



Bibliometric Analysis of Computer Science Literature of Pakistan

Adnan Faiz

Department of Computer Science and Information Technology, UET Peshawar;

Contact: +92332 9074081 email: dani20puk@gmail.com

Abstract:

Bibliometric analysis is the study of analysing the published works and to find trends in order to identify any weaknesses or different areas that are left unattended, and unexplored. It is also used to recognize and appreciate the importance of seminal works in various fields of sciences and humanities, and to identify future research directions in a field of study. In this work, we analyse the computer science literature published in Pakistan in the 21st century (2000 – 2017). We identify the top cited publications in the considered time span, and rank the publications based on cumulative citation count as well as normalized citation count. We analyse the distribution of citation count and compare it with other works. In addition, the top institutions in terms of publication counts are identified. Finally, we compared the computer science research output of Pakistan to that of Bangladesh, India, Malaysia, Saudi Arabia and Turkey. Our work is beneficial both to the research community as well as to the higher education authorities to assess the work of the computer science researchers, and its impact.

1 Introduction:

The term bibliometric analysis is commonly known as the mathematical and statistical techniques being used to know the quantity and quality of previously available scientific writings and to work on research drifts, journals and citation investigations, impact of publications, authorship and national and international impact in a particular field [1]. In other words, bibliometric analysis can provide an overall examination and quantitative viewpoint of a particular research topic supported by large amount of literature information. In particular, it can put in to reviewing the evolution and development trends of a scientific discipline, to identify hotspots and emerging ideas of a field, and to evaluate the performance and influence of different countries or journals. As a result, a better understanding of status quo can be achieved and future research directions can be identified from outcomes of bibliometric analysis [2].

The paper count and citation count are the main tools for assessment of the prominent work in a research field [3]. Paper count is mostly used for the identification of an institute or individual performance over time and citation count is its importance in a particular field as it is cited over time. There are different counting methods that could be employed to see the difference especially in the case of a university or a country performance such as in whole counting, straight counting or the fractional

counting [4]. The first method involves that each collaborating entity (institute or country) receives one full credit. In the second method, one full credit goes to the most prominent collaborator, and the third method shares the credit among all the collaborators. Other tools in bibliometric analysis include the co-citation analysis, bibliographic coupling and the direct citations [5].

The bibliometric analysis could be a very helpful tool for various countries in helping making decision for both the researchers and the institutions. Such as the studies conducted in Turkey [11, 23] for the Software Engineering (SE) Literature and in the Canada [6, 7] for the Electrical and Computer Engineering Institutes. These studies are helpful to the new researchers in a way that they get to choose their preferred area of interest and the institute which is working in that area. The institutes get to find the most proficient researchers and to give those researchers promotions or funding for their future work. Further, it would also assess the performance of the institute and the researcher on the national level and external funding could also be secured by the institute for a greater efficiency and to compete at the international level.

The objective of this work is to evaluate the computer science literature of the Pakistani institutes and researchers. For the purpose we used the bibliometric analysis to find out the most influential work, citations and authorship trends, most contributing institutions and their publications trend over time. We also compare the works of Pakistani authors with countries such as India, Turkey, Malaysia, Bangladesh and Saudi Arabia to see where Pakistani authors actually stand in the international grounds or at least in the Asian zone. To the best of our knowledge, there has been no study conducted yet in the field of computer science in Pakistan so it is with hope that it would benefit the institutes and the researchers. The rest of the article is organized as follows: Section 2 presents a review of the previous work in the field. Section 3 presents the research questions and the research methodology used for the extraction of data. Section 4 presents an analysis of the data. Section 5 presents the limitations of the study and Section 6 concludes the work.

2 Related Work

In this section we take a look at the previous work done for bibliometric analysis in different fields of science. The purpose is to take a general look at the different approaches adopted by different researchers to present their work in an efficient way possible.

Wohlin [8] in his work found the top articles in the domain of software engineering for the year 2000 taken from the ISI Web of Science. He extracted the top 20 publications based on their citations count. As there was no tool available to handle the Self-citations, hence they were removed manually. But it showed little change in the ranking. He then showed a list of top 20 articles from the last 20 years (1986-2005). Self-citations were not omitted as it was not possible to perform this task manually for a huge set of data. He showed his result in two lists. The first list showed the result based on citations count. As older publications accrue more citations, the authors calculated normalized citation count (number of citations/years since publication) to normalize the effect of older articles and showed the result in the second list. The top rated articles in both of the lists was about the SPIN model checker by Holzmann [9].

Ding et al. [10] performed the syntactic and semantic analysis of the citations in publications related to computer science. The syntactic part involves the identification of the location where the citations are found (i.e. in which section of the article). The semantic analysis finds the motivation of the citations through manual approach of predefined categorizations or semiautomatic approach of NLP. In predefined categorization the citations are divided into categories defined by words or phrases in a decision tree.

Garousi and Fernandes [11] computed and classified the top 100 papers in the field of software engineering. In their study they found the top papers based on the citation count and the average annual number of citations. They devised a GQM (Goal, Question, and Metric) methodology to find their goal by devising some questions in order to achieve it. They have also compared the top papers with the top papers of all areas of science. They also identified the fields to which the top cited papers are related. Further they also identified the venues in which the top papers are presented.

Fiala and Tutoky [12] in their study presented a bibliometric analysis of some 1.9 million articles related to computer science from 1945 to 2014 obtained from the ISI Web of Science. They investigated the number of papers & citations according to the document type. It also includes publications and citations of different subfields in computer science and their authors per page factor. They also identified the different countries publishing the articles along with the identification of different institutions in these countries which contributed the most in publishing articles. Different publication sources and the different timings of the year in when the conferences were held along with their destinations are also identified. Further, top citation references and the age of those citation references were also measured. An interesting finding is that the top most computing subfield is “Artificial Intelligence”. “Neural Networks” [13] seems to be losing its charm whereas “Cloud Computing” [14] is trending. Mostly the conferences were held from May to October at destinations such as Beijing, Orlando, Shanghai and San Diego.

The bibliometric work was not to be found in Pakistan in the field of computer science but in other fields of sciences such as the library sciences, biotechnology, pharmacy and medical sciences etc. Some of which is discussed in this section.

Ali and Richardson [15] in their study performed the bibliometric analysis of the research published by the LIS (Library and Information Sciences) scholars in Pakistan through a survey. A questionnaire was designed by the authors and distributed to the LIS scholars through email, yahoo groups and facebook to the representatives from all the provinces of Pakistan including Sindh, Punjab, Khyber Pakhtunkhwa, Baluchistan, Gilgit Baltistan, Azad Jammu and Kashmir & Federal Capital. The intention was to have a balanced approach towards the problem finding or analysis. The results obtained from the total of 104 respondents were then analysed in the SPSS version 21. The result shows that male respondents frequency was more (i.e. 74.03%) than their female counterpart (i.e. 25.97%). Among 104 respondents 74 were from the public sector institutions whereas 30 were from the private sector. Punjab has a more response rate (i.e. 26), closely followed by the Federal Capital and Sindh having 25 respondents each. So the three areas respondents accounted for about 73% of the total respondents. The distribution was not equal among the provinces despite an attempt to do so. Then the respondents both in the public and private sector were asked to identify when their first publication was published? The majority (i.e. more than 44%) identified their first publication in the session 2011-2015. An interesting part is also that 16.35% responded that they haven't yet published their work which could be because they recently passed their exams and waiting for their publication to get published or they are waiting for their publication to get published in any journal. Among the total of 354 publications the highest number i.e. 141 publications are single authored closely followed by the two-authored publications i.e. 138. Which shows that over the time the co-authorship trend is promoting contrary to the trend before that in which the LIS publications were mostly published with a single author. Punjab shows the highest number of publications i.e. 158, followed by Sindh and the Federal Capital. Also the authors were asked about their citations count for their publications so out of 357 publications 107 received citations which account for a mere 30% whereas a large number i.e. 247 which accounts for 70% didn't get any citations at all. Which

is an alarming factor as mostly the quality of the journal is assessed through its citations. Further a correlation was also found among the different job titles but it didn't showed any significant results.

