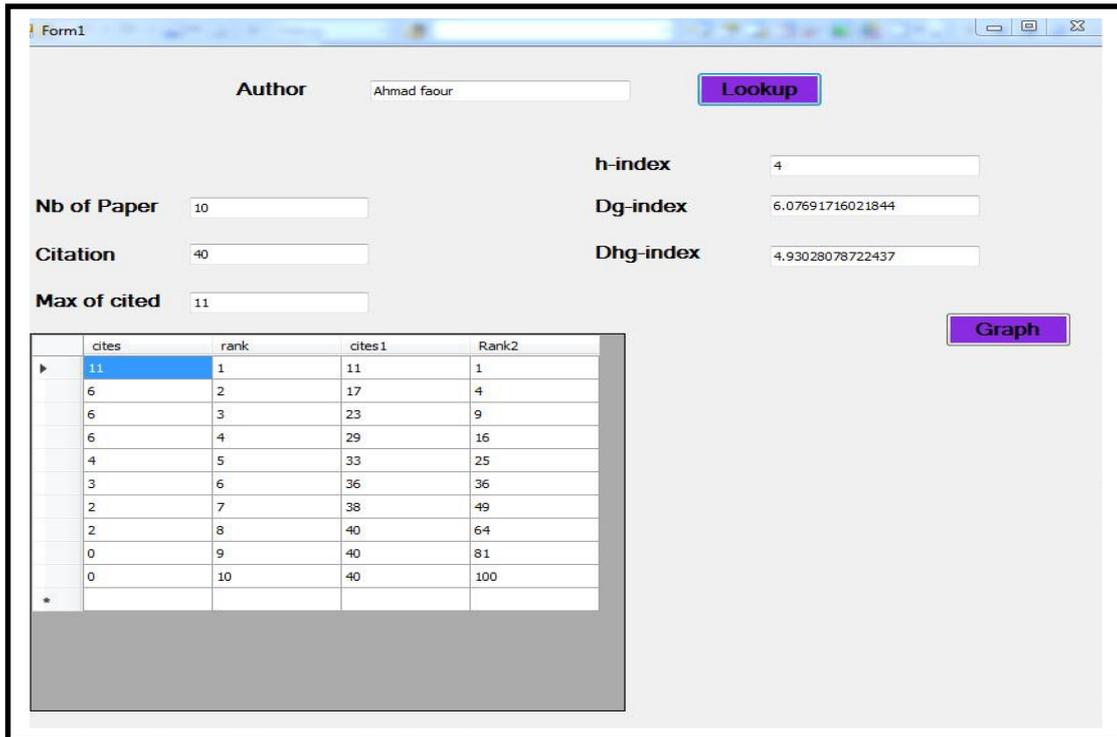
Figure 1.9: The interface of calculate Dgh-index by Dr. Ahmad faour
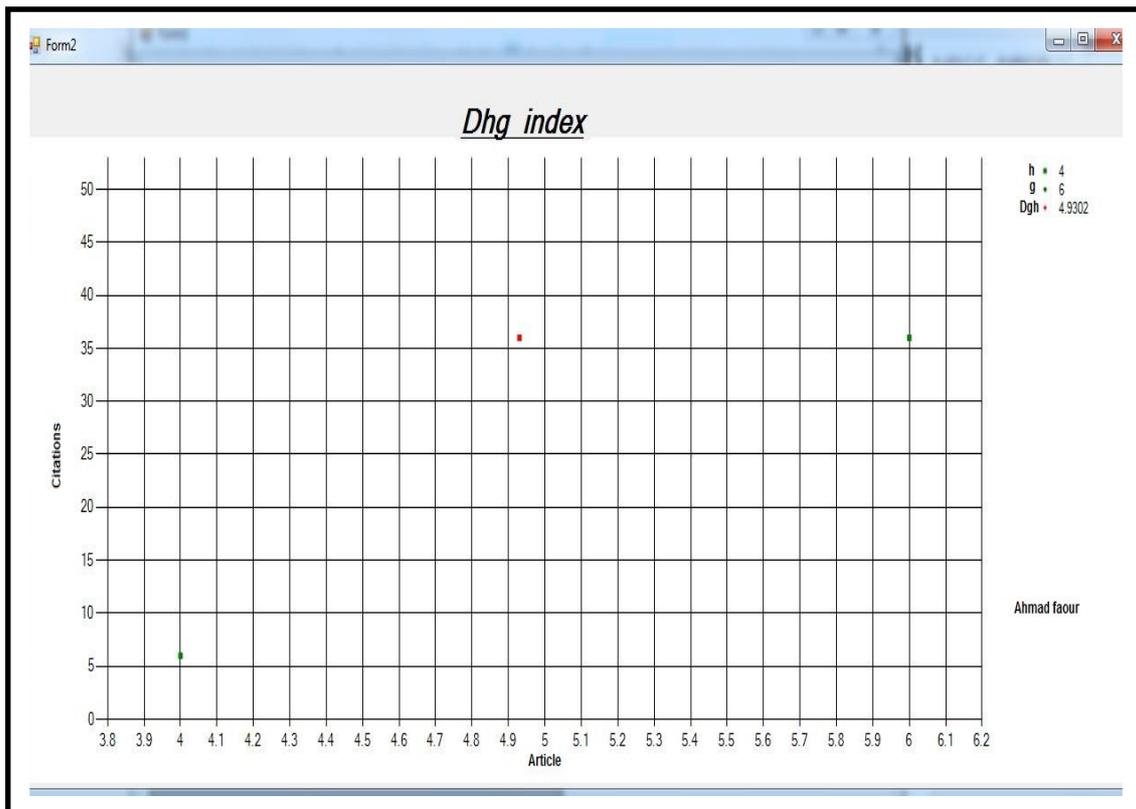
Figure 1.10 : The draw Dgh-index for Dr. Ahmad faour

