Scientometric Analysis of Computer Science Publications in Journals and Conferences with Publication Patterns

Priti Kumari* and Rajeev Kumar

School of Computer & Systems Sciences, Jawaharlal Nehru University, New Delhi 110 067, INDIA.

ABSTRACT

Prominence of Computer Science (CS) research articles in journals and conferences has been a topic of research. Major issues regarding these publications are pros. and cons. of review process and scientometric data which is sub-fields specific. Therefore, we adopt a different approach, we define metrics based on authors and their affiliations and study the patterns of different publications. We consider publication data of key CS journals and conferences in three sub-fields of past five years, during which that publication patterns have been stabilized. We calculate distinct institutions, occurrences of repeated authors and also examine the overlap of authors and institutions in consecutive years. Thus, we show diversity of CS publications in terms of authors and institutions. We have observed that in conferences 60-80% authors have participated from repeating institutions while this range is 40-60% for journals. Further, the maximum repeating frequency of authors is 28% in conferences, while it is 15% in journals. In addition, overlapping percentage of authors and institutions is also lesser in journals than that of conferences. Hence, journal publications are more diverse in nature in terms of participating authors and institutions. Moreover, cross comparison of conferences and journals in same sub-field indicates the mutual exclusiveness of these mediums.

Keywords: Computer science research, Conference, Journal, Research publications, Scientometrics.

INTRODUCTION

Research publications are key components for effective dissemination of new research findings. Peer–reviewed articles are considered as one of the well–grounded source of knowledge in most of the scientific domains. Such articles are published mostly in conferences and journals. Conferences publish usually short length papers while journals publish long length papers as well as extension of conference papers.[1] Since, journals follow rigorous review process hence journal's review process is considered as significantly sound. Therefore, articles of journals and transactions generally exhibit supremacy over all other articles published elsewhere. However, the case of Computer Science (CS) is quite different.

CS is comparatively a faster moving domain.[2] Since the CS technology is evolving very fast hence a totally new trend in CS can be observed in the past decade. Consequently, conferences in CS started to play a significant role. Thus, conferences in CS are being given more preference than that of journals. Numerous researches have been done to analyze the supremacy of CS publication venues.(We use the term publication venues for both conference and journal throughout the paper due to its significant use in literature for denoting both types of publications.[2-4])

Various factors have been studied so far regarding the conference and journal publications. Mainly the researchers have considered impact of the articles, review process and acceptance rates. Based on these parameters, researchers explored relevant facts and attempted to differentiate both the venues. However due to many reasons the claims made by such studies are debatable. For example, it has been analyzed that CS conferences have higher citation impact as compared to the journals,[5] while other studies have shown contradictory findings that says journal publications attract more citations.[6]

Another important factor in assessing the supremacy of CS publications, is the review process. Conference review process is deadline driven due to which the reviewers are unable to devote sufficient time for most of the papers. This results in comfortable acceptance of articles written by well-known authors while many innovative articles written by new comers may get their articles rejected.[6–8] This makes conference review process less effective. However, a 'rebuttal' stage is added in top-ranked conferences to overcome this short coming in which authors are given an opportunity to refute or address the reviewing comments.

Low acceptance rate of conference papers is also considered as an indicator of quality in CS publications e.g., top-tier conferences. However, researchers have analyzed that low acceptance rate of conferences could not be an indicator of quality,[9] since there exists the possibility of judgment error due to tight deadline and high heap of papers. As there are large number of papers at the same threshold hence selection of few papers from them might be arbitrary to some extent.[10,11] Thus, all such factors: impact factor, citations, review process, low acceptance rate etc., continue to remain a subject of further research and discussion over the supremacy of CS publications. Due to aforementioned issues, some authors have suggested possible alternatives. For an instance, Yann[1] has proposed a new open repository model of publications. Similarly, *arXiv* repository based publications are also gaining popularity.

Due to such shortcomings of abovementioned factors, we choose quite different dimensions for analyzing publication patterns and distinguishing CS conference and journal publications. For this purpose, we consider extended attributes of publications, namely, author names and their corresponding institutions. Since, bibliometric studies of CS publication venues are sub-field specific. Therefore, we carry-out combined study of few sub-fields of CS domain e.g., Theoretical Computer Science, Programming Language and Machine Learning. We collect data of few leading conferences and journals of selected CS sub-fields. We conduct analysis and find publication patterns of these venues with respect to selected CS sub-fields. We, in this work, do not attempt to analyze the supremacy of CS publication venues, rather we aim at finding distinct characteristics of both the venues.

The rest of the paper is organized as follows: In section 2, we carry-out literature review. Motivation of this work is given in Section 3. Section 4, describes the data used along with the detailed description of the methodology. Analysis, results and discussion are included in Section 5. Finally, we conclude our findings in Section 6.

