A Bibliometric Study of ACM SIGSOFT Software Engineering Notes from 2007 to 2016

Lov Kumar
NIT Rourkela (India)
lovkumar505@gmail.com

Ranjan Kumar Behera
NIT Rourkela (India)
jranjanb.19@gmail.com

Santanu Rath
NIT Rourkela (India)
skrath@nitrkl.ac.in

Ashish Sureka
ABB (India)
ashish.sureka@in.abb.com

DOI: 10.1145/3127360.3127369
http://doi.acm.org/10.1145/3127360.3127369

ABSTRACT

Bibliometric analysis is a commonly used technique to analyze scholarly publications to extract useful insights about research and scientific papers which can then be used for decision making by policy makers and administrators. Bibliometric analysis helps in understanding various aspects of scientific knowledge creation and dissemination such as author and institute productivity, impact of articles in terms of citations, university and industry collaboration, geographical contributions and ethnic and gender minority in authorship. ACM SIGSOFT Software Engineering Notes (SEN) is a non-refereed but a reputed and edited publication for informal writings and reports about Software Engineering (SE). ACM SIGSOFT SEN publishes various types of submissions such as paper, report, column, announcement and book review. These submissions are published in the ACM Digital Library (DL). We conduct a bibliometric analysis of articles published in ACM SIGSOFT SEN during a ten year period from 2007 to 2016. Our objective is to provide a historical overview (one decade) of ACM SIGSOFT SEN and reflect on the past so that the ACM SIGSOFT community and contributors can assess the strengths and shortcomings of the SEN. We believe that the bibliometric analysis presented in this paper can provide insights on the extent to which the SEN is meeting its desired objectives.

Keywords

ACM SIGSOFT SEN, Bibliometric Analysis, Software Engineering, Scientific Publication Mining, University Industry Collaboration, Citation Analysis

1. INTRODUCTION

ACM SIGSOFT Software Engineering Notes (SEN)[1] is an edited publication series (but not refereed) consisting of reports and informal writing on Software Engineering (SE) topics. ACM SIGSOFT SEN is being published in ACM Digital Library[2] since May 1976. The Volume 1 and Issue 1 was published in May 1976 and the January 2017 edition is Volume 42 Issue 1. There are several issues of ACM SIGSOFT SEN published in a year. ACM SIGSOFT SEN accepts several types of articles such as paper, report, column, announcement and book review. The number of years of existence of ACM SIGSOFT SEN, the number of contributing authors and articles getting published shows that the SEN is a widely disseminated and read publication. Our objective is to conduct a quantitative and historical analysis of last 10 years of ACM SIGSOFT SEN through a bibliometric study. A similar kind of analysis has been done by Agarwal et al. in the past for ACM SIGWEB Newsletter which is similar to ACM SIGSOFT SEN [3][4]. The bibliometric analysis on ACM SIGOSFT SEN will help the community in understanding how the SEN has evolved and what is the current status and performance. We conduct an analysis from the perspective of the following nine dimensions:

1. Author turnover
2. Average number of authors per paper
3. University and industry collaboration
4. Gender and authorship
5. Author affiliation - university or industry
6. Internal and external collaboration
7. Citation and impact metrics
8. Prolific authors
9. Geographical origin of publications

2. AUTHOR TURNOVER

Vasilescu et al. define author turnover as an indicator of a conference stability [17]. They mention that there is normally a stable group of core researchers who contribute towards several editions of a conferences and a large percentage of new authors contribute with respect to the previous editions of the conferences [17]. They conduct an analysis of several software engineering conferences and their findings reveal that the author turnover with respect to the previous four editions of the conference is about 50% [17]. Agarwal et al. conduct a similar analysis as Vasilescu et al on ACM SIGWEB sponsored conferences [5]. Their analysis reveals that conferences like CIKM and HT are highly dynamic and have a good author turnover [5].

Table 1: Total Number of Unique Authors (UNAUTH), Number of new authors (NUMNEW) and Percentage of New Authors Added Every Year (PERNEW)

Year	UNAUTH	NUMNEW	PERNEW
2007	118	0	0.00
2008	110	28	25.45
2009	204	45	22.06
2010	203	49	24.14
2011	244	62	25.41
2012	291	72	24.74
2013	278	80	28.78
2014	236	93	39.41
2015	260	80	30.77
2016	100	44	44.00

Table 1 displays the number of unique authors represented by

[1] https://www.sigsoft.org/SEN/
[2] http://dl.acm.org/citation.cfm?id=J728

UNAUTH, number of new authors represented by NUMNEW and the percentage of new authors with respect to previous editions of the conference in our dataset represented by PERNEW for the ACM SIGSOFT SEN for the year 2007 to 2016. Our analysis reveals that the author turnover (in-terms of percentage of new authors) varies from a minimum of 22.06% in the year 2009 to a maximum of 44% in the year 2016. Table 1 shows that the percentage of new authors with respect to the previous editions in our dataset is always above 20%. The number of unique authors varies from a minimum of 100 in the year 2016 to a maximum of 291 is the year 2012. Table 1 reveals that except for three years (2007, 2008 and 2016), the number of unique authors is always above 200.

