Received: 24 January 2021 | Revised: 28 July 2021 | Accepted: 23 August 2021

IET Software

DOI: 10.1049/sfw2.12046

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

Software engineering and 12 prominent sub-areas: Comprehensive bibliometric assessment on 13 years (2007–2019)

Sana Alam[1,2] | **Shehnila Zardari**[3] | **Muneera Bano**[4]

[1]Department of Computer Science and Information Technology, NED University of Engineering and Technology, Main University Road, Karachi, Pakistan

[2]Department of Computer Engineering, Sir Syed University of Engineering and Technology, ST-16 University Rd, Karachi City, Pakistan

[3]Department of Software Engineering, NED University of Engineering and Technology, Karachi, Pakistan

[4]Faculty of Science, Engineering and Built Environment, Deakin University, Melbourne, Victoria, Australia

Correspondence

Sana Alam, Department of Computer Science and Information Technology, NED University of Engineering and Technology, Main University Road, B-179, Block 15, Gulistan-e-Jauhar, Karachi , Sindh, 75290, Pakistan.
Email: sana.email43@gmail.com

Funding information

NED University of Engineering and Technology

Abstract

Software Engineering (SE) as a field has grown significantly in terms of the research publications in the last few years. Criterion-based assessment of research performance of scholars, countries along with collaboration networks and links among articles are better judged using bibliometric analysis. The results proposed in this work are based on quantitative analysis and data visualisation of 150,087 scholarly articles published in last 13 years (2007–2019) across 85 research areas based on seven categories by providing a detailed and distinctive time-frame-based comparison to observe a shift in research trend in SE and its 12 distinguished sub-areas. The results of the observations proposed in this work include analysis on types of documents, yearly publication trends with results suggesting a prominent increment for the period (2015–2017) in terms of research publications, languages of research publications with findings indicating that there are articles published in languages other than English language, also, publication count-based rankings of authors, collaboration networks of countries with findings showing the supremacy of USA and China and keyword statistics.

KEYWORDS

software engineering, software quality, systems analysis

1 | INTRODUCTION

It has been almost half a century since the inception of Software Engineering (SE), which is now considered to be one of the most well-established fields in computing [1]. It was the discussion on 'Software Crisis' in a conference held in 1968 that led to the emergence of the term 'Software Engineering' [1, 2]. SE is associated with the concepts, procedures and tools required for developing software in computer systems [2]. Johnson, P. in [3] laid emphasis on the theoretical approach in SE. Importance of the SE field is evident by witnessing the designing and development of large and complex systems, technological innovations and emergence of new sub-areas. SE comprises methods, processes and use of suitable tools to develop computer software of high quality in a timely manner [4]. Wasserman, A. I. in [5] defines SE as a collection of methods (process) for the management and creation of software-intensive systems either at an individual or at a cumulative (team) level. SE practices have a significant role in meeting the challenges of various complex modern-day systems such as health care systems, systems related to safety, environmental monitoring systems to name a few [6]. It is important to monitor the past, present and future trends in SE research while appreciating the depth and coverage of research conducted over decades around the world.

The need of bibliometric analysis in the field of SE is important to discover authors' productivity over time,

country-wise research collaboration network, and shift in publication trends in the SE domain along with its sub-areas of SE. The bibliometric studies are useful in providing an in-depth analysis to the researchers and students, thus enabling them to make better informed choices of institutions and organisations for professional or research affiliation, to examine different aspects of scientific collaboration.

Bibliometrics is an assessment procedure of published articles through the use of various quantitative parameters [7]. The contributions and influence of authors and publications are determined by using various bibliometric techniques [8]. The amount of notable and evolutionary contributions in the SE field indicates the maturity of the field since its origination in 1968. More than 70,000 scientific contributions have been published in the SE field since late 1960s [9, 10]. Bibliometric insight into the SE field based specifically on citation counts or citation analysis of the scientific work ([7, 9, 11–13]), contributions of the specific SE research community ([1, 14]), and assessment of institutions and scholars ([15–26]) demonstrates key aspects of a particular research area.

