CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Evaluation of Author Ranking Parameters based on Publications, Scientific Services and Affiliations in Computer Science Domain

by

Saman Saeed

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Computing
Department of Computer Science

2018

Dedicated to

My Parents,

My Teachers,

and

My Friends

who are always the source of Inspiration for me and their contributions are uncounted.

CAPITAL UNIVERSITY OF SCIENCE & TECHNOLOGY
ISLAMABAD

CERTIFICATE OF APPROVAL

Evaluation of Author Ranking Parameters based on Publications, Scientific Services and Affiliations in Computer Science Domain

by

Saman Saeed

MCS163022

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Ejaz Ahmed	FAST-NU, Islamabad
(b)	Internal Examiner	Dr. Muhammad Azhar Iqbal	CUST, Islamabad
(c)	Supervisor	Dr. Muhammad Tanvir Afzal	CUST, Islamabad

Dr. Muhammad Tanvir Afzal
Thesis Supervisor
November, 2018

Dr. Nayyer Masood
Head
Dept. of Computer Science
November, 2018

Dr. Muhammad Abdul Qadir
Dean
Faculty of Computing
November, 2018

Author's Declaration

I, **Saman Saeed** hereby state that my MS thesis titled **"Evaluation of Author Ranking Parameters based on Publications, Scientific Services and Affiliations in Computer Science Domain"** is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Saman Saeed)

Registration No: MCS163022

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Evaluation of Author Ranking Parameters based on Publications, Scientific Services and Affiliations in Computer Science Domain**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been dully acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Saman Saeed)

Registration No: MCS163022

Acknowledgements

In the name of ALLAH, the Most Beneficent and the Most Merciful, all worship and praise is for ALLAH (S.W.T), the creator of whole worlds. First and leading, I want to express profound gratitude to Him for providing me the courage, guidance and blessings to complete this research work. Secondly, special thanks to my respected supervisor Dr. Muhammad Tanvir Afzal for his assistance, valuable time and guidance. I sincerely thank him for his support, encouragement and advice in the research area. He enabled me to develop an understanding of the subject. He has taught me, both consciously and unconsciously, how good experimental work is carried out. Sir you will always be remembered in my prayers.

I am highly beholden to my parents, brothers, and my friends for their assistance, support (moral as well as financial) and encouragement throughout the completion of this Master of Science degree. This all is due to love that they shower on me in every moment of my life. No words can ever be sufficient for the gratitude I have for my parents. I hope I have met my parents' high expectations.

I pray to ALLAH (S.W.T) that may He bestow me with true success in all fields in both worlds and shower His blessed knowledge upon me for the betterment of all Muslims and whole Mankind.

Abstract

In the last few years, researchers' assessment has received a lot of attention from scientific community. The assessment of researchers plays an essential role in making the consequential decisions in industry such as: promotions, tenure, grants, renowned awards, fellowships and memberships of scientific societies and recruitment. The researchers' ranking community has proposed diversified parameters for researchers' assessment which include Publication count, Citation count, h-index and its 37 variants and extensions. In scientific literature, there is continuous debate on choice of parameters which provide appropriate ranking of researchers. In addition to the mentioned parameters, researchers make other contributions e.g. supervision of PhDs and volunteer scientific services to renowned journals. Furthermore, based on the performance and eminence of researchers in the field, they are granted with fellowships of scientific societies, Journal editorial memberships and institutions affiliations. Such **"Scientific Services and Affiliations"** based parameters have been exploited in the literature for different purposes. However, such parameters have not been comprehensively assessed and compared with each other for ranking the researchers. Therefore, this research is conducted to evaluate the impact of **"Scientific Services and Affiliations"** based parameters such as: **'Graduate Students Supervisions'**, **'Academic Institutions Affiliations'**, **'Scientific Societies Fellowships'**, **'Journal Editorial Memberships'** and **'Geographical Location'** and to compare with primitive parameters, citation intensity based parameters and age of publications based parameters for researchers' assessment.

To evaluate the proposed parameters with the existing parameters, the benchmark of international prestigious awardees in Computer Science domain have been utilized. We have considered 24 different renowned awards of 2 computer science societies yielding 1306 distinct awardees. Furthermore, we have added noise to this dataset which contains 1049 computer scientists who have not received any award and belong to the high, average, and low H-index values. The motive for this is to investigate whether the proposed parameters distinguishes scientific awardees from those researchers who have not received any award but belong to the all types of profiles. The aim of this research work is to investigate the: **1)** existence of international prestigious awardees in the top of ranked lists, **2)** correlation between proposed parameters and existing parameters and **3)** ranking of proposed parameters and existing parameters for classification.

The results demonstrate that the proposed parameters have outperformed existing parameters in distinguishing scientific societies awardees from those researchers who have not received any distinction.

Contents

List of Figures

List of Tables

Abbreviations

ACM	Association for Computer Machinery
IEEE	Institute of Electrical and Electronics Engineering
JUCS	Journal of Universal Computer Science
h-index	Hirsch Index
NB	Naive Bayes
KNN	K-Nearest Neighbor
SVM	Support Vector Machine
GSS	Graduate Student Supervisions
AIA	Academic Institutions Affiliations
SSF	Scientific Societies Fellowships
JEM	Journal Editorial Memberships
GL	Geographical Location

Chapter 1

Introduction

In recent years, the most striking aspect in academia was that how researchers are evaluated for promotions, recruitment, tenure, grants and renowned awards. Therefore, various parameters have been proposed by scientific community for different purposes such as for researchers' assessment to find domain expert, to grant fellowships, memberships and renowned awards, to hire talented faculty and to find reviewers of conferences and journals.

Researchers are continuously contributing in scientific field. To measure the researchers output in research field, various autonomous techniques are used in literature such as: publication count [1–3], citation count [4–6], h-index [7]. Furthermore, 37 variants of the h-index have also been proposed as highlighted by Bornmann et al [8] for example, g-index [9], A-index [10], R-index [10], Hg-index [11], m-quotient [7], AR-index [10] and m-index [12]. All of such parameters are totally dependent on publications and their citations.

For researchers' assessment, there is a continuous debate on choice of parameters. In addition to the mentioned parameters, researchers are also supervising PhDs and MS/MPhil thesis. Furthermore, based on the performance and eminence of researchers in the field, they are granted with fellowships of scientific societies, Journal editorial memberships and institutions affiliations. Such "Scientific Services and Affiliations" based parameters have been exploited in the literature for different purposes for example: - Rensburg et al suggest that the supervision of postgraduate students plays a vital role in the development of next generation practitioners who have the accurate educational skills to accomplish the future research desires [13]. Chan et al consider editorial board membership to rank institutions [14]. However, such parameters have not been comprehensively assessed and compared with each other for ranking the researchers. Therefore, this research is conducted to evaluate the impact of mentioned "Scientific Services and Affiliations" based parameters and to compare with primitive parameters, citation intensity based parameters and age of publications based parameters.

For this research, we need a comprehensive benchmark dataset which should consists of researchers acclaimed highly regarded by scientific community. Some previous studies have also considered award winners of scientific societies as benchmark to evaluate researchers ranking parameters [15–17]. We have adopted similar methodology and have considered 24 prestigious awards of two Computer Science Societies (ACM and IEEE) to evaluate and find which "Scientific Services and Affiliations" based parameters are associated with primitive parameters, citation intensity based parameters and age of publications based parameters.

1.1 Background

Researchers' assessment can be measured by various autonomous techniques. A very simple and earlier technique is publication count. According to this technique, a researcher is an expert of an area if he/she has more publications than other researchers in that area [2]. The main advantage of utilizing publication count is that it is easily available. However, publication count lacks to cover the true effect and eminence of one's work. The limitation of considering only the number of publications has been identified by Cameron et al in 2007 [2]. For instance, E.F. Codd and Hector Garcia Molina are known for their work in databases. E.F. Codd was the father of relational database and is considered more prolific than Hector Garcia Molina. But by only considering number of publications, Hector Garcia is considered more knowledgeable person than E.F. Codd. Because the publications count of Hector Garcia is larger than E.F. Codd publication count. E.F. Codd has 49 publications in DBLP, whereas Hector Garcia-Molina has 248 publications till the death of Codd. If we consider publication count to find expert, Hector Garcia-Molina would be ranked expert in the field of databases. However, E.F. Codd has a lot of contribution in the field of database and the impact and quality of his work is higher than Hector Garcia. This example depicts that a researcher with greater number of publications, by default may be considered more potential expert as compared to another researcher with fewer publications. However, publication count alone is not sufficient to find potential experts. Therefore, there is a need to find another parameter for finding experts.

As a substitute of publication count, citation count was proposed for the assessment of researchers. According to Yang et al study, a researcher is ranked higher if the citation count of his/her publications is higher than others [5]. Citation count has some limitations; some authors cite others' work only to criticize which may increase number of citations for him by West et al [6]. Furthermore, the survey papers normally get more citations which restricts the original work to be ranked higher than the survey papers [4].

To overcome the shortcomings of publication and citation count, J.E Hirsch introduced h-index which was the combination of two or more author ranking parameter such as publication count and citation count [7]. H-index for an author is defined to be h if and only if h of his publications have at least h citations each, and the other papers have less than or equal to h citations each. H-index has its own limitations: h-index mainly based on long time contribution of authors. Therefore, newcomers who have low publication and citation rate are less likely to be ranked expert beside their quality work [18]. Therefore, it is important to choose a parameter which incorporates better assessment of researchers.

Due to the disadvantage of the h-index, several h-index variants and extensions have been proposed i.e. g-index [9], A-index [10], R-index [10], Hg-index [11], AR-index [10], M-index [12] and M-quotient [7]. We can use all these parameters for researcher's assessment. These parameters have their own advantages and limitations.

In addition to the mentioned parameters, researchers have many other contributions such as supervisions of PhD's and are also granted with academic institution affiliations, scientific societies fellowships and journal editorial memberships. Such "Scientific Services and Affiliations" based parameters have been utilized for different purposes in literature.

[3] Employed editorial board memberships of academic journals to rank universities and also made comparison with primitive parameters such as publication count and citation count.

Chan and Torgler examine the association between scientific societies distinctions received by researcher's such as fellowships of economics societies and Nobel laureates [19]. They observed positive association between researchers' received society's fellowships and won Nobel Prize. Another interesting finding revealed by his study was that these fellows and Nobel laureates were also affiliated with world top universities such as Harvard and MIT.

Hall suggested to Australian mathematical society to utilize graduate student's supervisions as parameter for evaluation [20]. Variation exists in supervising number of graduate students from one area to another in mathematics. Consequently, supervisor should have both amount of scholarly research and the level of intellect to guide graduates. Most of Australia's top theoretical mathematicians have had relatively few graduate students during their careers. It requires an incredible level of knowledge for a supervisor to begin with doctoral research. On the other hand, research trends changes quickly in other areas of mathematics where significant results can be accomplished effectively by utilizing tools that are learned in advanced undergraduate courses.

There is no comprehensive research on aforementioned "Scientific Services and Affiliations" based parameters for particular domain to rank researchers. Therefore, these parameters should be evaluated for researchers' assessment.

1.2 Research Gap

As it has been identified from critical analysis, different h-index variants such as g-index, A-index, m-index, m-quotient, AR-index and hg-index have been proposed by scientific community and no one evaluated "Scientific Services and Affiliations" based parameters for particular domain comprehensively to rank researchers. Therefore, this thesis attempts to perform comprehensive analysis of "Scientific Services and Affiliations" based parameters with primitive, citation intensity and age of publications based parameters for Computer Science domain.

Secondly, primitive parameters, citation intensity based parameters and age of publications based parameters have been evaluated for finding correlation between different indices and utilizing for researchers' ranking. However, the effect of "Scientific Services and Affiliations" based parameters has not been comprehensively studied for finding association with primitive, citation intensity and age of publication based parameters in Computer Science domain. Therefore, we are going to use "Scientific Services and Affiliations" based parameters to find out the correlation of primitive parameters, citation intensity based parameters and age of publications based parameters with "Scientific Services and Affiliations" based parameters.

1.3 Research Questions

Based on the research gap mentioned in the previous section, this thesis evaluates the mentioned author ranking parameters against the list of international prestigious awardees as benchmark. The Computer Science domain was considered as a dataset in this thesis. This research has formulated following research questions:

1.3.1 Research Question 1

Which author ranking parameter out of above mentioned 15 parameters is able to bring international awardees in top ranking?

1.3.2 Research Question 2

Which parameter from scientific services and affiliations is best correlated with each parameter from other ranking parameter categories (primitive parameters, citation intensity based parameters, and Age of publications based parameters)?

1.3.3 Research Question 3

What is the ranking of all mentioned 15 parameters for classification?

1.4 Purpose

The aim of this research is to evaluate the impact of various "Scientific Services and Affiliations" based parameters with respect to traditional primitive parameters, citation intensity based parameters and age of publications based parameters and build up an argument that: Can "Scientific Services and Affiliations" based parameters replace primitive parameters, citation intensity based parameters and age of publication based parameters?

1.5 Scope

The scope of this thesis is the evaluation of "Scientific Services and Affiliations" based parameters with primitive parameters, citation intensity based parameters and age of publications based parameters in Computer Science Domain. The

"Scientific Services and Affiliations" based parameters include: "Graduate Students Supervisions", "Academic Institution Affiliations", "Scientific Societies Fellowships", "Journal Editorial Memberships" and "Geographical Location" and primitive, citation intensity and age of publications based parameters include publication count, citation count, h-index, g-index, A-index, R-index, hg-index, m-quotient, AR-index and m-index. The dataset contains researchers of prestigious international awards i.e. ACM and IEEE. We will exploit this dataset to assessed aforementioned parameter in Computer Science domain.

1.6 Application of Proposed Approach

The results obtained from this research has potential in various applications. Few of them are listed below:

- Decision makers can use these results to make important decisions related to the hiring of faculty in universities, giving awards, assigning memberships and fellowships of societies, hiring reviewers or editors for academic journals and government fund allocation.
- Scientific community can use these results to devise a more effective author ranking parameter for finding experts in different domains.
- Researchers can consider these results to increase their scientific contributions.

1.7 Limitations

- We have considered comprehensive dataset which contained Computer Scientists with high, average and low h-index values. It was comprehensive enough to evaluate existing parameters and proposed parameters. Though, limitation is that it does not cover all researchers of Computer Science domain.

- It was difficult to collect "Scientific Services and Affiliations" based parameters for researchers in our dataset because we have manually collected these parameters. However, it is possible that it might not covered all information of researchers.

Chapter 2

Literature Review

Today, a large number of researchers are contributing in scientific community. Therefore, scientific community needs to discover potential experts which can guide to make some decisions in industry. According to James [18] and Raheel et al [17] study, assessment of authors can help to make some decisions like 1) Who ought to get renowned award? 2) Who should serve as editor and reviewer for a scientific journal? 3) Who ought to get fellowship and membership of a scientific society? 4) Who should help universities in recruiting talented faculty? 5) Who has larger influence in the particular research field? Assessing researchers in specific field is often crucial for consulting in academia. An assortment of measures has been proposed to assess researchers in different areas. After critical analysis of related literature, we will discuss primitive parameters which include publication count [1–3, 21], citation count [4–6, 16], and h-index [7, 12, 18, 22, 23] in section 2.1. We will discuss citation intensity based parameters such as g-index [9, 11], A-index [10], R-index [10] and Hg-index [11] in section 2.2. We will discuss age of publications based parameters such as m-quotient [7, 24, 25], AR-index [10] and m-index [12] in section 2.3. We will discuss the "Scientific Services and Affiliations" based parameters such as: number of graduate students supervised by researchers [13, 26], number of academic institutions affiliations [27, 28], number of fellowships received from scientific societies [19, 29, 30], journal editorial memberships [14] and geographical location of researchers [31] in section 2.4. We will also discuss studies that evaluated primitive parameters, citation intensity based parameters, age of publications based parameters and "Scientific Services and Affiliations" based parameters in section 2.5.

2.1 Primitive Parameters

After careful analysis of related literature, various primitive parameters have been proposed to find valuable experts in different domains such as publication count [1–3, 21], citation count [4–6, 16] and combination of publications and citations called h-index [7, 12, 18, 22, 23]. This section provides details of primitive parameters by highlighting their strengths and weaknesses.

2.1.1 Publication Count

According to publication count based ranking, a researcher is an expert of an area if he/she has more publications than other researchers in that area [2]. According to Crowder et al study, a researcher is an expert if he/she has highest number of publications, greatest number of grants and extensive experience in particular field [1]. The main strength of utilizing publication count is that it is a well-known measurement and the strong evidence of his/ her knowledge in academic field [21]. However, there is continuous debate in scientific community that only publication count cannot be used for scientist's assessment because they give equal importance to the author work [2]. Publication count does not measure research quality [3].

2.1.2 Citation Count

To overcome the shortcomings of publication count, citation count measure was proposed for the assessment of researchers [4]. Citation count represents the number of citations received by a particular paper. In citation count, the author will be ranked on top if he/she has more citations than others [5, 16]. Citation count has maximum use in scientific community to rank researcher. Citation count has some flaws; some authors cite others' work only to criticize which may increase number of citations for him by West et al [6]. Moreover, survey papers normally get more citations than the original work [4].

2.1.3 H-index

Researchers proposed the new parameter to overcome the weaknesses of publication count and citation count (citation of publications). Hirsch proposed h-index which was a combination of publication count and citation count [7]. Hirsch assess authors according to their h-index. H-index of an author is calculated by considering the number of citations received by his/her most cited publications [7].

"A scientist has index h if h of his or her Np papers have at least h citations each and the other $(N_p - h)$ papers have fewer than ≤h citations each".

Table 2.1: H-Index Calculation

No. of Publications	Citations recieved
1	11
2	10
3	9
4	8
(h) 5	5
6	3
7	2

In Table 2.1, if a researcher has 7 publications and they were cited 11, 10, 9, 8, 5, 3 and 2 times then the researcher h-index will be 5 as the 6th paper has less than 6 citations when papers are sorted with respect to citations in descending order.

Considering the limitations of publication and citation count, Afzal et al adopted hybrid approach to find Computer Science experts of Journal of Universal Computer Science (JUCS) [22]. Their study combines multiple parameters to find experts. Their research only provided ranking of domain experts within JUCS which does not provide overall ranking of experts for an area.

H-index has several advantages: it is easy to compute and it takes both the quantity and the impact of the scientist's research publications [12].

H-index has its own weaknesses that number of citations can raise with time still if an author does not publish any new paper. The value of h-index does not reduce with the age of author. So, new researchers who have low publication and citation rate are less likely to be ranked expert beside their quality work [18]. Furthermore, number of citations can raise with time still if an author does not working and publishing any new paper. The value of h-index does not reduce with the age of author.

According to the Costas and Bordons study, h-index takes a measure of quantity (Publications) and impact (citations) [23]. H-index enable us to quantify the scientific output of a single researcher. But, they identified some hidden weaknesses such as researchers career length problem. For instance: a new researcher who have just started their career 1 or 2 years ago, will have shorter h-index because they will not have any chance to be cited more. Therefore, it is important to choose a parameter which incorporates better assessment of researchers.

2.2 Author Ranking Parameters based on H-index

After the discovery of h-index, some h-index variants came into existence to overcome the shortcomings of original h-index. These variants include g-index [9], A-index [10], R-index [10], Hg-index [11], m-quotient [7], AR-index [10] and m-index [12].

2.3 Hirsch Core

Some h-index variants include Hirsch Core in their calculation e.g. A-index, m-index, R-index and AR-index. Therefore, here the term Hirsch Core should be defined. Hirsch Core means "Scientist's best performance throughout his/her career in term of publications and their citations" [10].

2.4 H-index Variants based on Citation Intensity

Some h-index variants measure the citation intensity in the Hirsch core or give more weight to highly cited papers. Some of these variants are g-index, A-index, R-index and Hg-index.

2.4.1 G-index

To overcome the limitations of h-index by capturing more citations that was not covered by h-index, Egghe introduced g-index [9]. Egghe defines g-index as "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 $(g + 1)^2$ cites".

Table 2.2: G-Index Calculation

No. of Publications (g)	**Citations recieved (c)**	g^2	**Sum of Citations**
1	20	1	20
2	10	4	20+10=30
3	5	9	30+5=35
4	0	16	35+0=35
(g) 5	0	25	35+0=35
6	0	36	35+0=35

In Table 2.2, the researcher h-index is 3 and g-index is 5 such that the top 5 publications have at least 5^2=25 (35>25) citations and on rank 6 we have 6^2=36 (35<36) citations [9]. Therefore, we can say that g $\geq$ h. Although the g-index is much effective in assessing the researcher. Moreover, g-index has its own limitations.

