

# Finding Attractive Research Areas for Young Scientists

Nouman Malik<sup>1</sup>, Hikmat Ullah Khan<sup>2</sup>, Muhammad Ramzan<sup>3</sup>, Muhammad Shahzad Faisal<sup>4</sup>, Ahsan Mahmood<sup>5</sup>

Department of Computer Science, COMSATS University Islamabad, Attock, Pakistan<sup>1, 5, 4</sup>

Department of Computer Science, COMSATS University Islamabad, Wah Cantt, Pakistan<sup>2</sup>

School of Systems & Technology, University of Management and Technology, Lahore<sup>3</sup>

Department of Computer Science and IT, University of Sargodha, Sargodha, Pakistan<sup>3</sup>

**Abstract**—The selection of the research area is very vital for new researchers. One of the major issues for researchers is the selection of the domain of research on which he/she can carry out research. This case is very vital on the grounds that it decides the future of the researchers in that research area. Finding hot and attractive research areas is not considered in the relevant literature of Scientometrics. In this regard, the correct decision of the selection of the research domain helps the researchers to show better performance as well gain a good academic career. The main aim of this research study is to figure out the attractive research areas for the researchers, especially who are at the starting stage of their research life. To the best of our knowledge, this research area is still very limited due to limited work done in this area. So in order to distinguish the attractive research field for the new researchers, new rising fields are identified by applying the well-known g-index, which is widely used for finding the top authors in academic networks. In addition, we compute diverse, relevant features of the research fields which help us to identify top research area. The results demonstrate that the proposed methodology is capable to recommend the attractive research fields for potential future research work. An extensive empirical analysis has been carried out using the widely used academic database of DBLP.

**Keywords**—Research field; scientometrics; attractive areas; g-index

## I. INTRODUCTION

In the field of Research one of the basic thing is to choose a field of interest for research purpose that also determine the future of a scientist. In the initial stages of the research, the research finds a particular research field that is among particular other research field, but sometimes, researchers find it pretty hard to find a research area that is related to their interest and that is overall an attractive research field for the research community. Some topics are popular, but not attractive, some topics are very attractive, but doesn't guarantee more citations. While there is a race for better research topics by the research community, one of the most important issues of the emerging researchers is to choose a topic that is an attractive topic in the domain in which they are working.

While the research searches for a research topic, they want their research area to be attractive. Attractiveness is an important word and most of the research wants their area to be attractive. In this regard, researchers try to find attractive research areas. Usually papers with shorter titles are more

attractive than the paper with the longer titles [1]. Similarly, there are other ways that make a paper attractive. In this regard, researchers proposed many techniques to identify the attractiveness of the research papers [2]. In this regard, authors collaboration with the other authors also plays an important role in the attractiveness of the [3]. While author collaboration, paper titles, readability and other factors play an important role in the attractiveness of the paper our task in this paper is to identify yet another method that find attractive research areas. For this purpose, G-index is used. As G-index is an improved method and it computes the productivity of the science, it performs better and gives better results.

The ultimate objective of this paper is to provide the new researchers with some ideas of new attractive research fields. This way they will have a list of some attractive fields that relate to their interests so they can work in the field of their interest that is an attractive research area for the research community. Sometimes, if the chosen research area is not an attractive research idea it may result in a low number of citations and difficult publication by the quality journals because these days' journals also try to publish papers that are attractive for the research community. So we propose a method on the basis of G-index that finds the attractive research areas for the users. The findings will help the new scientists finding new attractive fields for their research. This in return sometimes provides more citations and easy publication in the quality journals.

The rest of the paper is divided as follows: Section II discusses the related work in the field of Scientometrics; Section III discusses the proposed Research Methodology; Section IV discusses the findings of the research paper while the paper is concluded in the next section.

## II. RELATED WORK

In this section, we discuss the important research work done in the field of Scientometrics to rank the journals, rank the important research areas and conferences.

Although in the recent research, researchers didn't highly focus on finding attractive research fields, there are some research works that are based on similar works. Lee proposed a method on the basis of co-word analysis to find out the trends in publications [4]. However, the research in this field is halted due to the non-availability of the keyword analysis approach available for this domain [5]. Therefore, having a proper

dataset available for this task is as important as the field itself because it's really hard to find relevant workable dataset for this research area.

