Empirical Study on Citation Count Prediction of Research Articles

Murali Krishna Enduri[1,*], V Udaya Sankar[2], Koduru Hajarathaiah[1]

[1]Department of Computer Science and Engineering, SRM University, Amaravati, Andhra Pradesh, INDIA.
[2]Department of Electronics and Communications Engineering, SRM University, Amaravati, Andhra Pradesh, INDIA.

ABSTRACT
Citation is a measure that quantifies the impact of the researcher, research article and journal's quality. Investigating the citation of articles and/or researchers is one of the important tasks in the research community. So, understanding and predicting citation patterns of research articles has become popular in scientific research fields. In this work, we give a machine learning approach to predict the citations of research articles using the keywords. We study the citation impact based on keywords motioned in the articles using the data set of publications which are published in the various physical review journals from 1985-2012. In this dataset, for each publication is allocated some PACS codes (keywords) by their authors which represent a sub-field of Physics. In this work, we are investigating the impact of PACS codes of article on article's citation. We are performing our analysis on the first (sub-field of physics), second (sub area of sub-field of physics) and third level of PACS codes. We observed that compared to the first level, every pair of citation patterns of the second level is highly correlated. We also obtained a universal approximation curve for the third level that matches with the average value of the first level. This curve looks like a shifted and scaled version of the Gaussian function and is right skewed. We can also predict the citations based on the keywords by using this universal curve.

Keywords: PACS codes, Citation, APS journals, Machine learning.

Correspondence
Murali Krishna Enduri
Department of Computer Science and Engineering, SRM University, Amaravati-522502, Andhra Pradesh, INDIA.
Email id: muralikrishna.e@srmap.edu.in
ORCID ID 0000-0002-9029-2187

Received: 18-12-2021
Revised: 28-03-2022
Accepted: 24-06-2022
DOI: 10.5530/jscires.11.2.17

INTRODUCTION

In recent years, bibliometric and scientometrics indicators have been applied in the context of research evaluation as well as research impact more generally. Citation is a measure that quantifies the impact of the researcher and the journal's quality. It plays a pivotal role in a journal's impact factor, researcher's *h*-index and various other measures. Recently this measure is paying more importance in the research community. Researchers and scholars contribute many publications, so they seek to make an impact in their corresponding scientific research communities during their academic careers. Reflect on their talents and create new collaboration opportunities if their articles have the highest possible research impact. Recently, publishing papers that are highly cited are increasing due to the competitiveness of research grants and collaboration based on the researcher's impact. As an output, identifying the variables that impact paper citations has been investigated by the publishers and the authors of research articles.

Related work

Research articles will receive a good number of citations if the papers include important research topics and/or are relevant, popular and useful.[1] However, the diversity of topics and influence of reference can also affect the citation of an article.[2-5] The effect of the bibliometrics on the citation has been investigated because they are available and do not change over time. Although bibliometrics does not spill out the entire impact of citations and does not claim to identify causal relationships.[6] So, predicting citations of research articles has become popular in scientific research fields. Researchers need to study or investigate various factors of articles involved and it affects citation counts of the articles. Recently, the research community has accepted citation counts as one of the main measures of the impact of research. Nowadays, many countries consider citation counts in their national research evaluation practices such as the United Kingdom,[7,8] Australia[9] and New Zealand.[10] Many researchers discovered so far various factors (bibliometric variables) affecting citation counts. Yu *et al*.[11] studied that the authors, the journal, the research area, and the papers themselves affect the citation. In the paper,[12] they observed the effect of citation patterns based on the number of publications and number of citations of each author in the articles. Based on the APS Physical review database, the authors,[14] identified the relationship between authorship and citation, and analyzed those individual researchers cite their co-authors work more frequently compared to others research work.

In the paper,[15] they studied medium diversity papers that receive more citation compared with very low and high diversity articles. Authors introduced a measure paper potential index which is defined based on inherently quality of scholarly paper and the scholarly paper impact decaying over time, early citations, and early citers' impact.[2,16,17] They observed that paper potential indexes better interpret the changes in citation, without the need to adjust parameters. Similarly, they are many others also look at different factors of articles or authors affect their citations.[18-21] In the paper,[22] they investigated whether multi/inter-disciplinary research activities are correlated to impact of research and number of publications. The researchers explored many other factors which plays key role in the impact of citation such as year of publication, number of pages, number of authors, number of references, abstract length, keyword repetitions in abstract.[6,18,21,23,24] Since citation counts have many usages within academia and other fields, it is very important to study why one article is cited more compared to another. In this work, we investigate whether the keyword mentioned in the publication affects the citations. This study helps researchers, readers and editors gain an insight into the intelligent use of keywords in research publications to gain the number of citations. We also give various statistical results on citations based on individual keywords mentioned in the article.

Our Contribution: In this work, we are investigating the impact of PACS codes of an article on the article's citation. The analysis we are doing on the first (sub-field of physics) and second (more sub area of physics) level of PACS codes. The maximum number of citations reached within two or three years of publishing time for every sub-field and these citations reduce over a period. Similarly in the second level PACS code, we observed that some sub areas of physics receive more citation compared to others. We also observed that compared to the first level, every pair of citation patterns of the second level is highly correlated. We find the universal approximation curve for the third level that matches with the average value of the first level. This universal curve is a shifted and scaled version of the Gaussian function, and it is right skewed. We can also predict the citations based on the keywords of the paper by using this universal curve.

LITERATURE REVIEW

To predict the citations of a paper with the keywords used in the paper. Keywords used in the paper usually from subfields of physics which is given in Table 1. Based on the keywords mentioned in the paper we can classify the paper and we can predict how many citations will be received by the paper.

Table 1: Sub-field of Physics based on PACS codes in first level.