Alonso et al compare the research output of two authors. The author A1 has published 30 papers, in one of them he/she got 500 citations but in all other papers he/she didn't get any citation. The author A2 has published 50 papers and he/she got 10 citations against each of his/her paper. The g-index of author A1 is 22 ($22^2 = 484 < 500$ [the citations of the best 22 papers], $23^2 = 529 > 500$ [the citations of the best 23 papers]) while the g-index of author A2 is 10 ($10^2 =$ 100 [the citations of the best 10 papers], $11^2 = 121 > 110$ [the citations of the best 11 papers]). As we can see that, the g-index of author A2 is less than the g-index of author A1 that only attained a big hit paper. Therefore, any big hit may affect the value of g-index [11].

2.4.2 A-index

Jin et al introduced A-index in 2007 to overcome the problem of h-index. He computed a-index by taking the "average number of citations of papers in the Hirsch Core" because h-index does not consider those publications that have huge number of citations [10].

$$A = \frac{1}{h}\sum_{j=1}^{h} cit_j \tag{2.1}$$

In equation 2.1, the numbers of citations (cit$_j$) are placed in descending order. Furthermore, a-index calculated on same data as h-index. Therefore, precision problem remains unsolved in a-index. Moreover, a-index has some flaws. If author A1 has 20 papers, one cited 10 times and all other ones just cited one time. Author A2 has 30 papers, one cited 10 times and all other ones exactly twice. The h-index of author A1 is 1 and author A2 is 2. But the A-index of author A1 is 20/1=20 and A-index of author A2 is 12/2=6. It is obvious that the author A2 is better than the author A1. As the A-index is calculated with the division of h, the author who has greater value of h, is punished due to his/her higher value of h which is a limitation [10].

2.4.3 R-index

As the a-index is computed with the division of h, the a-index punishes those researchers who has higher h-index. Therefore, to overcome the flaws of a-index, Jin et al introduced r-index in 2007 [10]. We can get r-index by taking the square root of sum of all citations in Hirsch core. Mathematically it can be described as follows:

$$R = \sqrt{\sum_{j=1}^{h} cit_j} \tag{2.2}$$

In equation 2.2, cit_j represents the citation count of jth paper and h represents h-index. The r-index is used as measure of h-core's citation intensity. But, it does not consider the age of publications in their calculation. So, we can't consider r-index for the researcher's assessment.

2.4.4 HG-index

To minimize the shortcomings of h-index and g-index, a new index was proposed by Alonso et al to assess researcher's productivity [11]. The hg-index is based on both h-index and g-index. The hg-index is calculated simply by taking square root of product of h-index and g-index.

$$hg = \sqrt{h * g} \tag{2.3}$$

The main strength of hg-index is that it is simple to compute and easy to understand [11]. But, hg-index is totally depending on h-index and g-index which forms an inherited limitation.

2.5 H-index Variants based on Publications Age

H-index has some flaws. It does not take age of publications in their calculation. Therefore, the parameters that resolve this problem were categorized under age of publications based parameters.

2.5.1 M-Quotient

Researchers publish their papers in different career stages. Therefore, it is very difficult to compare an h-index of old researcher with the h-index of new researcher e.g. an old researcher can have a lot of citations against his/her each publication and his/her h-index would also be high as compared to the new researcher [24]. To overcome this problem, J.E. Hirsch proposed m-quotient in his original paper [7]. We can get m-quotient by dividing h-index with the number of years since publishing the first paper.

$$m - quotient = \frac{h - index}{y} \tag{2.4}$$

But, m-quotient is unstable for new researchers as it considers publication year into account. Therefore, small changes in h-index can affect the large alterations in m-quotient [25].

2.5.2 AR-index

[10] state that taking into account the age of publications is a necessary condition to be able to evaluate performance changes. Thus, the AR-index can increase or

decrease over time, a property that is not shared by the previously mentioned indices. Formally, the AR-index is defined as:

$$AR = \sqrt{\sum_{j=1}^{h} \frac{cit_j}{a_j}} \tag{2.5}$$

Where h is the h-index, cit$_j$ is the number of citations of the jth most cited paper, a$_j$ is the number of years since the publication of the jth most cited paper. The main strength of ar-index is that it is simple to compute as it considers not only the total number of citations as well as age of publications. But, ar-index also suffers from some limitations, the ar-index measures publications and citations in the Hirsch Core and does not consider whole career of researcher. Therefore, the original problem remained unsolved by this index as in r-index.

2.5.3 M-index

The median number of citations received by publications in Hirsch Core contents is known as m-index but the value of m-index might be greater or smaller than the h-index. We know that citation distribution is mostly skewed; we cannot use average number of citations. Therefore, as a variation of the a-index, m-index was introduced [12].

2.6 Scientific Services and Affiliations based Parameters

Some researchers suggest that scientific community should consider other scientific contribution based parameters for researchers ranking such as number of graduate students supervised by researchers [13, 26], number of academic institutions affiliations [27, 28], number of fellowships received from scientific societies [19, 29, 30], journal editorial memberships [19, 32] and geographical location of researchers

[31]. This section provides details of such parameters and their potential to be used for researcher rankings.

2.6.1 Graduate Student Supervisions

An important measure of academic output is the supervision of graduate students. Lee and McKenzie present an evocative study to evaluate the quality of doctoral supervision [26]. They argued that most of the existing tools evaluate doctoral supervisions at departmental level rather than individual level and these tools measures the quantity of supervisions. Therefore, this study appraises the eminence of supervisions by utilizing an online student survey.

According to the Rensburg et al study, the supervision of postgraduate students plays a pivotal role in the development of next generation practitioners who have the accurate educational skills to accomplish the future research needs [13].

2.6.2 Academic Institutions Affiliations

In some studies, scientific community considers academic institution affiliation to assess researcher's progress and prestige. Pan and Chen extend [27] study and introduces author affiliation index to rank the quality of marketing journals [28]. This study also explores the cross-validation of journal assessments recognized by past researchers. The pros and cons are also discussed in this study.

2.6.3 Scientific Societies Fellowships

Hamermesh et al declared that a researcher who receives fellowship of scientific societies is considered a great achievement in the research career [29].

Chan and Torgler examine the association between scientific societies distinctions received by researcher's such as fellowships of economics societies and Nobel laureates [19]. They observed positive association between researchers' received society's fellowships and won Nobel Prize.

Chan et al examines whether the John Bates Clark Medal and fellowships of economic society is related with research productivity and status compared to a synthetic control group of non-recipient scholars with similar previous research performance [30]. After award receipt, their results propose positive publication and citation differences.

2.6.4 Journal Editorial Memberships

After analyzing literature, Chan et al ranked the institutions according to the representation of their faculty on editorial board of 30 leading international business journals [14]. They consider that the editorial board membership on high quality journals represents prestigious position of the institution.

Chan and Torgler explore the relationship between Nobel laureate and fellowship of Econometric society [19]. Author observed positive relationship between researchers' received fellowship and won Nobel Prize. They argue that fellowship may high rank than journal editorial membership. Wu et al also consider editorial board memberships to evaluate the scholarly impact of academic institutions [33].

2.6.5 Geographical Location

Many academic rankings have been presented to rank experts using publication and citation count. However, they do not consider geographical location where publications produced and citations are made for these publications. Mazloumian et al introduced a network-based index to rank major geographical locations based on number of publications produced and number of citations received by a specific region. Considering knowledge produced and consumed between 2000 and 2009, they identified global sources and sinks of this knowledge. Their results showed that Europe and North America were two major knowledge producing areas worldwide. However, South America and Asia were major knowledge consuming areas [31].

2.7 Studies on Evaluation of Researchers' Ranking

In the following paragraphs we critically review the papers in which primitive, citation intensity, age of publications and "Scientific services and Affiliations" based parameters were used to assess researchers and correlation is tested.

Recently, Raheel et al evaluate citation intensity and publication age based parameters for researchers' assessment. The dataset was consisting of 1060 international prestigious awardees of Civil Engineering domain. For evaluation, the correlation was tested between parameters and awardees occurrences was tested in all ranking lists. Correlational analysis revealed negative correlation among various parameters and none of the parameters succeeded in bringing 100% awardees to the top of ranked lists [17].

Ayaz and Afzal presents an empirical study to evaluate the h-index, g-index and complete-h on a real data set in the field of Mathematics for researchers' assessment [15]. This study considered award winners from the Mathematics field as a benchmark. Their result analysis shows that complete-h, which is defined to complete the definition of h-index is the best parameter in bringing award winners up in the ranking than h-index and g-index.

Cantin et al presents a descriptive and correlational study to measure the institutional productivity and institutions ranking on the basis of cumulative h-index, m-quotient and the total number of publications and citations [34]. This study accessed the Chilean Society of Anatomy Professor list for analysis of academic morphologists' h-indexes using Scopus database. According to this study, the top

3 institutions were Universidad de La Frontera, Universidad de Chile and Pontificia Universidad Catolica de Chile. This study also indicates that the h-index and the total number of publications were the best predictor of academic rank.

Schreiber presents a comparison study of the g-index for 26 physicists with the h-index, the A-index and the R-index [35]. This study analyzed the citation records of 26 physicists of the institute of physics at Chemnitz University of Technology. According to this study, the g-index is more suitable to illustrate the overall impact of the publications of a scientist than the h-index.

Hirsch gives an empirical study to predict the future achievement of an author using different Bibliometric indices such as publication count, citation count, citations per paper and h-index. According to this study, the h-index is the best indicator to predict the future achievement of an author [36].

According to Susarla et al study, they investigate the correlation between quantitative measures of academic productivity and academic rank among full-time academic plastic surgeons [37]. Bibliometric indices included the Hirsch index, I-10 index, publications count, citations count, and highest number of citations for a single publication. Correlation analysis revealed strong associations of the Hirsch index, I-10 index, number of publications, and number of citations with academic rank.

Erichsen et al examines the perception of PhD supervision and to comprehend the gratification of PhD student with graduate supervision where programs were conveyed using a variety of distance systems. Based on gender, significant differences in student reactions were discovered. While students whose programs were conveyed online were moderately gratified, students who were in the combined programs were more gratified [38].

Janney et al considers author affiliation parameter to evaluate journal quality [39] and compares their results with existing techniques. This research provides the mathematical calculation of affiliation parameter and ranks the quality of journal by analyzing the educational institutions which employ the researchers who publish in a specified journal.

Chan and Torgler examine the association between scientific societies distinctions received by researcher's such as fellowships of economics societies and Nobel laureates [19]. They observed positive association between researchers' received society's fellowships and won Nobel Prize. Another interesting finding revealed by his study was that Harvard and MIT are the main PhD granting institutions with respect to the generation of both Fellows and Nobel Laureates. Authors tested this finding using 483 fellows of economic society's dataset and ranked them based on their academic affiliation. Therefore, it is possible that a researcher who has fellowship of scientific society, Nobel laureate and affiliated with the top academic institution may be more expert in a specific field.

Wang examines the relationship between editorial board membership and universities research production in Computer Science domain by utilizing quantile regression models [40]. This research analyzed 447 journals and 14,442 editorial board memberships. According to their results, editorial board memberships are significantly associated with the publication count (quantity) and citation count (impact) of the research production from their corresponding universities. The quantile regression analysis showed the strong relationship between editorial board memberships and research production.

Table 2.3: Comparison studies of Author Ranking Parameters

Ref	Parameters	Methodology	Benchmark Dataset
[17]	Citation intensity and Publication age based Parameters	Finds the Correlation and Awardees occurences in all ranking lists	1060 International Prestigious Awardees of Civil Engineering Domain
[15]	H-index, g-index and Complete h	Evaluation of h-index, g-index and complete h for researchers' assessment.	671 International Prestigious Awardees of Mathematics Domain
[16]	Publication Count and Citation Count	A comparative analysis of graph-based ranking algorithm with citation count and publication count rank	Winners of 9 major Physics awards between 1960 and 2010 as benchmark
[35]	g-index, h-index, A-index, R-index	Presents a comparative study of g-index with h-index, A-index and R-index	26 Physicists of the institute of Physics
[19]	Scientific Societies Fellowships	Aassociation between scientific societies distinctions received by researcher's such as fellowships of economics societies and Nobel laureates.	Dataset consists of 483 fellows of economic society's.
[27]	Academic Institutions Affiliations	Employed author affiliation index to assess journals	41 finance journal
[30]	Editorial Board Memberships	Employed Editorial board memberships for Institutions ranking	Editorial board memberships of 30 international business journals
[31]	Geographical Location	Introduced a network based index to rank major Geographical Locations	Number of publications produced and number of citations received by a specific region

2.8 Prestigious Awards

In the previous section, we have discussed different techniques to find potential experts in the scientific community. Moreover, highly contributed researchers are granted with international prestigious awards in any scientific field. For example, in Computer Science domain, the prestigious awards include A.M. Turing Award, Millennium Technology Prize, Kyoto Prize in Advanced Technology, IEEE Medal of Honor and IEEE Internet Award. However, all these awards are given by considering the significant contributions of researchers in any scientific field.

Award is basically a certificate that is given to any group of people, individual, or organization due to excellent performance in any specific field. One most known example of award is "Noble Prize" which is awarded to those people who does excellent work in field of science (Peace, Economics, Medicine, Chemistry, Physiology, Physics and Literature). But in the field of Computer Science, any scientist was not awarded with Noble Prize. One most known example of award in Computer

Science is “A.M. Turing Award” which is also called as “Noble Prize of Computer Science”. There are some other awards that are awarded to computer scientists such as Software System Award, A.M. Turing Award and Gordon Bell Prize. The importance and selection criteria for these awards is discussed in following section.

2.8.1 A.M. Turing Award

Association for Computing Machinery (ACM) is one of the largest Computer Science organization which was established in 1947 in United States of America. ACM appreciates those computer scientists that works extraordinary in computer science field. One of the most prestigious computing award given by ACM is A.M. Turing Award, which is given for outstanding contributions in computer sciences and information technology. The A.M. Turing Award was named after the British Computer Scientist and Mathematician Alan Mathison Turing. He had extraordinary contributions in computer architecture, algorithms, artificial intelligence and formalization of computing technology. A.M. Turing Award is given on annual basis and given to the true inventors who are best scientist of that year in computing technology. Turing award is often known as “Nobel Prize of Computer Science”. Alan J. Perlis was the first computer scientist who was awarded with “Turing Award” in 1966 upon his outstanding performance. He built the compiler for ALGOL programming language. Frances E. Allen was the first woman who won the “Turing Award” in 2006 for her contribution in compiler optimization. She worked for development of parallel execution in multiprocessing. The “Turing Award” is being funded by Intel Corporation and Google Inc. and prize money of Turing award is $1,000,000 which was increased in last year, before this the prize money was $250,000. Patterson, David was the last year winner of A.M. Turing Award for his excellent performance in reliability and consistency of computer architecture.

For nomination of A.M. Turing award even though the lasting contributions of the scientist’s effort are taken into account, there should be a specific exceptional and popular technical achievement that nominates the key claim to the award. The scientists who are generally representative of nominee's research field that approve the nominee's lasting influence work.

2.8.2 ACM Fellow

The ACM Fellow is ACM's most prestigious membership acknowledgment to ACM members for their extraordinary performance and contribution in Computer Sciences and information technology. The ACM Fellows are nominated by ACM's most professional senior members. The ACM Fellow program was started in June 1993 by ACM committee to give appreciation to ACM members that have outstanding professional, technical and leadership inspiration. ACM Fellows are selected by ACM council according to following criteria:

- The member should have outstanding contributions in Computer Science field.
- The member should have done prominent service to ACM and approval from current ACM professionals senior members are required.
- The member should have technical experience that advance the field of computing.

2.8.3 Gordon Bell Prize

The Gordon Bell Prize is a prize given by ACM every year in the combination with Supercomputing Conference. This honor is achieved by specifically those scientists or innovators that work outstanding in high performance computing. The prime motive behind giving the Gordon Bell Prize is to appreciate those scientists who perform astonishing achievement in parallel computing, high performance applications in computer science and those who are large scale data analysts and engineers. It is awarded to those scientists who have extraordinary performance and those who are critical problem solvers in science and engineering. The qualifiers are invited to the Supercomputing Conference to receive their prize. This ceremony is organized every year in November. This award has prize money of \$10,000/- since 2011 which is provided by Gordon Bell itself. Any individual or group can apply for this award but there is a detailed procedure for the scientists which they have to follow in order to enlightening their innovations, or any specific high performance achievement in this global world. However, other measures are also taken by the selection team not only a specific achievement of scientists.

- Indication of essential algorithmic and implementation novelties.
- Pure novelties above the prior state-of-the-art.
- The system which can't be simulated by others or that can't be employed for solving a small problem.
- The extraordinary achievement in scalability of serious problems or any ultimate performance.
- Accomplishments that are generalizable, in the sense that other people can acquire and advantage from the novelties.

2.8.4 Software System Award

The software system award is given to individuals or organizations for developing software systems that have long lasting effect in computer sciences. The software system award is given by ACM and the prize money for software system award is \$35,000. The funding for software system award is provided by IBM.

The Nominees are evaluated for the confirmation they deliver of important theoretical outcome, general acceptance, impact on associated improvements, influence in infrastructure and actual alteration from theory to practice.

2.8.5 IEEE Fellow

On 1^{st} of January 1963, two institutes i.e. American Institute of Electrical Engineers and the Institute of Radio Engineers were merged to established the "Institute of Electrical and Electronics Engineers". The main reason of establishing IEEE was to appreciate the professional employees in all fields of electronics, computer sciences and other areas linked to science and technology. The scientists who bring outstanding advancements in any of the IEEE fields are acknowledged by the title of "IEEE Fellow". IEEE Fellow is awarded to one tenth percent of the total IEEE members every year. It is mandatory for the contestants to own extra ordinary achievements in electrical, electronic or computing fields. Any contender from some other field for example, academic world, government or industry can also be elected but he/she has to satisfy the following criteria at the time of nomination:

- The member should have served at least five years in any IEEE position.
- The nominee should be an older member of IEEE.
- The nominee should have endeavors that have donated to the progression or application of electronics, computing and technology.

2.9 Conclusion

In this chapter, we have critically reviewed the studies in which the author ranking parameters were proposed and utilized for different purposes. We have also reviewed the international prestigious awards in this chapter.

After reviewing the author ranking parameters, we find that the evaluation studies have utilized only small dataset and sometime imaginary dataset. Therefore, there is a need to assess author ranking parameters on comprehensive dataset. Secondly, "Scientific Services and Affiliations" based parameters have been utilized in different context. Therefore, this research is conducted to evaluate the impact of

“Scientific Services and Affiliations” based parameters with publication, citation intensity and age of publications based parameters.

Chapter 3

Methodology

A variety of methods have been employed by scientific community to assess researchers. The critical analysis of related literature has shown that most of the time publication based parameters are utilized for researcher's assessment such as publication count, citation count, h-index and its variants. However, it was highlighted that there are many other contributions of a researcher which could be applied to assess researchers such as: academic institutions affiliations, graduate student supervisions and journal editorial memberships. Few previous studies employed membership on editorial board of academic journals to rank academic institutions. Furthermore, some other studies only suggested parameters based on affiliations of researchers within scientific community. There is no comprehensive study on "Scientific Services and Affiliations" based parameters to assess researchers. Therefore, this thesis exploits the following parameters for researchers' assessment: 'Graduate Students Supervisions', 'Academic Institutions Affiliations', 'Scientific Societies Fellowships', 'Journal Editorial Memberships' and 'Geographical Location' of researchers.

Figure 3.1 shows the block diagram of adopted methodology. Every module of Figure 3.1 is discussed in detail in the upcoming sections.