3. Dr Imad issa

Figure 1.11: The software Pop for author (Dr. Imad issa)

NoC = 9 , NoA = 0 , g=6 ,h= 4 ,Dgh= 4.944

Figure 1.12: The interface of calculate Dgh-index for Dr. Imad issa
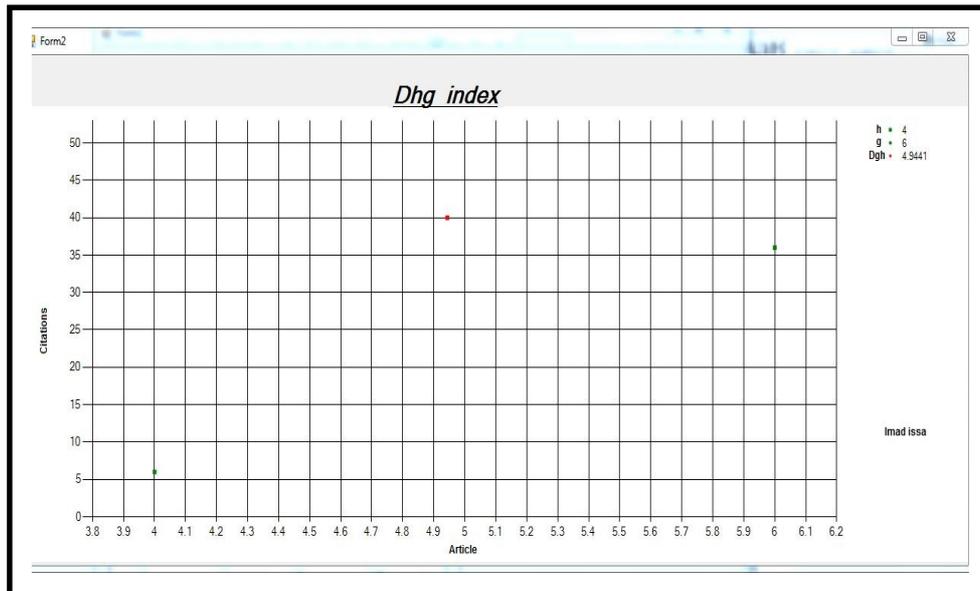
Figure 1.13 :The graph of Dgh-index for Dr. Imad issa

Result

We reach in simulation to solution to overcome problem g index and h-index by index is called Dgh-index Which is considered complementary for Dg-index and improvement to solve problem If more than scientist have the same value g-index but different number of citations and number of papers such as in table (1.1) and Fill the table after calculate Dgh-index, must be the result of the Dg less than h, such as in the table (1.2).

Table 1.4: The Dgh-index for the authors

Author's	Article	citations	h-index	g-index	Dg-index	Dgh-index
Ahmad Ghaddar	13	14	2	3	3.333	2.582
Hassan al-shalabi	4	12	2	3	3.249	2.549
Hussien chible	5	10	2	3	3.1666	2.517
Ali hamdoun	8	31	3	5	5.199	3.949
Oussama zein	26	35	5	5	5.124	5.061
Walid fahs	9	35	1	5	5.999	2.499
Abbas khalaf	98	48	4	6	6.0833	4.932
Ahmad faour	10	40	4	6	6.076	4.930
Imad issa	22	42	4	6	6.249	4.999
Abbas Hajj	21	56	2	7	7.124	3.775
Danielle Azar	16	62	5	7	7.249	6.123
Abu Salem, Fatima	23	67	4	7	7.499	5.476
Seifedine Kadry	76	102	6	8	8.076	6.961
Sanaa Sharafeddine	35	101	6	8	8.14	6.998
Hussein Hajj	24	67	6	8	8.071	6.958
Ali Kalakech	21	130	4	11	11.07	6.654
Azzam Mourad	34	140	8	11	11.05	9.402
Mcheick, Hamid	60	140	5	11	11.11	7.453
Ali hmaie	45	345	9	17	17.111	12.409
Haidar harmanani	49	309	8	17	17.0416	11.676
Keirouz , walid	25	298	9	17	17.0384	12.383

Conclusions

Evaluating research quality is an important but quite difficult issue in scientific communication. Particular attention must be paid to appropriate bibliometrics methods. The most well-known index is the h-index which is easy to calculate and to understand.