PACS CODE	Sub-field of Physics
00	General
10	The Physics of Elementary Particles and Fields
20	Nuclear Physics
30	Atomic and Molecular Physics
40	Electromagnetism, Optics, Acoustics, Heat Transfer, Classical Mechanics, and Fluid Dynamics
50	Physics of Gases, Plasmas, and Electric Discharges
60	Condensed Matter: Structure, Mechanical and Thermal Properties
70	Condensed Matter: Electronic Structure, Electrical, Magnetic, and Optical Properties
80	Interdisciplinary Physics and Related Areas of Science and Technology
90	Geophysics, Astronomy, and Astrophysics

Table 1 is level 1 classification of physics only. Here we give the decision tree in the machine learning technique. Let S be a set of samples, the percentage of class i is p_i. The entropy is

$$E(S) = -\sum_{i=1}^{m} p_i \log p_i$$

Where E is entropy. Partition the samples of S with the feature set F, and the information gain is:

$$igain(S,F) = E(S) - \sum_{P=1}^{mp} \frac{|S_p|}{|S|} E(S)$$

For each level, we do the analysis for citations which we will explain in next sections.

DATA SET

Physical Review journal-initiated publishing research articles from 1893 by the American Physical Society (APS). In subsequent years APS included journals according to the sub-field of physics which are shown in Table 3.

In this paper, we considered published research articles from APS Physical Review Journals that are mentioned in Table 1 from 1985 to 2012 to investigate the PACS codes impact on citations. For each article, the data set includes title, name of the authors, publication date, author's affiliation, unique digital object identifier (article ID) and PACS codes (keywords used in the article and it represents which area the article belongs to). We have considered citations of Physical review articles from ASP journals along with metadata of various journal articles (it does not include citations received from other than APS journals). These data sets are requested and received from. The basic details of the data are given in Table 2.

Table 2: Journal Name and Introduced Year.

Journal Name	Year
Modern Physics	1929
Physical Review Letters	1958
Physical Review A, B, C and D	1970
Physical Review E	1993
Physical Review X	2011

Pacs Classification

A hierarchical classification of the PACS scheme indicates various fields as well as sub-fields of Physics up to five levels. It contains two pairs of digits followed by a pair of alphabet characters, separated by special symbol dots. For a simple illustration, in PACS code 02.10.Ab, the first digit 0 is for General Physics, the second digit 2 for Mathematical methods in physics, 10 for Logic, set theory, and algebra and Ab represents Logic and set theory. These codes are updated frequently over a time by the American Institute of Physics (some codes introduced newly, and some are removed). In this paper, for our analysis, we restricted up to the third level of the hierarchy of PACS codes (first four digits) since these represent all sub-fields of physics and are reasonably stable. In our future work, we extend our analysis to a higher-level hierarchy of PACS codes. PACS codes were announced and introduced in 1975 and people are using these PACS in their articles. But most articles published from 1975 to 1984 have not allocated any PACS codes because PACS code was introduced recently, and people are not aware of it (Figure 1). We consider the period from 1985 onwards, as the amenability towards PACS codes jumped to more than 90% and has been consistently high since then.

In Figure 2, we show the fraction of researchers using various PACS codes in the articles published between 1985-2012. Most researchers have used only one to four PACS codes in their articles. Clearly, we can observe a power-law decay till PACS of 60, from thereafter, we see the graph is changing. The pattern of the plot follows multiple Pareto distribution.[25]

Fraction articles using the different PACS code in Figure 3. It shows most of the article using less than six PACS codes. The number of PACS used in an article is very small. Clearly, we can see a maximum 22 PACS used in very few numbers of articles.

In recent days most of the researchers studied,[26-28] the data analytic part in relation to temporal variations. Wang *et al.*[29] provided log-normal distribution about maximum citations that receive at age in (Wang, Song *et al.* 2013).[28] Enduri *et al.* analyze the citations of article's pattern over time and their correspondence with PACS codes.[15] An average number of citations of the papers during the year 1985-2012 is shown in Figure 4. Within the first two to four years a maximum number of citations was reached and later drastically decreased. After fifteen to twenty years the citations of the paper will be negligible. This analysis is done by the authors on the published articles from 1998–2006 in Spanish Psychology journals on the web of Science.[30] In this work, we investigate the impact of PACS codes (key word or sub-field of physics) on publication citation. We analyze and predict citations received for PACS code which is included in the research article.

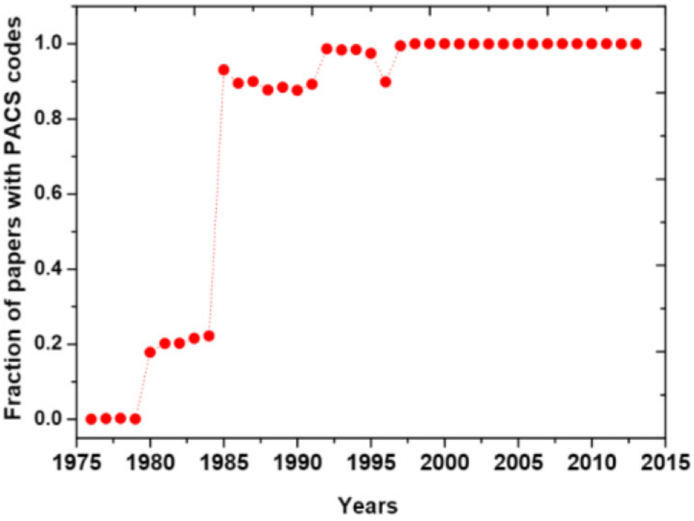

Figure 1: Pattern of articles with PACS codes over 28 years.

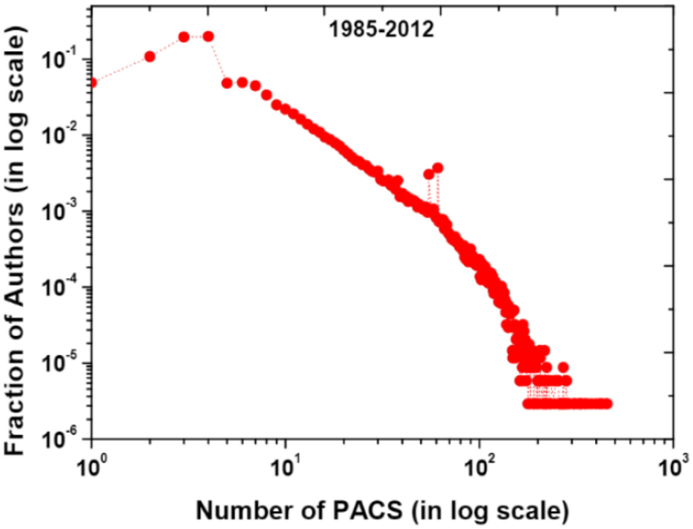

Figure 2: Fraction of authors including to various number of PACS codes.