Bajwa and Yaldram [16] performed a bibliometric study of the biotechnology research being conducted in Pakistan from 1980 to 2011. The dataset is obtained from Scopus. In their study they found the growth of the publications in the field of biotechnology in Pakistan by using Relative Growth Rate (RGR) and Doubling Time (D_t) as the identifying factors. A total of 18085 publications in 155 journals were found from Scopus in the field of biotechnology for a period of 30 years as mentioned above. Amongst them, 689 were present as conference proceedings and 129 were in the form of book series. About 25% of the total publications appeared in the local Journals of Pakistan with the "Pakistan Journal of Botany" being the preferred one with the highest number of publications i.e. 1710 having an impact factor (2011) of 0.836. The highest impact factor Journal from the list of top 20 appeared at number 12, titled "Journal of Hazardous Materials", and has an impact factor of 4.173 and 106 publications. It is worth noting that majority of the publications i.e. 17282 appeared in the journals with the impact factors in the range of 0-2 and 631 appeared in the range of 2-4 and 163 in the impact factor of range 4-6. Only 9 publications appeared in the range of 8-10 impact factor which is a negligible amount. The RGR is calculated as $RGR = (\ln N_2 - \ln N_1) / (t_2 - t_1)$. Where N_1 and N_2 are the cumulative publications in the two years t_1 and t_2 . Whereas D_t is the time required to double the publications in number for a particular value of RGR. D_t can be calculated as: $D_t = (t_2 - t_1) \ln 2 / (\ln N_2 - \ln N_1)$ or $D_t = \ln 2 / RGR$. A constant value of RGR in each year shows an exponential growth rate, whereas D_t is the characteristic time for this growth. The results with these values along with the citations for the publications and the citations per paper, shows that for the first 15 years i.e. from 1980-1995, there is no significant change during the period but from the next year onwards i.e. 1996, there is a sudden increase which is almost a 5.5 fold increase in the publications i.e. from 52 to 281. The reason behind this was not that the publications have increased but because of the inclusion of three Pakistani journals into the Scopus database. The ACPP (Average citations per paper) also shows an increasing trend from 1996 onwards till 2011 i.e. from 0.9 to 6.5. But to get a clearer picture other factors which could also be responsible to find the quality and quantity of the publications were calculated which are: h-index, g-index, hg-index and p-index. The results show that UoK (University of Karachi) is on top of the list with the most publications i.e. 2698, followed by the QAU (Quaid-i-Azam University) at the second place and AKU (Agha Khan University) at the third place according to the number of publications. The UAF (University of Agriculture, Faisalabad) and PU (Punjab University) came at third and fourth places respectively. The NIBGE (National Institute for Biotechnology and Genetic Engineering under the administration of Atomic Energy Commission of Pakistan) despite being at number 13 in terms of publications on the list shows the highest rating of ACPP i.e. 8.07. However, if we look at other indices i.e. h, g, hg & p, the AKU which is at the third place in terms of publications shows more impressive readings. Overall the growth rate for the last 15 years i.e. from 1997 to 2011 shows an impressive 22% increase.

3 Research questions and data set

3.1 Research questions

In the following, we present the research questions that form the basis of our research.

RQ1. Identify the top 10 most cited publications.

RQ2. Identify the top 2 publications for each year.

- RQ3. Identify how many publications have recorded 0, 1, 2 ...10 and 10+ citations.
- RQ4. Identify average citations per publication.
- RQ5. Identify the number of publications where no of authors are 1... 5 and 5+.
- RQ6. Identify average number of authors per publications.
- RQ7. Identify the top 10 contributing institutions.
- RQ8. Identify the trend over time of the contributing institutes.
- RQ9. How does Pakistan fare over the years in comparison to India, Turkey, Malaysia, Bangladesh and Saudi Arabia?

3.2 Data set extraction

To answer the research questions, we required a related data set. There are various online databases available that maintains the citation records such as Scopus, ISI Web of Knowledge, and Google Scholar. The data source that we have used for this research is from Scopus [17] online database which is quite comprehensive in term of research work records. Further, the advanced search feature can filter the large database for the specifics in need. The search query that we have used to extract the dataset from the Scopus database is given below:

Query
(SUBJAREA(COMP) AND (PUBYEAR > 1999) AND (PUBYEAR < 2018) AND AFFILCOUNTRY("Pakistan"))

Table 1 shows the query for the data extraction from Scopus

As is clear that we have searched the database for the research work published from 2000 to 2017 (both inclusive) which is a time frame in which the higher education policy of Pakistan is structured and we wanted to evaluate its impact in the research sector.

The query mentioned above returned a total of 15,025 publications including Conference Proceedings, Journals, Book series, Books and Trade Publications with majority publications in English (i.e., 14,960) with a few exceptions in other languages. The results were extracted and stored in the form of CSV format file and then further analysed for finding the research questions mentioned above. The data was downloaded on 10th April 2018.

4 Results

In this section, we will present the findings to the posed questions designed in section 3.1.

4.1 Identification of top 10research publications

Table 2 shows the publications based on the number of citations. Among the ten publications only two are the conference papers at position first and third respectively. The rest are all journal articles (shortly referred to as article in the table). The conference paper is about spatio-temporal maximum average correlation height filter for action recognition [18] came at the top with 651 citations. One important thing to mention is that the articles at rank 2, 6 and 8 are purely mathematical articles but since they are published in a journal related to computer sciences, therefore they made it to the list.

Rank	Title	Year	Source title	Citation Count	Document Type
1	Action MACH - A spatio-temporal maximum average correlation height filter for action recognition	2008	26 th IEEE Conference on Computer Vision and Pattern Recognition, CVPR	651	Conference Paper
2	On some new operations in soft set theory	2009	Computers and Mathematics with Applications	445	Article
3	Apex: Extending Android permission model and enforcement with user-defined runtime constraints	2010	Proceedings of the 5 th International Symposium on Information, Computer and Communications Security, ASIACCS 2010	312	Conference Paper
4	Consistent labeling of tracked objects in multiple cameras with overlapping fields of view	2003	IEEE Transactions on Pattern Analysis and Machine Intelligence	286	Article
5	Clock synchronization of wireless sensor networks	2011	IEEE Signal Processing Magazine	286	Article
6	Soft sets combined with fuzzy sets and rough sets: A tentative approach	2010	Soft Computing	266	Article
7	A deterministic particle swarm optimization maximum power point tracker for photovoltaic system under partial shading condition	2013	IEEE Transactions on Industrial Electronics	220	Article
8	On soft topological spaces	2011	Computers and Mathematics with Applications	200	Article
9	On the hybrid optimal control problem: Theory and algorithms	2007	IEEE Transactions on Automatic Control	197	Article
10	Swarm intelligence based routing protocol for wireless sensor networks: Survey and future directions	2011	Information Sciences	189	Article

Table 2 shows the top ten publications in the field of computer science journals based on their citation count.