This work presents a comprehensive bibliometric analysis to uncover the growth rate and major trends in the SE field from the year 2007 to 2019 over a pool of *150,087* scientific articles based on the type of documents being analysed, annual scientific production, productivity of authors, contributions of scholarly articles in various languages, research collaboration networks of countries and co-word analysis.

The rest of the paper is organised as follows: Since our aim is to statistically analyse published literature in the area of SE and its 12 sub-areas, therefore, in order to achieve the goal of providing comprehensive evaluation, detailed analysis of related work becomes necessary. Hence, Section 2 comprises a survey of related studies and the significant difference of our work from the already existing research work. Section 3 describes the methodology for data collection, refinement and analysis. Different subsections (Sections 3.3.1–3.3.6) of Section 3 are related to the detailed analysis of research questions presented in Section 3.2 established on various parameters. Section 4 includes a detailed discussion on the findings along with the limitations of our research work and the description of future work. Section 5 presents the conclusion.

2 | RELATED WORK

This section comprises the prominent relevant studies.

In 1994, Glass [15] presented his research work depicting leading scholars and institutions in the field of Systems and SE based on one-year study (1993). Anneliese von Mayrhauser and Elaine J. Weyuker topped the ranking of published scholars with each scoring 2.4. AT&T Bell Labs scored 10.58 and became the top-ranked institute for the year 1993. Glass assessment was limited by the journals surveyed to present the leading institutions. Based on the inclusion and exclusion criterion of the survey conducted in 1991 by the editorial board of the Journal of Systems and Software (JSS), six journals [Information and Software Technology (Butteworth-Heinemann, U.K.); JSS (Elsevier Science); Software Practice and Experience (SPE) (John Wiley & Sons, U.K.); IEEE Software (SW); ACM Transaction on Software Engineering and Methodologies (TOSEM); IEEE Transactions on Software Engineering (TSE)] were shortlisted to generate rankings of leading institutions. To calculate the score of each scholar, a single author of a published paper received a score of one, while each author of a multiple-authored paper initially received a score equal to their fractional representation on the paper. For author totals, the modification is performed in the initial scores of authors in multi-authored publications with a specific transformation (i.e. 0.5 becomes 0.7, 0.33 becomes 0.5, and those values that are less than or equal to 0.25 become 0.3). For institution ranking, it was decided to give same credit to the institution in case of the multi-authored paper as that of the single-authored research work. The ranking criterion is referred as JSS ranking.

Research studies [16–21, 23–25] covered different 5 years' time duration depicted as an ongoing study of original work presented in 1994 [15] by Glass, R. L. The frequency of publications or weighted scores with respect to publications are served as a metric to present their findings about most published scholars and institutions in the field of Systems and SE. The time duration covered by the above-mentioned studies are (1995–1999) [16] (1996–2000) [17] (1997–2001) [18] (1998–2002) [19] (1999–2003) [20] (2000–2004) [21] (2001–2005) [23] (2002–2006) [24], and (2003–2007 and 2004–2008) [25]. Publication-frequency-based assessment identified 15 leading scholars and 15 highly ranked institutions. Research studies [24, 25] also included Empirical Software Engineering (EMSE) journal to measure the productivity level of authors and institutions in seven journals.

Research studies [11–13] performed analysis of scholarly articles in ISI Web of Science (WoS) and presented 20 most cited articles as a result of evaluation. In research article [12], top 20 most cited articles for the time span (1986–2005) and (2000) were evaluated, whereas the research work [13] comprises the assessment of top 20 most productive SE researchers during (1988–2007) along with 20 highly cited articles for the time frame of 1 year (2001).

Manual calculation of publication-based ranking (JSS Ranking) possesses certain limitations on the time span and the number of journals surveyed. Ren and Taylor [22] proposed the concept of electronic bibliometric analysis to increase efficiency. The selection criterion for the analysis included two journals (ACM TOSEM; IEEE TSE) and two conferences (the International Conference on SE; ACM SIGSOFT International Symposium on the Foundations of SE) to identify top 50 scholars and institutions from 2000 to 2004 in the field of SE. They also analysed top 50 US computing graduate programs for the time span 1995–2003 on the criterion discussed in the research article [27].