RESULTS

In this section, we show different analyses on citations received by the different levels of the sub-field in physics over 28 years from the year of publication of the research article. We investigate the impact of PACS codes of the article on the article's citation. Our analysis is based on the first (sub-field of physics) and second (sub area of sub-field physics)

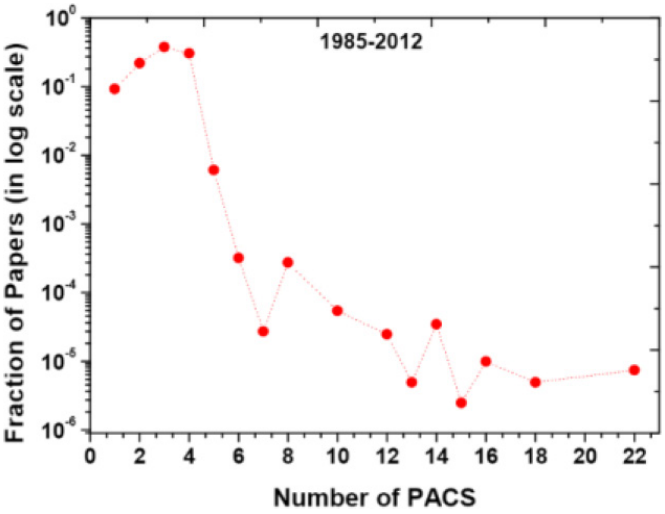

Figure 3: Fraction of articles including the various number of PACS codes.

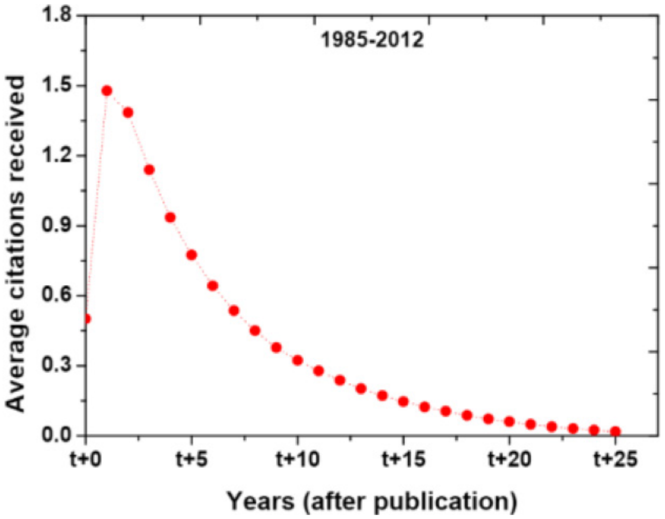

Figure 4: Average number of citations collected by an article at the tth year from its publication.

level of PACS codes. So, we consider first level PACS code information for each paper and find out its citations. So, we can analyze the impact of PACS on citation by analyzing citation information for each level of PACS code.

First Level

The First level of PAC code mapped with its Sub-fields for physics is given in Table 1. The first level of PAC code ranges from 0 (mapped to General physics) to 9 (mapped to Geophysics, Astronomy, and Astrophysics). The number of citations obtained per year wise for each PAC over a period of 28 years is given in Figure 6. More citations are received by the paper from condensed matter sub-fields and the paper on the Physics of gasses got the least number of citations. The maximum number of citations reached It is evident from the box plot shown in Figure 7, that the condensed matter sub-field starts with more number of citations and ends with a maximum number of citations among all other sub-fields of physics. IQR (Interquartile Range) is more for the Physics of elementary particles sub-field which specifies this field got a greater number of citations during the 7th – 21st year of

Table 3: Basic Overview of Data 1985-2012.

Number of authors	343055
Number of articles	399713
Number of articles (in Average) by an author	9.07
Number of authors (in Average) per article	7.59
Number of PACS codes (in Average) per author	10.04
Number of PACS codes (in Average) per article	2.92
Diversity of an author (in Average)	13.16
Diversity of an article (in Average)	3.59
Citation per article (in Average)	10.22

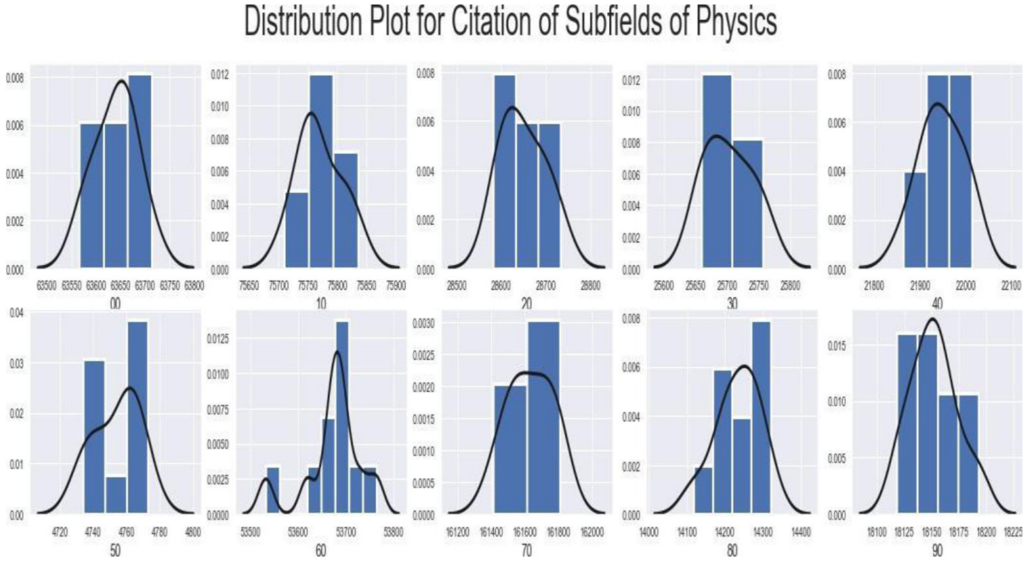

Figure 5: Distribution Analysis on citations of papers by considering second level of PACS codes.