In this regard, different focuses have been made by the researchers including co-citation approaches to find the trending research areas [6] and different sort of ranking algorithms to rank the important research areas [7]. Similarly, researchers also ranked Conferences on the basis of index and other properties [8] and also used other kind of indexes to rank the authors and journals like Ds index [9]. The researchers also proposed solutions for emerging researchers to help them find the useful topics that produce impactful research papers [2]. Researchers also used social network ranking measures to rank the authors [10].

In this field, many researchers focused on only journals and authors. However, there are some scientists who focused on the characteristics of the research its self like they ranked the journals, conferences, authors stats, publishers, etc. According to Bogdan et.al, Importance of different conferences also play an important role in the ranking of the research [11]. Similarly, the researchers also proposed the problems and discontent in the journal ranking system [12].

In this field as our focus is to extract the attractive research areas one of the main tasks is also to rank the journals and conferences. Different approaches have been used in this field. one of the recent approach is to rank journals on the basis of clustering and scaling techniques [13]. Other than that many researchers have worked to compare the difference between very known terms in the field of research. And how it impacts the actual outcome of the research and attractiveness of the topic [14]. With the help of these calculations it will become easy to predict the hot and attractive research areas and how research fields correlate well with actual scientific research trends[15].

In order to find the attractive research areas, we focus on using G-index for finding the attractive research fields among the papers. As G-index solves most of the problems in the H-index, G-index works better and usually results in better results.

### III. RESEARCH METHODOLOGY

The proposed research methodology is divided into four steps. Each step contains a number of sub-steps. In the first step, the data is extracted and processed. For this purpose, the dataset is prepared, the dataset is classified and data is preprocessed. In the second step, Taxonomy of the dataset is calculated using the IEEE and ACM taxonomy by extracting the computer science domain words and then extraction of all the papers related to the computer Science. For this purpose, 3 level hierarchical taxonomy is extracted. In the third step, attractive research areas are computed by using G-index and other features this step consist of four parts including recent article extraction, article impact extraction, journals reputation computation and author's influence computation. In the fourth step the research output is evaluated by finding the topic sensitive extensions and journals output is evaluated. The proposed framework for this research is given in Fig. 1.

#### A. Problem Statement

The problem statement consists of multiple steps, in the first steps research work is evaluated, then the journal's reputation is evaluated, then the articles are classified, then attractive research areas are identified, these steps are given in the form of a statement below:

On the basis of this analysis a subset  $A^{pi} \subseteq A$  and the researchers who have articles  $p^i$  where the research fields belonging to  $F^{pi} \subseteq F$ , and each author  $a^j$  has a number of papers  $P^{aj} \subseteq P$  while each of the research fields  $F^n$  contains a subset of papers  $P^{fn} \subseteq P$ . So the above scenario can be given in the form of a graph as shown in the Fig. 2.

So this shows that our ultimate goal is to identify research areas  $F$  that are actually attractive research areas for author's  $a^j$  for whom the metrics  $G_v^{aj}$  receive low values for this purpose, for this purpose, a special score is introduced  $S^{fn}$  for each field. Therefore, the problem consists of two sub-problems, first calculation of  $G_v^{aj}$  to evaluate the value of  $a^j$  and  $G_u^{bi}$  to evaluate the value of  $b^i$ .