Table 2: Statistics on Number of Authors Per Paper

Year	Min.	Max.	Mean	Median
2007	1	8	1.85	1
2008	1	7	1.97	1
2009	1	9	2.05	2
2010	1	6	2.02	2
2011	1	7	1.95	2
2012	1	8	2.14	2
2013	1	9	2.22	2
2014	1	9	2.32	2
2015	1	12	2.41	2
2016	1	7	2.12	2

3. NUMBER OF AUTHORS PER PAPER

Amin et al. mention that there is variance with respect to the average number of authors per paper across subject or topic areas [7]. For example, the number of collaborators per paper in fundamental life sciences is more than the number of authors per paper in social sciences [7]. Fundamental life sciences has more than four authors per paper and in social sciences there are about two authors per paper. Researchers study number of authors per paper to answer several research questions such as the correlation between the average impact factor and the average number of authors per paper [7]. Elmacioglu et al. conduct an analysis of 19 journals and 81 conferences closely related to the database community from the year 1968 to 2003 [10]. Their hypothesis is that there is a publish or perish pressure which sometimes drives researchers to form collaborations and increase their publication count. Their findings reveal that there is an increasing trend towards the average number of authors per paper and reached 2.8 co-authors per paper in the year 2003 [10]. A study by Sharma et al. on research papers published in the Requirements Engineering (RE) conference reveals that 7.07% of the articles are solo-authored [14]. Their findings shows that 27.04%, 30.30% and 19.41% articles are co-authored by two, three and four authors respectively [14].

Table 2 shows the statistics on the number of authors per paper. Table 2 displays the minimum, maximum, mean and median authors per paper from the year 2007 to 2016. We observe that the average number of authors per paper for ACM SIGSOFT SEN increased from the year 2011 to 2015 and then declined in the year 2016. We observe that during the 2007 to 2016 period, the average number of authors per paper varies from 1.85 to 2.41. The maximum values ranging from 6 to 12 shows that there are large collaborations but the median value of 1 to 2 shows that the middle value in the list is 1 or 2 which means that for at-least 50% of the papers the average number of authors per paper is less than the respective median value.

Table 3: Number and Percentage of Papers having All Authors from University (AU), All Authors from Industry (AI) and Authors from both University and Industry (UI)

Year	NUM	AU	AI	UI
2007	88	64 (72.73%)	19 (21.59%)	5 (5.68%)
2008	70	47 (67.14%)	19 (27.14%)	4 (5.71%)
2009	133	99 (74.44%)	24 (18.05%)	10 (7.52%)
2010	129	83 (64.34%)	34 (26.36%)	12 (9.30%)
2011	166	112 (67.47%)	50 (30.12%)	4 (2.41%)
2012	161	114 (70.81%)	34 (21.12%)	13 (8.07%)
2013	157	106 (67.52%)	34 (21.66%)	17 (10.83%)
2014	126	85 (67.46%)	25 (19.84%)	16 (12.70%)
2015	120	80 (66.67%)	25 (20.83%)	15 (12.50%)
2016	60	38 (63.33%)	14 (23.33%)	8 (13.33%)

4. UNIVERSITY AND INDUSTRY COLLABORATION

Analyzing university and industry scientific collaboration through co-authorship in scholarly publications is a common technique to study interaction and technology transfer between the two types of organizations. The collaboration between academia and industry is important for a healthy research ecosystem and innovation. Butcher et al. conduct a bibliometric analysis of scientific publications in the field of membrane use for water treatment [8]. They study 8 major journals during the year 1967 to 2001 and found that the field is highly collaborative [8]. Abramo et al. conduct a bibliometric examination to study university and industry collaboration in Italy [1]. They study co-authorship in international journals and one of their findings that researchers who collaborate (university and industry collaboration) demonstrate a higher performance than researchers who do not collaborate between the two types of organizations [1]. Sharma et al. extract data about author affiliation from 205 scientific papers published at the ISEC (India Software Engineering Conference) over a period of 9 years and observed that majority of the papers are from authors affiliated to University [15]. Their research reveals that there is on noticeable upward or downward trend with regards to university and industry collaboration [15].