1 http://yann.lecun.com/ex/pamphlets/publishing-models.html

LITERATURE REVIEW

Several research studies have been dedicated for analyzing the supremacy of conference and journal publications in CS. Researchers have included, mainly, the following parameters in their studies such as, time-bound conference review process, impact of the articles, low acceptance rate of conferences, etc.[4–6,9–11,14–16] Based on these parameters, researchers have highlighted important issues and also compared both the publication venues of CS. In addition, several researchers have differentiated both the venues based on bibliometric indicators e.g., citations, impact factor, *h*-index etc. Rahm and Thor[3] examined that the conferences have higher citations impact as compared to journals. Same researchers further concluded that conferences also have higher impact factors than that of journals.[17] On the contrary, Franceschet[18] found that conferences publish research articles on recent and hot topics than those of Journal publications, yet the journal publications received higher citations. Vrettas and Sanderson[19] mentioned the reasons behind such contradictory findings. Authors attributed these contradictions to the use of different datasets by the researchers in their study.

Researchers have also compared publication venues on the basis of impact factor and concluded that A* journals gained more citations as compared to their counterpart conferences, however many of the A grade conferences performed equally well as A grade journals.[5] Authors concluded such findings on behalf of two popular disciplines of CS. Vrettas and Sanderson[19] also compared conference and journals however they used average citation count for this purpose. They found that A* conferences have significantly higher average citations count than that of similar journals. However, such difference is not statistically significant for Grade A journals and conferences. They further observed that Grade B and C journals have more citation impact than conferences of these categories.

In addition, Kim[20] differentiated the conferences and journals of CS based on publishing pattern of authors. Researcher concluded that authors publish more papers in conferences than journals. Moreover, they found that authors collaboration is more in conferences with respect to journals. Overall, the studies conducted so far have drawn several conflicting claims. Although, we do not aim to discard these studies completely, on the contrary, we argue that their findings are sub-field specific and may not be generalized across all of the sub-fields of CS domain.

Motivation and Study Parameters

Researchers have considered several parameters e.g., review process, impact factor, citations, etc. to differentiate both publication venues of CS. Studies show that views are conflicting and supremacy of CS publication venues based on such factors, may not be determined unambiguously.

We assert that citations based metrics e.g., impact factor is domain-specific. The impact factor largely depends on the sub-fields of CS research. We compiled impact factors of some arbitrarily selected CS Journals of Quality I (Q1) ranking in Table 1. From Table 1 it can be observed that impact factor varies widely across the CS sub-fields. It can be seen that most of the journals in core CS fields, like Algorithms and theoretical aspects of CS, have very low impact factor sometimes lower than 1.0. Needless to say, that such journals are on top of the quality scale of CS research. It can also be observed from Table 1, that impact factor of top-journals of CS sub-fields, e.g., Programming Languages, Architectures etc. is quite low, while for other CS sub-fields like, Neuro-Computing, Evolutionary Computing, Image Processing etc., it is sufficiently high. This variation in impact factors is due to the nature and readership of the sub-fields. The main point, we are attempting to make, is that impact factors do not necessarily attribute to the quality. Though, quality of journals within a sub-field can be approximated by their impact factors up to some extent, however, this may not be generalized across all the sub-fields of CS research.

Ioannidis[21] have also argued that citations based metrics are misleading at times. The journals with low impact factors do not necessarily tend to of low quality. In addition, citations based findings of several researchers have contradictory claims.[3,5,17,18,21] Thus, we cannot generalize impact of articles from conferences and journals based on such metrics. In this study, we do not consider bibliometrics. Moreover, this study is not aimed at finding supremacy of CS publication venues. Rather, we wish to analyze other characteristics of both the venues of CS research.

In some of the previous studies, authors have mentioned some influencing factors that play a key role in paper selection process such as institutions' name and reputational standings of the authors.[9] Thus, in this work, we consider authors and their institutions and analyze multiplicity and overlaps in between the publication venues. In addition, we argue that most of the bibliometric studies of CS publication venues are limited to a particular sub-area of CS research.[3,5,17,18] Therefore, it is essential to conduct a combined study of different CS sub-domain. Hence, in this paper, we include publication venues of three sub-domains, namely, Theoretical Computer Science, Programming Language and Machine Learning. We consider publication profile data of a few leading conferences and journals for each of the selected sub-domains. Additionally, it is believed usually that a group of authors publish their papers in particular conferences and journals. Once a paper has been accepted in any of these venues, author(s) become more comfortable and understand the trends of such publication venues. In such scenario, author(s) start submitting papers frequently in such venues. To assess these type of patterns, this study also explores the frequency of such authors who have been publishing repeatedly in particular venues.