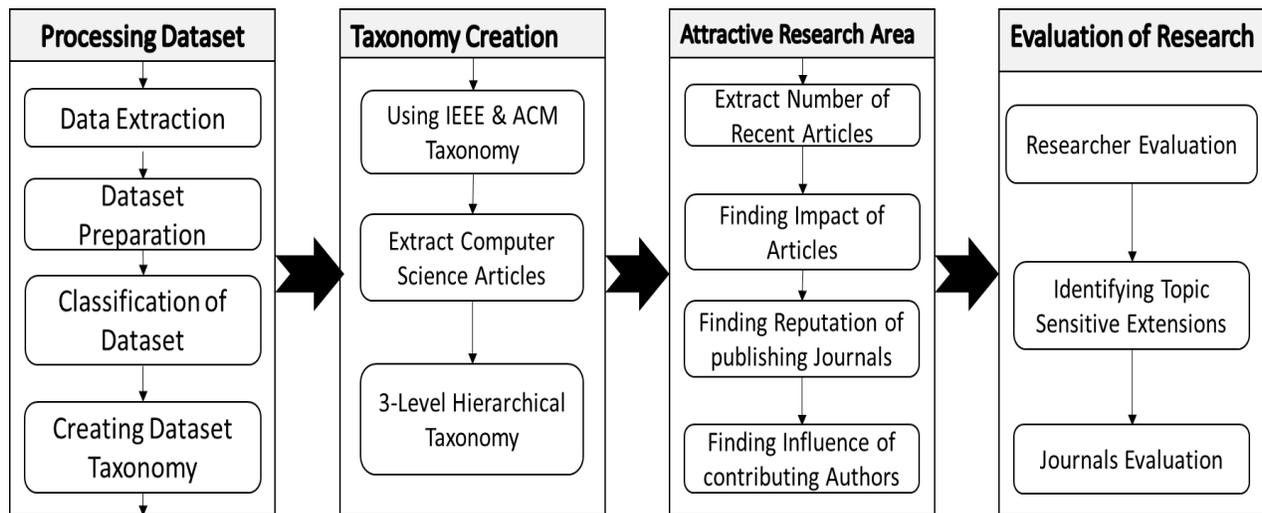


Fig. 1. Proposed Framework of Research.

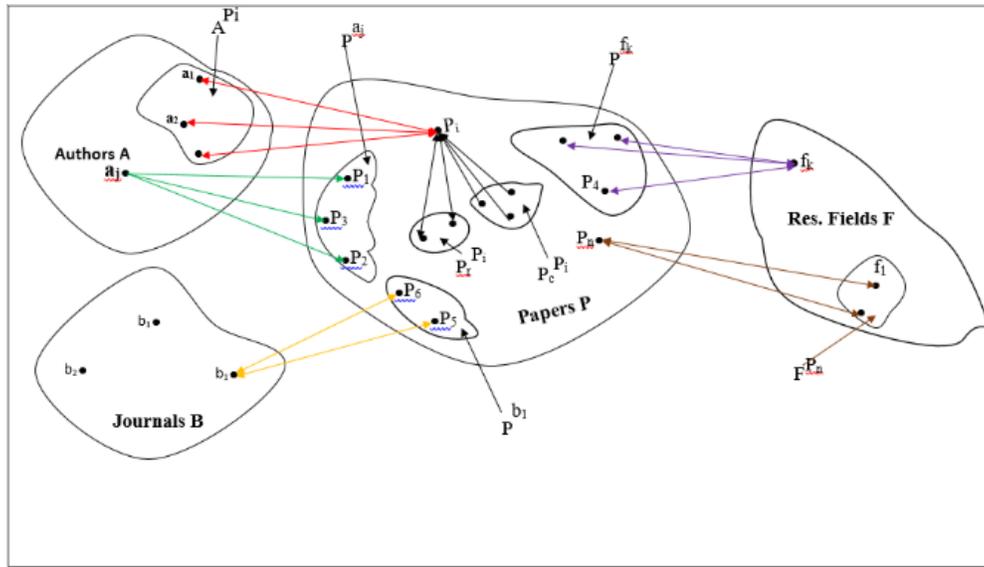


Fig. 2. Graph Representation of Connected Universe.

### B. Dataset and Taxonomy

In the first step data set is collected, extracted, preprocessed and taxonomies are computed on the basis of IEEE and ACM taxonomy.

### C. Research Evaluation

After classifying the dataset, reputed scientists are computed. For this purpose, G-index is used. In Table 1 contains the summarized information of G-index.

### D. Identify Prestigious Journals

The next important step is to find the prestigious journal. This will help in finding the particular topics published in their journals and their importance. The main reason behind finding, identifying Prestigious Journal's so that if we are able to calculate that and if the number of articles on a particular topic is published in those journals. We noted that using h-index metrics for journal is not a good way due to its limitations. Therefore, G-index is used to evaluate the importance of a journal. This also computes if the impact of the journal is increasing with time or not.