Table 3 shows the number and percentage of papers having all authors from university represented by AU, all authors from industry represented by AI and authors from both university and industry represented by UI. Table 3 shows that the university and industry collaboration or the articles having co-authors from both university and industry varies from a minimum of 2.41% to a maximum of 13.33%. Our analysis reveals that the percentage of papers having co-authors from both university and industry is less than 10% from the year 2007 to 2012. The percentage of papers having co-authors from both university and industry is above 10% from the year 2013 to 2016. Table 3 shows that the percentage of papers having all authors from university varies from a minimum of 63.33% in the year 2016 to a maximum of 74.44% in the year 2009. Table 3 shows that the percentage of papers in which all the authors are from university is above 60% across all the 10 years. The percentage of papers for which all the authors are from industry vary from a minimum of 18.05% in the year 2009 to a maximum of 30.12% in the year 2011. Figure 1 shows the stacked bar chart for the data in Table 3.

5. GENDER AND AUTHORSHIP

Agarwal et al. conduct a study on women in computer science research [2] and analyze thousands of publications in computer

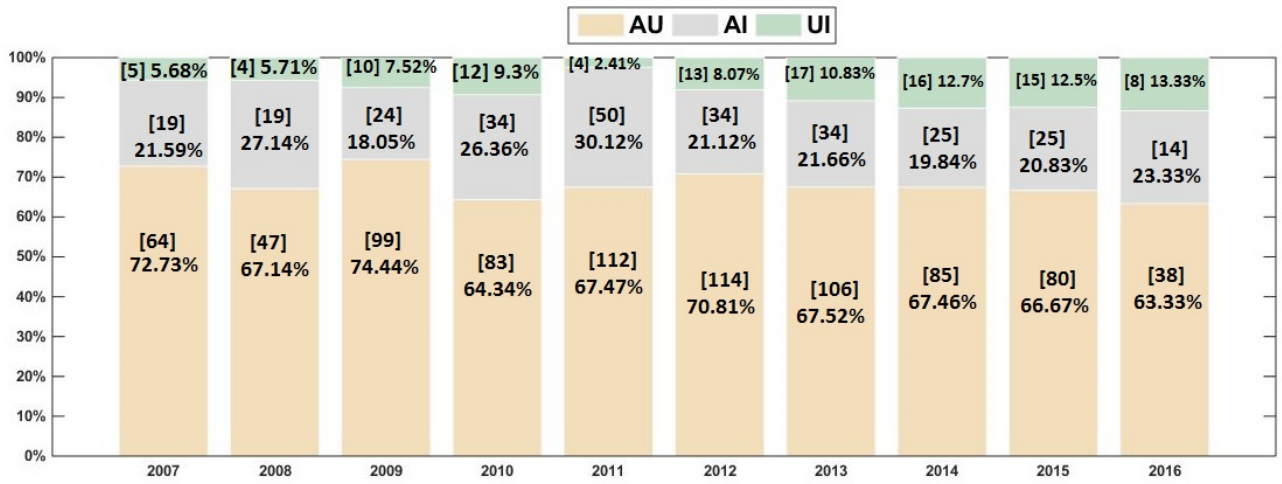

Figure 1: Stacked Bar Chart for the Data in Table 3

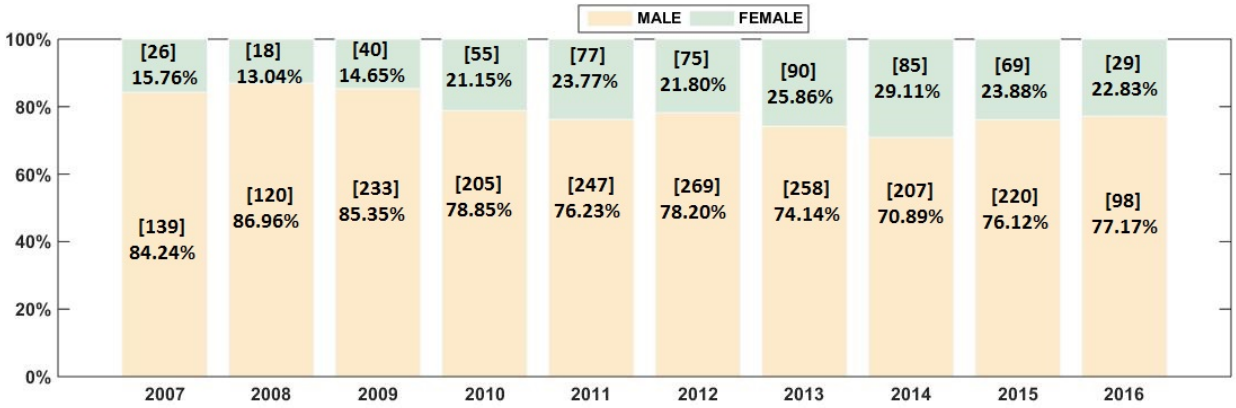

Figure 2: Percentage of Male and Female Authors (Gender Imbalance)

science as well as sub-fields within computer science such as data engineering, theory and software engineering. Their analysis of several years of publications revealed that in computer science, 79% of the authors were male and 21% were female [2]. We use the Genderize.io[3] API to determine the gender of a given name. Genderize.io API is free and as of 19 April 2017 contains 216286 distinct names across 79 countries and 89 languages in its database.