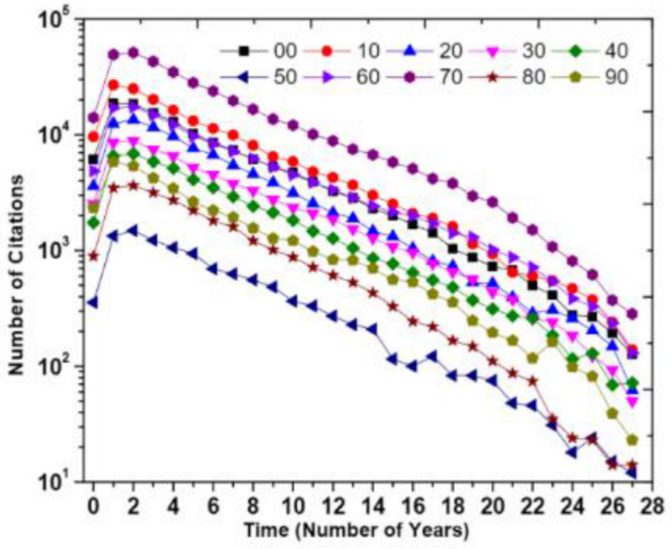

Figure 6: Citations of papers by considering the first level of PACS codes.

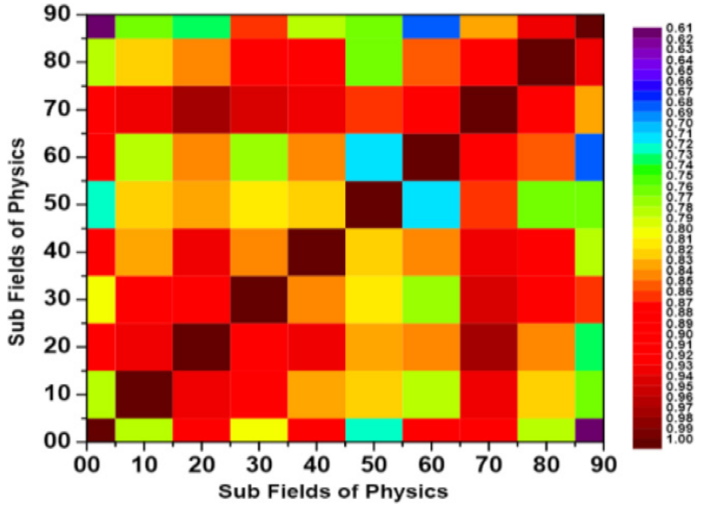

Figure 8: Correlation between of the citation pattern of every pair of first level PACS codes.

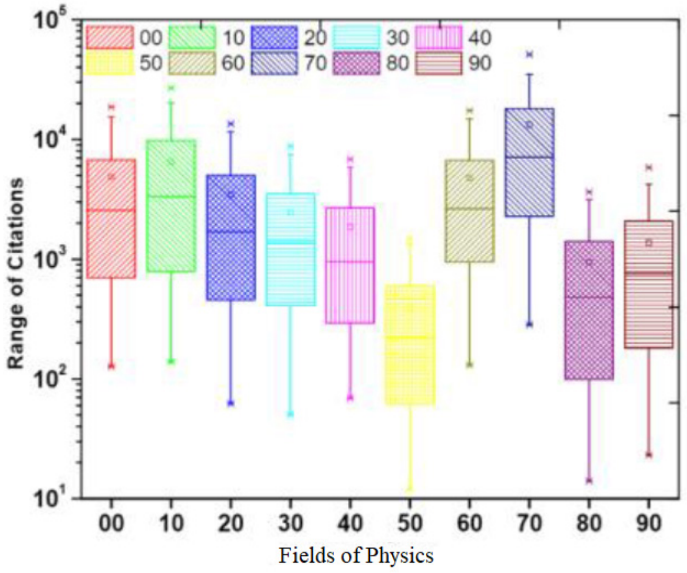

Figure 7: Analysis on citations of papers by considering the first level of PACS codes.

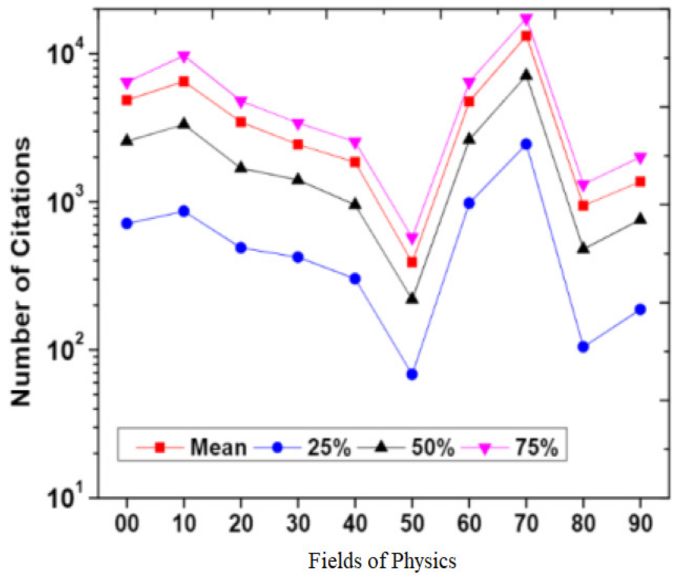

Figure 9: Average on citations of papers by considering first level of PACS codes.

publications. within two or three years of publishing time for every sub-field and these citations reduce over a period.