To ensure the fairness in ranking of the publications (i.e. the older publications tend to have more time to get cited compared to the newer ones), we calculated an average value of the publications to get a normalized view of the publications not entirely by the published year (i.e. old versus recent publications). This can be summarized as:

$$\text{Normalized citations} = \text{Total Citations} / (\text{2017} - \text{Publication Year} + 1)$$

The results are shown in Table 3.

Rank	Title	Year	Source title	Citation Count	Source Title	Normalized Citations	Change in Rank w.r.t Table 1
1	Action MACH: A spatio-temporal maximum average correlation height filter for action recognition	2008	26th IEEE Conference on Computer Vision and Pattern Recognition , CVPR	651	Conference Paper	65.10	--
2	Security in cloud computing: Opportunities and challenges	2015	Information Sciences	183	Article	61.00	+9
3	Numerical simulation for melting heat transfer and radiation effects in stagnation point flow of carbon–water nanofluid	2017	Computer Methods in Applied Mechanics and Engineering	61	Article	61.00	+88
4	On some new operations in soft set theory	2009	Computers and Mathematics with Applications	445	Article	49.44	-2
5	Analysis of flow and heat transfer in water based nanofluid due to magnetic field in a porous enclosure with constant heat flux using CVFEM	2017	Computer Methods in Applied Mechanics and Engineering	47	Article	47.00	+148
6	A deterministic particle swarm optimization maximum power point tracker for photovoltaic system under partial shading condition	2013	IEEE Transactions on Industrial Electronics	220	Article	44.00	+1

7	Numerical simulation for magneto Carreau nanofluid model with thermal radiation: A revised model	2017	Computer Methods in Applied Mechanics and Engineering	44	Article	44.00	+168
8	On magnetohydrodynamic flow of nanofluid due to a rotating disk with slip effect: A numerical study	2017	Computer Methods in Applied Mechanics and Engineering	44	Article	44.00	+168
9	A Linear Assignment Method for Multiple Criteria Decision Analysis with Hesitant Fuzzy Sets Based on Fuzzy Measure	2017	International Journal of Fuzzy Systems	43	Article	43.00	+175
10	Clock synchronization of wireless sensor networks	2011	IEEE Signal Processing Magazine	286	Article	40.86	-5

Table 3 shows Publications with Normalized Citations

As is evident from the result that the first publication retained its position and the articles at position 2,4,6 and 10[19,20,21,22] have seen little change in their positions but the rest are all replaced by recent articles from 2017 which shows a trend that the recent publications are now cited the most as compared to older articles.

4.2 Most cited articles for each year

The results for the top two publications are shown here in order to find out the trend in the citations that whether the publications over the years shows a steady increase or if the publications who made it to the list of top publications in Table 1 retained their positions over the years. The results are shown in Table 4. For the sake of simplicity, we have shown only two publications for each year.

Year	Title	Citation Count
2000	Internet use in university libraries of Pakistan	23
	Prediction-based Iterative Learning Control (PILC) for uncertain dynamic nonlinear systems using system identification technique	10
2001	Camera calibration and three-dimensional world reconstruction of stereo-vision using neural networks	44
	Visualization of shaped data by a rational cubic spline interpolation	38
2002	MHD flows of an Oldroyd-B fluid	42
	Sign language recognition using sensor gloves	40

2003	Consistent labeling of tracked objects in multiple cameras with overlapping fields of view	286
	Software clustering techniques and the use of combined algorithm	42
2004	Deformation and fracture behavior of high manganese austenitic steel	86
	Libra: A computational economy-based job scheduling system for clusters	84
2005	The relationship between system usage and user satisfaction: A meta-analysis	78
	Detecting single-feature polymorphisms using oligonucleotide arrays and robustified projection pursuit	64
2006	Medium access control issues in sensor networks	72
	Effects of excessive Internet use on undergraduate students in Pakistan	59
2007	On the hybrid optimal control problem: Theory and algorithms	197
	Hierarchical clustering for software architecture recovery	172
2008	Action MACH: A spatio-temporal maximum average correlation height filter for action recognition	651
	Strategic advantages of interoperability for global manufacturing using CNC technology	117
2009	On some new operations in soft set theory	445
	Nonrigid structure from motion in trajectory space	105
2010	Apex: Extending Android permission model and enforcement with user-defined runtime constraints	312
	Soft sets combined with fuzzy sets and rough sets: A tentative approach	266
2011	Clock synchronization of wireless sensor networks	286
	On soft topological spaces	200
2012	Future internet: The internet of things architecture, possible applications and key challenges	168
	Series solutions of non-Newtonian nanofluids with Reynolds' model and Vogel's model by means of the homotopy analysis method	137
2013	A deterministic particle swarm optimization maximum power point tracker for photovoltaic system under partial shading condition	220
	A survey on mobile data offloading: Technical and business perspectives	175
2014	A review of wireless sensors and networks' applications in agriculture	150
	A study of natural convection heat transfer in a nanofluid filled enclosure with elliptic inner cylinder	111
2015	Security in cloud computing: Opportunities and challenges	183
	Study of Natural Convection MHD Nanofluid by Means of Single and Multi-Walled Carbon Nanotubes Suspended in a Salt-Water Solution	114
2016	Applications of wireless sensor networks for urban areas: A survey	80
	Hesitant fuzzy linguistic arithmetic aggregation operators in multiple attribute decision making	70
2017	Numerical simulation for melting heat transfer and radiation effects in stagnation point flow of carbon-water nanofluid	61
	Analysis of flow and heat transfer in water based nanofluid due to magnetic field in a porous enclosure with constant heat flux using CVFEM	47

Table 4 shows Top 2 Publications for Each Year (2000-2017)

Here we can see an increase in number from the beginning reaching to max in the year 2010 and then a decline is recorded for all the way till 2017 from 312 in the year 2010 to 61 in the year 2017. Further, there is quite a difference among the top two publications in each year.

© GSJ

4.3. Distribution of Citation Count

In this section, the number of publications were recorded against their citation counts based on numbers from 0 to 10 and 10+ (for the rest of the count). The results are shown in a graph in Figure 1. As the graph shows the highest number of publication count i.e., 5,720 (~38% of the total pool of 15,025 publications) did not get any citations. This value is slightly less than the values compared to the ones indicated by Garousi and Fernandes [11] and in a similar work by Garousi and Matyla [23] who performed the bibliometric analysis on the Turkish Software Engineering community and shows that almost 43% of the pool of 71,668 publications did not get any citations. The second largest number of publications i.e., 2,588 have only one citation. Next the values for the count is decreasing with the number of citations on rise except in 10+ the number of counts is 1551 having multiple citations greater than 10.