Hamadicharef [28] performed the scientometric study for the time duration (1980–2010) on IEEE TSE. This research study constitutes the data set gathered from ISI WoS to perform evaluation on authors, citations and keywords, collaboration networks of authors and countries, and amount of references.

The main objective of study in [29] was to determine trends in authorship in the field of SE. About 70,000 research articles from DBLP over a time span (1971–2012) were analysed. An evident increase is observed in the number of published articles in SE with an average increase of around +0.40 authors/decade. A clear shift in authorship indicates that until 1980, majority of the publications were written by sole authors, whereas three or four authors contributing to a publication in recent times is a common practice.

Prominent research studies [1, 14] based on specific research communities have been carried out in conducting bibliometric assessments. The research study [14] aimed at performing analysis on researchers and institutions to identify top 50 influential researchers and top 50 institutions in Canada for the time duration 1996–2006. The evaluation was based on impact factors and h-index of the published articles in 12 highest ranked conferences and journals. The study [1] presented bibliometric analysis on Turkish SE community or the time duration of 1992–2014. Important findings as a result of the survey were given as follows: (i) negligible contribution of Turkey in SE research, (ii) SE research community in Turkey lacks diversity in SE area, (iii) good level of international collaborations, (iv) level of collaboration between industry and academia in SE domain is low, (v) low level of citations of Turkish SE papers, (vi) suggestions for an enhancement in the quality and quantity of SE papers produced by Turkey.

The recent work of [26] incorporates analysis of top 20 institutions, depiction of emerging directions in the SE field, categorisation of top 20 authors (novice, consolidated and experienced scholars) based on the frequency of publications, impact of research publications and research areas of authors over a time span of 8 years (2010–2017). However, the study [26] does not include assessment for language-based research publication analysis, co-word analysis, and evaluation of collaboration networks amongst countries. Our study is more recent and comprehensive, which includes 85 interdisciplinary research areas and software practices in 12 prominent sub-areas of SE for the time duration of 13 years (2007–2019), whereas the work in [26] has included papers published for 8 years only.

Table 1 lists down the parameter analysed, significant difference of our research study and our contributions. Our research study would help not only the researchers but also the students to analyse variation in productivity trends, research productivity of top-ranked authors, and relatedness of published studies in terms of co-word. Thus, students and researchers can make a wise decision regarding collaboration and understand the necessity of collaboration for enhanced productivity in less time. Researchers and students can further analyse the productivity rate of SE sub-areas to know the emerging sub-areas of SE, for example, Software Mining turns out to be one of the emerging sub-areas of SE. Moreover, our research study also affirms the fact that quality studies are published in languages other than English language thus, encouraging students and researchers to perform research in other languages also. Distribution of 13 years into various time frames provides an opportunity to analyse variations in the results of research questions (RQs) (Table 6) over different time rather than analysing all the RQs commutatively for 13 years. Hence, research trends especially in terms of productivity is better judged.

3 | RESEARCH METHODOLOGY

Our research design is inspired by the framework of the systematic mapping study [30]. We used ISI WoS database during the search process for the articles published during (2007–2019) in SE and its 12 sub-areas. The ISI WoS is one of the most renowned electronically available databases provided by Thomson Scientific's Institute for Scientific Information in terms of research publications [31]. Citation reports and data in WoS are prepared after careful investigation; thus, WoS provides citation service at the highest level [32]. Thus, we can say that WoS indexes high-quality journals, proceedings papers, reviews, editorial materials and books.

3.1 | Data collection and refinement

Our data collection process involved defining the right criterion based on an appropriate search term to collect relevant data required for performing comprehensive bibliometric assessment.