Figure 2 displays a stacked bar chart showing the distribution of male and female authors from the year 2007 to 2016. We extract all the co-author names for each paper and then compute the gender using the genderize.io API. The graph in Figure 2 reveals that there is a gender imbalance in authorship. The percentage of female authors varies from a minimum of 13.04% in the year 2008 to a maximum of 29.11% in the year 2014. We do not observe any consistently increasing or decreasing trend in-terms of percentage of women authors. Our analysis reveals that the percentage of female authors is above the 25% mark for only 2 out of 10 years. The percentage of female authors showed a steady increase from 21.80% to 29.11% from the year 2012 to 2014 but then declined subsequently in year 2015 and 2016.

[3]`https://genderize.io/`

6. AUTHOR AFFILIATION - UNIVERSITY OR INDUSTRY

We extract the affiliation of each author and classify it into one of the two classes: University or Industry. We create only two classes and any organizations which is a non-profit research labs but not teaching universities are classified under industry. Figure 3 shows a stacked bar chart showing the percentage of authors from university and industry for the ten year period from 2007 to 2016. The graph reveals that the percentage of industry authors varies from a minimum of 13.19% in the year 2013 to a maximum of 23.19% in the year 2008. Our analysis shows that the percentage of industry authors have always been below 25%. We do not observe any consistently or steady increasing or decreasing trend and the percentage value of industry authors fluctuates within a band of 13.19% and 23.19%. However, in comparison to the year 2015, there is a slight increase in the percentage of industry authors in the year 2016.

7. INTERNAL AND EXTERNAL COLLABORATION

Collaborations between co-authors in a paper can be categorized into multiple types. One type of collaboration is university and industry collaboration in which co-authors from a paper are from both university and industry. Another type of collaboration is internal or external collaboration. External collaboration is cross-