### E. Popular Research Areas

After computing the prestige journals, the popular research areas are computed using the popularity measures. For this purpose, well known journals and their articles are used. Equation 1 shows the method of computing the popular research areas. In this equation  $S^{fn}$  is special score,  $F^n$  is an arbitrary research area,  $P^{fn}$  are the papers belong to  $f^n$ ,  $P^i$  is an arbitrary paper,  $P^{pi}$  are papers referred by  $P^i$  authors who created  $P^i$ .

TABLE I. SUMMARY OF THE METRICS FOR EVALUATING THE WORK OF A SCIENTIST

V	Symbol	Meaning
1	$G_v^{aj}$	G-index for authors
2	$G_\mu^{bi}$	G-index for journals

$$S_{1,y}^{fn} = |P_Y^{Fn}| + \sum_{i=1}^{|P^{Fn}|} |P_{C,Y}^{Pi}| + \sum_{i=1}^{|P^{Fn}|} |A^{Pi}| \quad (1)$$

### F. Attractive Research Areas for New Scientists

Finally, the attractive research areas are calculated from the data. In this regard one important thing to note is that not all attractive fields are popular and not all the popular fields are attractive. In other words, popularity and attractiveness are two different things.

The attractiveness of a research is computed by using the equation as shown in Equation 2. According to the equation 2,  $S^{fn}$  is special score,  $F^n$  is an arbitrary research area,  $P^{fn}$  are the papers belong to  $f^n$ ,  $P^i$  is an arbitrary paper,  $P^{pi}$  is the set of the papers referred by  $P^i$  authors who created  $P^i$ ,  $B^i$  is an arbitrary Journal,  $h^{bi}$  is the metrics evaluating the prestige of  $b^i$ ,  $Y^i$  is the year of publication,  $A^j$  is an arbitrary Author,  $D_s^{aj}$  is a metric evaluating work of an author then the attractiveness of a research area can be given as shown in the Equation 2.

$$S_{2,v,\mu}^{fn} = \sum_{i=1}^{|P^{fn}|} \frac{|P_c^{Pi}| G_\mu^{bi}}{(\Delta Y_i)^\delta} \left( \sum_{j=1}^{|A^{Pi}|} \frac{\lambda}{G_v^{aj}} \right) \quad (2)$$

From Equation 1 and Equation 2 a scoring formula to evaluate the criteria can be given as shown in the Equation 3.

$$S_{3,v,\mu}^{fn} = \sum_{i=1}^{|P^{fn}|} \frac{G_\mu^{bi}}{(\Delta Y_i)^\delta} \left( \sum_{x=1}^{|P_c^{Pi}|} \frac{1}{(\Delta Y_x)^\delta} \sum_{j=1}^{|A^{Pi}|} \frac{\lambda}{G_v^{aj}} \right) \quad (3)$$

## IV. RESULTS AND DISCUSSIONS

This section discusses the results and findings of the research. According to our findings we computed the top journals, authors, etc. We discuss the details of each field.

To the best of our knowledge, there is no publically available dataset for the taxonomy. Therefore, in order to manage the records, after extraction and preprocessing of the dataset it was required to manage the dataset in terms of domains and sub-domains. Therefore, we used IEEE taxonomy for this purpose. This helps us identifying the papers according to their titles and relevant matches in the IEEE taxonomy. We

have expelled repeated articles and those which were surrendered by the ideal meta-data, e.g. authors, journal or date of publication). At the end, 429,398 Articles were filtered out to form one data set. For the classification of data set, we will use the taxonomy of IEEE and will segregate articles in 1168 research fields.

Table II shows Top 30 authors using G-index. According to the results, Xiaoou Tang is the top author according to the G-index while the results clearly shows that g-index is a more productive way to find the top authors because of its structure of computing the productivity. Similarly, Jian Sun, Athanasios V. Vasilakos and other are among the top authors according to the G-index results.

Similarly, Table III shows the Top 15 journals according to the G-index. The results show that “Computer Vision and Pattern Recognition” is the top journal according to its G-index while some more popular journals are not among the top journals. The results show that in some cases, more citations doesn’t guarantee the quality of a journal.