3.1.1 | Description and refinement of appropriate search terms

To achieve our objective of comprehensive assessment on scholarly articles published in SE and its related sub-areas for 13 years (2007–2019), we conducted an analysis on **150,087** ISI WoS Core Collection publications for **Research Query 5**. Refinement and analysis of multiple criteria are discussed below. All the five research queries are typed for ISI WoS Core Collection under 'Topic' dropdown.

Research Query 1: We initially completed our inspection on ISI WoS Core Collection by the term ***'Software Engineering'*** However, the results comprised above 4,000 scientific materials that lacked relevance to the field of SE.

Validity Threat of Research* Query *1: The findings showed that an exhaustive examination of specific SE research topics is not possible by using the phrase 'Computer Software'.

We therefore enhanced our search criterion by using the search string based on criterion 2.

Research Query 2: Research outcomes of ***Research* Query** 1 compelled us to opt for ***'Software Engineering'*** as a search term.

Validity Threat of Research* Query *2: An in-depth evaluation resulted in the fact that many SE materials not necessarily include the terminology 'Software Engineering'.

We therefore decided to perform a broad-spectrum analysis in the field of SE by including several related keywords in multiple search queries.

Research Query 3: We performed analysis by using individual queries for individual sub-areas names (such as first

TABLE 1 Parameters analysed, significant differences and contributions

Time duration	Data source	Parameters analysed
2007-2019	Web of Science	Following parameters are analysed with respect to the research questions in Table 6. • Types of documents, • Annual scientific publication, • Languages of research articles, • Active researchers, • Collaboration amongst countries, • Link among research studies in terms of co-word

Significant difference of our research study

1. **Sub-areas of Software Engineering (SE):** Our study targets a much broader aspect of Software Engineering with limited analysis on its 12 prominent sub-areas *(Mining, Evolution, Quality, Development, Architecture, Maintenance, Testing, Requirements,* ***Modeling*** *Or Modelling, Automation, Design and System)*. Details of these 12 sub-areas are presented in Section 3.1.1. under Research query 5.
2. **Review timeline and corresponding data sets**: In our research work, we performed the most recent trend-based assessment on SE and its 12 significant sub-areas. The research study includes an analysis on a total of 150,087 records existing in ISI Web of Science for the time duration of 13 years (2007–2019) in two phases. The two phases are discussed in Research Methodology (Section 3).
3. **Inclusion of multi-disciplinary research areas**: Inspection based on 7 categories (Figure 1b.) across 85 research areas for 13 years is one of the primary features of our research work.
4. **Languages**: Our research study is based on the quality research papers in the area of SE and its 12 prominent sub-areas published not only in English language but also in other languages. The inclusion of research studies conducted in other languages is considered to avoid language bias and to perform a comprehensive bibliometric assessment.

Contributions of our research study

1. **Inclusion of most knowledge areas covered by SWEBOK** [33]: Our research work includes 10 out of the 15 knowledge areas specified by SWEBOK [33] under 12 sub-areas of SE. Further details are provided in Section 3.1.1. under Research Query 5.
2. **Research trend analysis over different time duration:** To give an insight into research trend, we perform analysis over different time durations for a period of 13 years. The productivity trend in SE and its 12 sub-areas are clearly depicted in our analysis of 3 time periods of 3 years and 1 time period of 4 years.
3. **Inclusion of relevant studies that are under the sub-areas of SE**: The studies so far have used 'Software Engineering' as their research queries. However, many research studies do not necessarily include the terminology 'Software Engineering'.
4. **Authors' productivity in terms of frequency of publications**: Top 10 authors for each time duration provides a greater understanding of how authors have performed over different time durations (3-year period and 4-year period) rather than analysing a cumulative 13-year period. The emergence of new most active authors is evident from our analysis.
5. **Focus on collaboration amongst countries to increase the research productivity**: Our analysis suggests that Research productivity is significantly increased when countries collaborate with each other. Most of the top-ranked active researchers belong to countries having large collaborative networks.
6. **Relatedness amongst research studies**: To demonstrate the relatedness amongst research studies, co-word assessment is performed showing how the research studies are related with one another.
7. **Application of SE**: Emergence of 85 multi-disciplinary research areas affirms the fact that SE plays a vital role in other fields also.
8. **Analysis demonstrating the productivity of research studies in languages other than English Language**: We also include the productivity in terms of number of published articles in languages other than English language to demonstrate the fact that although English being a primarily language, quality studies are present in other languages also.

using search string 'Software Engineering', then using search string 'software mining'and so on) related to the main SE field and its 12 sub-areas.