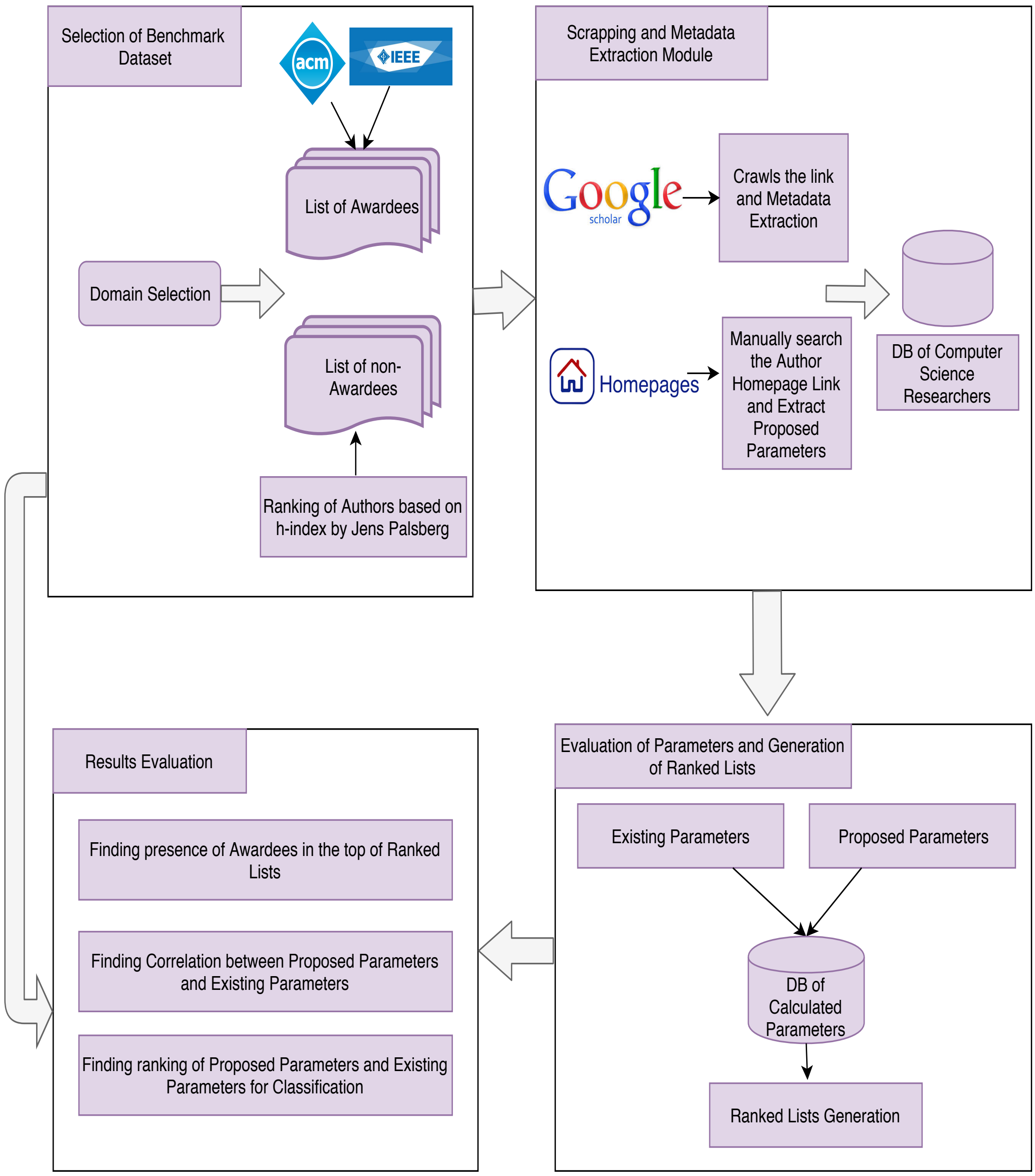

Figure 3.1: Block Diagram of Methodology

3.1 Selection of Benchmark Dataset

In first module, we will select benchmark dataset to evaluate the impact of "Scientific Services and Affiliations" based parameters with primitive, citation intensity and age of publications based parameters.

3.1.1 Domain Selection

We have selected the Computer Science domain to evaluate the impact of "Scientific Services and Affiliations" based parameters with primitive, citation intensity and age of publications based parameters for researchers' assessment. Furthermore, "Scientific Services and Affiliations" based parameters have been compared with publications based parameters. The reason behind selecting Computer Science domain is that it has extensive applications in all field of science. Moreover, none of the comprehensive and evaluating study has been performed in this field based on "Scientific Services and Affiliations" of researchers.

3.1.2 Benchmark Dataset

In this thesis, "Scientific Services and Affiliations" based parameters have been proposed to assess researchers. Three research questions have been formulated to evaluate "Scientific Services and Affiliations" based parameters with primitive, citation intensity and age of publications based parameters. To answer these research questions, we required a comprehensive dataset which comprises of researchers highly regarded by scientific community. Highly regarded researchers are generally acknowledged by scientific societies to give scientific awards. Therefore, we have considered award winners of some major International Scientific Societies which are Association for Computing Machinery (ACM) and The Institute of Electrical and Electronics Engineers (IEEE) to evaluate whether the suggested parameters acknowledge these researchers or not. Some previous studies have also considered award winners of scientific societies as Gold standard to evaluate authors ranking parameters [15–17]. We have also considered 24 awards of the mentioned international computer science award giving societies.

Moreover, we have added some noise to this dataset. We have considered a dataset built by Jens Palsberg [41] which contain about 1000 Computer Scientists ranked based on h-index. Jens Palsberg is a professor of Computer Science in University of California, Los Angeles. Jens Palsberg rank the Computer Science researchers every year based on h-index of 40 or higher according to Google Scholar. We have considered this dataset as non-awardees in this thesis. The reason behind this is to investigate whether the proposed "Scientific Services and Affiliations" based parameters distinguishes awardees from those researchers who have not received any award but belong to the all type of profiles (High, Average and Low). For this purpose, we have used a dataset of 1992 computer science researchers as award winners in our thesis.

3.2 Scrapping and Metadata Extraction

Next step was to ascertain web source to gather "Scientific Services and Affiliations", primitive, citation intensity and age of publications based parameters of computer science researchers in our dataset.

The first web source identified was "Home Page" to gather "Scientific Services and Affiliations" based parameters of an individual researcher. Because, the best source to collect accurate "Scientific Services and Affiliations" based parameters of researcher are available in the form of their homepages as shown in the Figure 3.2:

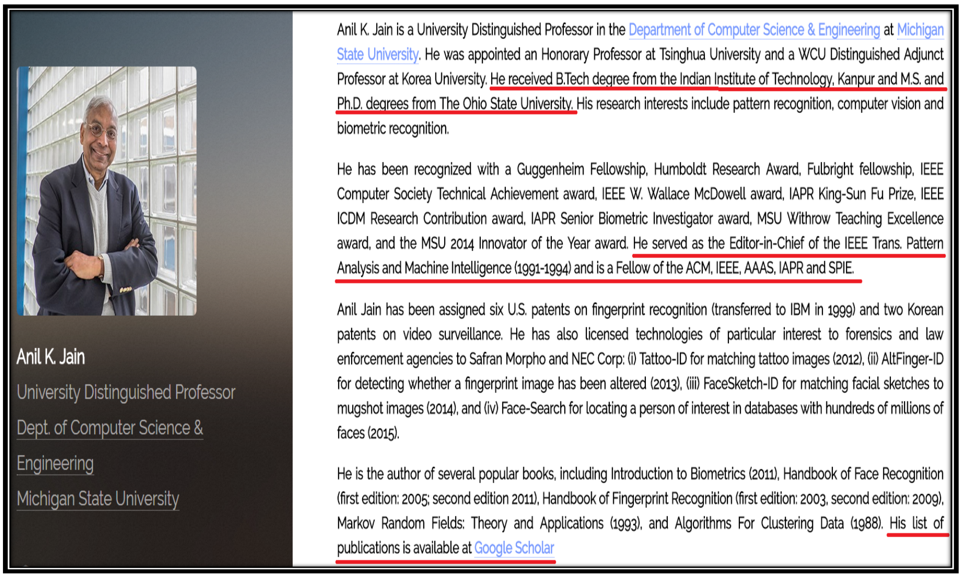

FIGURE 3.2: Anil K. Jain Homepage

As we want to evaluate "Scientific Services and Affiliations" based parameters with primitive, citation intensity and age of publications based parameters. Therefore, we have to implement primitive, citation intensity and age of publications based parameters. For implementation, we need publications, their citations and the year of publication. But the question is that how to obtain these publications, their citations and publication year. Therefore, we have considered Google Scholar to collect publications, their citations and publication year [15]. There are many other search engines such as Scopus, Web of Science, Digital Bibliographic and Library Project (DBLP), CiteSeerX, Summon and PubMed. All of these have their own pros and cons. Some of them are subscription based online sources. Their access is not openly available. In contrast to these resources, Google Scholar is an open access online resource and contains comprehensive data in almost every field of science [17]. Moreover, Google Scholar has huge coverage than aforementioned search engines [42].

As we know that Google Scholar is an open access online resource. Therefore, we can search and retrieve various research articles, find all the appropriate research papers and acquire parameters about related fields. For example, if we search Anil K. Jain on Google Scholar then the results will be shown in the Figure 3.2:

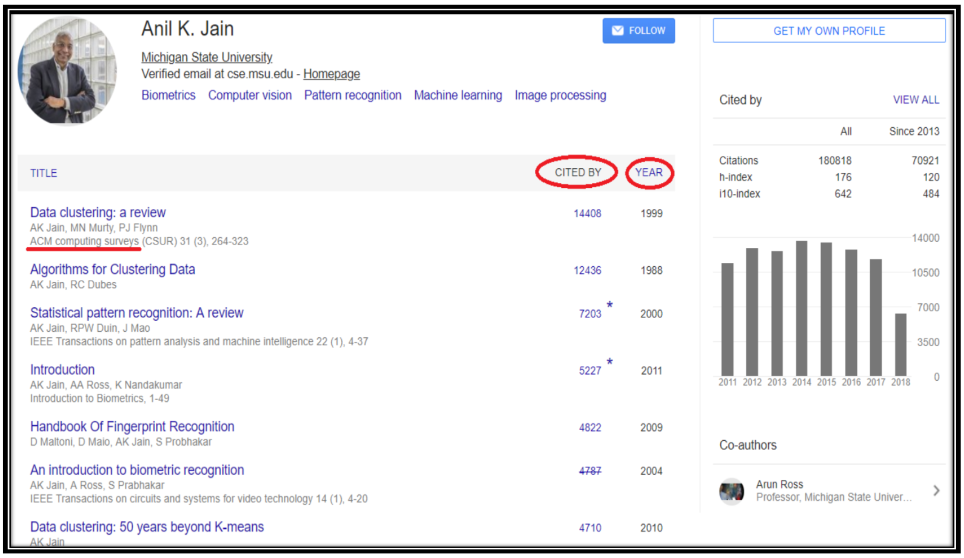

FIGURE 3.3: Anil K. Jain Profile on Google Scholar

To collect data manually from Google Scholar is a time consuming task. Therefore, a dedicated crawler was built to acquire researcher's publications, their number of citations and publishing year. The dedicated crawler was developed in such a way that it manually searches the researcher profile link and verify the link and then download the link and fetch the desired data using Dom (Document Object Model). Crawler gets the HTML of the given link and parses the HTML using DOM parser. The DOM parser makes it possible to process XML documents in PHP. The algorithm written in the crawler gets the given information of researcher such as publication count, citation count and publishing year.

3.2.1 Tools and Technologies Used

- Programming Language PHP
- XAMPP Server

3.2.2 Crawler for Scholars Profile

In algorithm, we have created web crawler function to extract number of publications, their total citations and publishing year. We use DOM (Document Object Model) to process the html document. We get the child elements with the specific tag name. When we get the tag "tr" then we get the total citations and h-index from this tag tr. The metadata that is being crawled by web crawler function also includes title of research papers, no-of-citations recieved from each article and publishing year. After extracting researcher's records, this data is used to calculate the existing parameters such as g-index, A-index, R-index, HG-index, AR-index, m-index and m-quotient. The steps to calculate all these parameters is discussed in each function as shown below in Algorithm 1. The mathematical calculation of all existing parameters also discussed in section 3.3.

Once the data has been collected, it is necessary to verify the data and clean it wherever required. To verify the dataset, author disambiguation was performed on

Algorithm 1 Scholars Profile Crawler

```
function WEB_CRAWLER(String R[0....n-1], String URL[0...n-1])
//Input:Array of Researchers List and Array of URL's
//Output:Extract Researchers' Publications, their no of Citations and Publishing Years


String tableclassname, propertyclassname, propertyValue_classname, ProfileInfoTable_classsname;
int PaperPublications[][], CitationID;
dom profiles, doc;
int g − index, m − index;
float mQ, A − index, R − index, HG − index, AR − index;
for(i ← 0 to n − 1)
dom = getDOM(URL[i]) //Parse the html into a DomDocument
   doc = newDOMXPath(dom) // XPath object to query DOM
   table_classname = "gsc_rsb_st"; //classname used for table such as citations and h-index
   property_classname = "gsc_rsb_sc1"; //classname used for citations name and
   h-index name
   propertyValue_classname = "gsc_rsb_std"; //classname used for total citations value and h-index value
   ProfileInfoTable_classname = "gsc_a_b"; //classname used for researchers information such as
   publications, their citations and publishing year
   Profile_Table ← doc → query(contains[table_classname]);
   FOREACH(Profile_tr ← Profile_Table → query(contains[tr])
      if(Profile_tr → query(contains[property_classname] = "Citations"))then
         return total_citations ← Profile_tr → query(contains[propertyValue_classname]);
      endif
      if(Profile_tr → query(contains[property_classname] = "h − index"))then
         return h-index ← Profile_tr → query(contains[propertyValue_classname]);
      endif
      AuthorsInfoTable ← doc → query(contains[ProfileInfoTable_classname]);
      Size ← AuthorsInfoTable → count(tr);
      Paper_Publications ← newString[size][3];
      FOREACH(ProfileInfo_tr ← AuthorsInfoTable → query(next[tr])
         CitationID ← ProfileInfo_tr → query(contains["TitleName"]);
         PaperPublications[i][0] ← CitationID → getElementByTagName[a];
         CitationID ← ProfileInfo_tr → query(contains["CitedBy"]);
         PaperPublications[i][1] ← CitationID → getElementByTagName[a];
         CitationID ← ProfileInfo_tr → query(contains["Year"]);
         PaperPublications[i][2] ← CitationID → getElementByTagName[a];
    mQ=MQuotient(h-index,year);
    g-index=G-INDEX(publications[][],size);
    A-index=A-INDEX(h-index,publications[][],size);
    R-index=R-INDEX(publications[][],size);
    HG-index=HG-INDEX(h-index,g-index);
    AR-index=AR-INDEX(publications[][],size);
    M-index=M-INDEX(publications[][],size);
end function//WEB_CRAWLER function ends here

function DOM GETDOM(url)
//Input: URLs
//Output:Content retrieved against this url by a variable dom
links = arr();
htm ← makerequesttotargeturlusingcURL;
   if(html is empty)
  show error message;
  Exit;

  else

  //Parse the html into a DOMDocument.
  dom = new DOMDocument(); // Create new DOMDocumet with no value.
  dom− > loadHTML(html);
  return dom;
end function//getDom function ends here
```

```
function G-INDEX(int publications[][],int size)
int sorted_publications;
sorted_publications ← SortDescending_w.r.t_Citations(Publications);
//Compute sum of citations
sum_of_citations[]=new int[size];
for(i ← 0tosize)
if(i=0) then
    sum_of_citations[i] ← sorted_publications[i][1];
     else
    sum_of_citations[i] ← sum_of_citations[i − 1] + sorted_publications[i][1];
     endif
for(i ← size − 1to0)
if((i*i)<sum_of_citations) then
return i;
end function//g-index function ends here

function A-INDEX(int h-index,int publications[][],int size)
float temp;
int sorted_publications;
sorted_publications ← SortDescending_w.r.t_Citations(Publications);
temp=1/h-index;
SumTillh-index=0;
for(i ← 0toh − index)
SumTillh-index ← SumTillh − index + sorted_publications[i][1];
return SumTillh-index*temp;
end function//A-index function ends here

function HG-INDEX(int h-index,int g-index)
HG-index ← sqrt(h − index ∗ g − index);
return HG-index;
end function//HG-index function ends here

function R-INDEX(int publications[][],int size)
int sorted_publications, SumTillh − index;
sorted_publications ← SortDescending_w.r.t_Citations(Publications);
//Compute sum of citations
SumTillh-index=0;
for(i ← 0toh − index)
SumTillh-index ← SumTillh − index + sorted_publications[i][1];
return sqrt(SumTillh-index);
end function//R-index function ends here

function AR-INDEX(int publications[][],int size)
int sum = 0, no_of_citations, age_of_Publication;
for(i ← 0toh − index)
no_of_citations ← sortedPublications[i][1];
age_of_Publication ← currentyear − sortedPublications[i][2];
sum ← sum + (no_of_citations/age_of_Publication);
return sqrt(Sum);
end function//AR-index function ends here

function M-INDEX(int publications[][],int size)
int sorted_publications, citationArr[];
sorted_publications ← SortDescending_w.r.t_Citations(Publications);
for(i ← 0toh − index)
citationsArr[i] ← sortedPublications[i][1];
return median(citationsArr);
end function//M-index function ends here

function MQUOTIENT(int h-index,int year)
mQ=h-index/year;
return mQ;
end function//MQuotient function ends here
```

the dataset [43]. This problem arises when multiple authors share same first name and last name. Therefore, the researchers' who have same names were identified and verified manually from the collected dataset.

3.3 Evaluation of Parameters and Generation of Ranked Lists

In this module, researchers were ranked on the basis of primitive, citation intensity, age of publications, Scientific Services and Affiliations based parameters.

3.3.1 Authors Ranking by Publications based Parameters

This step explains the calculation of publications based indices. After this step, all the parameters scores will be stored in the database and based on the calculated values, we will get 10 rankings which will be further evaluated with proposed "Scientific Services and Affiliations" based parameters.

3.3.1.1 Publication Count

An author with greater number of publications is considered to be an expert [1, 2, 44]. Therefore, the publication count is defined as:

$$P = \sum_{j=1}^{n} P_j \tag{3.1}$$

In equation 3.1, this formula calculates the total number of publications of a particular researcher.

3.3.1.2 Citation Count

Citation count represents the impact of an author work in scientific community [4, 5]. Citation count is defined as:

$$C = \sum_{j=1}^{n} Cit(P_j) \tag{3.2}$$

In equation 3.2, this formula calculates the total number of citations received against each publication of a particular researcher.

3.3.1.3 H-index

H-index of an author is calculated by considering the number of citations received by his/her most cited publications. "A scientist has index h if h of his or her Np papers have at least h citations each and the other $(N_p - h)$ papers have fewer than ≤h citations each" [7].

TABLE 3.1: H-Index Calculation

No. of Publications	**Citations recieved**
1	33
2	30
3	20
4	15
5	7
(h) 6	6
7	5
8	4

To calculate h-index of a researcher, organize all of the publications with respect to citations in descending order. Therefore, if any researcher has 8 publications and they were cited 33, 30, 20, 15, 7, 6, 5 and 4 times then the researcher h-index will be 6. The 7th paper has citations (5) less than 7. So, 7th paper does not have any influence in significance of h-index.

3.3.1.4 G-index

G-index was introduced to overcome the limitations of h-index by Egghe [9]. "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 $(g+1)^2$ cites" [9].

TABLE 3.2: G-Index Calculation

No. of Publications (g)	**Citations recieved (c)**	g^2	**Sum of Citations**
1	20	1	20
2	10	4	20+10=30
(h) 3	5	9	30+5=35
4	0	16	35+0=35
(g) 5	0	25	35+0=35
6	0	36	35+0=35

In Table 3.2, the researcher h-index is 3 and g-index is 5 such that the top 5 publications have at least 5^2 (g^2) =25 (35>25) citations and on rank 6 we have 6^2 $(g+1)^2$=36 (35<36) citations (Egghe, 2006). Therefore, we can say that g $\geq$ h.

3.3.1.5 A-index

A-index was introduced by jin et al in 2007 which overcomes the limitations of h-index [10]. He computed a-index by taking the "average number of citations of articles in the Hirsch Core".

$$A = \frac{1}{h}\sum_{j=1}^{h} cit_j \tag{3.3}$$

In equation 3.5, papers are sorted with respect to citations (cit_j) in descending order. If an author has 6 publications as in Table 3.2 with 20,10,5,0,0,0 citations and the h-index is 3. The sum of citation in h-core will be 20+10+5=35, the value of a-index = 35/3=11.6 which is greater than h-value.

6) R-index

To overcome the flaws of A-index, R-index was introduced by jin et al in 2007[10].

We can get r-index by taking the square root of sum of citations in Hirsch core. R-index is stand for root index. R-index is defined as:

$$R = \sqrt{\sum_{j=1}^{h} cit_j} \tag{3.4}$$

3.3.1.6 HG-index

To compensate the shortcomings of h-index and g-index, a new index was proposed by Alonso et al to assess researcher's productivity [11]. The hg-index is based on both h-index and g-index. The hg-index is calculated simply by taking square root of product of h-index and g-index.

$$hg = \sqrt{h * g} \tag{3.5}$$

3.3.1.7 M-quotient

M-quotient first introduced by Hirsch [7]. We can get m-quotient by dividing h-index with the number of years since publishing the first paper.

$$m - quotient = \frac{h - index}{y} \tag{3.6}$$

3.3.1.8 AR-index

The AR-index can increase or decrease over time, a property that is not shared by the previously mentioned indices [10]. Formally, the AR-index is defined as:

$$AR = \sqrt{\sum_{j=1}^{h} \frac{cit_j}{a_j}} \tag{3.7}$$

Where h is the h-index, cit_j is the number of citations of the jth most cited paper, a_j is the number of years since the publication of the jth most cited paper.