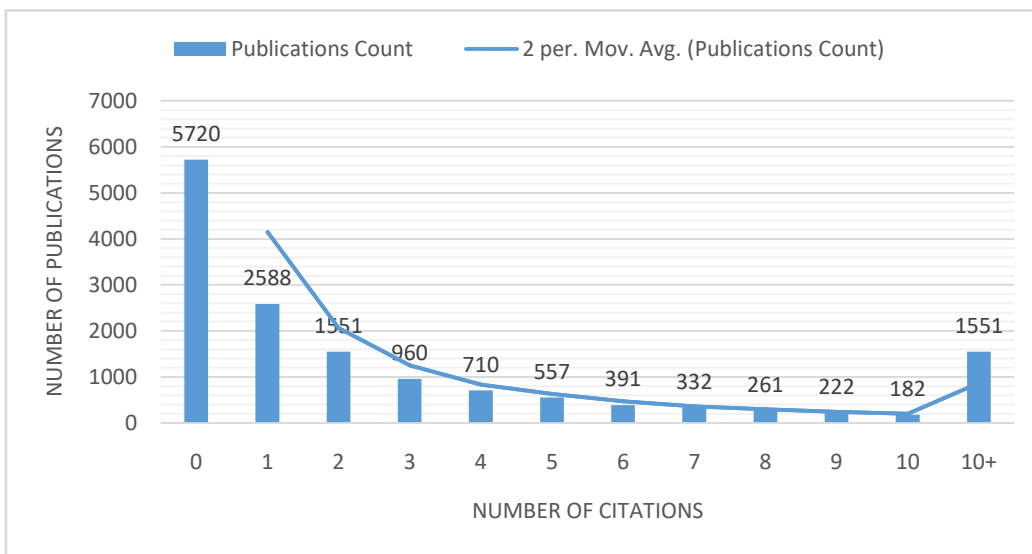


Figure 1 shows the distribution of citations

However, this is an accumulated value indicating the number of all publications which has received more than 10+ (10, 11, 12 .. etc) citations.

4.4. Identifying the average citations per publication

The total number of the publications were counted to be 15,025. The total number of citation count is 67,058. So to identify the average number of citations per publication we have to divide the two quantities as given below:

$$\begin{aligned}
 \text{Average number of citations per publication} &= \frac{\text{Number of citations}}{\text{Number of Publications}} \\
 &= \frac{67,058}{15,025} \\
 &= 4.48
 \end{aligned}$$

This means that on average each publication has received 4.48 citations.

4.5 Identifying the number of authors per publication

In this section, we find out the number of authors which shows the trend for the authors that whether in the field of computer science the authors are mostly interested in working by themselves or in collaborations as the case with the most other fields. The results are shown in a graph in Figure 2.

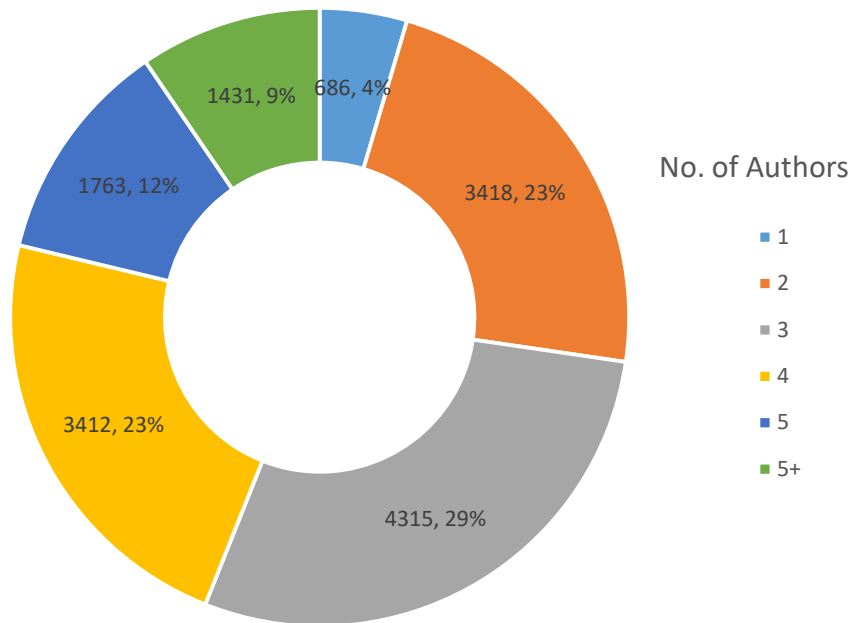


Figure 2 shows the distribution of author's trend

The number of publications for single author is the lowest count of 686 (only 4%) amongst the pool, whereas for the two authored publications, the value is 3,418 accounting for 23% of the total pool of papers, and similarly for three authored publications, the value continues to rise to 4,315 (about 29%). After which a decline in publications is noted with the increase in number of authors. This clearly shows that multiple authored publications with two or three authors (combined which accounts for more than 50% of the pool of the papers as shown in Fig 2) is the preferred number for the researchers to undergo the research work with ease and comfort versus the single and more than 3 authors case. The results are shown with the help of a chart in Figure 2.

4.6 Identification of the average number of authors per publication

For the complete duration (2000-2017) the average number of authors per publication is found to be 3.57. We also found that 43.35% of articles have more than the average number of authors, i.e., 6,514 publications have at least 4 authors.

4.7 Identification of the most contributing institutions

In this section, the top ten institutes are identified based on their contributions over the span of eighteen years i.e., from 2000 to 2017. The results are shown in Table 5 with National University of Sciences and Technology (NUST) on top having 3305 publications followed by COMSATS having 2246 publications and then National University of Computer & Emerging Sciences (NUCES) and QeA (Quaid-e-Azam) University obtaining 837 and 707 publications respectively.

S. No.	Institute	No. of Publications
1	National University of Sciences and Technology (NUST) Pakistan	3305
2	COMSATS Institute of Information Technology	2246
3	National University of Computer and Emerging Sciences (NUCES), Islamabad	837
4	Quaid-i-Azam (QeA) University	707
5	Lahore University of Management Sciences (LUMS)	578
6	University of Engineering and Technology (UET) Lahore	565
7	International Islamic University (IIU) Islamabad	558
8	Mohammad Ali Jinnah University	470
9	Bahria University	443
10	Pakistan Institute of Engineering and Applied Sciences (PIEAS)	425

Table 5 shows Top 10 Institutions

The data of the institutes obtained from Scopus was further refined by removing the repetitions. The publications by the university and its sub-campuses were listed separately, so, we identified these institutes and merged their record. The university record was obtained from the database on a separate yearly basis so as to ensure accuracy. For example, the results for the Center of Excellence in Science and Applied Technologies (CSEAT) appeared five times in a list for the year 2015. The name appeared in succession four times and once appeared as acronym for the institute i.e., CSEAT. We combined all these records to remove redundancies and consistency of the results obtained. Further, the institutes such as COMSATS Institute of Information Technology and its sub-campus at Lahore were listed as separate research institutes which technically are the same, so we merged the record of the two under the name of COMSATS Institute of Information Technology. The same is done for the NUST under which the Military College of Signals and the College of Electrical and Mechanical Engineering works. All that linking and merging work is done manually as there was no such tool or functionality to achieve the objective.

4.8 Trend over time of the contributing institutes

In this section, we identified the research contribution trends of various institutions over the 18 years. For this purpose, we divided the whole span of 18 years into sessions of three years. The results are shown in Table 6.

As is clear from the results that NUST keeps its position at the top over all the periods except for the first session (2000-2002) where GIKI tops the list. COMSATS initiated a surprise rise from the third session i.e., 2006-08 into the second place and then retained that position till the end. NUCES also made it to the list in the third session in the 3rd place and retained it for 3 sessions except the last one i.e., 2015-17. Quaid-i-Azam University which is at the fourth place in Table 5 is not among top 3 except for the last

session in 3rd place. Lahore University of Management Science which appeared at the fifth place in Table 5, appeared only once in the list in Table 6 i.e., for the session 2003-05 at the second place.