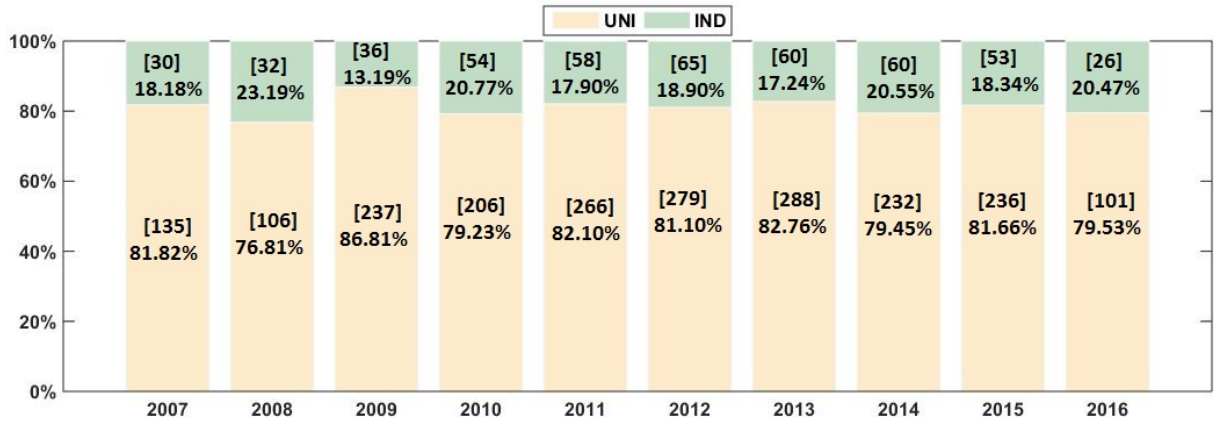

Figure 3: Percentage Distribution of authors across Industry and University

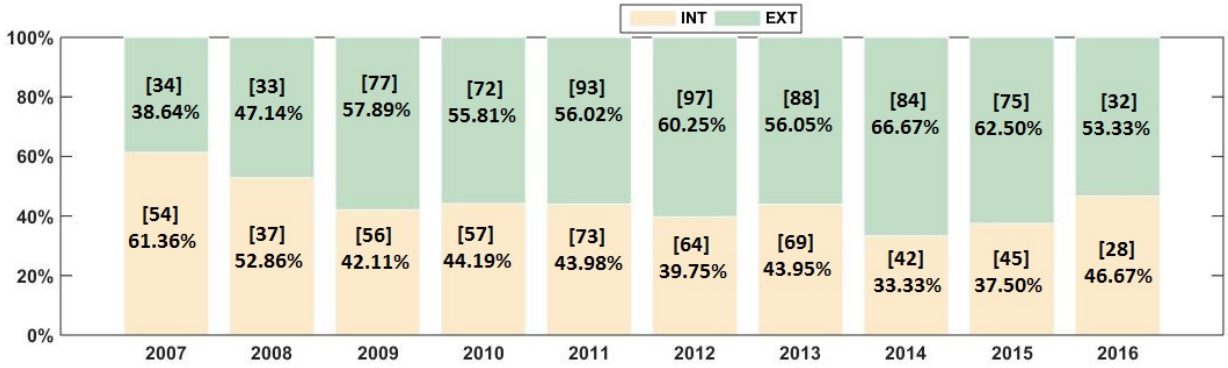

Figure 4: Percentage Distribution of Authors across Internal and External Collaboration

organizational or cross-institutional collaboration irrespective of whether the two collaborating organizations are industry or university. Internal collaboration is a type of collaboration in which all the contributing co-authors are from a single institution (industry or university). Our objective is to measure statistical indicators for external collaboration as external collaboration involves participation of people from multiple institutions and is good for the ecosystem. Figure 4 shows a stacked bar chart displaying the internal and external collaboration for the 10 year period from 2007 to 2016. We observe that the external collaboration varies from a minimum of 38.64% in the year 2007 to a maximum of 66.67% in the year 2014. We observe that the external collaboration is quite common as the external collaboration percentage of more than 50% is found in 8 out of 10 years. Figure 4 reveals an increasing trend towards external collaboration as in general the percentage of external collaboration for the years 2014 to 2016 is higher in comparison to the years 2007 and 2008.

8. CITATION AND IMPACT METRICS

Google scholar[4] is a freely accessible web application which is widely used and a popular source of bibliometric data. Google scholar indexes a large number journals, conference proceedings, workshop papers, books, thesis and technical reports. Google scholar citations is an accepted metric to measure the scholarly impact of an author, article, journal or conference or institution [6][11]. We compute the Google scholar citations of all the papers published in ACM SIGSOFT SEN for the ten year period from 2007 to 2016. We compute the citation count on 23 April 2017.

[4]http://scholar.google.com/

Table 4: Descriptive Statistics for ACM 2007 TO 2016 Google Scholar Citations

	Min.	Max.	Mean	Med	Sum	Skew	Kurt
2007	0	103	9.03	2.5	795	3.20	14.67
2008	0	63	7.34	1	514	2.54	9.76
2009	0	48	6.01	2	799	2.13	7.81
2010	0	124	7.26	1	936	4.78	29.07
2011	0	73	5.86	2	972	3.48	18.60
2012	0	90	5.07	1	817	5.22	36.13
2013	0	40	3.54	0	556	3.16	13.56
2014	0	24	2.14	0	270	2.93	14.76
2015	0	16	1.70	0	204	2.57	9.88
2016	0	15	0.97	0	58	4.21	20.71

Table 4 shows the descriptive statistics for the citation count. We compute the minimum, maximum, mean, median, sum, skewness and kurtosis. Table 4 reveals that the mean and maximum number of citations for articles published in the year 2007 is 9.03 and 103 respectively. We observe that the mean number of citations decreases gradually from 2007 to 2016 except for the year 2010. This is expected as recently published article will acquire citations gradually. We infer that the articles published in the year 2010 have the best impact in-terms of number of years since existence and in-terms of the maximum and mean number of citations. The sum of the citations have not crossed the 1000 mark for any of the years from 2007 to 2016. The top 3 years in-terms of the sum of the citation count is 2010 (936), 2011 (972) and 2012 (817). We also measure the skewness and kurtosis for the citation count.

Table 5: Most Cited ACM 2007 TO 2016 Papers (Citation Count $\geq$ 50) based on Google Scholar Metrics (Citation Metrics Collected on 23 Aril 2017)

Rank	Year	Paper Title	First Author	Volume	Issue	Citations
1	2010	The overture initiative integrating tools for VDM.	Peter Gorm Larsen	35	1	124
2	2010	Software development lifecycle models.	Nayan B. Ruparelia	35	3	121
3	2007	A systematic review of software process tailoring.	Oscar Pedreira	32	3	103
4	2012	Testing android apps through symbolic execution.	Nariman Mirzaei	37	6	90
5	2012	Engineering a compiler, second edition by Keith D. Cooper and Linda Torczon.	Chris Hathhorn	37	1	88
6	2007	Architectural knowledge and rationale - issues, trends, challenges.	Paris Avgeriou	32	4	75
7	2011	Usability testing essentials - ready, set...test! by Carol M. Barnum.	William Del Ra	36	5	73
8	2007	Podcasts about software engineering.	Jorg Rech	32	2	67
9	2008	A gentle introduction to OSGi.	Andre Luiz Camargos Tavares	33	5	63
10	2011	TEPE - a SysML language for time-constrained property modeling and formal verification.	Daniel Knorreck	36	1	61
11	2008	Estimation of quality for software components - an empirical approach.	Arun Sharma	33	6	60
12	2010	Why do we need personality diversity in software engineering?	Luiz Fernando Capretz	35	2	58
13	2010	Test case prioritization using ant colony optimization.	Yogesh Singh	35	4	58
14	2010	Software cost estimation using fuzzy logic.	Anish Mittal	35	1	51