3.3.1.9 m-index

The median number of citations received by publications in Hirsch Core contents is known as m-index but the value of m-index might be greater or smaller than the h-index. We know that citation distribution is mostly skewed; we cannot use average number of citations. Therefore, as a variation of the a-index, m-index was introduced [12].

3.3.2 Authors Ranking by Proposed “Scientific Services and Affiliations” based Parameters

In this step, we will rank the Computer Science researchers on the basis of “Scientific Services and Affiliations” based parameters such as Graduate Student supervisions, Academic Institutions Affiliations, Scientific Societies Fellowships, Journal Editorial Memberships and Geographical location. These parameters have been explained with the help of equation in the following paragraphs:

3.3.2.1 Graduate Student Supervisions

A computer scientist is an expert if he/she has supervised the highest number of students throughout his research career. This parameter counts the total number of MS and PhD students supervised by a particular computer scientist. It can be defined with the help of following equation:

$$S(i,j) = \sum_{j=1}^{n} S_{ij} \tag{3.8}$$

In equation 3.8, 'S(i,j)' calculates the sum of supervisions of ith researcher where 'i' represents the number of researchers from 1 to n and j represents all graduate students supervised by a particular researcher. Then, we will sort all the student supervisions of each researcher in descending order and select the maximum supervisions of researchers' according to their positions. The researcher having greater number of student supervisions will be considered an expert as compared to the researcher having smaller number of student supervisions. The researchers having equal number of student supervisions will be ranked at the same position.

3.3.2.2 Academic Institutions Affiliations

A computer scientist is an expert if he/she has more affiliations of world top universities than other computer scientists. We have selected QS World universities ranking of 2018. QS assesses universities using six performance indicators such as academic reputation, employer reputation, student-to-faculty ratio, research citations per faculty member, proportion of international faculty, and proportion of international students. These rankings will assign score to each university as shown in the Figure 3.4.

The equation for this parameter is defined as follows:

$$A(i,j) = \sum_{j=1}^{n} (W(A_{ij})) \tag{3.9}$$

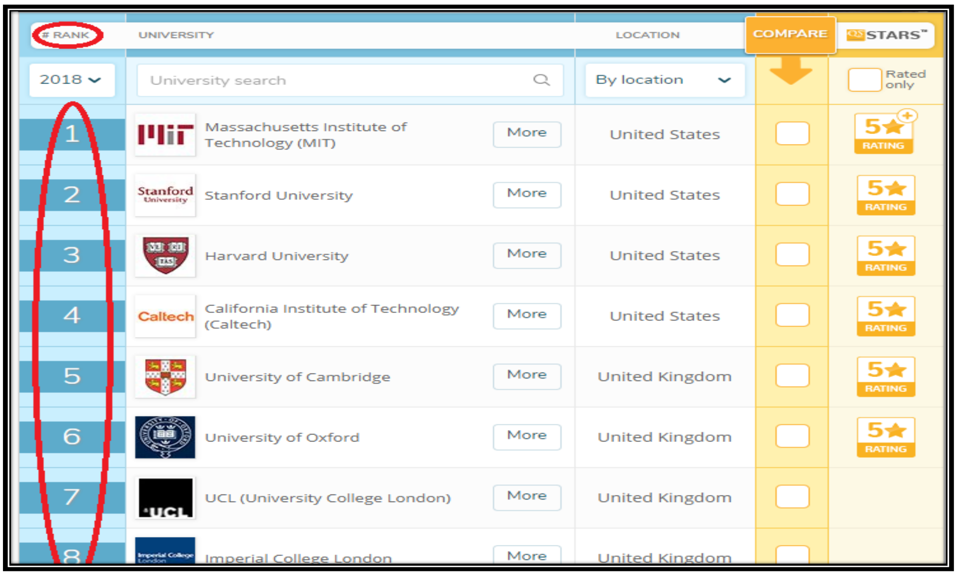

FIGURE 3.4: QS World University Ranking

In equation 3.9, 'W(A_{ij})' takes rank score which is 'W' of a university from QS ranking and assigns to a researcher affiliation. 'A(i,j)' sum up all scores of all academic institutions for ith researcher where 'i' represents the number of researchers from 1 to n and j represents the total score of all academic institutions for a particular researcher. Then, we will organize all the affiliations scores of individual researcher in ascending order and select the minimum affiliation score of researchers' according to their positions. The reason behind selecting minimum score is that the researcher belongs to the world top universities. The researchers having smaller affiliations score will be considered more expert than the researcher having greater affiliations score. The researchers having equal affiliations score will be ranked at the same position.

3.3.2.3 Scientific Societies Fellowships

A computer scientist is an expert if he/she has received the greater number of fellowships from scientific societies throughout his research career. This parameter counts the total number of scientific fellowships received by an individual scientist.

The equation for this parameter is as follows:

$$F(i, j) = \sum_{j=1}^{n} F_{ij} \tag{3.10}$$

In equation 3.10, 'F(i,j)' calculates the sum of fellowships of ith researcher where 'i' represents the number of researchers from 1 to n and j represents all fellowships of a particular researcher. Then, we will sort all the scientific societies fellowships of each researcher in descending order and select the maximum societies fellowships of researchers' according to their positions. The researcher having greater number of scientific societies fellowships will be considered an expert as compared to the researcher having smaller number of scientific societies fellowships. The researchers having equal number of scientific societies fellowships will be ranked at the same position.

3.3.2.4 Journal Editorial Memberships

This parameter considers editorial board memberships of computer science journals where a computer scientist has served throughout his research career. A researcher is an expert if he/she has served on editorial boards of many high impact factor (IF) computer science journals. We have collected editorial membership of journals for all computer science researchers in our dataset. Impact factor is calculated on yearly basis for those journals indexed in Journal Citation Reports (JCR). The ISI Web of Knowledge indexes more than 8,000 science and social science journals. We have considered only ISI web of knowledge indexing of journals for this parameter. This parameter can be defined with the help of following equation:

$$J(i, j) = \sum_{j=1}^{n} (IF(J_{ij})) \tag{3.11}$$

In equation 3.11, 'IF(J_{ij})' represents the impact factor of each journal 'J_{ij}' retrieved from Thomson Routers. 'J(i,j)' sum up impact factor (IF) for all journals of ith researcher where 'i' represents the number of researchers from 1 to n and j represents the total impact factor (IF) for all journals of a particular researcher. Then, we will sort all the impact factor scores of each researcher in descending order and select the maximum impact factor score of researchers' according to their positions. The researcher having greater impact factor score will be considered an expert as compared to the researcher having smaller impact factor score. The researchers having equal impact factor score will be ranked at the same position.

3.3.2.5 Geographical Location

According to Mazloumian et al [31] study, an area which produces publications is known as knowledge producing whereas knowledge consumers are those who cite these publications. Therefore, a computer scientist is an expert if he/she belongs to the knowledge producing area otherwise he/she belongs to the knowledge consuming areas. Their results showed that Europe and North America are two major knowledge producing areas worldwide whereas, South America and Asia are major knowledge consuming areas. We have used 0-10 range of values to assign weight 'W' to major geographical location. W>5 assign to knowledge producing areas and W<5 assign to knowledge consuming areas.

3.4 Awards Benchmark

List of internationally recognized awards like ACM Fellow, Turing Award, IEEE Technical Achievement Award, Software System Award, Gorden Bell Prize, Computer Pioneer Award, IEEE John von Neumann Medal, Infosys Foundation Award, Seymour Cray Award, Programming Languages Achievement Award, Distinguished Service Award, Doctoral Dissertation Award, IEEE CS Eckert-Mauchy Award, Grace Murray Hopper Award etc. were taken with the awardees for the evaluation of all mentioned author ranking parameters. We have considered all these awardees lists as a benchmark in our thesis.

3.5 Results Evaluation

After acquiring ranked lists of authors according to each ranking parameters, a comprehensive evaluation is performed. This module explains process of evaluation of results using three research questions and benchmark dataset. The evaluation of "Scientific Services and Affiliations" based parameters with primitive, citation intensity and age of publications based parameters is presented in Chapter 4.

3.5.1 International Prestigious Awardees Trends in Authors' Ranking

The first research question checks presence of award winners in the top of ranked lists obtained from "Scientific Services and Affiliations", "Primitive", "Citation Intensity" and "age of Publications" based parameters. For this purpose, we will check percentage occurrences of international prestigious awardees within top 1-10%, 11-20%, 21-30%, 31-40% and 41-50% of ranked lists [17]. This evaluation process is performed in chapter 4 of experimental results.

3.5.2 Measure Correlation for Proposed Parameters and Existing Parameters between Ranked Lists

The second research question is defined to measure correlation between "Scientific Services and Affiliations" based parameters and primitive parameters, citation intensity based parameters, age of publications based parameters. The purpose of finding correlation is to find similarity between all these parameters. Considering the ranked nature of the data, the best suitable correlation measure is spearman [45]. The detail outcome of this research question is discussed in chapter 4.

3.5.3 Ranking of Proposed Parameters and Existing Parameters for Classification

The 3rd research question will rank a parameter which can best classify international prestigious awardees and non-awardees. The main aim of classification is to check the effect of parameters for classification of award winners and non-awardees and it will identify which parameter perform well. A lot of classification algorithms have been proposed for binary classes. Wu et al surveyed top 10 data mining algorithms identified by the IEEE international conference on Data Mining (ICDM) such as: C4.5, k-Means, SVM, Apriori, EM, PageRank, AdaBoost, kNN, Naive Bayes, CART and found out that these top 10 algorithms are among the most significant data mining algorithms in the research community [46]. A detailed survey has been presented between different classification algorithms in [47, 48]. To rank the parameters accurately, we will use binary classifiers because our dataset has binary class. Therefore, we have utilized Naive Bayes (NB), K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) classifiers. For this purpose, WEKA tool has been used. This tool is freely available on the web and is generally utilized for classification.

3.5.3.1 Naive Bayes Classifier

Naive Bayes algorithm generally used to predict the class of unknown dataset. Naive Bayes algorithm works on assumptions to label an item whose features are known but the name is unknown. Naive Bayes algorithm is a comparatively fast algorithm in terms of classification. It works faster on huge datasets and achieves high accuracy. In Naive Bayes classifier, the error rate is minimum as compared to other classifiers. However, the problem is that it assumes the independence between features. It cannot learn the interactions between features.

3.5.3.2 K-Nearest Neighbors Classifier

K-Nearest Neighbor is considered a lazy learning algorithm that classifies data sets based on their similarity with neighbors. K represents the count of dataset items that are considered for the classification. K-Nearest Neighbor is easy to understand and implement classification technique. It also performs well in many situations regardless of its simplicity. But it requires more work when making a classification and it depends on the K-value.

3.5.3.3 Support Vector Machine Classifier

Support Vector Machine (SVM) is the most well-known and examined machine learning algorithms. It remains in mainstream around the time they were created in the 1990s and keep on being the go-to technique for a high-performing algorithm with a little tuning. SVM converts training data into a higher dimension, where it finds a hyperplane that isolates the data by class utilizing important training tuples called support vectors.

3.5.3.4 Metrics for Evaluating Classifier Performance

Precision, Recall and F-measure were calculated to evaluate the performance and accuracy of each classifier. The formula of Precision, Recall and F-measure is shown in equation 3.12, 3.13 and 3.14:

$$Precision = \frac{True \quad Positive}{True \quad Positive + False \quad Positive} \tag{3.12}$$

Precision measures that how many results are correctly classified for a class.

$$Recall = \frac{True \quad Positive}{True \quad Positive + False \quad Negative} \tag{3.13}$$

Recall measures that how much relevant results have been returned out of total relevant results existed for that class.

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3.14}$$

F-measure combines Precision and Recall into one measure. The detailed results of this research question is discussed in chapter 4.

Chapter 4

Experimental Results

This chapter explains the results achieved by the adopted methodology discussed in chapter 3. It evaluates "Scientific Services and Affiliations" based parameters with primitive, citation intensity and age of publications based parameters using three research questions. The section 4.1 describes results obtained from our dataset. The section 4.2 represents ranking criteria of researchers based on author ranking parameters. Section 4.3 describes results evaluations using three research questions.

4.1 Dataset Results

In this thesis, "Scientific Services and Affiliations" based parameters have been proposed for researchers' assessment. We have considered prestigious awards of two International Computer Science Societies which include Association for Computing Machinery (ACM) and The Institute of Electrical and Electronics Engineers (IEEE) as benchmark to evaluate three research questions. These highly regarded researchers can help to find whether the proposed "Scientific Services and Affiliations" based parameters also discovers these highly regarded researchers or not. We have considered 24 awards of the afore-mentioned two international computing societies. The list of Computer Science awards and their total number of awardees have been shown below in Table 4.1:

Table 4.1: Computer Science Societies and their awards with total number of awardees

Societies and Their Awards	Total Awardees
ACM	
ACM Turing Award	60
ACM-Fellow	958
Software System Award	120
Gordon Bell Prize	161
Karl V. Karlstrom Outstanding Educator Award	24
Outstanding Contribution to ACM Award	45
Doctoral Dissertation Award	92
Grace Murray Hopper Award	40
Paris Kanellakis Theory and Practice Award	40
Distinguished Service Award	44
Infosys Foundation Award	08
Programming Languages Achievement Award	19
Eugene L. Lawler Award	10
AAAI Allen Newell Award	21
ACM Presidential Award	17
ACM Prize in comp. science and Engg.	14
Programming system and language Award	31
ACM-W Athena Lecture Award	09
ACM-IEEE CS Eckert-Mauchly Award	36
ACM-IEEE CS George Michael HPC Fellowships	31
Total	1780
IEEE	
IEEE Seymour Cray Award	16
IEEE Computer Pioneer Award	94
IEEE technical achievement Award	75
IEEE John von Neumann Medal	27
Total	212
Grand Total	1992

In the Table 4.1, benchmark dataset consists of 1992 award winners of two international computer science societies. However, this benchmark dataset contained repetitions of researchers' name and unavailability of researchers on the web. After removing ambiguities, our benchmark dataset reduced to 1306 unique awardees out of 1992 awardees.

Moreover, we have added some noise to this dataset. We have considered a dataset built by Jens Palsberg (Jens Palsberg, Jan 31, 2018) which contain 1134 Computer Scientists ranked based on h-index. Jens Palsberg is a professor of Computer Science in University of California, Los Angeles. Jens Palsberg rank the Computer Science researchers every year based on h-index of 40 or higher according to Google Scholar. We have considered this dataset as non-awardees in this thesis. The reason behind this is to investigate whether the proposed "Scientific Services and

Affiliations" based parameters distinguishes highly regarded researchers from those researchers who have not received any distinction.

"Scientific Services and Affiliations" based parameters of a particular researcher have been collected from their home pages as identified in chapter 3. Some problems have been faced during the collection of dataset such as researchers' name ambiguities and unavailability of researchers profiles on the web. After resolving all afore-mentioned problems, our dataset was reduced to 1306 awardees and 1049 non awardees. This dataset is comprehensively sufficient to demeanor experimental study and to evaluate "Scientific Services and Affiliations" based parameters for researchers' assessment.

4.2 Ranked Lists Generation based on Author Ranking Parameters

After collecting and establishing the data, we have achieved the ranking of authors based on author ranking parameters. The authors having greater parameter value were considered more expert than the authors having smaller parameter value. The author having equal parameter value were ranked at the same position. After acquiring 15 ranking lists from author ranking parameters, research questions will be answered.

4.3 Results Evaluation

This section evaluates "Scientific Services and Affiliations" based parameters, primitive parameters, citations intensity based parameters and age of publication based parameters based on three research questions. Section 4.3.1 deals with the first research question i.e. "Which author ranking parameter out of above mentioned 15 parameters is able to bring international awardees in top ranking". Section 4.3.2 deals with the second research question to find which parameter from scientific services and affiliations is best correlated with each parameter from other ranking parameter categories (primitive parameters, citation intensity based parameters, and age of publications based parameters). Section 4.3.3 deals with the 3rd research question i.e. "What is the ranking of all mentioned 15 parameters for classification".

4.3.1 Evaluation of Awardees Occurrences in Ranked Lists

To answer the first research question, we checked the awardees occurrences in top 10% of the ranked lists obtained from each author ranking parameter. Benchmark list of 1306 awardees have been considered in this thesis after duplication. Figure 4.1 shows percentage occurrences of awardees in top 10% of ranked lists obtained from author ranking parameters.

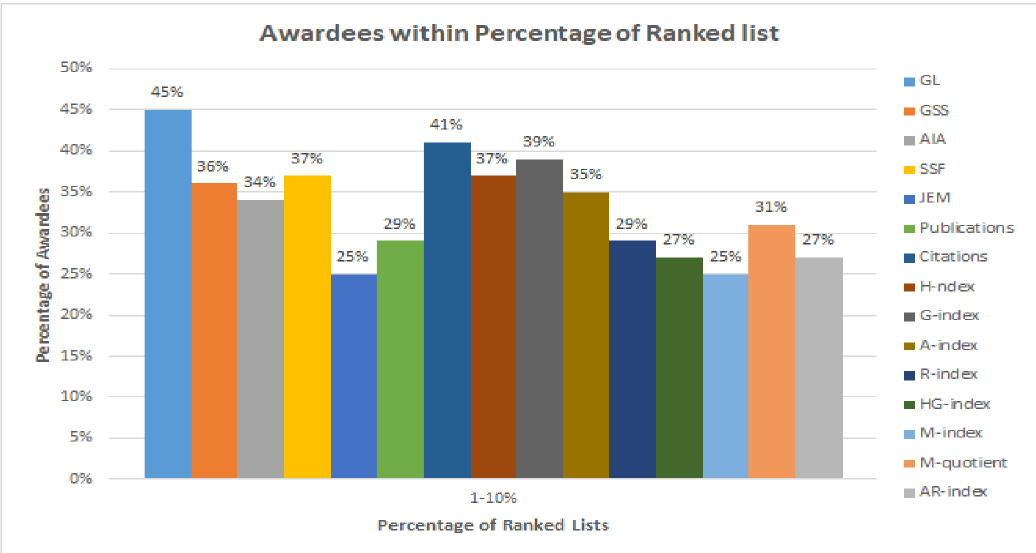

FIGURE 4.1: Percentage Awardees Occurrences within top 10% of ranked lists

Figure 4.1 shows percentage occurrences of awardees in the top 10% of ranked lists obtained from the proposed "Scientific Services and Affiliations", primitive, citation intensity and age of publications based parameters. Our dataset contained 2355 computer science researchers which consisted of both award winners and non-awardees. It can be clearly seen from Figure 4.1 that all the parameters performed well in order to bring international prestigious awardees on top 10% of the ranked lists. On average 25% of awardees exists in top 10% of ranked lists for each parameter. From Figure 4.1, we can see that Geographical Location and Citations have ranked higher awardees within 10%. However, Journal Editorial Memberships and m-index have least percentage of awardees in top 10% of ranked lists. Graduate Student Supervisions, Scientific Societies Fellowships, h-index, g-index and a-index performed almost equivalent than other ranking parameters. Overall, all parameters showed increased percentage of occurrences of award winners in top 10% of ranked lists for each author ranking parameter.