Session	Institutes	Number of Publications
2000-2002	Ghulam Ishaq Khan Institute of Engineering Sciences and Technology	25
	National University of Sciences and Technology Pakistan	23
	University of Engineering and Technology Lahore	15
2003-2005	National University of Sciences and Technology Pakistan	113
	Lahore University of Management Sciences	46
	Ghulam Ishaq Khan Institute of Engineering Sciences and Technology	45
2006-2008	National University of Sciences and Technology Pakistan	455
	COMSATS Institute of Information Technology	122
	National University of Computer and Emerging Sciences	92
2009-2011	National University of Sciences and Technology Pakistan	647
	COMSATS Institute of Information Technology	323
	National University of Computer and Emerging Sciences	307
2012-2014	National University of Sciences and Technology Pakistan	818
	COMSATS Institute of Information Technology	623
	National University of Computer and Emerging Sciences	202
2015-2017	National University of Sciences and Technology Pakistan	1249
	COMSATS Institute of Information Technology	1158
	Quaid-i-Azam University	321

Table 6 shows the Trend over Time of Contributing Institutes

4.9 Comparison of Pakistan over the years with India, Turkey, Malaysia, Bangladesh and Saudi Arabia

In this section, the results of Pakistan are compared with other countries. In order to compare Pakistan's performance in this region of the world (i.e., South Asia), we considered India, Bangladesh, Saudi Arabia, Turkey and Malaysia as these are the countries who have invested recently in their higher education promotion considerably.

A number of metrics can be used to compare the research output of country. A simple and commonly used approach is to consider the absolute count of research publications [24]. However, the method is flawed, as it favours countries with large population (such as USA, China and India etc.). To offset the large population factor, publications per million people is an alternative metric. The metric normalizes the publication count based on the population, and thus provides a level benchmark for comparing the research productivity of a country. Other factors such as number of higher education institutes, and country's GDP can also be considered [25]. For simplicity, we consider the absolute number of publication versus the GDP, and the publications per million population as our criterion for research performance of a country. The population and GDP data of a country are taken from the World Bank online database [26].

Figure 3 shows the world wide count of the computer science literature being published during the years 2000 to 2017. It shows that the top four countries including United States, China, Germany and United Kingdom accounts for almost 43% of the total pool of papers published. Next in line are the India, Canada, Taiwan and Australia in terms of the publication count. The percentage share of India, Malaysia, Turkey, Saudi Arabia, Pakistan and Bangladesh accounts for 3.66%, 0.84%, 0.78%, 0.36%, 0.29% and 0.14% respectively of the total pool of the publications which amounts to a total of 5126070 publications.

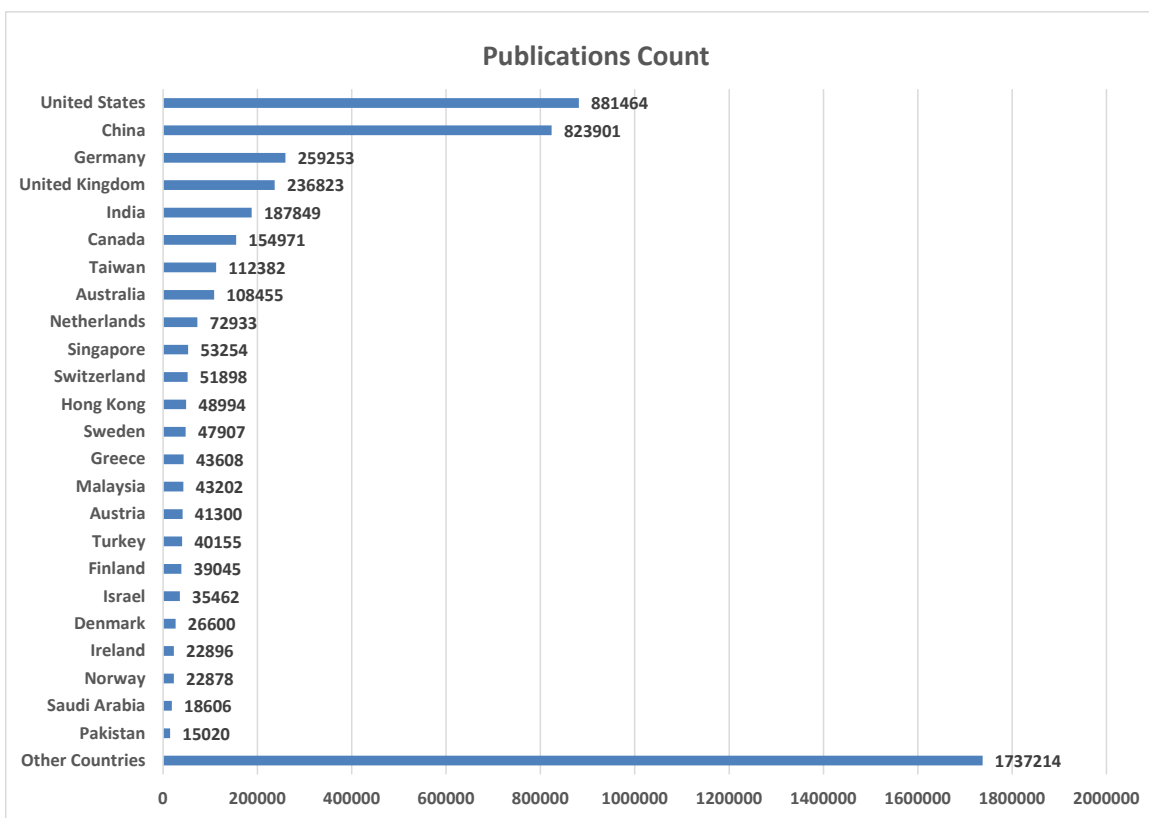


Figure 3 shows the world wide contribution of computer science publications

In order to present a more rationale view of the performance of these countries, we normalized the results by dividing the publications count over their population (in millions). This shows us a different result as smaller countries such as Singapore, Finland, Hong Kong, Switzerland and Taiwan came to the top such (see Figure 4). The countries at the top in Figure 3 shows a much lower performance as shown in the normalized graph in Figure 4. An important point to mention here is that the top 5 countries which

make it to the list in Figure 4 are shown in Figure 3 but the rest of the countries in the list are omitted. As it was not possible to show the countries in the graph for which we were comparing the results i.e., Pakistan, Bangladesh, India, Turkey, and Malaysia. Further the countries which were omitted, their publications count was added up to the “Other countries” entry in the graph as shown in Figure 3.