The mean, median, Q_1, Q_3 values of citations for all the sub-fields are plotted in Figure 9. It is observed that mean of citations is always more than the median of citations for all sub-fields and hence histogram of citations for all sub-fields are right skewed. Hence, over the years citations will reduce for all sub-fields. From the Physics of gasses (50) sub-field to Interdisciplinary Physics (80) the values of mean and Q_3 are almost close to each other that specifying a more significant number of citations after the 21st year of publication.

We have obtained the following Distribution plots as Figure 8 shows the correlation among the citations of physics sub-field using heat-map. The high positive value of correlation indicates that there is a high degree of linear variation in the positive direction. It is observed that physics sub-fields with PACS 20 (≥90%) (Nuclear Physics) and PACS 70 (Condensed matter: Electronics structure) are highly correlated. It is also observed that the Condensed matter sub-field is highly correlated (≥85\%) with Atomic and Molecular, Electromagnetism, Physics of gasses, Condensed matter, Interdisciplinary Physics sub-fields of physics (with PACS codes are 30,40,50,60 &80). The Geophysics (PACS 90) sub-field is low correlated with the condensed matter sub-field (PACS 60) (around 67 %) and with General Physics (PACS 00) (around 60%). We can infer that most of the people who cite sub-fields with PAC codes

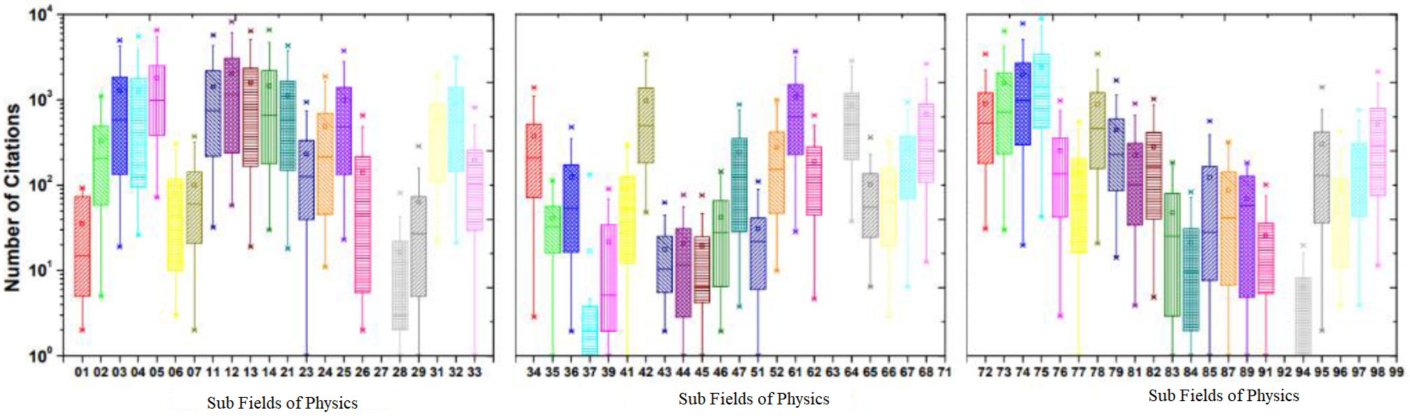

Figure 10: Analysis on citations of papers by considering second level of PACS codes.

other than 70 also cite the sub-field with PAC code 70 shown in Figure 5 for the sub-fields of physics. It is observed that sub-fields General (PACS 00), Electromagnetism, Optics (PACS 40), and Condensed matter (PACS 70) have almost looked like symmetric (Gaussian curve) but with different means and spreads (variance). A similar inference we can obtain from box plots is shown in Figure 7. All other PACS have skewed distributions in either positive or a negative direction. It is observed that condensed matter sub-field have a larger mean as compared with other sub-fields. Interestingly it is observed that for Condensed matter:

Structure, Thermal and Mechanical properties (PACS 60), the distribution seems to be bimodal with a lower peak appearing around 53550 and a higher peak occurring around 53700. Also, it is observed that the spread around higher peaks is larger than that of smaller peaks.

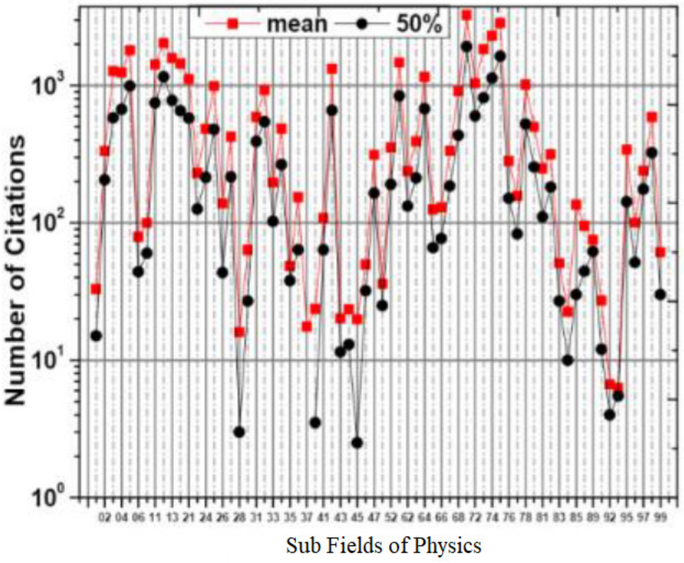

Figure 11: Analysis on citations of papers by considering second level of PACS codes.

Second Level

The second level of PACS code mapped to the sub area of the physics which can be seen in. The second level means we need to consider up to the second digit in the PACS code and it ranges from 00 to 99. We have a maximum of 100 sub areas and the really assigned areas in physics is 68. Some two digits out of 100 were not assigned or a negligible number of citations in 28 years. So, for 68 sub areas, we observed the citations of 28 years of publications which these subareas (PACS codes) mentioned in their publications. Here also we can observe that the same behavior of citations reached a maximum within two or three years of publishing time for every sub-field of physics which is shown in Figure 5. The intention of going to the second level can investigate the impact of citations based on the sub areas mentioned in the research article.