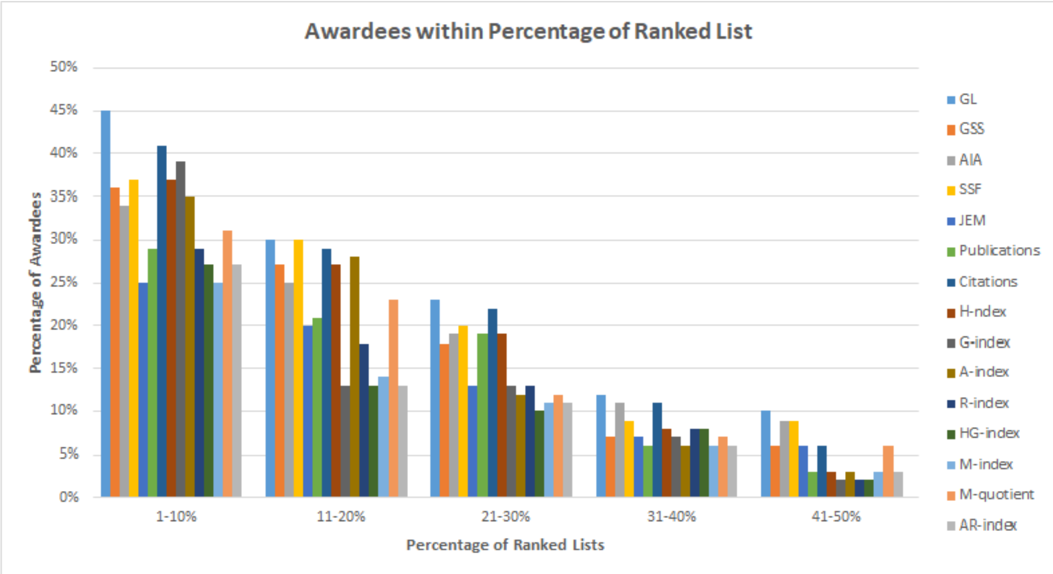

FIGURE 4.2: Percentage Awardees Occurrences within Ranked Lists

Figure 4.2 represents the percentage occurrences of awardees in 1-10%, 11-20%, 21-30%, 31-40%, and 41-50% of the ranked lists of authors. It can be seen clearly from Figure 4.2 that most of the awardees exist in top 10% of the ranked lists of authors. It is clear from the Figure that Geographical Location and Citations performed much better than all other ranking parameters in top 10% of the ranked lists. While in 11-20%, Geographical Location, Scientific Societies Fellowships, Citations and A-index have better performance than other ranking parameters. In 21-30%, Geographical Location, and Citations have maximum number of awardees. In 31-40% and 41-50%, the performance of all ranking parameters remained equal except Geographical Location, Academic Institution Affiliations, Scientific Societies Fellowships, and Citations.