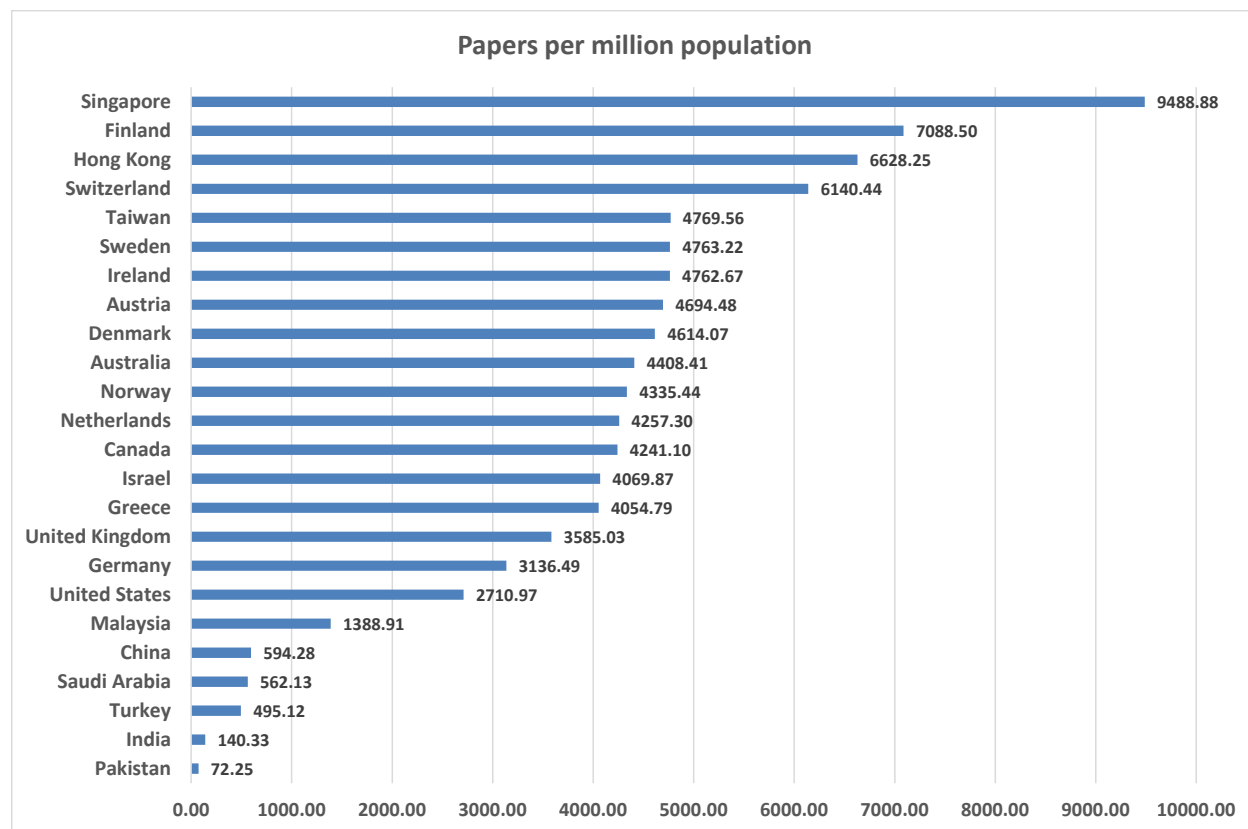


Figure 4 shows the normalized count of computer science literature published world-wide during the period 2000-2017

In order to show a clear picture of the performance of all the countries over time we plotted a graph showing the percentage share of the selected set of countries (Bangladesh, India, Malaysia, Saudi Arabia and Turkey) along with Pakistan. Figure 5 shows the percentage distribution of the publications over the years from 2000 to 2017. The figure clearly shows that the Bangladesh started at 1.29% in 2000, and over the years reached at 3.05% in 2017. India starting with 61.37% in 2000 ended up with the value 61.10% which is a slight decrease but over the years it has shown fluctuations. Such as after 2000, the next year the contribution dropped to 58.27% and again next year i.e., 2002 it reached the maximum at 65.36%. Malaysia performance shows an increase by starting at 7.61% and ending at 14.13% which is almost the double of the starting value. It shows a steady increase over the years but a few are interesting to note where the improvement is major such as in 2003 the value switched from 7.77% (in 2002) to 13.72% which shows a great difference but immediately the next year it dropped back to 7.75%. Similarly, in 2007 starting at 11.48% and the next year i.e. in 2008, it reached a maximum of 17% which consequently moved on to next year of 18.77% but after that it shows a decline and ended up at 14.13% in 2017. Pakistan overall showed an increase from 1.29% to 5.61% but there are a lot of ups and downs along the way. Such as immediately after 2000 i.e. in 2001, the value increased from 1.29% to 3.10% but after that in 2003 it again moved up to 3.88% and so on up to 6.05% for 2005 and the next year it dropped back to 4.64%. In 2009 it reached a peak value of 6.47% but it then went on declining the years onwards and in 2017 it stopped at 5.61%. Saudi Arabia however showed an overall decline by starting at 6.78% in

2000 and ended up at 6.35% in 2017. The peaked value was observed immediately after the next year from the starting year i.e., 2001 which is 7.80% but the following year it dropped to a low value of 4.56% from where onward it shows a decline in values over the years till 2009, after which it jumped back to 5.02% the next year from 3.55% in 2009. Then onward it shows a rise till 2015 for which the value is at 7.50%, for which the next year i.e. in 2016 we noted a downfall to 6.13% and then in 2017 it finally reaches 6.35%. Turkey overall showed a much worse performance amongst all the countries by starting at 21.65% in 2000 and ended up at 9.76% in 2017 which is a huge loss in numbers. For the first three years it shows a decrease in numbers reaching to 19.13% in 2002, but the next year i.e. in 2003, it performs well at 20.28% and continue to escalate the next year forward. But immediately after this it starting declining till the very end reaching to 9.76% in 2017.

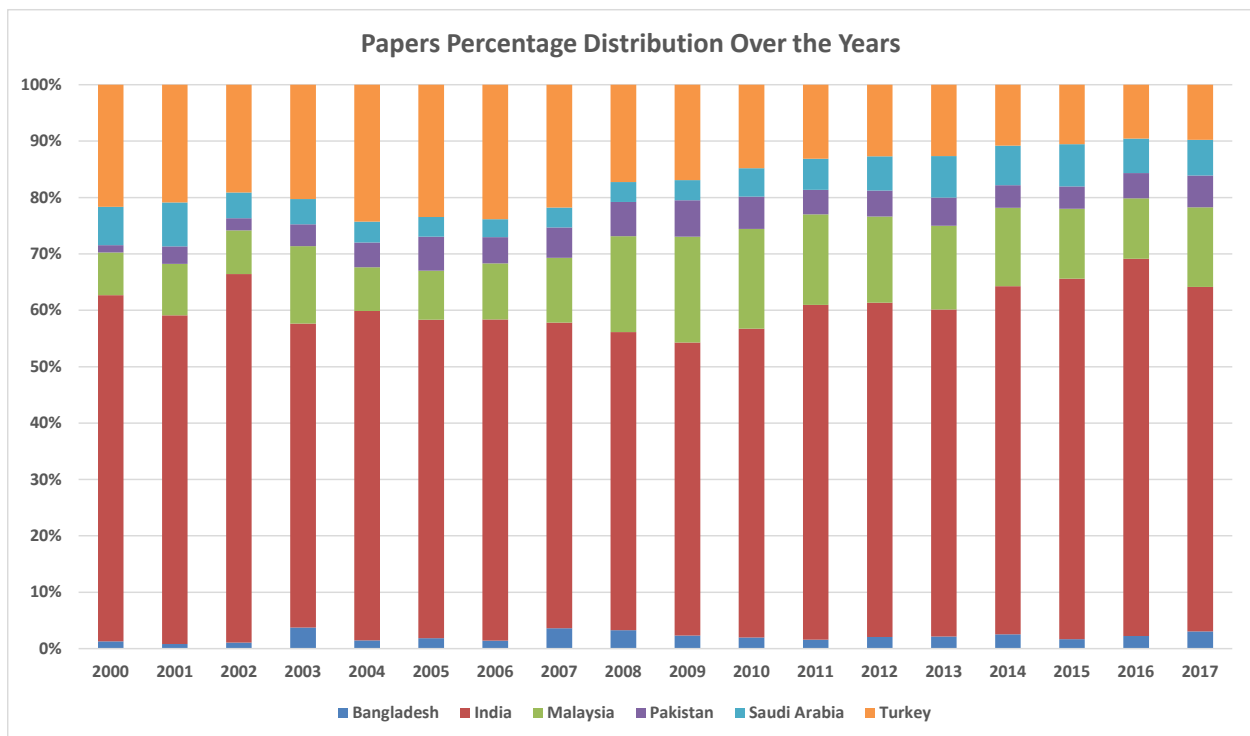


Figure 5 shows the percentage count of the computer science literature published over the years from 2000-2017