Validity Threat of Research* Query *3: Findings include duplicate studies.

Research Query 4: Careful analysis shows that the same study appears in more than one sub-areas, therefore resulting in duplicate studies. To solve the problem of duplicate studies, we formulate the query ***('software engineering' OR 'software mining' OR 'software evolution' OR 'software quality' OR 'software development' OR 'software architecture' OR 'software maintenance' OR 'software testing' OR 'software requirements' OR 'software modeling' OR 'software modelling' OR 'software automation' OR 'software design' OR 'software system').***

Validity Threat of Research* Query *4: Detailed inspection reveals that many relevant papers use 'software' and 'sub-areas' (engineering/evolution/mining/quality/development/architecture/maintenance/testing/requirements/modeling or modelling/automation/design/system)' as two separate words rather than as a string (as used in the Research Query 4). However, the results of the **Research Query** 4

return only those articles where exact phrase matching of the aforementioned query takes place without including lemmatization. To overcome this issue, Research Query 5 is defined. We have used 'software modeling' and 'software modelling' as the assessment of research studies suggest that in some papers, both the spellings are used. So not to miss out on eligible papers, different spelling notations were used.

Research Query 5: To eliminate the issue of Research **Query** 4, the final formulated query is designed as ***(software AND (engineering OR mining OR evolution OR quality OR development OR architecture OR maintenance OR testing OR requirements OR modeling OR modelling OR automation OR design OR system))***.

Pros: Our findings cover quiet a number of areas within SWEBOK [33]. The five knowledge areas covered from SWEBOK V3.0 (downloaded from SWEBOK portal) are 'Software Requirements', 'Software Design', 'Software Testing', 'Software Maintenance', and 'Software Quality' as a separate research term mentioned in WoS, whereas five other knowledge areas of SWEBOK v3.0. namely 'Software Engineering Models and Methods', 'Software Engineering Management', 'Software Engineering Process', 'Software Engineering Professional Practice', 'Software Engineering Economics' are all covered under the term software AND engineering. The inclusion of other research terms was done as WoS places research articles related to SE under various terminologies. It is to be noted that remaining knowledge areas of SE will be studied as our future work. In our research study, the terminology 'sub-areas' represents knowledge areas as identified by SWEBOK as well as different terminologies used by WoS.

Cons: An enormous data set is generated resulting in **456,323** research studies. Hence, refining of retrieved documents as per Inclusion Criterion 1 (Figure 1a.) and Inclusion Criterion 2 (Figure 1b.) requires substantial time and effort.

3.1.2 | Refining retrieved documents

We refined our results to include only the relevant scientific materials. This section (Figure 2a.) summarises the methodological process followed in the refinement of extracted document based on the Research Query 5 discussed in Section 3.1.1 for cumulative analysis. For individual analysis, a general systematic refinement approach applied on 12 sub-areas is depicted in Figure 2b. The inclusion criteria for the publications are discussed in Figure 1a,b.

Phase I: Selection criterion for the designing of four distinguished time periods for the 13 years referred in Table 2 is defined as follows:

$$\text{Selection Criterion of Time Frames for cumulative analysis} = \text{Publication count of relevant research studies} \leq 42,000 \text{ records (published articles)} / \text{ time frame}$$

The WoS data source used in our analysis has a limitation of maximum 50,000 records to be saved as 'Marked List' for analysis. This limitation allows us to distribute our enormous data record into multiple chunks of almost even time frames (one 4-year time period and three 3-year-time periods) for assessment.

Moreover, the 42,000 limit is set, when we tried to analyse data having records greater than 42,000; then 'out of memory' error occurred in VOSviewer, so it was needed to allocate more memory. Therefore, for ease of analysis, we perform distribution of dataset in such a way as to keep the records in each dataset less than 42,000.