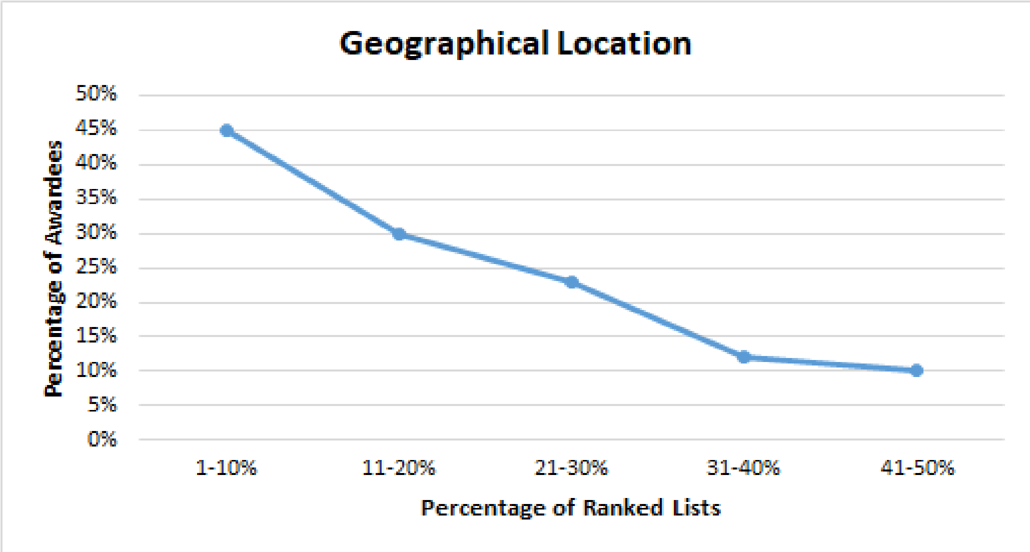

FIGURE 4.3: Trend of Geographical Location

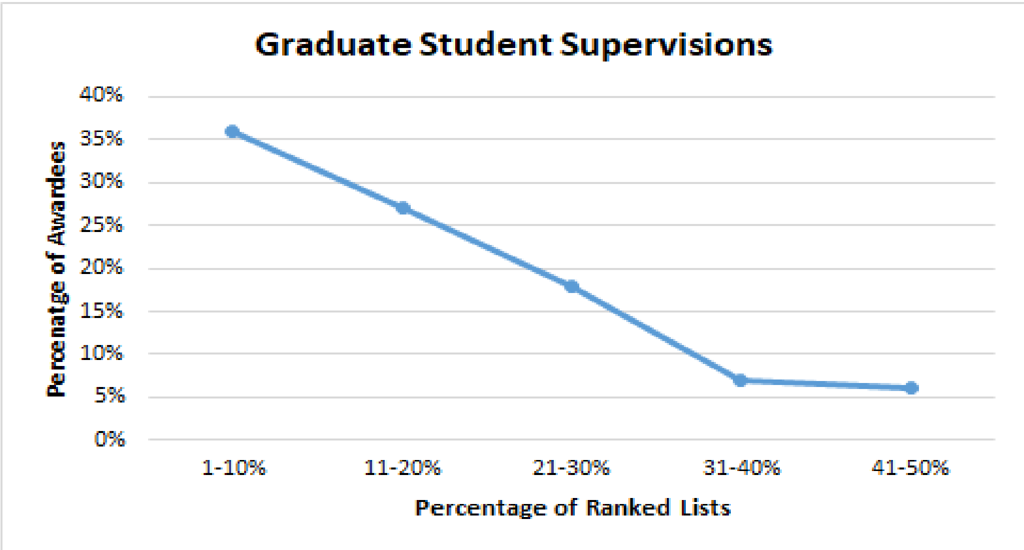

FIGURE 4.4: Trend of Graduate Student Supervisions

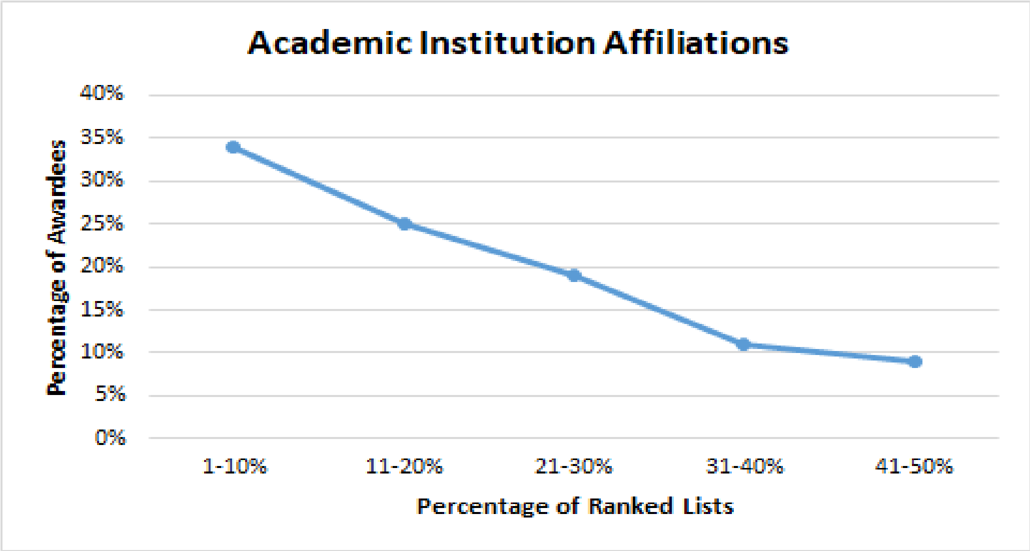

Figure 4.5: Trend of Academic Institution Affiliations

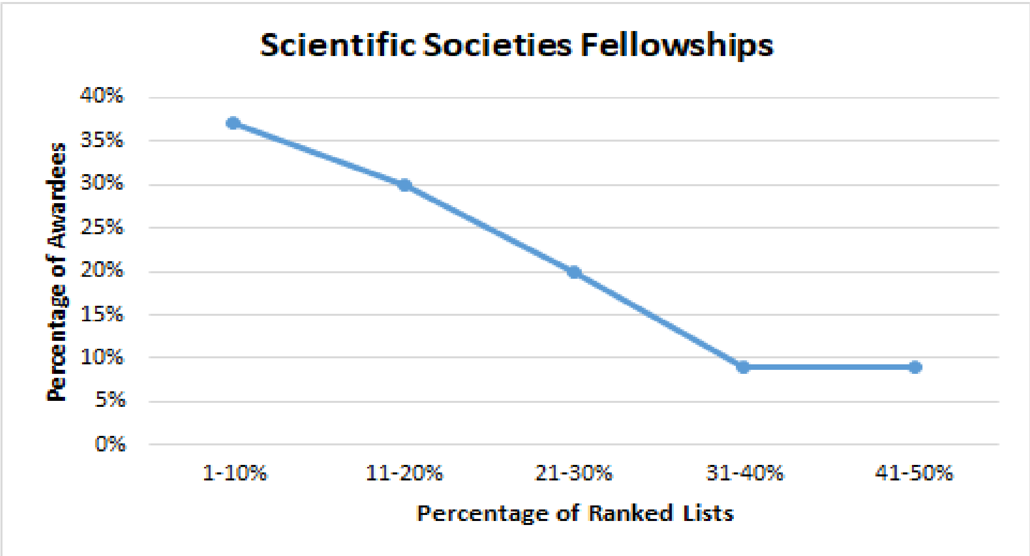

Figure 4.6: Trend of Scientific Societies Fellowships

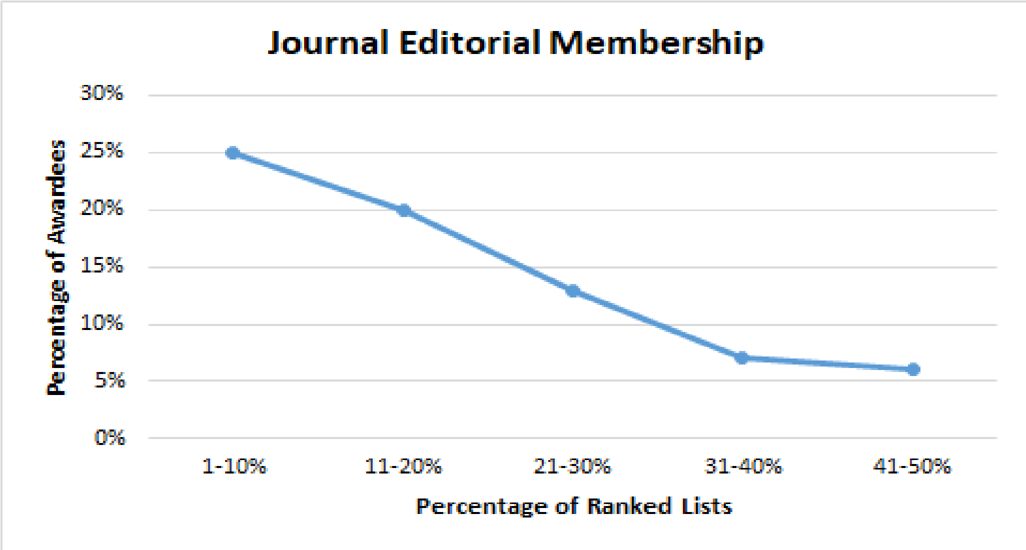

FIGURE 4.7: Trend of Journal Editorial Memberships

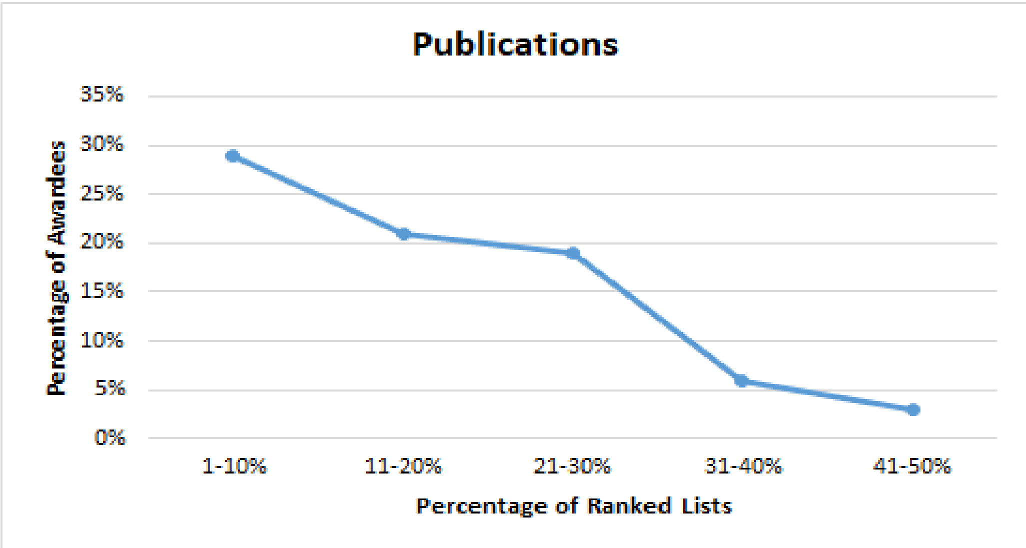

FIGURE 4.8: Trend of Publications

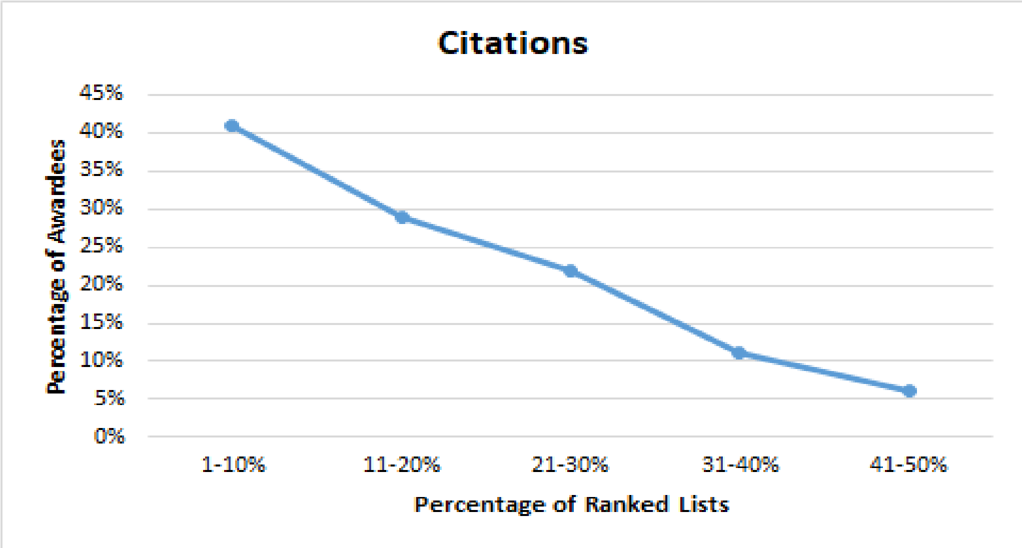

FIGURE 4.9: Trend of Citations

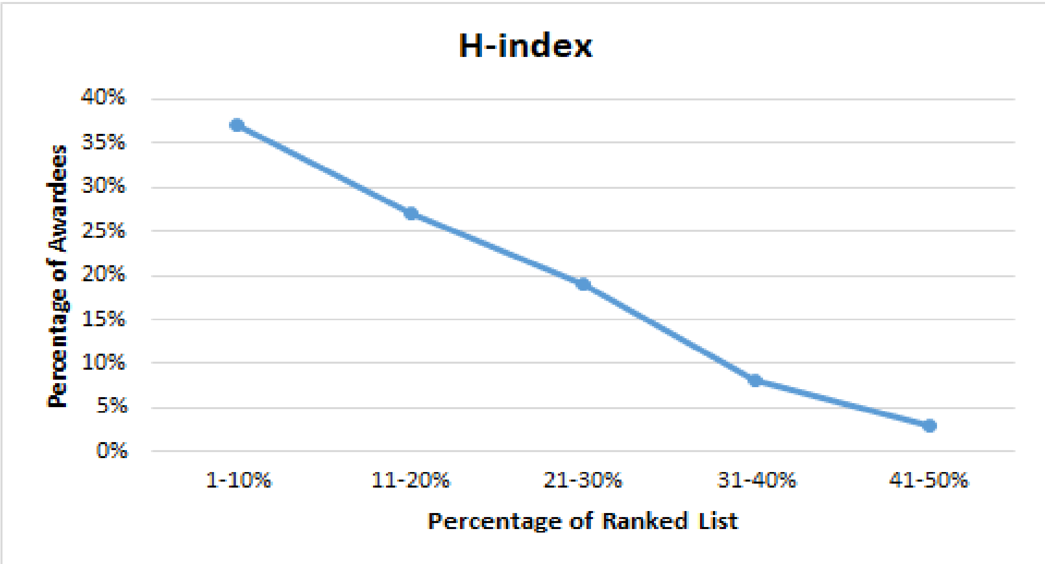

FIGURE 4.10: Trend of H-index

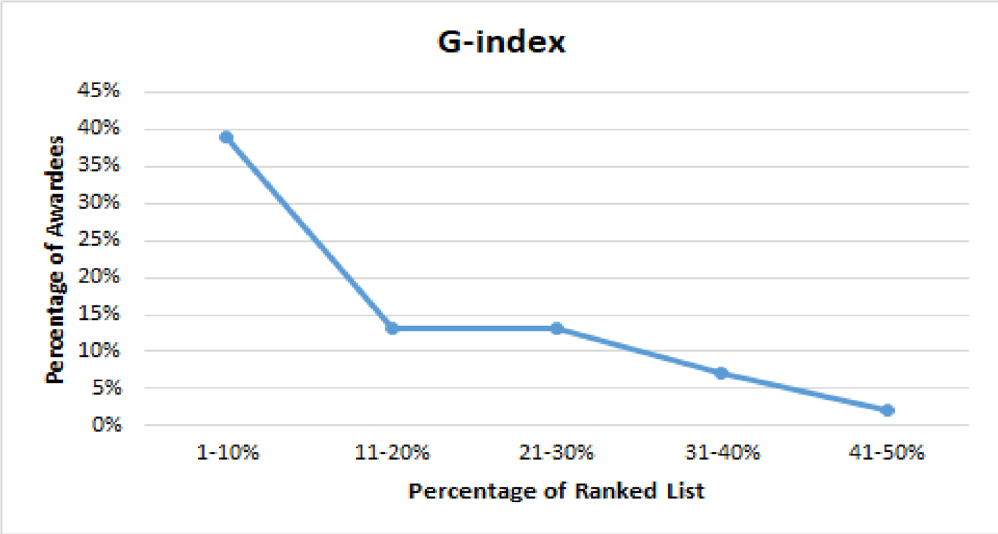

FIGURE 4.11: Trend of G-index

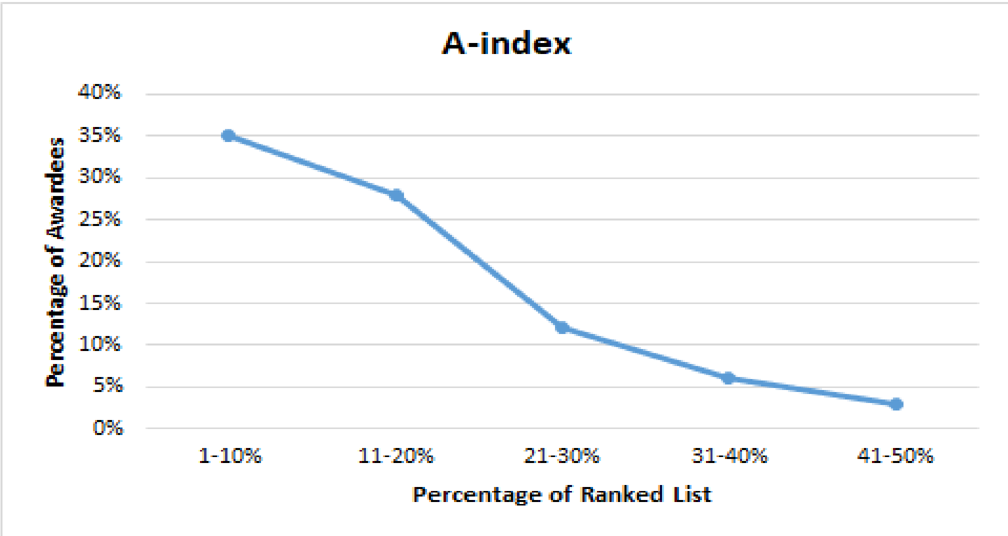

FIGURE 4.12: Trend of A-index

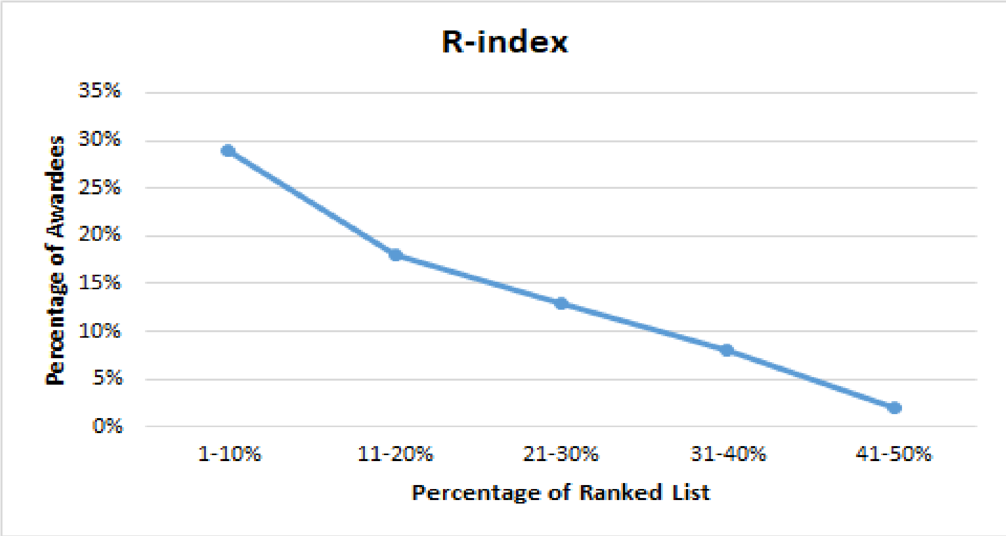

FIGURE 4.13: Trend of R-index

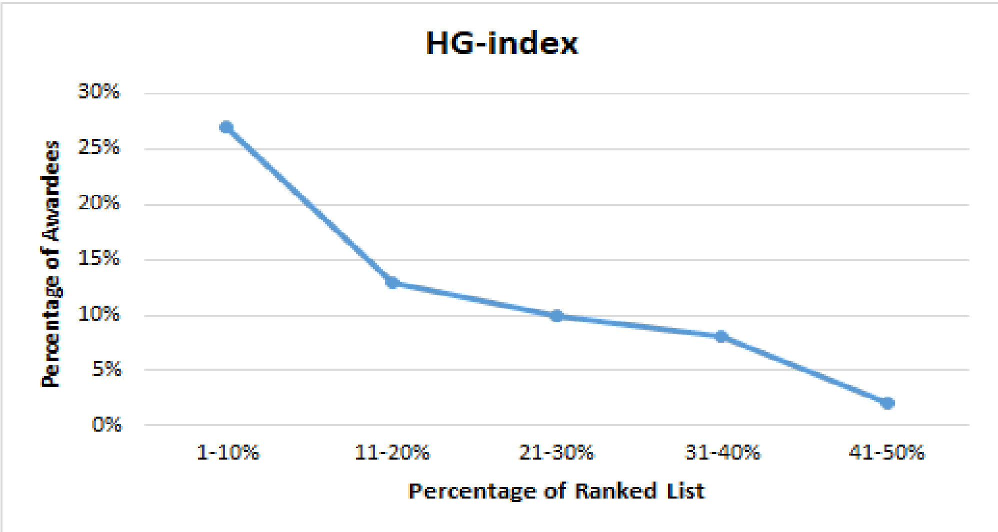

FIGURE 4.14: Trend of HG-index

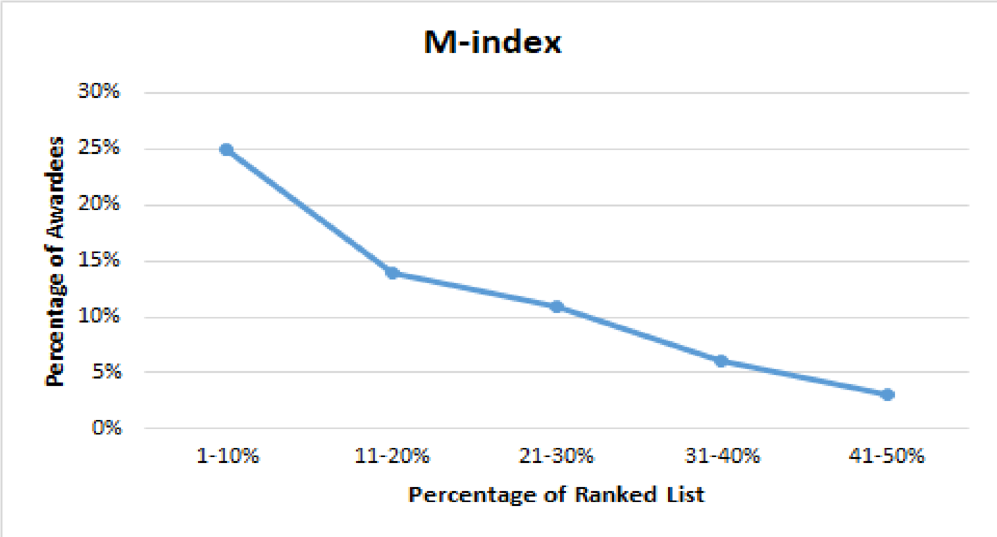

FIGURE 4.15: Trend of M-index

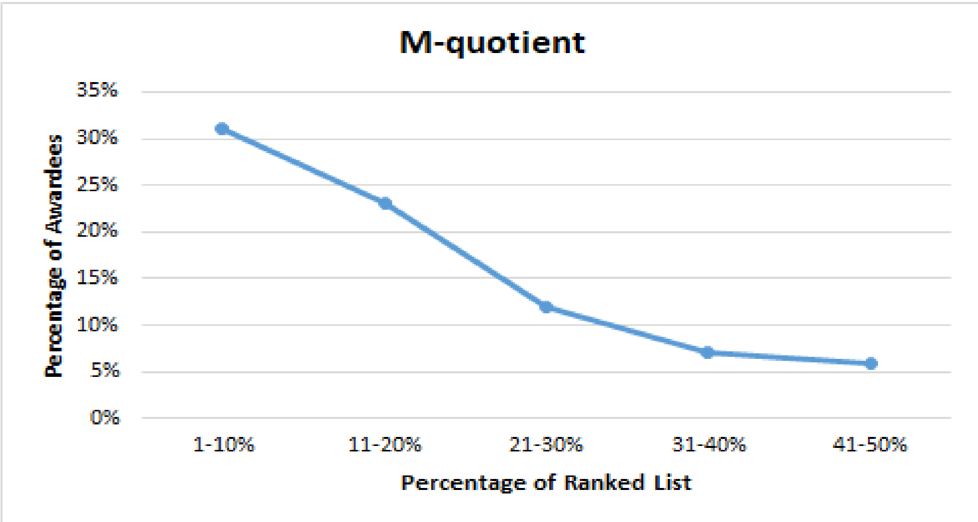

FIGURE 4.16: Trend of M-quotient

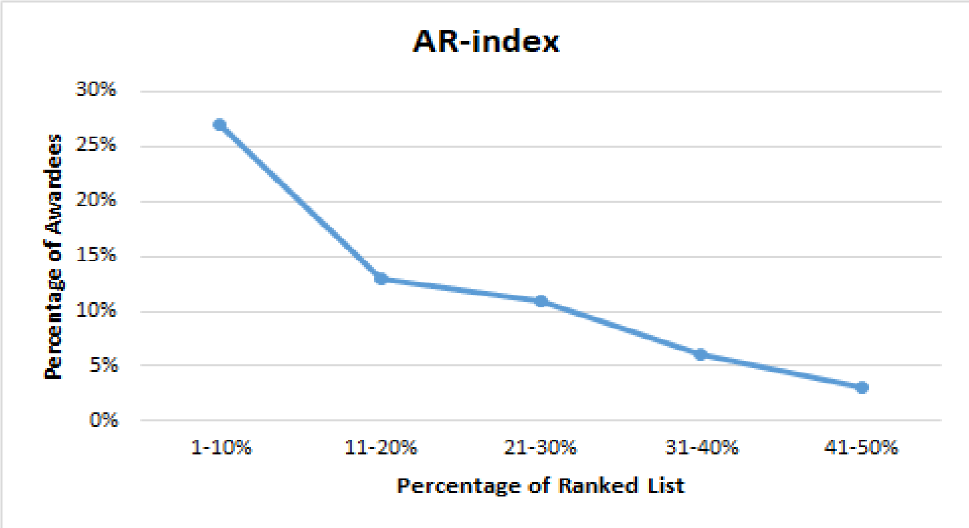

FIGURE 4.17: Trend of AR-index

Figure 4.3 to 4.17 represent the trend of awardees occurrences in ranked list of each ranking parameter. This shows that all the parameters brought the awardees in top 10% of ranked lists and as we go downward in the ranking, the awardees occurrences decrease smoothly. This is logical that each parameter brings most of the awardees in the top ranking, however, one important finding is that the maximum number of awardees in the top 10% are just 45% by any of the evaluated parameter.

4.3.2 Evaluation of Correlation Between Ranked Lists

Our second research question is defined to measure correlation between proposed "Scientific Services and Affiliations" based parameters and primitive, citation intensity and age of publications based parameters. The main purpose of finding correlation is to find similarity between these parameters. Correlation coefficient values greater than 0 shows positive correlation which indicates that two parameters have similar behavior whereas, correlation coefficient values less than 0 shows negative correlation which means two parameters have different behavior [49]. There is considered a strong correlation if the correlation coefficient is greater than 0.8 and a weak correlation if the correlation coefficient is less than 0.5 [50].

Formally, Spearman correlation is defined as:

$$\rho = \frac{6\sum di^2}{n(n^2-1)} \quad \text{Whereas,} \quad di = xi - yi \tag{4.1}$$

4.3.2.1 Spearman Correlation Between Ranked Lists Obtained from Primitive Parameters and Proposed Parameters

Figure 4.18 to 4.20 presents spearman correlation of ranked lists generated from Primitive parameters and "Scientific Services and Affiliations" based parameters. Figure 4.18 shows spearman correlation of "Publication" with proposed "Scientific Services and Affiliations" based parameters. It can be clearly seen that 'Publication' has positive correlation with all proposed parameters such as 'Graduate Student Supervisions', 'Academic Institutions Affiliations', 'Scientific Societies Fellowships', 'Journal Editorial Memberships', and 'Geographical Location'. It means that 'Publication' based ranking has similar behavior with proposed parameters based ranking. 'Publication' has highest correlation with 'Scientific Societies Fellowships' that is 0.89 and least correlation with 'Journal Editorial Memberships' which is 0.46.

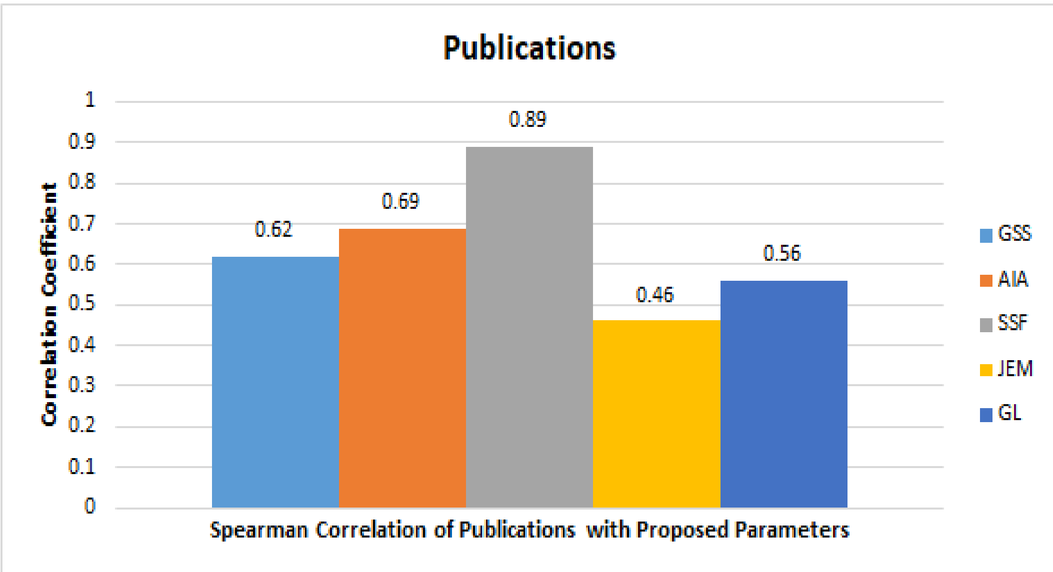

FIGURE 4.18: Correlation of Publication Parameter with Proposed Parameters

Figure 4.19 presents spearman correlation of "Citation" with proposed "Scientific Services and Affiliations" based parameters. 'Citation' has also positive correlation with all proposed parameters such as 'Graduate Student Supervisions', 'Academic Institutions Affiliations', 'Scientific Societies Fellowships', 'Journal Editorial Memberships', and 'Geographical Location'. 'Citation' has highest correlation with 'Scientific Societies Fellowships' that is 0.94 and least correlation with 'Geographical Location' which is 0.43.

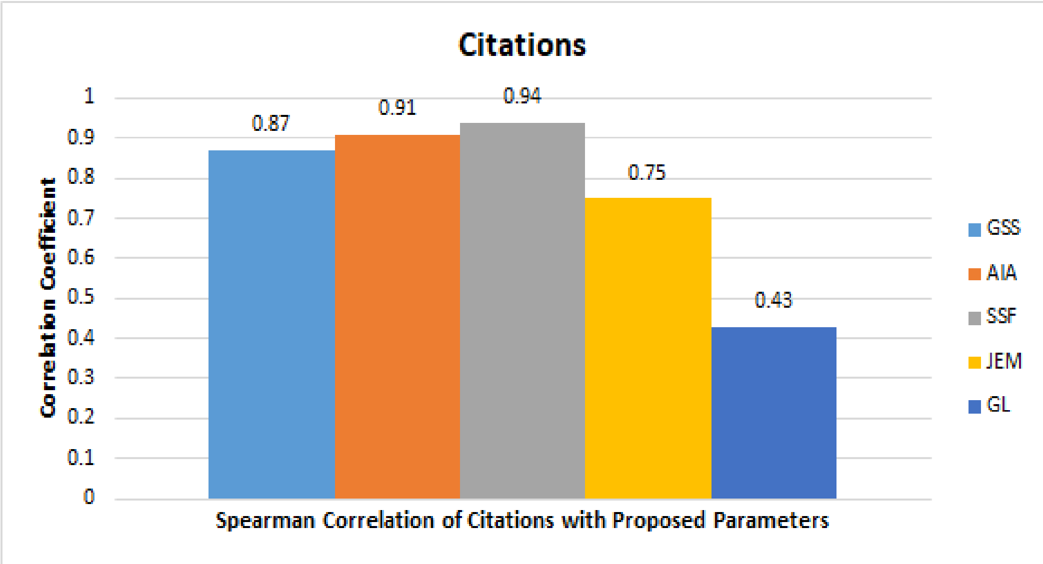

FIGURE 4.19: Correlation of Citation Parameter with Proposed Parameters

Figure 4.20 depicts spearman correlation of "H-index" with proposed "Scientific Services and Affiliations" based parameters such as 'Graduate Student Supervisions', 'Academic Institutions Affiliations', 'Scientific Societies Fellowships', 'Journal Editorial Memberships', and 'Geographical Location'. 'H-index' has also positive correlation with all proposed parameters. 'H-index' has highest correlation with 'Scientific Societies Fellowships' that is 0.90 and least correlation with 'Journal Editorial Memberships' which is 0.49.

From Figure 4.18-4.20, it can be concluded that 'Primitive' parameters have higher positive correlation with 'Scientific Societies Fellowships' than other parameters which reveals that if we want to quickly identify the prestige of an author, we can conclude it by just looking on his/her "Scientific Societies Fellowships" rather than calculating complex values on publications and citations. It is not an easy task to get all publications and citations of a researcher and then performing complex formulae on them. Moreover, all 'Primitive' parameters have positive correlation

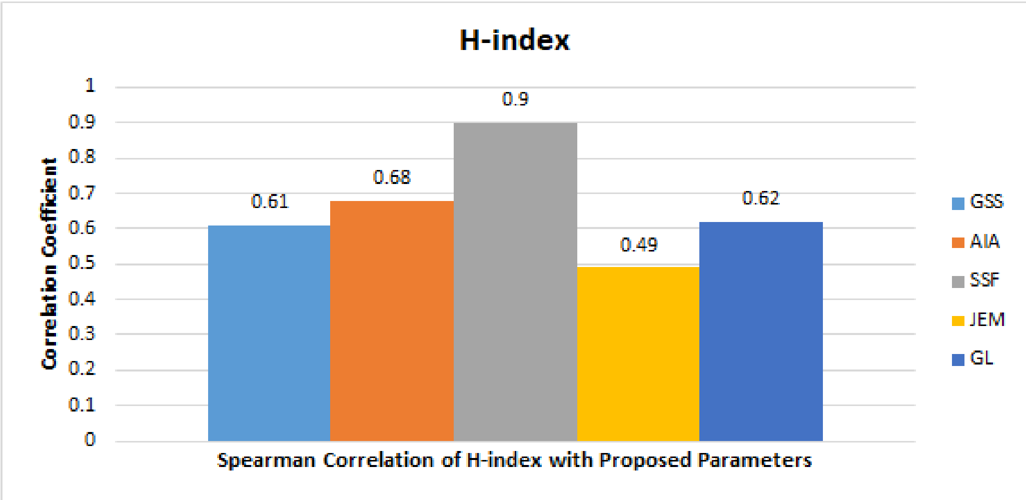

FIGURE 4.20: Correlation of H-index Parameter with Proposed Parameters

with 'Scientific Services and Affiliations' based parameters as shown in the Figures 4.18-4.20.

4.3.2.2 Spearman Correlation Between Ranked Lists Obtained from Citation Intensity based Parameters and Proposed Parameters

Figure 4.21 to 4.24 shows spearman correlation between ranked lists generated from "Citation Intensity" based parameters and "Scientific Services and Affiliations" based parameters. Figure 4.21 presents spearman correlation of "G-index" with proposed "Scientific Services and Affiliations" based parameters. It can be clearly seen that 'G-index' has positive correlation with all proposed parameters such as 'Graduate Student Supervisions', 'Academic Institutions Affiliations', 'Scientific Societies Fellowships', 'Journal Editorial Memberships', and 'Geographical Location'. It means that 'G-index' has greater similarity with proposed parameters based ranking. 'G-index' has highest correlation with 'Scientific Societies Fellowships' that is 0.92 and least correlation with 'Journal Editorial Memberships' which is 0.54.

Figure 4.22 shows spearman correlation of "A-index" with proposed "Scientific Services and Affiliations" based parameters such as 'Graduate Student Supervisions',

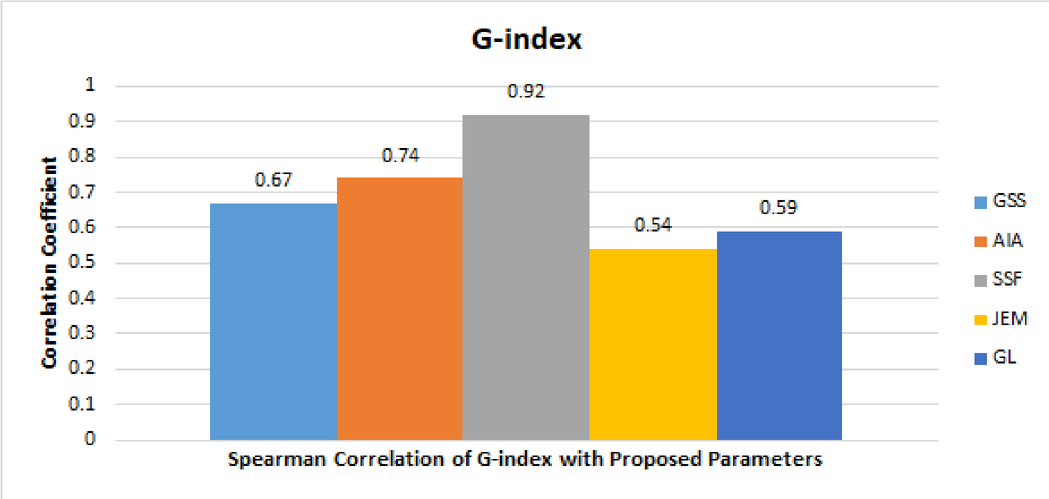

FIGURE 4.21: Correlation of G-index Parameter with Proposed Parameters

'Academic Institutions Affiliations', 'Scientific Societies Fellowships', 'Journal Editorial Memberships', and 'Geographical Location'. It can be clearly seen that 'A-index' has positive correlation with all proposed parameters. It means that 'A-index' has also greater similarity with proposed parameters based ranking. 'A-index' has highest correlation with 'Scientific Societies Fellowships' that is 0.95 and least correlation with 'Geographical Location' which is 0.49.