Table 6: List of Most Prolific Authors (# PR denotes Number of Papers)

		Citations Analysis				
Name	# PR	Min.	Max.	Mean	Med	Sum
Peter G. Neumann	59	0	0	0.00	0	0
Mark Doernhoefer	57	0	15	0.56	0	32
Robert Schaefer	32	0	8	0.84	0	27
Peter B. Henderson	30	0	19	0.63	0	19
Mark A. Ardis	23	0	19	0.87	0	20
Larry Bernstein	20	0	5	0.25	0	5
Michael Wing	18	0	2	0.22	0	4
Vasile G. Teodorovici	17	0	0	0.00	0	0
Alok Mishra	16	0	15	1.44	0	23
Alex Groce	16	0	0	0.00	0	0
Yogesh Singh	14	0	58	12.07	6	169
Deepti Mishra	13	0	15	1.77	0	23
Will Tracz	12	0	1	0.08	0	1
Hardeep Singh	12	0	20	5.83	4	70

Skewness measures lack of symmetric in the dataset. Kurtosis measures the variability in the dataset in-terms of whether it is light-tailed or heavy-tailed with respect to a normal distribution. We observe that the skewness varies from 2.13 to 5.22 and hence some of the distributions are moderately skewed towards the left whereas some are moderately skewed towards the right.

Table 5 displays the most cited papers having a citation count of more than 50. Table 5 shows the title of the paper, year of publication, volume, issue number and the citation count. Our analysis reveals that there are 14 articles having a citation count of more than 50. Table 5 shows that there are 3 articles with a citation count of more than 100. The publication year of all the articles in Table 5 varies from 2007 to 2012.

9. PROLIFIC AUTHORS

Bibliometric studies analyzing author productivity and frequency of publications is conducted by several researchers across various disciplines. Shabani et al. study most prolific authors in the behavior analysis discipline [13]. They identify most prolific authors as well as institutions from 1992 to 2001 [13]. Garg et al. conduct a bibliometric study in the laser science and technology field and discover that prolific authors produce 25% of the papers in the scientific journals in their dataset [12].

Table 6 shows the list of most prolific authors and the number of articles published by them from the year 2007 to 2016. There are several submission types of ACM SIGSOFT SEN. We do not differentiate between the submission types and consider all of them as an article. The different types of submission types are: paper, report, column, announcement and book review. There are several contributions called as risks to the public by edited by Peter G. Neumann. Similarly, there are several columns by Mark Doernhoefer called as Surfing the Net for Software Engineering Notes. There are more than 30 articles by Robert Schaefer on various topics.

Table 6 reveals that there are 14 authors who have contributed more than 10 articles. We also conduct a Google scholar citation analysis of the 14 most prolific authors. From Table 6, we infer that the highest citation impact is by Yogesh Sinha who has published 14 articles and received a total citation of 169. Hardeep Singh has published 12 articles during the 2007 to 2016 period and received 70 citations.

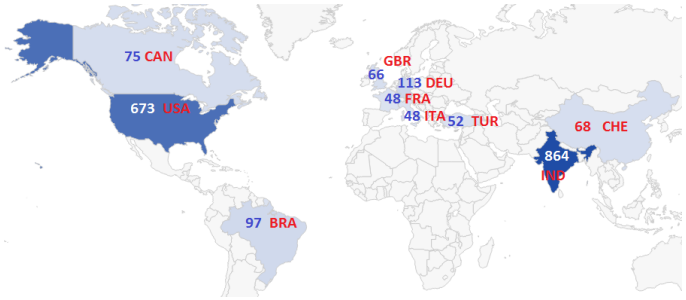

Figure 5: Top 10 Countries with the Highest Scholarly Output and Contributions

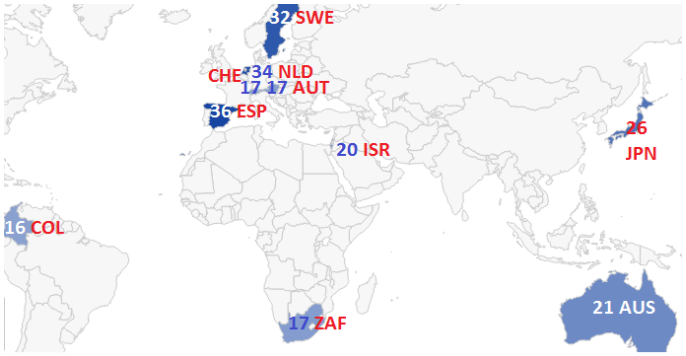

Figure 6: Top 11 to 20 Countries with the Highest Scholarly Output and Contributions