Thus, we include articles' authors and their affiliations in both the venues. Based on these primitives, we define following parameters in our study:

- The number of research institutions and distinct institutions to whom publishing authors belong to. We count these parameters for both the venues per year basis during the five year period of our study, i.e., 2012-2016.

- Authors' participation frequency (number of repeating authors with respect to total number of authors) per year during the research duration.

- Publication profile of conferences and journals with respect to overlap and mutual exclusiveness (in terms of overlapping authors and institutions in consecutive years and across the venues).

Based on these parameters, we attempt to explore the patterns of conferences and journals and try to differentiate both the venues.

MATERIALS AND METHODS
Data Sources and Processing

This study is conducted based on data (authors' names and their affiliations) of published articles of three conferences and three journals in the duration of 2012–16. Table 2 shows the

Table 1: CS journals and corresponding impact factor.	
Journals	**Impact Factor**
J. Logic Computation (Oxford)	0.509
ACM Trans. Algorithms	0.400 (2013)
Theoretical Computer Sc. (Elsevier)	0.718
J. Functional Prog. (Cambridge)	0.837
J. Applied Logic (Elsevier)	1.13 (2017)
ACM Trans. Database Sys. (TODS)	1.517 (2017)
J. Very Large Databases (Springer)	1.973
J. AI Review (AAAI Press)	2.284
J. ACM	2.353 (2011)
Machine Learning J. (Springer)	2.809
ACM Trans. Prog. Lang. Sys. (TOPLAS)	3.033 (2017)
IEEE Trans. Computers	3.131
Evolutionary Comp. J. (MIT Press)	3.469
Neurocomputing (Elsevier)	4.072
Neural Computing Appl. (Springer)	4.664
Neural Network (Elsevier)	5.785
IEEE Trans. Image Processing	6.79
IEEE Trans. Evolutionary Comp.	8.508
IEEE Trans. Neural Network Learning Sys.	11.683

Table 2: CS Sub-Fields and the corresponding Conferences and Journals taken for study.

Area	Publication Venue
Journals		
Theoretical Computer Science	Journal of the ACM (JACM)
Programming Language	ACM Trans. Prog. Lang. Sys. (TOPLAS)
Machine Learning	Machine Learning Journal (ML)
Conferences		
Theoretical Computer Science	ACM Symp. Theory of Computing (STOC)
Programming Language	ACM Symp. Principles of Prog. Lang. (POPL)
Machine Learning	Int. Conf. Machine Learning (ICML)

list of sub-fields of CS that we have considered. Conferences and journals are taken from three different CS sub-fields. Selected conferences are ICML[2], POPL[3] and STOC[4] and selected journals are ML[5], TOPLAS[6] and JACM[7]. These venues are considered as seed publication venues of CS in these chosen sub-fields.

We have crawled data of these venues from their respective websites. However, in some cases the information was not available, so we have collected such information from publicly available data sources, e.g., Google Scholar, DBLP and ACM Digital Library. We have covered a period of five years for each venue. For authors belonging to more than one institute, we have included both the institutions. However, we collated information of an institution into one location which works in multiple locations.

Data Processing: We first retrieved the year–wise list of published articles of each venue during the period from 2012 to 2016. Next, we manually crawled author name (s) of each paper and their research institutes. Based on this information we created year–wise dataset of each conference and each journal. Finally, we categorized these datasets into two subsets based on: (i) Year–wise authors' institutions and their frequency distribution and (ii) Year–wise authors and their frequency distribution. This procedure is repeated for journal publications as well. (Since some authors are associated with more than one institutions so number of authors and number of their institutions may differ in some cases.)

Methodology

In this section, we define and discuss metrics that we used for analyzing our data.

2 https://icml.cc
3 https://www.sigplan.org/Conferences/POPL/
4 http://acm-stoc.org/
5 https://link.springer.com/journal/volumesAndIssues/10994
6 https://toplas.acm.org/
7 https://dl.acm.org/journal/jacm/about

Ranking methods: It is observed that authors belonging to some institutions repeatedly publish their papers in specific conferences and journals. To depict this trend, we mine top 10 and top 5 institutions based on their frequency of occurrence. We consider two types of ranking: Ordinal and Dense.[23]

Overlapping: Overlap metrics are designed to measure similarity and dissimilarity between two or more than two datasets. Such metrics help to determine the patterns in between two different datasets. Overlapping simply specify the entities that appear in the list of published papers in consecutive years. This denotes the proportion of repetitive authors and institutions. This suggests that some institutions and authors prefer some particular publication venues. Since Jaccard method is used to find overlapping percentage of two datasets, hence, we used Jaccard similarity[24] to compute the overlap of top ten and top five institutions in consecutive years. Further, for showing the relative overlap of authors in consecutive years, we use relative overlap method.