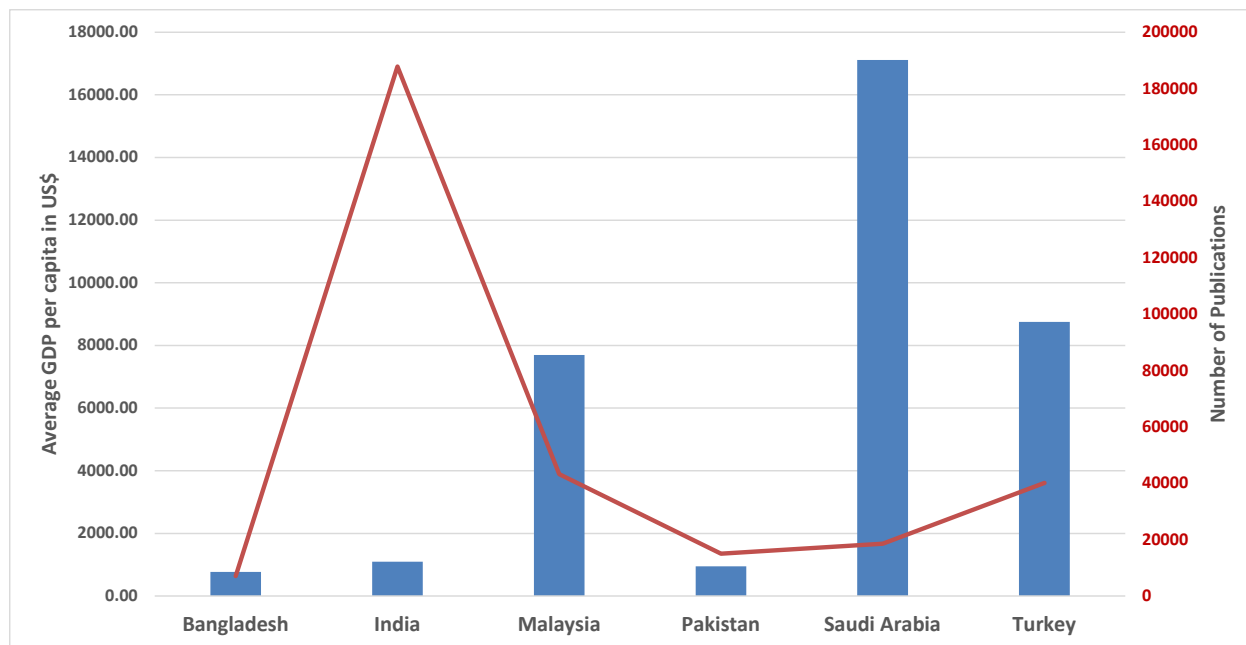


Figure 6 shows the comparison of the GDP per capita against the Number of publications over the years from 2000-2017 by Bangladesh, India, Malaysia, Pakistan, Saudi Arabia and Turkey.

GDP shows the growth of the country in terms of how much a person can generate given the suitable environment by the government in that particular country. GDP per capita shows the average GDP generated by each individual in a particular year. We take a look at the GDP per capita in comparison to the number of publications among the selected countries. The results are shown in a graph in Figure 6. From the figure it is evident that the Bangladesh with the lowest GDP per capita has the lowest outcome in terms of publications. Bangladesh in comparison to India and Pakistan with the almost the same levels of GDP per capita, Pakistan shows some improvement as its publications mark surpasses its GDP per capita mark. In contrast India shows a huge peak if we look at its publications line in the graph almost in level to the GDP per capita mark of Saudi Arabia. Malaysia shows an average performance as its publications are somewhat in the middle of its GDP per capita readings. Saudi Arabia on the other hand having the highest GDP per capita among the selected countries shows a very low value in terms of publications. Turkey does show average performance but if we compare it to Malaysia, its GDP per capita is high but its publications show a lower value than Malaysia. From the comparison it is concluded that the GDP per capita does not necessarily result in higher publication count.

There is another factor that might be interesting to take such as the GDP spending on the R&D of a country. This could give us more insight into the spending at the research area that could yield the development. The data is taken from the UNESCO online database [27]. We have taken the values for each year for the eighteen years i.e., from 2000 to 2017 and then we calculated the average of these values. The results are shown in Figure 7. The result shows that India shows an exceptional case where it spends a mere of 25.98\$ per year on the R&D which results in a massive amount of publications. Malaysia and Saudi Arabia spends a huge amount on R&D i.e. 205.01\$ and 167.33\$ respectively but their results are not satisfactory at all in publications. Pakistan on average have a spending of the lowest among the selected countries i.e. 12.75\$ but the publication results are promising. Turkey also despite having a large spending on the R&D with 127.44\$, does not show promising results in its publications. Bangladesh is not shown here in the figure as there were no given values for R&D spending of Bangladesh on the UNESCO site.