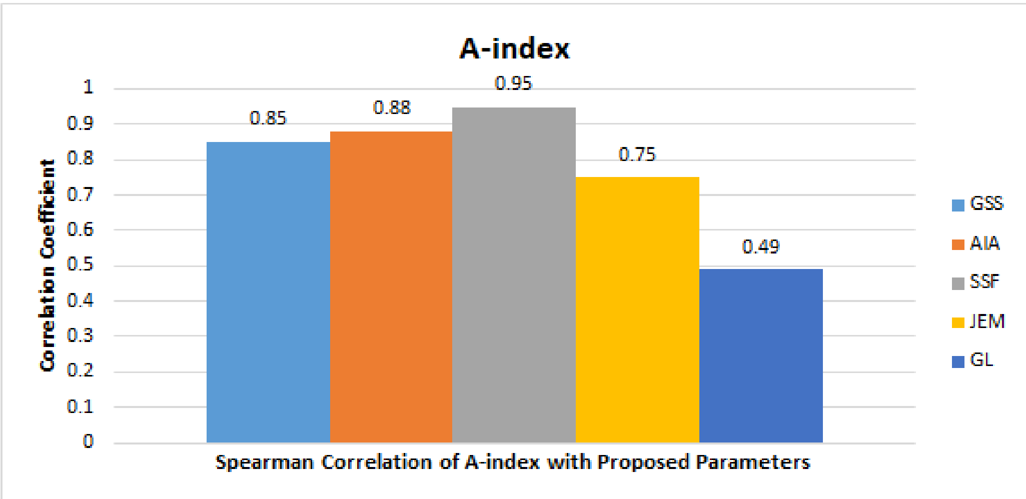

FIGURE 4.22: Correlation of A-index Parameter with Proposed Parameters

Figure 4.23 presents spearman correlation of "R-index" with proposed "Scientific

Services and Affiliations" based parameters. 'R-index' has also positive correlation with all proposed parameters such as 'Graduate Student Supervisions', 'Academic Institutions Affiliations', 'Scientific Societies Fellowships', 'Journal Editorial Memberships', and 'Geographical Location'. 'R-index' has highest correlation with 'Scientific Societies Fellowships' that is 0.92 and least correlation with 'Journal Editorial Memberships' which is 0.55.

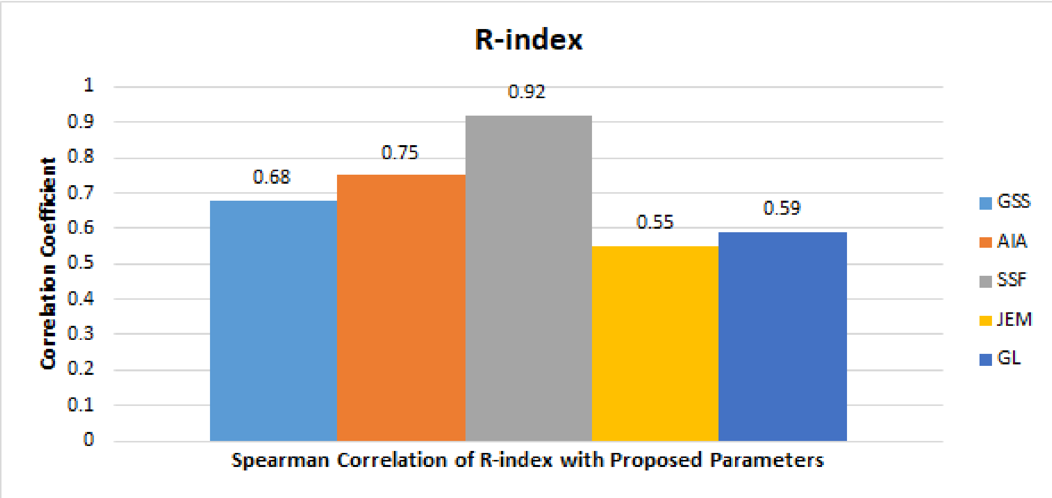

FIGURE 4.23: Correlation of R-index Parameter with Proposed Parameters

Figure 4.24 shows spearman correlation of "HG-index" with proposed "Scientific Services and Affiliations" based parameters such as 'Graduate Student Supervisions', 'Academic Institutions Affiliations', 'Scientific Societies Fellowships', 'Journal Editorial Memberships', and 'Geographical Location'. It can be clearly seen that 'HG-index' has positive correlation with all proposed parameters. It means that 'HG-index' has also greater similarity with proposed parameters based ranking. 'HG-index' has highest correlation with 'Scientific Societies Fellowships' that is 0.81 and least correlation with 'Journal Editorial Memberships' which is 0.47.

From figure 4.21-4.24, it can be concluded that 'Citation Intensity' based parameters have higher positive correlation with 'Scientific Societies Fellowships' than other parameters which reveals that if we want to quickly identify the prestige of an author, we can conclude it by just looking on his/her "Scientific Societies Fellowships" rather than calculating complex values on publications and citations.

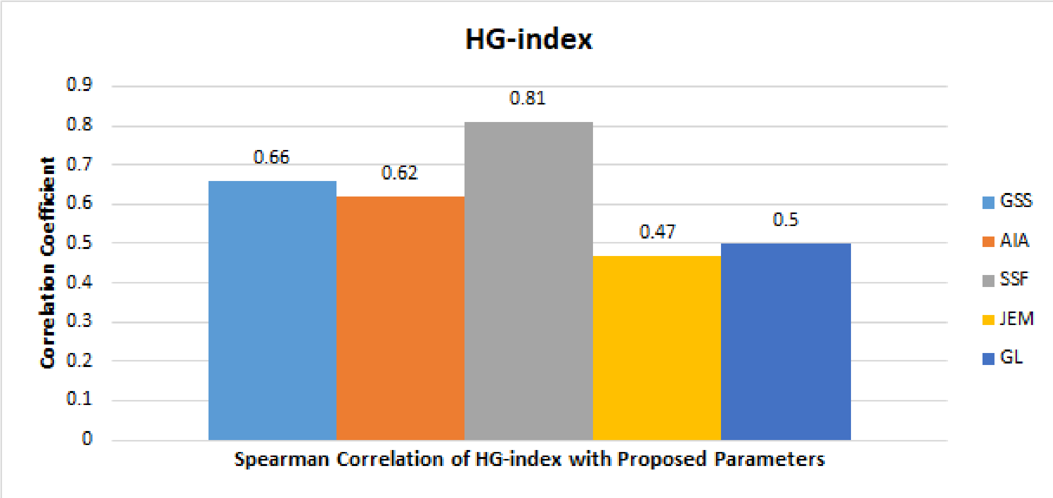

FIGURE 4.24: Correlation of HG-index Parameter with Proposed Parameters

It is not an easy task to get all publications and citations of a researcher and then performing complex formulae on them. Moreover, all 'Citation Intensity' based parameters have positive correlation with 'Scientific Services and Affiliations' based parameters as shown in the Figures 4.21-4.24.

4.3.2.3 Spearman Correlation between Ranked Lists Obtained from Age of Publications based Parameters and Proposed Parameters

Figure 4.25 to 4.27 presents spearman correlation of ranked lists generated from "Age of Publications" based parameters and "Scientific Services and Affiliations" based parameters. Figure 4.25 shows spearman correlation of "M-index" with proposed "Scientific Services and Affiliations" based parameters. It can be clearly seen that 'M-index' has positive correlation with all proposed parameters such as 'Graduate Student Supervisions', 'Academic Institutions Affiliations', 'Scientific Societies Fellowships', 'Journal Editorial Memberships', and 'Geographical Location'. It means that 'M-index' based ranking has similar behavior with proposed parameters based ranking. 'M-index' has highest correlation with 'Graduate Student Supervisions' that is 0.68 and least correlation with 'Geographical Location' which is 0.23.

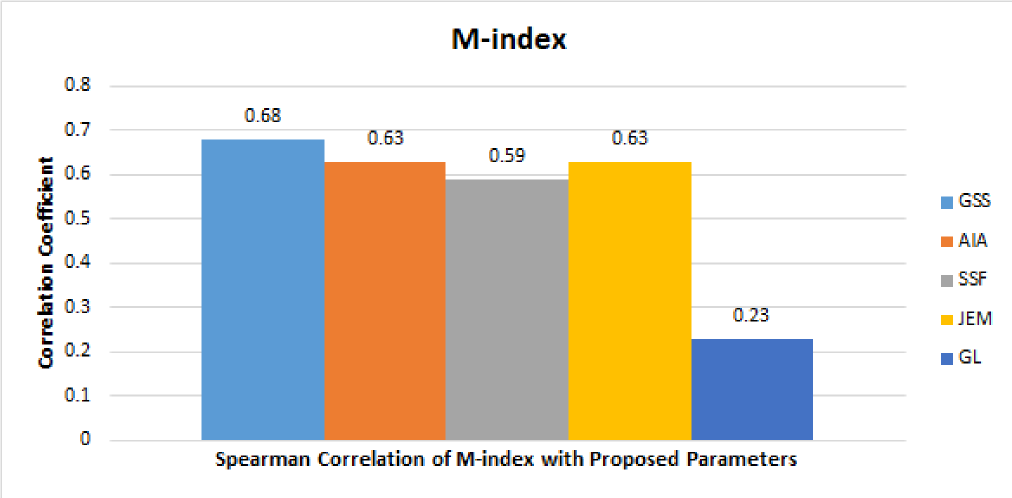

FIGURE 4.25: Correlation of M-index Parameter with Proposed Parameters

Figure 4.26 shows spearman correlation of "M-quotient" with proposed "Scientific Services and Affiliations" based parameters. It can be clearly seen that 'M-quotient' has positive correlation with all proposed parameters such as 'Graduate Student Supervisions', 'Academic Institutions Affiliations', 'Scientific Societies Fellowships', 'Journal Editorial Memberships', and 'Geographical Location'. 'M-quotient' has highest correlation with 'Scientific Societies Fellowships' that is 0.89 and least correlation with 'Journal Editorial Memberships' which is 0.43.

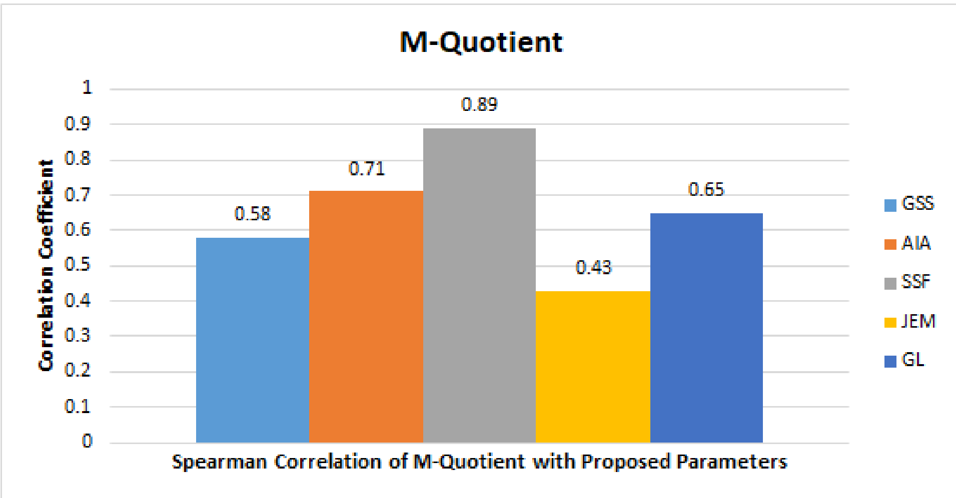

FIGURE 4.26: Correlation of M-Quotient Parameter with Proposed Parameters

Figure 4.27 shows spearman correlation of "AR-index" with proposed "Scientific Services and Affiliations" based parameters. It can be clearly seen that 'AR-index' has positive correlation with all proposed parameters such as 'Graduate Student Supervisions', 'Academic Institutions Affiliations', 'Scientific Societies Fellowships', 'Journal Editorial Memberships', and 'Geographical Location'. 'AR-index' has highest correlation with 'Journal Editorial Memberships' that is 0.66 and least correlation with 'Geographical Location' which is 0.16.

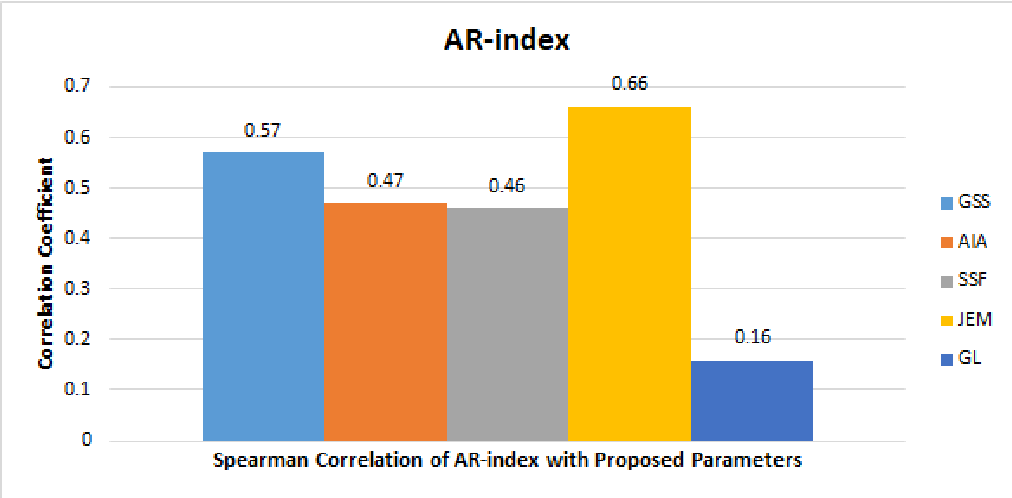

FIGURE 4.27: Correlation of AR-index Parameter with Proposed Parameters

From figure 4.25-4.27, it can be concluded that 'M-quotient' has higher positive correlation with 'Scientific Societies and Fellowships' which reveals that these parameters produced similar rankings. Moreover, all 'Age of Publications' based parameters have positive correlation with 'Scientific Services and Affiliations' based parameters as shown in the Figures 4.21-4.24. This means if we want to quickly identify the prestige of an author, we can conclude it by just looking on his/her "Scientific Societies Fellowships" rather than calculating complex values on publications and citations. It is not an easy task to get all publications and citations of a researcher and then performing complex formulae on them. The conclusion from the results is the "Scientific Societies Fellowships" proposed parameter is equally good than the other evaluated parameters.

4.3.3 Ranking of Proposed and Existing Parameters for Classification

Our 3rd research question is defined to rank "Scientific Services and Affiliations", primitive, citation intensity and age of publications based parameters for classification. Three classifiers Naive Bayes, K-Nearest Neighbors and Support Vector Machine have been identified to rank "Scientific Services and Affiliations", primitive, citation intensity and age of publications based parameters for classification. The formula of Precision, Recall and F-measure are used to evaluate the accuracy of a classifier.

Figure 4.28 and Figure 4.29 shows results of Naive Bayes classifier when applied to rank the effect of "Scientific Services and Affiliations", primitive, citation intensity and age of publications based parameters to classify researchers. Precision, Recall and F-measure score is calculated against all three classifiers. It can be clearly seen from Figure 4.29 that 'Graduate Student Supervisions' is the proposed parameter that is ranked on the first position. It means that this parameter has returned more correct classification results. Then 'Geographical Location' is ranked on the second position according to the Naive Bayes classifier and 'H-index' is ranked on the 3rd position. This means that proposed "Scientific Services and Affiliations" based parameters outperforms the existing primitive, citation intensity and age of publication based parameters according to Naive Bayes classifier. Furthermore, Naive Bayes classifier has achieved the precision of 0.76, Recall of 0.97 and F-measure of 0.85.

Figure 4.30 shows results of K-Nearest Neighbor classifier when applied to rank the effect of "Scientific Services and Affiliations", primitive, citation intensity and age of publications based parameters to classify researchers. It can be clearly seen from Figure 4.30 that K-Nearest Neighbor also produce same features ranking as in Figure 29. Moreover, K-Nearest Neighbor classifier has achieved the precision of 0.803, Recall of 0.805 and F-measure of 0.804.

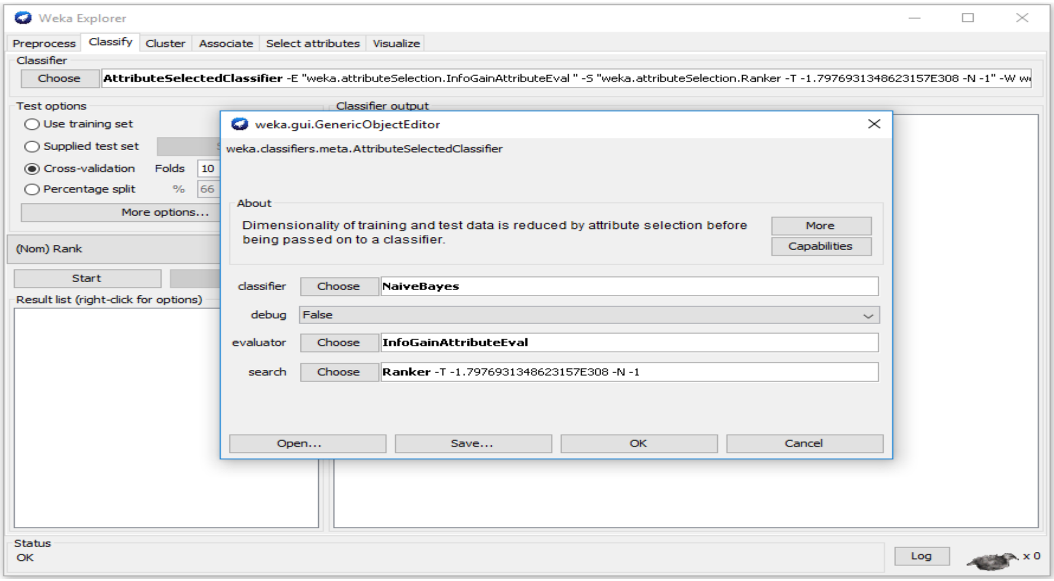

FIGURE 4.28: Ranking of Parameters using Naïve Bayes Classifier

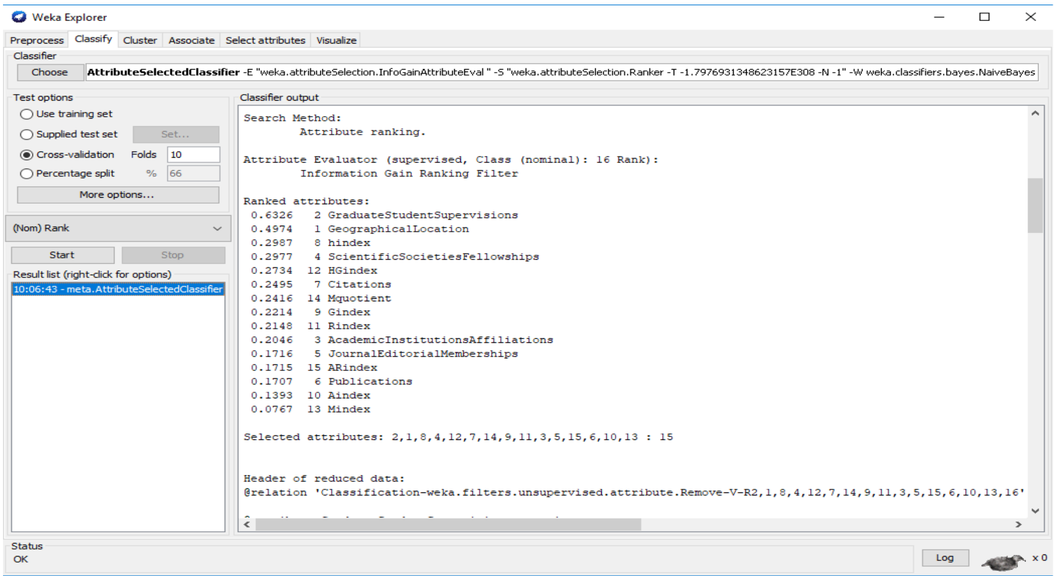

FIGURE 4.29: Ranking of Parameters using Naive Bayes Classifier

Figure 4.31 shows results of Support Vector Machine classifier when applied to rank the “Scientific Services and Affiliations”, primitive, citation intensity and age of publications based parameters to classify researchers. It can be clearly seen from Figure 4.31 that Support Vector Machine also produce same features ranking as in Figure 29 and Figure 30. Moreover, Support Vector Machine classifier has achieved the precision of 0.79, Recall of 0.85 and F-measure of 0.82.