Relative overlap: Relative overlap[25] is used to show relative closeness of two datasets. For example, consider two datasets D1 and D2. Relative overlap here specifie show close D2 is with respect to D1. Hence, we use relative overlap to specify the active participation of authors in each consecutive year.

$$\text{Percentage overlap in D1} = 100 \times \frac{|D1 \cap D2|}{|D1|} \tag{1}$$

$$\text{Percentage overlap in D2} = 100 \times \frac{|D1 \cap D2|}{|D2|} \tag{2}$$

Proposed Metrics

Percentage of Distinct institutions (PDIs): The institution that appears without repetition is considered as a distinct institution. Distinct institutions percentage is calculated mathematically as:

$$\text{PDIs} = 100 \times \frac{\text{DIs}}{\text{TIs}} \tag{3}$$

Where:

DIs: distinct institutions;

TIs: total institutions.

Total Frequency of Repeating Authors (*TFRA*): In a particular year, if any author is associated with more than one paper then it is taken as a repeated author. We have shown a few examples of such metric in the Table 3. In a particular year, frequency of repeating authors is calculated as:

$$\text{TFRA} = \sum_{i \geq 1, \ j > 1}^{n} ij \tag{4}$$

Where:

i: denotes number of authors;

j: denotes repeating times.

Table 3: Sample examples.

i	j	TFRA
1	1	No repetition
1	2	2(one author is repeating two times)
2,3……,30…..	2, 3…., 20..	4(two authors are repeating two times), 6(two authors are repeating three times or vice versa)…..

Percentage of Frequency of Repeating Authors (PFRA): We calculate the percentage of frequency of repeating authors as follows:

$$PFRA = 100 \times \frac{TFRA}{TA} \qquad (5)$$

Where:

    TFRA: total frequency of repeating authors;

    TA: total authors.

Overlap Percentage (OP): We calculate the overlap percentage of entities (top institutions or authors) with the help of Jaccard method[24], we define overlap percentage as follows:

$$OPI = 100 \times \frac{NCEDD}{TNEBD} \qquad (6)$$

Where:

    NCEDD: number of common entities into different datasets;

    TNEBD: total number of entities in both datasets.

We explore publication patterns of research articles of conference and journal publications of CS with respect to the above defined metrics and measures.

RESULTS AND DISCUSSION

In this section, we explore publication patterns of three selected conferences and three journals of CS in the period of 2012-2016. The first sub-section explores research institutions in each year in these conferences and journals. The second sub-section analyzes repetitions of authors each year in these conferences and journals. Third sub-section analyzes overlapping patterns in the datasets of institutions and authors for both of the venues.

Research institutions in conferences and journals

Collected institutional data is described in two categories: total institutions and distinct institutions. Percentage of distinct institutions can be seen in Table 4.

Distinct Institutions in Conferences: In STOC conference, every year, the number of institutions is in the range of 235-288, out of which distinct institutions vary in the range 76-90. From

Table 4a, we can infer that 60–62% authors participated from repeating institutions. Based on the aggregated distribution of five years, we find that total participation is 1286, out of which only 205 institutions are distinct. This means distinct institutions are only 15% with respect to the overall participation.

In POPL conference, number of institutions is in the range 151-226 each year and distinct institutions lie in the range 58-82. Table 4b shows percentage of distinct institution. Results indicate that 61-64% authors have participated from repeating institutions in each particular year. However, aggregated participation of institutions in 5 years is 904, out of which only 194 are distinct institutions. This implies that total percentage of distinct institutions is only about 21%.

In ICML conference, the number of institutions lies between 642-1116 per year where only 154-221 institutions are identified as distinct. Table 4c suggests that 76-81% authors have participated from repeating institutions. We find a total of 4097number of participation from 614 distinct institutions.

Table 4: Percentage of distinct institutions in three conferences and journals.

(a) STOC Conference

Year	Distinct Institutions
2012	39%
2013	39%
2014	40%
2015	38%
2016	39%
TOTAL	15%

(b) POPL Conference

Year	Distinct Institutions
2012	38%
2013	39%
2014	37%
2015	38%
2016	36%
TOTAL	21%

(c) ICML Conference

Year	Distinct Institutions
2012	23%
2013	24%
2014	23%
2015	22%
2016	19%
TOTAL	14%

(d) JACM Journal

Year	Distinct Institutions
2012	60%
2013	58%
2014	55%
2015	46%
2016	56%
TOTAL	**33%**

(e) TOPLAS Journal

Year	Distinct Institutions
2012	55%
2013	50%
2014	40%
2015	47%
2016	59%
TOTAL	**36%**

(f) ML Journal

Year	Distinct Institutions
2012	45%
2013	50%
2014	46%
2015	51%
2016	48%
TOTAL	**36%**

Therefore, only 14% distinct institutions are found in total participation.