Phase 2: Table 3 presents the review timeline for 12 sub-areas (mentioned in Research Query 5 (Section 3.1.1)) based on two distinct time intervals of 7 years (2007–2013) and 6 years (2014–2019) along with the relevant number of studies. It is to be noted that single research study might come under various sub-areas; hence, it is to be ensured to include that particular research study for each sub-area. For cumulative analysis, duplicate studies are removed.

Phase 1 constitutes the analysis in accordance to the RQs (RQ1–RQ6) articulated in Table 6. Three RQs (RQ1, RQ2 and RQ3) designed in Table 6 are also assessed during Phase 2 of

(a)

Inclusion Criterion 1

Applying Research Query 5 (sec 3.1.1) on the following distinct time spans:
For Cumulative Analysis defined as per Research Query 5 (sec 3.1.1), the four distinguished time frames are as follows:

- 3 years (2007 – 2009)
- 4 years (2010 – 2013)
- 3 years (2014 – 2016)
- 3 years (2017 – 2019)

For Individual Analysis on 12 sub-areas defined as per Research Query 5 (sec 3.1.1), the two non-identical time frames are as follows:

- 7 years (2007 – 2013)
- 6 years (2014 – 2019)

(b)

Inclusion Criterion 2

After applying Inclusion Criterion 1, the selected documents across **85** research areas (mentioned in Table 3) are refined according to the following **7** categories.

- Computer Science Software Engineering
- Computer Science Theory Methods
- Computer Science Information Systems
- Computer Science Artificial Intelligence
- Computer Science Interdisciplinary Applications
- Computer Science Hardware Architecture
- Computer Science Cybernetics

FIGURE 1 (a) Inclusion Criterion 1 applied for documents' selection. (b) Inclusion Criterion 2 applied on selected documents

our research work. Table 4 lists down the research areas included in our analysis. These 85 research areas are calculated by WoS database.

Research Areas constitute a subject categorisation scheme that is shared by all WoS product databases. Research Area terms are article-based. The Research Areas' field displays the research areas found within an article. As a result, one can identify, retrieve and analyse documents from multiple databases that pertain to the same subject.

The WoS Categories, on the other hand, are journal-based. Journals and books covered by WoS Core Collection are assigned to at least one WoS category. Each WoS category is mapped to one Research Area.