However, a disadvantage is that it loses some valuable information about scientific contribution and so the g-index was proposed. Although the g-index can provide a more comprehensive measure of scientific contribution, we believe that some important information is still wasted and it may not discriminate sufficiently well between researchers.

We have proposed Dgh-index is amendments to the g-index and h-index which includes the need of citations and need of articles. Theoretically, the calculations of these indices were explained in details, and practically, the results of the comparison study show the researchers contain the same g-index and different of number of article and number of citations this is consider disadvantage of g-index, It can be clearly seen that the g-index depend to number of citation.

References

Alonso, S., Cabrerizo, F., Herrera-Viedma, E., Herrera, hg-index: a new index to characterize the scientific output of researchers based on the h- and g-indices, 2010.

BALL, P, Index aims for fair ranking of scientists. Nature, 2005.

Bar-Ilan, Judit, "Informetrics at the beginning of the 21st century A review",2008.

BATISTA, P. D., CAMPITELI, M. G., KINOUCHI, O., MARTINEZ, A. S, It is possible to compare researchers with different scientific interests?,2006.

Egghe , Theory and practice of the g-index, 2006.

Garfield, E, Citation Analysis as a Tool in Journal Evaluation, 1972.

Garfield, E, Citation Indexing: Its Theory and Application in Science, Technology, and Humanities.1979.

GLANZEL, W, On the h-index – A mathematical approach to a new measure of publication activity and citation impact. Scientometrics, 2006.

Glänzel, W. On the opportunities and limitations of the H-index. Science Focus, 2006.

Harzing , A.-W. Reflections on the h-index. Retrieved 23 April 2008, from http://www.harzing.com/pop_hindex.htm.

Hess, D. J. Science Studies: An advanced introduction. New York: New York University Press,1997.

Hirsch JE , An index to quantify an individual's scientific research output.2005.

Jin, B.H., Liang, L.M., Rousseau, R., Egghe, L., The R- and AR-indices: Complementing the h-index, 2007.

Lennart Björneborn and Peter Ingwersen, "Toward a basic framework for webometrics",2004.

LEWISON, G., COTTRELL, R., DIXON, D, Bibliometric indicators to assist the peer review process in grant decisions.1999.

Liu, YX. and Rousseau, R.. Hirsch-type indices and library management: the case of Tongji University Library, 2007.

MARTIN, B. R, The use of multiple indicators in the assessment of basic research. Scientometrics, 1996.

Meho, L. I., & Yang, K, Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. Journal of the American Society for Information Science and Technology, 2007.

NSIDIROPOULOS, A., KATSAROS, D., MANOLOPOULOS, Y, Generalized h-index for disclosing latent facts in citation networks, 2007.

R. Karpagam, S. Gopalakrishnan and M. Natarajan :Scientific measures and tools for research literature output,Indian Journal of Science and Technology, July 2011.

Rousseau, "Simple models and the corresponding h-and g-index", Science Technology Development, 2006.

Sidiropoulos , Dimitrios Katsaros , B Yannis Manolopoulos," Generalized Hirsch h-index for disclosing latent facts in citation networks",2007.

Tague-Sutcliffe, Jean , "An introduction to informetrics",1992.

VAN LEEUWEN, T. N., VISSER, M. S., MOED, H. F., NEDERHOR, T. J., VAN RAAN, A. F. J, The Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence , 2003.

VAN RAAN, A. F. J, Comparisons of the Hirsch-index with standard bibliometric indicators and with peer judment for 147 chemistry research groups, 2006.

Wouters , The Citation Culture. Amsterdam: Unpublished Ph.D. Thesis, University of Amsterdam,1999.

Author map, http://project.cis.drexel.edu/authorlink/.

Bibcouple,http://users.fmg.uva.nl/lleydesdorff/software/bibcoupl/index.htm.

Cites pace ,http://cluster.cis.drexel.edu/%7Ecchen/citespace/.

Clean Pop , http://cleanpop.ifris.org/guide.html.

Co-auth, http://users.fmg.uva.nl/lleydesdorff/ software/coauth/index.htm.

Fulltext , http://users.fmg.uva.nl/lleydesdorff/software/fulltext/index.htm

HistCite , http://www.histcite.com/index.htm.

ISI , http://users.fmg.uva.nl/lleydesdorff/software/isi/index.htm.

Patent Pictures, http://www.researchinformation.info/ rijanfeb 04patents.html.

Publish or perish, http://www.harzing.com/pop.htm.

RefViz ,http://www.refviz.com/.

TI ,http://users.fmg.uva.nl/lleydesdorff/software/ti/index.htm.