Distinct Institutions in Journals: In JACM journal, only 55–75 distinct institutions are found in a total of 91–138 of institutional participation each year. It can be concluded from Table 4d that a total of 40–54% authors have participated from repeating institutions. The aggregated distribution of 5 years suggests that there are only 201 distinct institutions (33%) from a total participation of 600.

In TOPLAS journal, we find a total of 642–740 institutional participation while only around 153–170 distinct institutions are found in the list. Table 4e shows that 41–60% authors have participated from repeating institutions. Further, in aggregated distribution, we observe 105 distinct institutions in a total of 289 institutional participations which is around 36%.

ML journal shows 77–104 distinct institutions in a total of 168–215participation per year. Table 4f implies 50–55% authors participated from repeating institutions. Distribution of five year data depicts 355 distinct institutions in a total of

979 participation of institutions. Thus, 36% of the overall distinct institutional participation is found.

Results suggest that in conferences, approximately 40% distinct institutions participated each year while it is about 60% in journals. Same trends can be seen from aggregated data which shows approximately 21% distinct institutions whereas they are 36% in journals. These findings indicate that journal papers are published by authors from a more diverse group of institutions. This shows higher diversity of journals than that of conferences in terms of participating institutions.

Frequency of Participating Authors in Conferences and Journals

Figure 1 lists the year-wise authors' participation frequency in three conferences and journals. We have shown authors' data into the following groups (i) Total authors and (ii) Frequency of repeating authors and their percentage. For calculating the frequency of repeating authors, we consider equation 4.

Participating Authors in Conferences: In STOC conference, the frequency of repeating authors is 13–29% in the chosen time span. In POPL conference, frequency of repeating authors during this period is in the range of 1–11%. Whereas, in ICML conference, this percentage is 16–24. This implies that POPL conference has lesser frequency of repeating authors as compared to other conferences.

Participating Authors in Journals: From Figure 1, we notice that in JACM journal, the frequency of repeating authors ranges from 1–12%. Whereas in TOPLAS, no repetition was found in years 2012, 2013 and 2015, there was negligible overlap in years 2014 and 2016. In ML maximum frequency of repeating authors is 15%.

Overall, the results show that the repeating frequency of authors is quite lesser in journals than that of conferences for chosen timespan.

Overlapping Patterns

As mentioned in Section 4, we analyze the overlap among datasets to study the trends of publications. We use Jaccard method to show the overlap in top ten and top five institutions[8] and compare the results. The authors' overlap is shown by using relative overlap in consecutive years. These results show diversity of institutions as well as authors who have participated in these conferences and journals.

Overlapping among Institutions in Conferences: With the help of equation 6, we calculate overlapping percentages for conferences which are represented in Figure 2. We find that the maximum and minimum overlapping is 50% and 34%,

---

8 With the help of two ranking methods (ordinal and Dense), we obtain these top 10 and top 5 institutions based on the distribution of data.
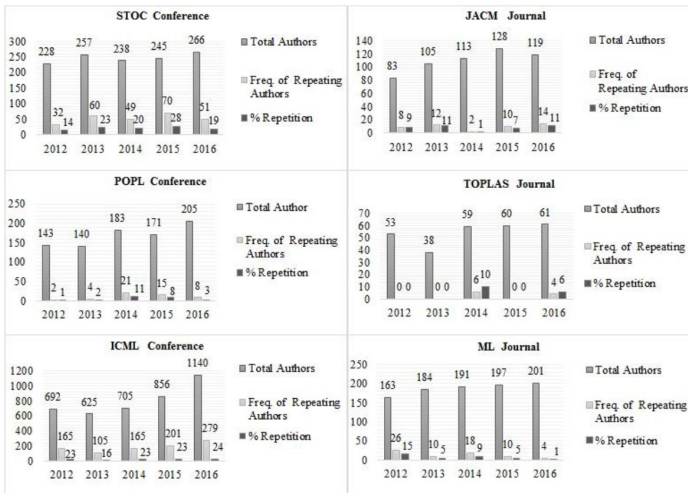
Figure 1: Participating author in conferences and journals: Total Authors, Total Frequency of Repeating Authors.
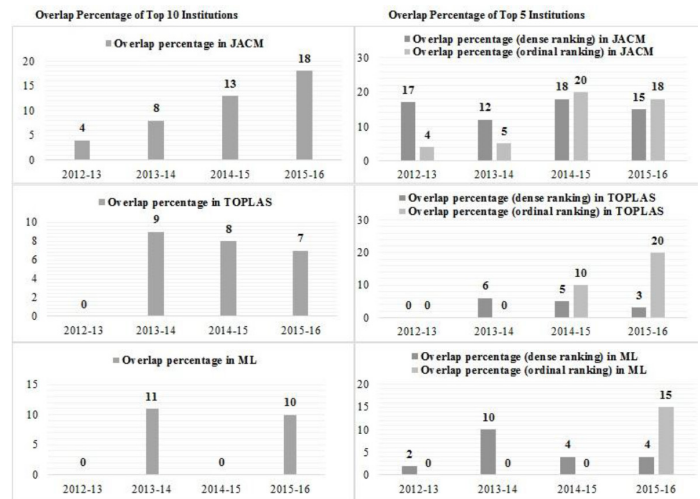