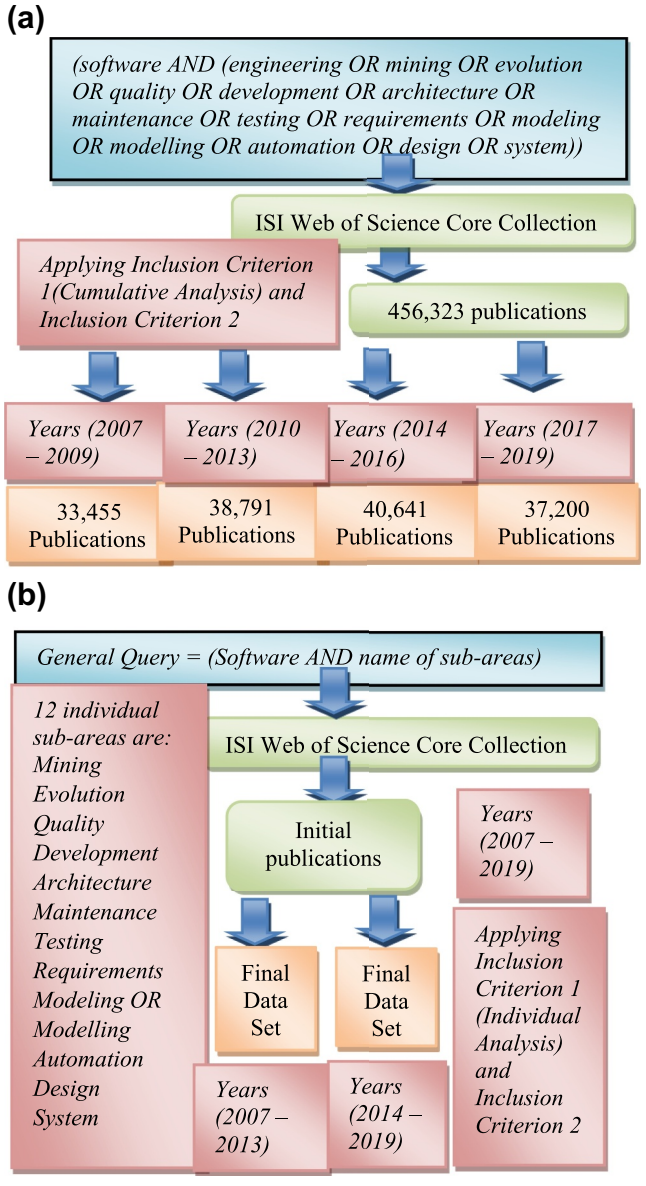

FIGURE 2 (a) Summary of Selection and Refinement procedures for cumulative analysis of publications as per Research Query 5 (Section 3.1.1) during Phase I. (b) Summary of Selection and Refinement procedures of publications for each of the 12 individual sub-areas during Phase 2

Top 10 Research Areas for each time duration is demonstrated in Table 5. For each time duration, 100% of the studies belong to Computer Science. So, Computer Science is the first Research Area in each time period as shown in Table 5.

TABLE 2 Review timeline for Phase 1 of research study as per Research Query 5 (Section 3.1.1)

Search criterion (research query 5, Section 3.1.1)	Review time span	Relevant number of research studies
(Software AND (engineering OR mining OR evolution OR quality OR development OR architecture OR maintenance OR testing OR requirements OR modeling OR modelling OR automation OR design OR system))	2007–2009	33,455
	2010–2013	38,791
	2014–2016	40,641
	2017–2019	37,200

TABLE 3 Review timeline for Phase 2 of research study for 12 sub-areas of SE (the sub-areas are in the same order as discussed in Research Query 5 (Section 3.1.1))

Sub-areas	Review time span	Relevant number of research studies
software AND mining	2007–2013	1881
	2014–2019	2673
software AND evolution	2007–2013	3005
	2014–2019	3484
software AND quality	2007–2013	9039
	2014–2019	11,497
software AND development	2007–2013	21,347
	2014–2019	21,219
software AND architecture	2007–2013	13,476
	2014–2019	13,757
software AND maintenance	2007–2013	2617
	2014–2019	2808
software AND testing	2007–2013	12,547
	2014–2019	15,075
software AND requirements	2007–2013	9981
	2014–2019	10,513
software AND (modelling OR modeling)	2007–2013	30,358
	2014–2019	31,702
software AND automation	2007–2013	1554
	2014–2019	2114
software AND design	2007–2013	27,750
	2014–2019	28,473
software AND system	2007–2013	42,350
	2014–2019	42,565

Abbreviation: SE, software engineering.

TABLE 4 Multi-disciplinary research areas

Multi-disciplinary research areas					
Computer science	Engineering	Telecommunications	Automation control systems	Robotics	Mathematics
Physics	Materials science	Education educational research	Business economics	Toxicology	Architecture
Geology	Environmental sciences ecology	Operations research management science	Chemistry	Rehabilitation	Mechanics
Optics	Remote sensing	Mathematical computational biology	Agriculture	International relations	Arts humanities other topics
Music	Health care sciences services	Information science library science	Transportation	Crystallography	Linguistics
Orthopaedics	Social sciences other topics	Public environmental occupational health	Physical geography	History philosophy of science	Social issues
Art	Energy fuels	Biochemistry molecular biology	Pharmacology pharmacy	Water resources	Geography
Food science technology	Instruments instrumentation	Science technology other topics	Public administration	Government law	Psychology
Behavioural sciences	Astronomy astrophysics	Imaging science photographic technology	Life sciences biomedicine other topics	Cardiovascular system cardiology	Communication
Fisheries	Neurosciences neurology	Biotechnology applied microbiology	Urban studies	Geriatrics gerontology	Biophysics
Nursing	Construction building technology	Radiology nuclear medicine medical imaging	Medical informatics	Mathematical methods in social sciences	Acoustics
Archaeology	General internal medicine	Criminology penology	Cultural studies	Philosophy	Area studies
Sociology	Sport sciences	Nuclear science technology	Surgery	Mining mineral processing	Literature
Plant sciences	Electrochemistry	Infectious diseases	Research experimental medicine	Marine freshwater biology	Forestry
Legal medicine					