TABLE II. TOP 30 AUTHORS USING G INDEX

Authors Name	g-index	Citations
Xiaoou Tang	19	321
Jian Sun	18	301
Athanasios V. Vasilakos	16	202
Pushmeet Kohli	16	137
Rob Fergus	15	116
Yoshua Bengio	15	163
Yi Ma	15	149
Jeffrey G. Andrews	15	137
Jitendra Malik	15	136
Andrea Vedaldi	15	118
Xuelong Li	14	178
Fabrizio Benevenuto	14	119
Florent Perronnin	14	121
Patrick Pérez	14	135
Shuicheng Yan	13	105
Axel Legay	13	103
Xiaogang Wang	13	101
Ajith Abraham	13	114
Hervé Jégou	13	107
Thomas Pock	13	97
Ruslan Salakhutdinov	13	110
Shahram Izadi	13	108
Jiawei Han	12	111
Guiwu Wei	12	99
Guanrong Chen	12	103
Min Chen	12	106
Constantine Caramanis	12	102
Nanning Zheng	12	119
David Parker	12	100
Carsten Rother	12	106

TABLE III. TOP 15 JOURNALS USING G-INDEX

Venue	Citations	G index
computer vision and pattern recognition	3862	40
European conference on computer vision	3237	38
Expert Systems With Applications	2136	36
Information Sciences	1536	39
IEEE Transactions on Information Theory	1350	35
Neuroimaging	1324	33
international conference on machine learning	1159	35
soft computing	1125	34
national conference on artificial intelligence	1080	33
Pattern Recognition	1063	31
Neurocomputing	1056	28
IEEE Transactions on Signal Processing	1048	27
IEEE Transactions on Image Processing	1037	25
human factors in computing systems	1024	22
international conference on robotics and automation	1018	21

After finding the top authors and journals, the most attractive fields are calculated according to the  $S_{2,u,v}^{fn}$ . According to the results, Information systems are the most attractive research area. When we take a look at the current research trends we find a number of the top papers and a huge number of authors work on these top attractive fields. Similarly, topics like Cloud computing, Computer vision, clustering, etc. are also among the top attractive fields according to the computed results. The results are given in Table IV.

TABLE IV. RANKING OF THE MOST ATTRACTIVE FIELDS OF RESEARCH ACCORDING TO THE S SCORE

FIELDS	$S_{2,u,v}^{fn}$
Information systems	3876.67
Design	3757.71
Clustering	3502.77
Metrics	2994.54
Cloud computing	2898.46
Architectures	2495.53
Reliability	1695.12
Modes of Computation	1529.73
Computer vision	755.45
Ontologies	726.69
Process management	444.84
Use cases	407.24
Principal component analysis	349.26
E-learning	314.84
Optimization algorithms	288.79
Visual analytics	286.93
Virtual reality	238.05
Object detection	236.50
Network structure	182.36
Semi-supervised learning	140.32

TABLE V. RANKING OF THE MOST ATTRACTIVE FIELDS OF RESEARCH ACCORDING TO THE  $S^{fn}$  SCORE

FIELDS	$S_{3,u,v}^{fn}$
Design	317140175.4
Information systems	221320024.1
Clustering	149408805.1
Scheduling	147660053.1
Machine learning	136784635
Cloud computing	117475456.7
Metrics	76569737.89
Reliability	70164301.07
Cryptography	62359568.67
Smartphones	40330966.19
Virtual reality	38987674.02
Computer vision	36799319.45
E-learning	33101525.04
Computational geometry	31443517.04
Architectures	26015802.42
Ontologies	21223968.58
Network security	20525337.71
Multimedia information systems	13862564.98
Reconfigurable computing	12633623.18
Real-time systems	10901137.37

Table V shows results according to  $S_{3,u,v}^{fn}$ . According to these results, Design is the top research field. Similarly, the fields like Information system, clustering, scheduling is also among the top fields computed by the system. All these fields are attractive research fields according to the modern research areas.

## V. CONCLUSION

In this paper, we examined the issue of identifying attractive research areas for new scientists. Since in the modern scientific age, with the presence of a huge number of research topics, this is an issue, we first identified each of the fields one by one. We recognize the top authors by G-index and then find the top journals according to G-index. Similarly, by using these results, the Top most popular research areas are computed. Similarly, the attractive research areas are computed by further processing of the previous results. For this purpose, we introduced two scoring patterns and combined multiple different factors into these patterns. In this work, the strategy for evaluating the work of the scientists, we presented the usage of G-index. This index is performing better than the H-index that only check the number of citations and number of

papers. This G-index computes the productivity of the papers and journals by using a much better procedure.