Table 7: Country of Author and Respective Frequency Data

2007		2008		2009		2010		2011	
Country	**Frequency (%)**	**Country**	**Frequency (%)**	**Country**	**Frequency (%)**	**Country**	**Frequency (%)**	**Country**	**Frequency (%)**
USA	46 (27.88%)	USA	42 (30.43%)	INDIA	107 (39.19%)	INDIA	101 (38.85%)	INDIA	125 (38.58%)
INDIA	26 (15.76%)	INDIA	34 (24.64%)	USA	70 (25.64%)	USA	64 (24.62%)	USA	80 (24.69%)
BRAZIL	18 (10.91%)	BRAZIL	11 (7.97%)	CHINA	12 (4.40%)	BRAZIL	13 (5.00%)	FRANCE	24 (7.41%)
GERMANY	15 (9.09%)	GERMANY	11 (7.97%)	IRAN	9 (3.30%)	UK	10 (3.85%)	GERMANY	15 (4.63%)
UK	14 (8.48%)	CHINA	8 (5.80%)	UK	8 (2.93%)	GERMANY	10 (3.85%)	CANADA	8 (2.47%)
2012		**2013**		**2014**		**2015**		**2016**	
Country	**Frequency (%)**	**Country**	**Frequency (%)**	**Country**	**Frequency (%)**	**Country**	**Frequency (%)**	**Country**	**Frequency (%)**
INDIA	126 (36.63%)	INDIA	133 (38.22%)	INDIA	117 (40.07%)	USA	77 (26.64%)	USA	37 (29.13%)
USA	91 (26.45%)	USA	82 (23.56%)	USA	84 (28.77%)	INDIA	66 (22.84%)	INDIA	29 (22.83%)
GERMANY	15 (4.36%)	BRAZIL	20 (5.75%)	ITALY	12 (4.11%)	GERMANY	15 (5.19%)	SWEDEN	7 (5.51%)
FRANCE	14 (4.07%)	CANADA	17 (4.89%)	GERMANY	10 (3.42%)	BRAZIL	12 (4.15%)	TURKEY	6 (4.72%)
CHINA	12 (3.49%)	ITALY	12 (3.45%)	SPAIN	9 (3.08%)	CHINA	10 (3.46%)	GERMANY	6 (4.72%)

10. GEOGRAPHICAL ORIGIN OF PUBLICATIONS

Analyzing geographic origin of scholarly publications is important to understand the worldwide scientific outcome of a particular field across countries or regions. Christopher et al. use bibliography data in the field of veterinary medicine to understand geographical differences in research outcome [9]. They study several journal data indexed in the Scopus database from the year 1996 - 2011 and found that the number of articles from authors from North America (USA and Canada) and Western Europe accounts for 60.9% of papers [9]. Similar studies are conducted in other fields also. For example, Van et al. study 18717 papers from 15 major surgical journals in the Medline or PubMed database over a period of six years from 2000 to 2005 [16]. Their study revealed that 88.8 of the output is generated by 15 countries [16]. Their findings revealed that USA, Japan and UK produced 42.1%, 9.1% and 7.6% of the output respectively [16].

We extract the country of affiliation for each co-author for every paper in our dataset. Figure 5 and 6 shows the list of top $1-10$ and top $11-20$ countries in-terms of their scholarly contribution in ACM SIGSOFT SEN from 2007 to 2016. The top 10 countries are: India (864), USA (673), Germany (113), Brazil (97), Canada (75), China (68), UK (66), Turkey (52), France (49) and Italy (48). The number in brackets denotes the number of authors. For example, the number of authors from India and USA are 864 and 673 respectively. The countries from rank $11-20$ are: Spain (36), Netherland (34), Sweden (32), Japan 26, Australia (21), Israel (20), Switzerland (17), Austria (17), South Africa (17) and Colombia (16). Our analysis reveals that authors belong to a total of 67 countries. The top 2 countries constitutes 30% of the authors. The top 5 countries constitutes 71% of the authors. We observe a long tail phenomenon and notice that there are 39 countries below 10 authors each.

Table 7 shows the country of authors who have published in ACM SIGSOFT SEN from 2007 to 2016 and their respective frequency. We observe that USA, Germany and India are in general amongst the top 5 across several years. We observe that the top 3 countries contribute more than 50% of the articles. There are significant percentage of papers from several other countries which are in top five in addition to USA, India and Germany: Brazil, UK, China, Iran, France, Canada, Italy, Spain, Turkey and Sweden.