It is clear from classification results that Naive Bayes, K-Nearest Neighbor, and

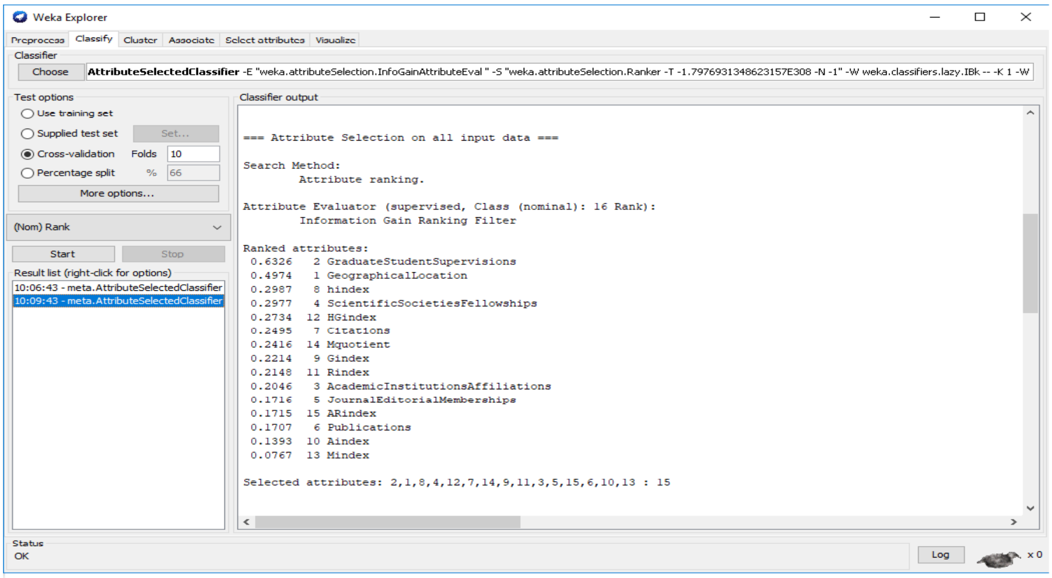

FIGURE 4.30: Ranking of Parameters using K-Nearest Neighbor Classifier

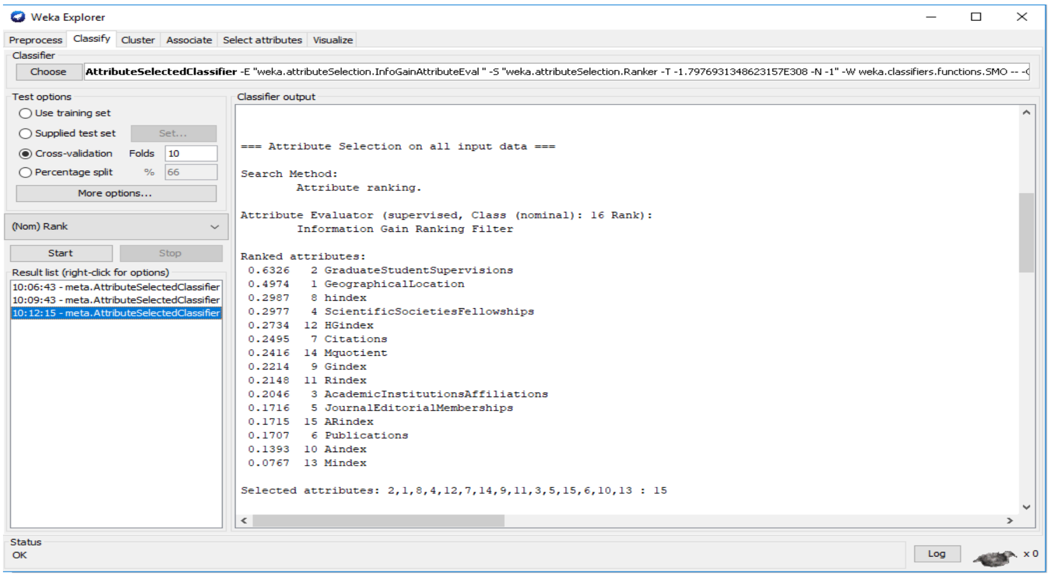

FIGURE 4.31: Ranking of Parameters using Support Vector Machine Classifier

Support Vector Machine produced similar features ranking. This means that the behavior of ranking features remained same by applying different classifier. Furthermore, Naive Bayes classifier has achieved the best Precision, Recall, and F-measure scores among other classifiers.

Chapter 5

Conclusion and Future Work

The aim of current chapter is to conclude the results, which are reported earlier in the previous chapter. It will also explain the significance of the research and future work.

5.1 Conclusion

Researchers' assessment has received increased interest in recent years. Researchers' assessment is useful in different scenarios such as: to find domain experts, to find reviewer of conferences and journals, to grant fellowships, memberships and awards, to hire talented faculty in universities and for promotions etc. In literature, different measures have been proposed to assess researcher's productivity such as: publication count, citation count, and h-index. Publication parameter measures the quantity of an individual researcher in scientific community. Limitation of using publications count is that publication count does not cover the true impact and quality of one's work. Later citation count proposed which presents total number of citations received by a researcher. It considers both quantity and quality of one's work. However, it also has some limitations. For example, some authors cite others work only to criticize which may increase number of citations for a paper. To overcome this problem, Hirsch proposed a new parameter called h-index. H-index combines the positive effect of both parameters which can help to assess researchers more appropriately. H-index has some hidden weaknesses such as number of citations received by a paper may increase over time even if an author is not publishing any new paper. However, h-index does not change for a researcher. Furthermore, 37 variants of the h-index have also been proposed which are g-index, A-index, R-index, Hg-index, m-quotient, AR-index and m-index. All of such parameters are totally dependent on publications and their citations.

After study these parameter, it has been identified that researchers have many other contributions such as supervisions of PhD's and are also granted with academic institution affiliations, scientific societies fellowships and journal editorial memberships. Such "Scientific Services and Affiliations" based parameters have been utilized for different purposes in literature. However, the effect of "Scientific Services and Affiliations" based parameters such as: "Graduate Students Supervision", "Academic Institution Affiliations", "Scientific Societies Fellowships", "Journal Editorial Memberships" and "Geographical Location" of researchers have not yet proposed for researchers' assessment. Therefore, this thesis is conducted to evaluate the impact of mentioned "Scientific Services and Affiliations" based parameters and to compare with primitive parameters, citation intensity based parameters and age of publications based parameters. All these ranking parameters have been evaluated on small dataset and sometime imaginary dataset. Therefore, these parameters need to be evaluated on comprehensive dataset.

In this thesis, we evaluate the impact of "Scientific Services and Affiliations" based parameters with primitive, citation intensity and age of publications based parameters for researchers' assessment in Computer Science domain. The reason behind selecting Computer Science domain is that it has extensive applications in all field of science. Moreover, none of the comprehensive and evaluating study has been performed in this field based on "Scientific Services and Affiliations" of researchers. We have formulated three research questions in this thesis i.e.

RQ1: Which author ranking parameter out of above mentioned 15 parameters is able to bring international awardees in top ranking?

RQ2: Which parameter from scientific services and affiliations is best correlated with each parameter from other ranking parameter categories (primitive parameters, citation intensity based parameters, and Age of publications based parameters)?

RQ3: What is the ranking of all mentioned 15 parameters for classification?

To answer these research questions, we required a comprehensive dataset which comprises of researchers highly regarded by scientific community. We have considered award winners of some major International Scientific Societies which are Association for Computing Machinery (ACM) and The Institute of Electrical and Electronics Engineers (IEEE) to evaluate whether the suggested parameters acknowledge these researchers or not. Moreover, we have added some noise to this dataset. We have considered a dataset built by Jens Palsberg (Jens Palsberg, Jan 31, 2018) which contain about 1000 Computer Scientists ranked based on h-index. Jens Palsberg rank the Computer Science researchers every year based on h-index of 40 or higher according to Google Scholar. We have considered this dataset as non-awardees in this thesis. The reason behind this is to investigate whether the proposed "Scientific Services and Affiliations" based parameters distinguishes awardees from those researchers who have not received any award but belong to the all type of profiles (High, Average and Low). For this purpose, we have used a dataset of 1992 computer science researchers as award winners in our thesis.

'Home pages' of researchers have been used to gather "Scientific Services and Affiliations" based parameters. Some problems such as researcher name duplication have been faced during collection of the "Scientific Services and Affiliations" of researchers' from web sources. Unavailability of "Scientific Services and Affiliations" of researchers was another problem which was faced during parameters collection. To evaluate "Scientific Services and Affiliations" based parameters with primitive, citation intensity and age of publications based parameters we needed publications, their citations and year of publications. The required data was collected from Google Scholar with the help of dedicated crawler. The reason to choose

Google Scholar is that it has huge coverage of data. Our final dataset contained 1306 award winners and about 1000 non awardees.

To answer the first research question, we checked the awardees occurrences in top 1-10%, 11-20%, 21-30%, 31-40%, and 41-50% of the ranked lists obtained from each author ranking parameter. On average 25% of awardees exists in top 10% of ranked lists for each parameter. Geographical Location and Citations performed much better than all other ranking parameters in top 10% of the ranked lists. While in 11-20%, Geographical Location, Scientific Societies Fellowships, Citations and A-index have better performance than other ranking parameters. In 21-30%, Geographical Location and Citations have maximum number of awardees. In 31-40% and 41-50%, the performance of all ranking parameters remained equal except Geographical Location, Academic Institution Affiliations, Scientific Societies Fellowships, and Citations.

To answer the second research question, we find correlation between ranked lists obtained from Primitive, Citation Intensity, age of Publications and "Scientific Services and Affiliations" based parameters. It has been observed from results analysis that Primitive, Citation Intensity and age of Publications have positive correlation with all "Scientific Services and Affiliations" based parameters. Furthermore, Primitive, Citation Intensity and age of Publications have higher positive correlation with "Scientific Societies Fellowships". This means if we want to quickly identify the prestige of an author, we can conclude it by just looking on his/her "Scientific Societies Fellowships" rather than calculating complex values on publications and citations. It is not an easy task to get all publications and citations of a researcher and then performing complex formulae on them. The conclusion from the results is the "Scientific Societies Fellowships" proposed parameter is equally good than the other evaluated parameters.

Our 3rd research question ranks the "Scientific Services and Affiliations", Primitive, Citation Intensity, and age of Publications based parameters for classification using three binary classifiers Naive Bayes, K-Nearest Neighbors, and Support Vector Machine. It is clear from classification results that Naive Bayes, K-Nearest

Neighbor, and Support Vector Machine produced similar features ranking. This means that the behavior of ranking features remained same by applying different classifier. Furthermore, Naive Bayes classifier has achieved the best Precision, Recall, and F-measure scores among other classifiers.

At the end, it can be concluded from above mentioned results analysis that the proposed "Scientific Services and Affiliations" based parameters have outperformed traditional "Publications" based parameters most of the time for researchers' assessment. Researchers' assessment can guide decision makers to make important decisions such as hiring faculty in universities, giving awards, assigning memberships and fellowships of societies, hiring reviewers or editors for academic journals and for researchers to focus on increasing their scientific contributions rather than only publishing articles.

5.2 Future Work

This research work can explore other scientific contributions such as: number of current projects, number of patents and number of lectures of researchers etc. Additionally, it can also evaluate proposed "Scientific Services and Affiliations" based parameters for other datasets and in different domains. Furthermore, decision makers can use these results to promote researchers on higher positions.

Bibliography

[1] R. Crowder, G. Hughes, and W. Hall, "An agent based approach to finding expertise," pp. 179–188, 2002.

[2] D. H. L. Cameron, B. Aleman-Meza, S. Decker, and I. B. Arpinar, "Semef: A taxonomy-based discovery of experts, expertise and collaboration networks," Ph.D. dissertation, University of Georgia, 2007.

[3] B. S. Frey and K. Rost, "Do rankings reflect research quality?" *Journal of Applied Economics*, vol. 13(1), pp. 1–38, 2010.

[4] T. Bogers, K. Kox, and A. van den Bosch, "Using citation analysis for finding experts in workgroups," pp. 21–28, 2008.

[5] Z. Yang, L. Hong, and B. D. Davison, "Topic-driven multi-type citation network analysis," pp. 24–31, 2010.

[6] R. West, K. Stenius, and T. Kettunen, "Use and abuse of citations," *Addiction Science: A Guide for the Perplexed*, p. 191, 2004.

[7] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National academy of Sciences*, vol. 102(46), pp. 16 569–16 572, 2005.

[8] L. Bornmann, R. Mutz, S. E. Hug, and H.-D. Daniel, "A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants," *Journal of Informetrics*, vol. 5(3), pp. 346–359, 2011.

[9] L. Egghe, "Theory and practise of the g-index," *Scientometrics*, vol. 69(1), pp. 131–152, 2006.

[10] B. Jin, L. Liang, R. Rousseau, and L. Egghe, "The r-and ar-indices: Complementing the h-index," *Chinese science bulletin*, vol. 52(6), pp. 855–863, 2007.

[11] S. Alonso, F. Cabrerizo, E. Herrera-Viedma, and F. Herrera, "hg-index: A new index to characterize the scientific output of researchers based on the h-and g-indices," *Scientometrics*, vol. 82(2), pp. 391–400, 2009.

[12] S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, and F. Herrera, "h-index: A review focused in its variants, computation and standardization for different scientific fields," *Journal of Informetrics*, vol. 3(4), pp. 273–289, 2009.

[13] G. H. van Rensburg, P. Mayers, and L. Roets, "Supervision of post-graduate students in higher education," *Trends in Nursing*, vol. 3(1), 2016.

[14] K. C. Chan, H.-G. Fung, and P. Lai, "Membership of editorial boards and rankings of schools with international business orientation," *Journal of International Business Studies*, vol. 36(4), pp. 452–469, 2005.

[15] S. Ayaz and M. T. Afzal, "Identification of conversion factor for completing-h index for the field of mathematics," *Scientometrics*, vol. 109(3), pp. 1511–1524, 2016.

[16] X. Jiang, X. Sun, and H. Zhuge, "Graph-based algorithms for ranking researchers: not all swans are white!" *Scientometrics*, vol. 96(3), pp. 743–759, 2013.

[17] M. Raheel, S. Ayaz, and M. T. Afzal, "Evaluation of h-index, its variants and extensions based on publication age & citation intensity in civil engineering," *Scientometrics*, vol. 114(3), pp. 1107–1127, 2018.

[18] J. D. Wells, "Completing hirsch's h-index measuring scholarly impact by james d. wells (university of michigan, ann arbor)," 2014.

[19] H. F. Chan and B. Torgler, "Econometric fellows and nobel laureates in economics," *Economics Bulletin*, vol. 32(4), pp. 3365–3377, 2012.

[20] P. Hall, "Measuring research performance in the mathematical sciences in australian universities," *Australian Mathematical Society Gazette*, vol. 34(1), p. 26, 2007.

[21] K.-H. Yang, C.-Y. Chen, H.-M. Lee, and J.-M. Ho, "Efs: Expert finding system based on wikipedia link pattern analysis," pp. 631–635, 2008.

[22] M. T. Afzal and H. A. Maurer, "Expertise recommender system for scientific community." *J. UCS*, vol. 17(11), pp. 1529–1549, 2011.

[23] R. Costas and M. Bordons, "The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level," *Journal of informetrics*, vol. 1(3), pp. 193–203, 2007.

[24] D. F. Thompson, E. C. Callen, and M. C. Nahata, "New indices in scholarship assessment," *American Journal of Pharmaceutical Education*, vol. 73(6), p. 111, 2009.

[25] S. Froghi, K. Ahmed, A. Finch, J. M. Fitzpatrick, M. S. Khan, and P. Dasgupta, "Indicators for research performance evaluation: an overview," *BJU international*, vol. 109(3), pp. 321–324, 2012.

[26] A. Lee and J. McKenzie, "Evaluating doctoral supervision: Tensions in eliciting students' perspectives," *Innovations in Education and Teaching International*, vol. 48(1), pp. 69–78, 2011.

[27] C. R. Chen and Y. Huang, "Author affiliation index, finance journal ranking, and the pattern of authorship," *Journal of Corporate Finance*, vol. 13(5), pp. 1008–1026, 2007.

[28] Y. Pan and C. R. Chen, "Author affiliation index: A new approach to marketing journal ranking," *Marketing Education Review*, vol. 21(3), pp. 275–188, 2011.

[29] D. S. Hamermesh and P. Schmidt, "The determinants of econometric society fellows elections," *Econometrica*, vol. 71(1), pp. 399–407, 2003.

[30] H. F. Chan, B. S. Frey, J. Gallus, and B. Torgler, "Academic honors and performance," *Labour Economics*, vol. 31, pp. 188–204, 2014.

[31] A. Mazloumian, D. Helbing, S. Lozano, R. P. Light, and K. Börner, "Global multi-level analysis of the 'scientific food web'," *Scientific reports*, vol. 3, p. 1167, 2013.

[32] K. C. Chan and R. C. Fok, "Membership on editorial boards and finance department rankings," *Journal of Financial Research*, vol. 26(3), pp. 405–420, 2003.

[33] D. Wu, J. Li, X. Lu, and J. Li, "Journal editorship index for assessing the scholarly impact of academic institutions: An empirical analysis in the field of economics," *Journal of Informetrics*, vol. 12(2), pp. 448–460, 2018.

[34] M. Cantín, M. Muñoz, and I. Roa, "The h-index in academic morphology." *International Journal of Morphology*, vol. 33(2), pp. 706–711, 2015.

[35] M. Schreiber, "An empirical investigation of the g-index for 26 physicists in comparison with the h-index, the a-index, and the r-index," *Journal of the American Society for Information Science and Technology*, vol. 59(9), pp. 1513–1522, 2008.

[36] J. E. Hirsch, "Does the h index have predictive power?" *Proceedings of the National Academy of Sciences*, vol. 104(49), pp. 19 193–19 198, 2007.

[37] S. M. Susarla, J. Lopez, E. W. Swanson, D. Miller, D. O'Brien-Coon, J. E. Zins, J. M. Serletti, M. J. Yaremchuk, P. N. Manson, and C. R. Gordon, "Are quantitative measures of academic productivity correlated with academic rank in plastic surgery? a national study," *Plastic and reconstructive surgery*, vol. 136(3), pp. 613–621, 2015.

[38] E. A. Erichsen, D. U. Bolliger, and C. Halupa, "Student satisfaction with graduate supervision in doctoral programs primarily delivered in distance education settings," *Studies in Higher education*, vol. 39(2), pp. 321–338, 2014.

[39] J. J. Janney, S. Gove, and B. P. Matherne, "The author affiliation index (aai) as an assessment of 25 entrepreneurship journals, 2007-2011," vol. 2014(1), p. 17162, 2014.

[40] X. Wang, "The relationship between sci editorial board representation and university research output in the field of computer science: A quantile regression approach," *Malaysian Journal of Library & Information Science*, vol. 23(1), pp. 67–84, 2018.

[41] J. Palsberg, "The h index for computer science," *University of California, Los Angeles, available at http://www. cs. ucla. edu/˜ palsberg/h-number. html*, 2009.

[42] H. F. Moed, J. Bar-Ilan, and G. Halevi, "A new methodology for comparing google scholar and scopus," *Journal of informetrics*, vol. 10(2), pp. 533–551, 2016.

[43] I. F. Aguillo, "Is google scholar useful for bibliometrics? a webometric analysis," *Scientometrics*, vol. 91(2), pp. 343–351, 2012.

[44] H. Singh, R. Singh, A. Malhotra, and M. Kaur, "Developing a biomedical expert finding system using medical subject headings," *Healthcare informatics research*, vol. 19(4), pp. 243–249, 2013.

[45] A. Richardson, "Nonparametric statistics for non-statisticians: A step-by-step approach by gregory w. corder, dale i. foreman," *International Statistical Review*, vol. 78(3), pp. 451–452, 2010.

[46] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, S. Y. Philip *et al.*, "Top 10 algorithms in data mining," *Knowledge and information systems*, vol. 14(1), pp. 1–37, 2008.

[47] C. S. Tejavath and R. Singh, "Compressive study on various classification techniques used in data mining," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 2(6), pp. 2456–3307, 2017.

[48] S. S. Nikam, "A comparative study of classification techniques in data mining algorithms," *Oriental Journal of Computer Science & Technology*, vol. 8(1), pp. 13–19, 2015.

[49] E. Szmidt and J. Kacprzyk, "The spearman and kendall rank correlation coefficients between intuitionistic fuzzy sets." pp. 521–528, 2011.

[50] S.-D. Bolboaca and L. Jäntschi, "Pearson versus spearman, kendall's tau correlation analysis on structure-activity relationships of biologic active compounds," *Leonardo Journal of Sciences*, vol. 5(9), pp. 179–200, 2006.