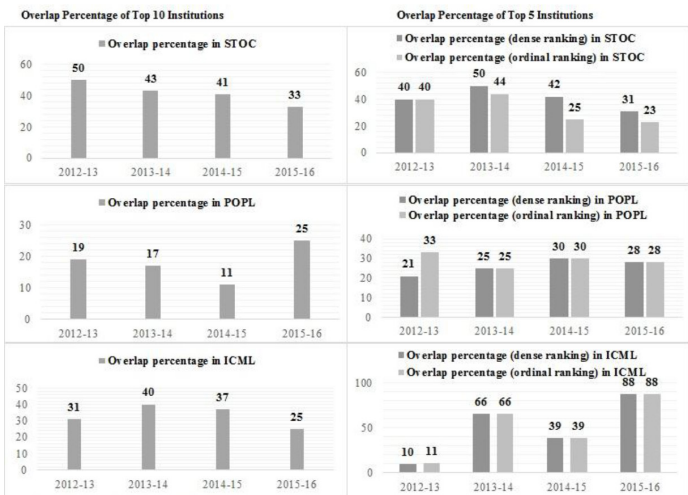**Figure 3:** Overlap Percentage of Top 10 and Top 5 Institutions in Three Journals.

Figure 2: Overlap Percentage of Top 10 and Top 5 Institutions in Three Conferences.
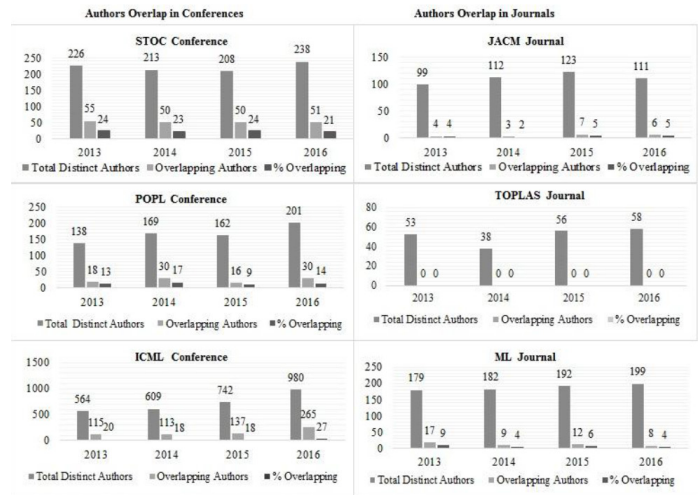

Figure 4: Overlapping Authors in Conferences and Journals in Consecutive Years.

respectively, in top 10 repetitive institutions of STOC conference. For top 5 repetitive institutions, overlapping percentage lies between 22-51. In POPL conference, the overlapping percentage decreases gradually except in the last years 2015 and 2016. The maximum overlap found is 25% for top 10 institutions. Further, overlap percentage, in top 5 institutions ranges from 21-33. In ICML conference, maximum overlapping found is 40% in the years 2013 and 2014 and minimum overlap is 25% in the years 2015 and 2016 for top 10 repeating institutions. It can be seen in Figure 2 that overlapping percentage is in the range of 10-88 in case of top 5 institutions.

Overlapping among Institutions in Journals: Figure 3 represents the overlap percentage of institutions in selected journals. Maximum overlapping found in JACM journal for top 10 repeating institutions is 18% and the minimum overlap is

4%. The overlap for top 5 institutions ranges from 4-20 in JACM. In TOPLAS journal, based on top 10 institutions, the maximum overlapping is obtained 9% in the years 2013 and 2014 and the minimum 7% in the years 2015 and 2016. We also notice that there is no overlapping in top 5 institutions in the years 2012 and 2013. The maximum overlap is 20% in top 5 institutions. In ML journal the maximum overlap is 11% for top 10 institutions while it is 15% in case of top 5 institutions.