The mean and median values of citations for all the sub-fields of the second level are plotted in Figure 11. Most of 50% of citations are below the average for this second level and very few times, 50% of citations are equal to the mean. The PACS codes 71 (Electronic structure of bulk materials in physics), 74 (Superconductivity) and 75 (Magnetic properties and materials) receiving more citations compared to other PACS codes. Other than this sub-field of condensed matter, 12 (Specific theories and interaction models) and 05 (Statistical physics, thermodynamics, and nonlinear dynamical systems) are capturing more citations. In the sub-field of physics, 28 (Nuclear engineering and nuclear power studies), 39 (Instrumentation and techniques for atomic and molecular physics), 45 (Classical mechanics of discrete systems), and 92 (Hydrospheric and atmospheric geophysics) received very few citations. This is due to a smaller number of articles or authors published in this area and very less attraction of these articles within these areas.

We have shown box plots of citations for all the sub-fields of physics of the second level in Figure 10. The sub-field of physics 75 (Magnetic properties and materials) receives more citation compared to other PACS codes. The sub-field of physics 04 (General relativity and gravitation), 26 (Nuclear astrophysics), 83 (Rheology), and 89 (applied and interdisciplinary physics) received a wide range of citations over a 28-year period. The sub-field of physics 35 (Atomic and molecular collision processes and interactions), 43 (Acoustics), and 67 (Quantum fluids and solids; liquid and solid helium) captured a small range of citations in a 28-year span. The sub-field of physics 28 (Nuclear engineering and nuclear power studies), 39 (Instrumentation and techniques for atomic and molecular physics), and 45 (Classical mechanics of discrete systems) are having citation distribution left skewed. The sub-field of physics 77 (Dielectrics, piezoelectrics, and ferroelectrics and their properties), 94 (Aeronomy and magnetospheric physics), 97 (Stars) are having citation distribution right skewed. Most distributions of the sub-field of physics are in a normal distribution.

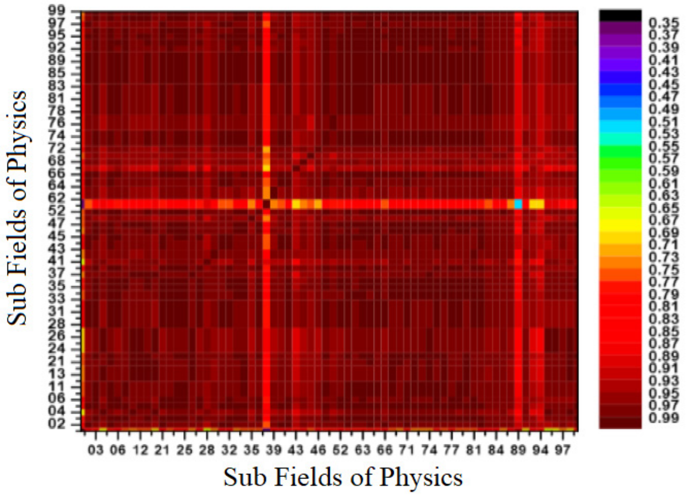

Figure 12: Correlation between of the citation pattern of every pair of second level PACS codes

Table 4: Universal curve details for every sub-field of physics.

Field	Reduced χ^2	Adj. R^2	y_0	xc	w	A
00	1558.31	0.868	26.47	2.58	6.74	2340.00
10	3409.75	0.869	36.59	2.25	7.06	3626.82
20	555.96	0.878	14.95	2.89	6.36	1390.42
30	345.34	0.862	12.86	2.86	7.18	1121.62
40	159.03	0.866	8.65	2.96	6.62	726.83
50	64.17	0.862	4.75	3.04	7.11	480.79
60	390.24	0.852	15.26	2.77	6.53	1069.21
70	2356.50	0.861	36.27	2.73	5.92	2521.59
80	44.45	0.875	3.79	2.97	6.34	386.84
90	374.67	0.878	14.81	2.11	6.08	1107.96

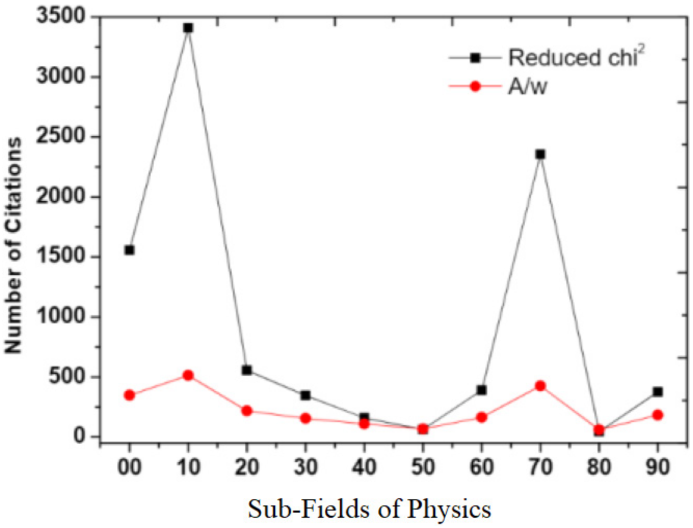

Figure 14: Reduced χ2, A/w versus Sub-fields.