Our strategies have been confirmed tentatively by utilizing a vast set of self-crawled research articles. The examinations gave some significant findings: The first is that there are exist some research areas which in spite of their popularity, they are not attractive for researchers who are presently beginning of their research career. Then again, some research fields are unpopular be that as it may; they give fantastic open doors at these researchers.

## REFERENCES

- [1] F. Habibzadeh and M. Yadollahie, "Are Shorter Article Titles More Attractive for Citations? Cross-sectional Study of 22 Scientific Journals," *Croat. Med. J.*, vol. 51, no. 2, pp. 165–170, Apr. 2010.
- [2] L. Akritidis, D. Katsaros, and P. Bozanis, "Identifying attractive research fields for new scientists," *Scientometrics*, vol. 91, no. 3, pp. 869–894, Jun. 2012.
- [3] H. Hou, H. Kretschmer, and Z. Liu, "The structure of scientific collaboration networks in Scientometrics," *Scientometrics*, vol. 75, no. 2, pp. 189–202, Dec. 2007.
- [4] W. H. Lee, "How to identify emerging research fields using scientometrics: An example in the field of Information Security," *Scientometrics*, vol. 76, no. 3, pp. 503–525, Sep. 2008.
- [5] R. L. Ohniwa, A. Hibino, and K. Takeyasu, "Trends in research foci in life science fields over the last 30 years monitored by emerging topics," *Scientometrics*, vol. 85, no. 1, pp. 111–127, Oct. 2010.
- [6] S. P. Upham and H. Small, "Emerging research fronts in science and technology: patterns of new knowledge development," *Scientometrics*, vol. 83, no. 1, pp. 15–38, Apr. 2010.
- [7] A. Sidiropoulos and Y. Manolopoulos, "Generalized comparison of graph-based ranking algorithms for publications and authors," *J. Syst. Softw.*, vol. 79, no. 12, pp. 1679–1700, Dec. 2006.
- [8] M. Farooq, H. U. Khan, T. Iqbal, and S. Iqbal, "An index-based ranking of conferences in a distinctive manner," *Electron. Libr.*, Feb. 2019.
- [9] M. Farooq, H. U. Khan, S. Iqbal, E. U. Munir, and A. Shahzad, "DS-Index: Ranking Authors Distinctively in an Academic Network," *IEEE Access*, vol. 5, pp. 19588–19596, 2017.
- [10] F. Bibi, H. U. Khan, T. Iqbal, M. Farooq, I. Mehmood, and Y. Nam, "Ranking Authors in an Academic Network Using Social Network Measures," *Appl. Sci.*, vol. 8, no. 10, p. 1824, Oct. 2018.
- [11] B. Vasilescu, A. Serebrenik, T. Mens, M. G. J. van den Brand, and E. Pek, "How healthy are software engineering conferences?," *Sci. Comput. Program.*, vol. 89, Part C, pp. 251–272, Sep. 2014.
- [12] J. R. F. Arruda et al., "The Journal Impact Factor and its discontents: steps toward responsible metrics and better research assessment," *Open Scholarsh. Initiat. Proc.*, vol. 1, no. 0, Apr. 2016.
- [13] C. J. A. Bradshaw and B. W. Brook, "How to Rank Journals," *PLOS ONE*, vol. 11, no. 3, p. e0149852, Mar. 2016.
- [14] H. Kianifar, R. Sadeghi, and L. Zarifmahnoudi, "Comparison between impact factor, Eigenfactor metrics, and SCImago journal rank indicator of pediatric neurology journals," *Acta Informática Médica*, vol. 22, no. 2, p. 103, 2014.
- [15] N. Yamashita, M. Numao, and R. Ichise, "Predicting Research Trends Identified by Research Histories via Breakthrough Researches," *IEICE Trans. Inf. Syst.*, vol. E98.D, no. 2, pp. 355–362, 2015.