Hence, above results state that overlapping in top 10 as well as top 5 repeating institution in journals is less than that of conferences.

Overlapping Authors in Conferences and Journals: Figure 4 contains the percentage of overlapping authors in consecutive years. We calculate the authors' overlap in each conference and each journal with the help of relative overlap method (equation 2

Table 5: Sub-field-wise overlap of authors and institutions across the conferences and journals.

Sub-field	Across conferences viz. vis journals (of same field)	Authors' Overlap (%)	Overlap among Top 10 Institutions (%)
Theoretical CS	STOC-JACM	14	38
Programming Language	POPL-TOPLAS	9	22
Machine Learning	ICML-ML	5	8

and 3). We present the results with respect to the previous year. In STOC conference, the maximum overlap of authors is 24%, while in POPL and ICML the maximum authors' overlap is 17% and 27% respectively. Whereas, in case of journals, we find that in JACM journal, have the maximum overlap that is 5% and it is 9% for ML. TOPLAS journal shows no such overlap. Thus, results infer that journals have less overlapping authors in consecutive years.

Overlapping Authors and Institutions across the Conferences and Journals of Same Sub-field: Table 5 shows the percentage of overlapping authors and institutions across the conferences and journals of the same sub-field with the help of equation 6. We have compared aggregated data of five years taken from each of the leading conferences and journals of the same sub-field. From Table 5, it can be observed that overlapping authors across the conference and journal range between 4-15%. This indicates that a very few groups of common authors have published in both, conferences as well as journals. Further, overlapping among top 10 institutions across the conferences and journals lies in the range 8-38%. Such overlapping patterns, thus, indicate that conferences and journals are mutually exclusive to some extent.

DISCUSSION

We assert that bibliometrics studies of CS publication venues are sub-field specific and prominence of publication venues depends on several issues. Similarly, review process and the use of low acceptance rate as quality measure are still the matters of discussion. Therefore, in this study we neither considered review based nor the impact based indicators, instead we considered several other key attributes of published articles to distinguish the publication patterns in CS domain.

We included the attributes such as authors and their corresponding institutions of published article for conferences as well as journals. Such data is not readily available, therefore, we manually crawled and created datasets for each venue. The main focus of our study is to assess the diversity of publication venues with respect to authors and their institutions. For this purpose, we have defined a few parameters based on previous studies and made certain assumptions. These parameters are

then used to carry out a comparative analysis of conferences and journals.

With respect to our parameter *"number of distinct institutions in conferences and journals"*, the results show that authors from some of the institutions have published more frequently over the years. From Table 4, we can infer that in conferences 19-40% authors are from distinct institutions in each year. However, in journals, this percentage is 40-60. Moreover, based on the aggregated data we found that in conferences, 14-21% authors are from distinct institutions, while in journals 33-37% are from distinct institutions. Therefore, we can infer that journals have more participation from distinct institutions than conferences which indicates journals are more diverse than conferences.

Based on the parameter *"authors' participation frequency in conferences and journals"*, it is observed that some authors publish more than one paper in a particular year, either individually or in collaboration. Authors' repetition pattern conveys that some prominent group of authors or senior researchers publish more frequently rather than newcomers or young researchers. Thus, we present the frequency of repeating authors in conferences and journals. From Figure 1, we can notice that the maximum repeating frequency is 28% in conferences, whereas it is only 15% are in journals. This implies that in conferences, the same group of authors publish more as compared to journals.

The parameter *"overlapping analysis of authors and institutions in consecutive years",* shows the overlap percentage of authors and institutions among top 10 and top 5 repeating institutions, which helps to understand the trends of publication venues in selected duration. From Figure 2, we observe that in conferences, overlapping is 11% to 50% among top 10 institutions in consecutive years, while in the top 5 institutions, it is 10-88%. However, in journals (Figure 3), overlapping among top 10 institutions is 0-18% and among top 5 institutions is 0-20%. Thus, journals have lesser overlapping percentage as compared to conferences. In addition, we have also examined the authors' overlapping patterns in consecutive years. From Figure 4, it can be observed that in conferences 9% to 27% authors are repeatedly publishing in consecutive years, whereas, for journals it is 2-9%. In TOPLAS journal, there is no overlap of authors during the selected period. Thus, journals have lesser overlapping percentage of authors than that of conferences. These observations conclude that trends in journal publications are more diverse than those of conferences with respect to publishing authors as well as institutions. Finally, based on sub-field-wise comparison across the conferences and journals (Table 5), we deduce that conferences and journals of CS publications are mostly mutually exclusive.