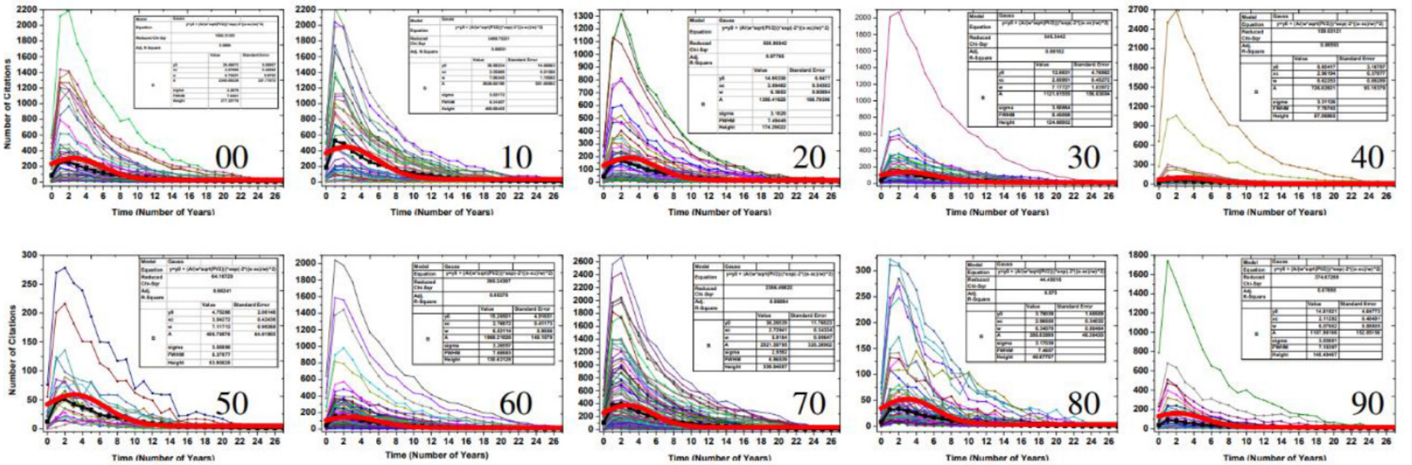

Figure 13: Third level comparison with first level and universal approximation (red color) along with an average value of citation of the first level (block color).

We have shown the correlation between the citation pattern of every pair of second level PACS codes in Figure 12. Compared to the first level, every pair of citation patterns in the second level is highly correlated. Most of the PACS are extremely correlated with their citations and very few of PACS are not. The second level PACS 61 (Structure of solids and liquids; crystallography) and 89 (Other areas of applied and interdisciplinary physics) are having very low correlation. We can clearly observe that 61 (Structure of solids and liquids; crystallography), 89 (Other areas of applied and interdisciplinary physics), and 37 (Instrumentation and techniques for atomic and molecular physics) have less correlation with all other PACS codes.

Third Level

The third level of PACS code is mapped to sub areas of sub-fields of the physics which can see in [24]. Third level means we need consider up to four digits in the PACS code and it ranges from 00.00 to 99.99. We have a maximum of 1000 sub areas and really assigned areas in physics is 944 in the third level. Some two digits out of 1000 were not assigned or a negligible number of citations in 28 years. So, for 944 sub areas, we observed the citations of 28 years of publications which these subareas (PACS codes) mentioned in their publications. For each PACS code, we look at citations of the paper in which the paper has the PACS code. Consider the first 28 years of citation of this paper including it to the PACS code. For each PACS code, we can have a citation pattern. Here also we can observe that the same behavior of citations reached to maximum within two or three years of publishing time for every sub-field of physics. The Intention of going to the third level can investigate the universal pattern of citations based on the sub areas mentioned in the research article. Figure 13 describes plots of citations with respect to years for all the sub-fields starting with codes 00–90. Each Figure within the Figure contains a universal approximation (red color) along with an average value of citation of the first level (block color).

The universal curve is obtained using a shifted and scaled version of the Gaussian curve (given in the below equation) which is right skewed Gaussian. Standard deviation w species the width of the universal curve. Table 4 gives details about the universal curve for every sub-field of physics. Reduced χ2 specifies the goodness of fit and it should be low. It is also observed from Figure 14, that Reduced χ2 and the ratio are proportional to each other.

$$y = y_0 + \frac{A}{W\sqrt{\frac{\pi}{2}}} e^{-2\left(\frac{x - x_c}{W}\right)^2}$$

From Figure 14, it is observed that sub-fields 10 and 70 have a maximum value of reduced χ2, A/w indicates that even though these sub-fields have many dominant sub-fields, the universal curve approximates that of an average value of the first level. Hence, this approximation has less good off it as compared with other sub-fields. For the case of sub-fields 30,40,90 very few third levels citations are dominating hence they have almost the same level of A/w. For the case of sub-field 30, the dominant sub-field is 32.80 with 2071 corresponding citations, and the sub-field 90 has the dominant sub-field 98.80 with 1742 citations. Similarly, for the case of sub-field of 40, it has two dominant sub-fields 42.50 and 42.65 with corresponding citations as 2689 and 1059. It is observed from the Figure 14 that a universal curve (sub-field 70) with a lower standard deviation (w= 5.92) has narrow width (that reduces quickly from maximum) as compared with that of a higher standard deviation (w= 0.75) has wider width (sub-field 30) where the curve reduces very slowly from the maximum value. The reduced χ2 and A/w is lower for the sub-fields 50, 80 specifying that universal curves for these sub-fields have better goodness of fit than compared with other sub-fields. If we get any paper in the sub-field, we can predict the first level citation with the help of the corresponding universal curve for that sub-field derived in the 13.

CONCLUSION AND DISCUSSION

In this work, we are investigating the impact of PACS codes of the article on the article's citation. The analysis we are doing on the first (sub-field of physics) and second (more sub area of physics) level of PACS codes. More citations are received by the paper from condensed matter sub fields and the paper on the Physics of gasses got the least number of citations. The maximum number of citations reached within two or three years of publishing time for every sub-field and these citations reduce over a period. Similarly, in the second level PACS code we observed that some sub areas of physics receive more citations compared to others.

We observed that Condensed Matter: Structure, Mechanical and Thermal Properties (PACS 60) has bi-modal distribution, General (PACS 00), Electromagnetism, Optics (PACS 40), and Condensed matter (PACS 70) has distribution almost looks like Gaussian with different means and spreads (variance).

We also observed that compared to the first level, every pair of citation patterns of the second level is highly correlated. Most of 50% of citations are below the average for this second level and very few times, 50\% of citations are equal to the mean. We are investigating third level PACS codes with respective citations and how it is correlated. Our future goal is if we give an input PACS code or set of PACS codes then what is the citation pattern for these PACS codes or set of PACS codes. We can also investigate the pattern of the correlation between the third level to the first level. We also obtained a universal approximation curve for the third level that validates with the average value of the first level. We can also predict the citations based on the keywords of the paper by using this universal curve.