11. CONCLUSION

Our analysis reveals a good author turnover. The percentage of new authors with respect to the previous editions is always above 20%. The number of unique authors are above 100 every year. Hence the contribution is from a large number of diverse people. The average number of authors per paper varies from 1.85 to 2.41. We observe both large collaborations as well as solo-authored papers. We observe that the university and industry collaboration is not much. The percentage of papers having co-authors from both university and industry is never more than 14% during the 10 year period. Majority of the papers (more than 60%) are from university. We observe a low participation from women and a gender imbalance in authorship as the percentage of female authors is always below 30% across the 10 year period. We observe that the percentage of industry authors are generally low and below 25% for every year. There is a good percentage of external collaboration and our analysis reveals that it is common for authors from more than one organization to co-author a paper. More than 38% of the papers every year have external collaboration. There are 14 articles in the 10 year period having a Google scholar citation of more than 50. There are few papers with more than 100 citations also. We identify prolific authors and our analysis shows that there are 14authors who have contributed more than 10 articles. The top 10 countries in-terms of the number of author affiliations are: India (864), USA (673), Germany (113), Brazil (97), Canada (75), China (68), UK (66), Turkey (52), France (49) and Italy (48) where the number in the bracket denotes the number of authors.

12. REFERENCES

[1] Abramo, G., DÂĂŹAngelo, C. A., Di Costa, F., and Solazzi, M. University–industry collaboration in italy: A bibliometric examination. *Technovation 29*, 6 (2009), 498–507.

[2] Agarwal, S., Mittal, N., Katyal, R., Sureka, A., and Correa, D. Women in computer science research: what is the bibliography data telling us? *ACM SIGCAS Computers and Society 46*, 1 (2016), 7–19.

[3] Agarwal, S., Mittal, N., and Sureka, A. A glance at seven acm sigweb series of conferences. *ACM SIGWEB Newsletter*, Summer (2016), 5.

[4] Agarwal, S., Mittal, N., and Sureka, A. A scientometric analysis of 9 acm sigweb cooperating conferences. *ACM SIGWEB Newsletter*, Autumn (2016), 6.

[5] Agarwal, S., Mittal, N., and Sureka, A. How healthy are acm sigweb sponsored conferences? *ACM SIGWEB Newsletter*, Spring (2017), 4.

[6] Aguillo, I. F. Is google scholar useful for bibliometrics? a webometric analysis. *Scientometrics 91*, 2 (2011), 343–351.

[7] Amin, M., and Mabe, M. Impact factors: use and abuse. *International Journal of Environmental Science and Technology:(IJEST) 1*, 1 (2004), 1.

[8] Butcher, J., and Jeffrey, P. The use of bibliometric indicators to explore industry–academia collaboration trends over time in the field of membrane use for water treatment. *Technovation 25*, 11 (2005), 1273–1280.

[9] Christopher, M. M., and Marusic, A. Geographic trends in research output and citations in veterinary medicine: insight into global research capacity, species specialization, and interdisciplinary relationships. *BMC veterinary research 9*, 1 (2013), 115.

[10] Elmacioglu, E., and Lee, D. On six degrees of separation in dblp-db and more. *ACM SIGMOD Record 34*, 2 (2005), 33–40.

[11] Franceschet, M. A comparison of bibliometric indicators for computer science scholars and journals on web of science and google scholar. *Scientometrics 83*, 1 (2010), 243–258.

[12] Garg, K., and Padhi, P. Scientometrics of prolific and non-prolific authors in laser science and technology. *Scientometrics 49*, 3 (2000), 359–371.

[13] Shabani, D. B., Carr, J. E., Petursdottir, A. I., Esch, B. E., and Gillett, J. N. Scholarly productivity in behavior analysis: The most prolific authors and institutions from 1992 to 2001. *The Behavior Analyst Today 5*, 3 (2004), 235.

[14] Sharma, R., Aggarwal, P., and Sureka, A. Insights from mining eleven years of scholarly paper publications in requirements engineering (re) series of conferences. *ACM SIGSOFT Software Engineering Notes 41*, 2 (2016), 1–6.

[15] Sharma, R., and Sureka, A. A nine year story of the india software engineering conference from 2008 to 2016. *ACM SIGSOFT Software Engineering Notes 41*, 5 (2016), 31–44.

[16] Van Rossum, M., Bosker, B., Pierik, E., and Verheyen, C. Geographic origin of publications in surgical journals. *British Journal of Surgery 94*, 2 (2007), 244–247.

[17] Vasilescu, B., Serebrenik, A., Mens, T., van den Brand, M. G., and Pek, E. How healthy are software engineering conferences? *Science of Computer Programming 89* (2014), 251–272.