CONCLUSION

This study focuses on comparative analysis of publication patterns in the conferences and journals with respect to article authors and their affiliations. Since the scientometrics data vary widely across sub-fields of CS research, we have considered three distinct sub-fields. Then, we have taken one leading publication venue from each of the conferences and journals from three sub-fields.

We have explored the trends of publishing authors and their affiliations. Based on these trends, our main findings are: (*i*) Journal publications show more diversity than those of conferences in terms of distinct institutions; (*ii*) Authors' repetition frequency is lesser in journals than of conferences; (*iii*) Overlapping percentage in top 10 and top 5 repeating institution is comparatively lower in journals than in conferences; (*iv*) Overlapping percentage of authors in consecutive year is also lesser in journals as compared to conferences; and (*v*) Journal and conference authors are mostly mutually exclusive.

However, we do not wish to generalize these findings, as our study is constrained by three sub-domains of CS within limited period. Yet, we may hope that the same holds good over the years as well as for other sub-fields of CS. We emphasize that this study was not undertaken for finding the prominence of CS publication venues.

We further hope that findings of this study may provide valuable insights to the researchers and motivate them to study other related characteristics and distinctions of CS publications. This study could be extended to utilize voluminous sub-field specific CS publication data and to extracting variants of CS publication patterns so that larger research questions can be addressed in future.

ACKNOWLEDGEMENT

REFERENCES

1. Meyer B, Choppy C, Staunstrup J, Leeuwen JV. Research evaluation for computer science. Communications of the ACM. 2009;52(4):31-4.
2. Franceschet M. The skewness of computer science. Information Processing and Management. 2011;47(1):117-24.
3. Rahm E, Thor A. Citation analysis of database publications. ACM Sigmod Record. 2005;34(4):48-53.
4. Fortnow L. Time for computer science to grow up. Communications of the ACM. 2009;52(8):33-5.
5. Freyne J, Coyle L, Smyth B, Cunningham P. Relative status of journal and conference publications in computer science. Communications of the ACM. 2010;53(11):124-32.
6. Birman K, Schneider FB. Viewpoint Program committee overload in systems. Communications of the ACM. 2009;52(5):34-7.
7. Patterson DA. The health of research conferences and the dearth of big idea papers. Communications of the ACM. 2004;47(12):23-4.
8. Cabot J, Izquierdo JL, Cosentino V. Are CS conferences (too) closed communities?. Communications of the ACM. 2018;61(10):32-4.
9. Parhami B. Low Acceptance Rates of Conference Papers Considered Harmful. Computer. 2016;49(4):70-3.
10. Anderson T. Conference reviewing considered harmful. ACM SIGOPS Operating Systems Review. 2009;43(2):108-16.
11. Wallach DS. Rebooting the CS publication process. Communications of the ACM. 2011;54(10):32-5.
12. Korth HF, Bernstein PA, Fernandez M, Gruenwald L, Kolaitis PG, McKinley K, Ozsu T. Paper and proposal reviews: is the process flawed?. ACM Sigmod Record. 2008;37(3):36-9.
13. Vardi MY. Conferences vs. journals in computing research. Communications of the ACM. 2009;52(5):5-5.
14. Shamir L. The effect of conference proceedings on the scholarly communication in Computer Science and Engineering. Scholarly and Research Communication. 2010;1(2).
15. Halpern JY, Parkes DC. Journals for certification, conferences for rapid dissemination. Communications of the ACM. 2011;54(8):36-8.
16. Terry D. Publish now, judge later. Communications of the ACM. 2014;57(1):44-6.
17. Rahm E. Comparing the scientific impact of conference and journal publications in computer science. Information Services and Use. 2008;28(2):127-8.
18. Franceschet M. The role of conference publications in CS. Communications of the ACM. 2010;53(12):129-32.
19. Vrettas G, Sanderson M. Conferences versus journals in computer science. Journal of the Association for Information Science and Technology. 2015;66(12):2674-84.
20. Kim J. Author-based analysis of conference versus journal publication in computer science. Journal of the Association for Information Science and Technology. 2019;70(1):71-82.
21. Ioannidis JP, Baas J, Klavans R, Boyack KW. A standardized citation metrics author database annotated for scientific field. PLoS Biology. 2019;17(8).
22. Wainer J, Eckmann M, Goldenstein S, Rocha A. How productivity and impact differ across computer science subareas. Communications of the ACM. 2013;56(8):67-73.
23. Cichosz P. Data mining algorithms: Explained using R. John Wiley and Sons. 2014.
24. Jaccard P. The distribution of the flora in the alpine zone. New Phytologist. 1912;11(2):37-50.
25. Gluck M. A review of journal coverage overlap with an extension to the definition of overlap. Journal of the American Society for Information Science. 1990;41(1):43-60.