Limitations

In the data set, we studied up to the third level of PACS codes. If want to apply the same techniques to other data sets then we need to have the data for each paper with a specific correlation with keywords similar these data sets in our paper. Getting the data is one of the challenging tasks for this kind of research work. We have predicted the citations for each sub-field of physics with maximum accuracy of 88%. One of the challenging tasks is to improve this accuracy by exploring the depth of the keyword analysis and more data. We recommend for this kind of research work data sets with more correlation of papers and keywords so that we can get more accurate results.

ACKNOWLEDGEMENT

The authors would like to acknowledge and thank the American Physical Society for giving us the data and details about all Physical Review journal papers and their citations.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. Garfield E. The history and meaning of the journal impact factor. Jama. 2006;295(1):90-3.
2. Gazni A, Sugimoto CR, Didegah F. Mapping world scientific collaboration: Authors, institutions, and countries. Journal of the American Society for Information Science and Technology. 2012;63(2):323-35.
3. Didegah F, Thelwall M. Determinants of research citation impact in nanoscience and nanotechnology. Journal of the American Society for Information Science and Technology. 2013;64(5):1055-64.
4. Uddin S, Hossain L, Rasmussen K. Network effects on scientific collaborations. PloS One. 2013;8(2):e57546.
5. Baba T, Baba K, Ikeda D. Citation Count Prediction using Abstracts. Journal of Web Engineering. 2019.
6. Robson BJ, Mousquès A. Can we predict citation counts of environmental modelling papers? Fourteen bibliographic and categorical variables predict less than 30% of the variability in citation counts. Environmental Modelling and Software. 2016;75:94-104.
7. Mryglod O, Kenna R, Holovatch Y, Berche B. Comparison of a citation-based indicator and peer review for absolute and specific measures of research-group excellence. Scientometrics. 2013;97(3):767-77.
8. Madani F, Zwick M, Daim T. Keyword-based patent citation prediction via information theory. International Journal of General Systems. 2018;47(8):821-41.
9. Crowe SF, Prado C. Excellence in research in Australia: The souffle keeps on rising. Australian Psychologist. 2020;55(5):468-87.
10. Anderson DL, Smart W, Tressler J. Evaluating research–peer review team assessment and journal based bibliographic measures: New Zealand PBRF research output scores in 2006. New Zealand Economic Papers. 2013;47(2):140-57.
11. Yu T, Yu G, Li PY, Wang L. Citation impact prediction for scientific papers using stepwise regression analysis. Scientometrics. 2014;101(2):1233-52.
12. Fu L, Aliferis C. Using content-based and bibliometric features for machine learning models to predict citation counts in the biomedical literature. Scientometrics. 2010;85(1):257-70.
13. Redner S. Citation statistics from 110 years of physical review. arXiv preprint physics/0506056. 2005 Jun 7.
14. Martin T, Ball B, Karrer B, Newman ME. Coauthorship and citation in scientific publishing. arXiv preprint arXiv:1304.0473. 2013 Apr 1.
15. Enduri MK, Reddy IV, Jolad S. Does diversity of papers affect their citations? Evidence from American Physical Society Journals. In 2015 11th International Conference on Signal-image Technology and Internet-based Systems. 2015;505-11. IEEE.
16. Bai X, Zhang F, Lee I. Predicting the citations of scholarly paper. Journal of Informetrics. 2019;13(1):407-18.
17. Janavi E, Mansourzadeh MJ, Samandar Ali Eshtehardi M. A methodology for developing scientific diversification strategy of countries. Scientometrics. 2020;125(3):2229-64.
18. Bornmann L, Schier H, Marx W, Daniel HD. What factors determine citation counts of publications in chemistry besides their quality?. Journal of Informetrics. 2012;6(1):11-8.
19. Stegehuis C, Litvak N, Waltman L. Predicting the long-term citation impact of recent publications. Journal of informetrics. 2015;9(3):642-57.
20. Brizan DG, Gallagher K, Jahangir A, Brown T. Predicting citation patterns: Defining and determining influence. Scientometrics. 2016;108(1):183-200.
21. Alimoradi F, Javadi M, Mohammadpoorasl A, Moulodi F, Hajizadeh M. The effect of key characteristics of the title and morphological features of published articles on their citation rates. Annals of Library and Information Studies. 2016;63(1):74-7.
22. Jamali HR, Abbasi A, Bornmann L. Research diversification and its relationship with publication counts and impact: A case study based on Australian professors. Journal of Information Science. 2020;46(1):131-44.
23. Falahati Qadimi Fumani MR, Goltaji M, Parto P. The impact of title length and punctuation marks on article citations. Annals of Library and Information Studies. 2015;62(3):126-32.
24. Khan A, Choudhury N, Uddin S, Hossain L, Baur LA. Longitudinal trends in global obesity research and collaboration: A review using bibliometric metadata. Obesity Reviews. 2016;17(4):377-85.
25. Reed WJ, Jorgensen M. The double Pareto-lognormal distribution-a new parametric model for size distributions. Communications in Statistics-theory and Methods. 2004;33(8):1733-53.
26. Guns R, Rousseau R. Real and rational variants of the *h*-index and the *g*-index. Journal of Informetrics. 2009;3(1):64-71.
27. Hirsch JE. Does the *h*-index have predictive power?. Proceedings of the National Academy of Sciences. 2007;104(49):19193-8.
28. Radicchi F, Castellano C. Rescaling citations of publications in physics. Physical Review E. 2011;83(4):046116.
29. Wang D, Song C, Barabási AL. Quantifying long-term scientific impact. Science. 2013;342(6154):127-32.
30. Contreras EJ, López-Cózar ED, Pérez RR, Rodriguez-Garcia G, De-la-Moneda-Corrochano M. Spanish psychology journals: Demography, editorial tendencies and impact. In Proceedings of the Workshop on European Psychology Publication Issues; Supplement I-2009. ZPID (Leibniz Institute for Psychology Information).