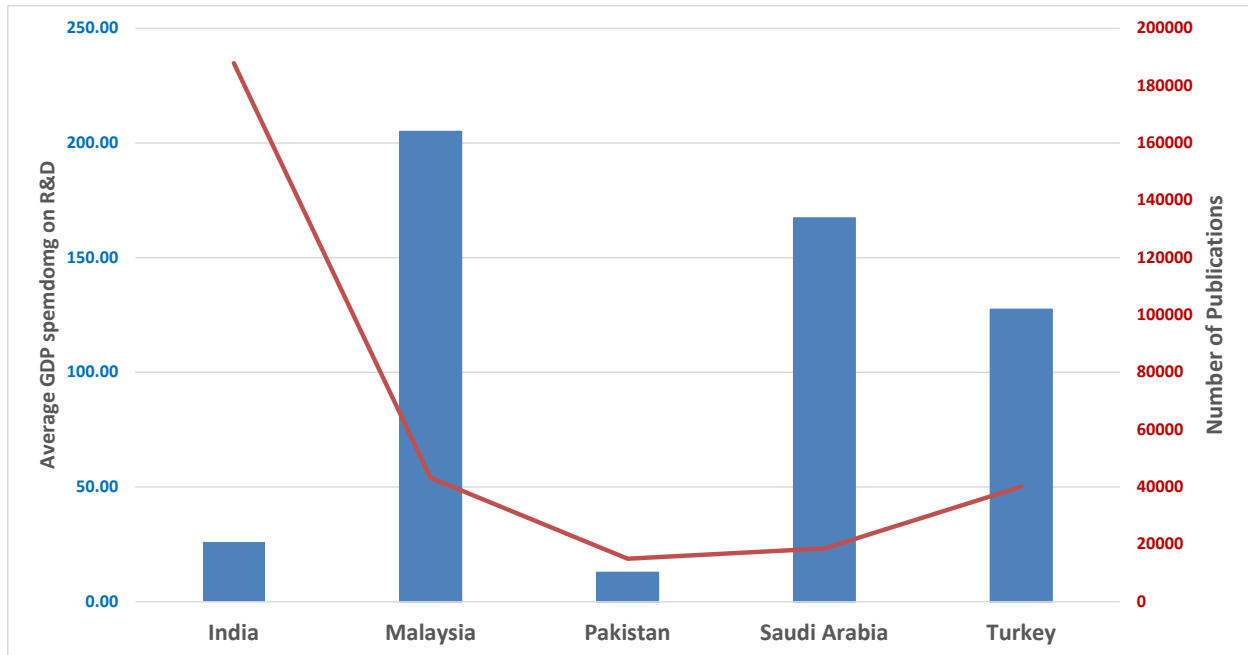


Figure 7 shows the comparison of the Average GDP spending on the R&D against the number of publications for India, Malaysia, Pakistan, Saudi Arabia and Turkey (except Bangladesh).

5 Limitations of the study

Like many scientific studies, our study has its limitation. For example, replicating the study might show slightly different results as the number of citations vary from source to source. Citation count for a paper in Scopus might vary than the one reported by Google Scholar or ISI Web of Knowledge. Further the data was taken from Scopus on 10th April 2018 which might be subject to change if searched afterwards.

6 Conclusion

This work presents a systematic analysis of the research work performed by the Pakistani researchers in the field of computer science over a span of eighteen years i.e., from 2000 to 2017. We found out that almost 38% of the publications get zero citations and on average each publication gets 4.48 citations. Moreover, single authored papers account for a mere 4% whereas 2 and 3 authored papers account for more than 50% of the total pool with an average of 3.57 authors per publication. The leading institute with most publications is the NUST followed by COMSATS.

This work can be extended to form a citation network and its various types can also be used to calculate the citation relations such as direct citation, co-citation analysis and bibliographic coupling relation. Further keyword co-occurrences of the publications and co-authorship relationships can also be discovered. The historiography can be used to illustrate the publication network of publications that are published and cited over time.

References:

- [1] W. M. Sweileh, S. W. Al-Jabi, A. S. AbuTaha, H. Z. Sa'ed, F. M. Anayah, and A. F. Sawalha, "Bibliometric analysis of worldwide scientific literature in mobile-health: 2006–2016," *BMC medical informatics and decision making*, vol. 17, p. 72, 2017.
- [2] S. R. Kolle, T. H. Shankarappa, M. Arun, and T. Manjunatha Reddy, "Characteristics and trends in global lead removal research: A science citation index expanded-based analysis," *Desalin. Water Treat*, vol. 80, pp. 164-173, 2017.
- [3] F. Narin, *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*: Computer Horizons Cherry Hill, NJ, 1976.
- [4] C.-S. Lin, M.-H. Huang, and D.-Z. Chen, "The influences of counting methods on university rankings based on paper count and citation count," *Journal of Informetrics*, vol. 7, pp. 611-621, 2013.
- [5] K. W. Boyack and R. Klavans, "Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?," *Journal of the American Society for Information Science and Technology*, vol. 61, pp. 2389-2404, 2010.
- [6] V. Garousi and T. Varma, "A bibliometric assessment of canadian software engineering scholars and institutions (1996-2006)," *Computer and Information Science*, vol. 3, p. 19, 2010.
- [7] V. Garousi and T. Varma, "A Bibliometrics Analysis of Canadian Electrical and Computer Engineering Institutions (1996-2006) Based on IEEE Journal Publications," *Computer and Information Science*, vol. 5, p. 1, 2012.
- [8] C. Wohlin, "An analysis of the most cited articles in software engineering journals-2000," *Information and Software Technology*, vol. 49, pp. 2-11, 2007.
- [9] G. J. Holzmann, "The model checker SPIN," *IEEE Transactions on software engineering*, vol. 23, pp. 279-295, 1997.
- [10] Y. Ding, G. Zhang, T. Chambers, M. Song, X. Wang, and C. Zhai, "Content-based citation analysis: The next generation of citation analysis," *Journal of the Association for Information Science and Technology*, vol. 65, pp. 1820-1833, 2014.
- [11] V. Garousi and J. M. Fernandes, "Highly-cited papers in software engineering: The top-100," *Information and Software Technology*, vol. 71, pp. 108-128, 2016.
- [12] D. Fiala and G. Tutoky, "Computer Science Papers in Web of Science: A Bibliometric Analysis," *Publications*, vol. 5, p. 23, 2017.
- [13] J. E. Dayhoff, *Neural network architectures: an introduction*: Van Nostrand Reinhold New York, 1990.
- [14] P. Mell and T. Grance, "The NIST definition of cloud computing," 2011.
- [15] J. Richardson and M. Y. Ali, "Research Publishing by Library and Information Science Scholars in Pakistan: A Bibliometric Analysis."
- [16] R. Bajwa and K. Yaldram, "Bibliometric analysis of biotechnology research in Pakistan," *Scientometrics*, vol. 95, pp. 529-540, 2013.
- [17] J. F. Burnham, "Scopus database: a review," *Biomedical digital libraries*, vol. 3, p. 1, 2006.

- [18] M. D. Rodriguez, J. Ahmed, and M. Shah, "Action mach a spatio-temporal maximum average correlation height filter for action recognition," in *Computer Vision and Pattern Recognition*, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8.
- [19] M. Ali, S. U. Khan, and A. V. Vasilakos, "Security in cloud computing: Opportunities and challenges," *Information sciences*, vol. 305, pp. 357-383, 2015.
- [20] M. I. Ali, F. Feng, X. Liu, W. K. Min, and M. Shabir, "On some new operations in soft set theory," *Computers & Mathematics with Applications*, vol. 57, pp. 1547-1553, 2009.
- [21] K. Ishaque and Z. Salam, "A deterministic particle swarm optimization maximum power point tracker for photovoltaic system under partial shading condition," *IEEE transactions on industrial electronics*, vol. 60, pp. 3195-3206, 2013.
- [22] Y.-C. Wu, Q. Chaudhari, and E. Serpedin, "Clock synchronization of wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 28, pp. 124-138, 2011.
- [23] V. Garousi and M. V. Mäntylä, "Citations, research topics and active countries in software engineering: A bibliometrics study," *Computer Science Review*, vol. 19, pp. 56-77, 2016.
- [24] C. R. Carpenter, D. C. Cone, and C. C. Sarli, "Using publication metrics to highlight academic productivity and research impact," *Academic emergency medicine*, vol. 21, pp. 1160-1172, 2014.
- [25] S. A. Meo, A. A. Al Masri, A. M. Usmani, A. N. Memon, and S. Z. Zaidi, "Impact of GDP, spending on R&D, number of universities and scientific journals on research publications among Asian countries," *PLoS One*, vol. 8, p. e66449, 2013.
- [26] Available at <https://www.worldbank.org/>
- [27] UNESCO Institute for Statistics [Online]. Available: <http://data.uis.unesco.org/>