3.2 | Designing appropriate Research Questions

Our research study is guided by 6 research questions that are presented in Table 6. The results obtained by answering these RQs represent variations in different bibliometric features of SE along with its 12 selected sub-areas. In total, **150,087** research studies based on Research Query 5 discussed in Section 3.1.1 were retrieved with 'Topic' dropdown menu using Web of Science Core Collection database for all six RQs. The connecting thread among our research question is our overarching research objective of exploring and understanding research out in SE. That included our analysis of the productivity of research outputs in the domain such as types of research documents, annual rate of publication of research papers, the languages used as a medium to publish research studies, top 10 authors based on these research studies, how are the documents of these research studies related to one another in terms of co-word, and what is the impact of collaboration amongst countries on the frequency of publications.

3.3 | Data analysis

Bibliometric analysis involves several complex steps to correctly collect, evaluate, and interpret data. Mostly, the tools available for mapping are commercially licensed software [34]. The outcomes of the search query were stored and graphically represented in Excel.

Bibliometric assessment enables the investigators to present the outcomes of empirical studies based on co-citation analysis, co-word analysis and bibliographic couplings. VOSviewer is a comprehensive software visualisation tool for analysing bibliometric maps [35]. We used VOSviewer to perform co-word analysis and research collaboration network analysis in terms of countries.

3.3.1 | Types of documents analysed

We categorised documents in **19** types namely proceedings paper, article, editorial material, book chapter, review, book, letter, meeting abstract, correction, news item, book review,

TABLE 5 Top 10 research areas in four distinct time durations

Year 2007–2009		Year 2014–2016	
Top 10 research areas	**No. of publications**	**Top 10 research areas**	**No. of publications**
Computer science	33,455	Computer science	40,641
Engineering	12,335	Engineering	13,699
Telecommunications	3600	Telecommunications	3057
Automation control systems	2028	Automation control systems	1448
Mathematics	1575	Robotics	946
Business economics	1052	Mathematics	831
Education educational research	1043	Education educational research	623
Operations research management science	967	Operations research management science	610
Imaging science photographic technology	865	Mathematical computational biology	574
Information science library science	856	Information science library science	526
Year 2010–2013		**Year 2017–2019**	
Top 10 research areas	**No. of publications**	**Top 10 research areas**	**No. of publications**
Computer science	38,791	Computer science	37,200
Engineering	13,170	Engineering	12,555
Telecommunications	2127	Telecommunications	4804
Automation control systems	1406	Automation control systems	1293
Mathematics	1363	Mathematics	889
Operations research management science	1039	Education educational research	738
Materials science	765	Robotics	669
Mathematical computational biology	763	Mathematical computational biology	576
Education educational research	711	Biochemistry molecular biology	470
Information science library science	613	Operations research management science	442

early access, retracted publication, biographical item, software review, bibliography, reprint, retraction and data paper for analysis. All 19 classifications of documents (where applicable) are considered for evaluation to avoid bias in our research study for Tables 2 and 3. It is clear from Table 7 that a research study belongs to more than one category. Therefore, not only the primary research studies are considered, but other studies are also a part of our research work.

3.3.2 | Annual scientific production

In this work, the ISI WoS database is used to extract the pool of published studies for the four distinct time spans (2007–2009), (2010–2013), (2014–2016) and (2017–2019) as per Research Query 5 (Section 3.1.1). Year-wise publications shown in (Figure 3) provide a comprehensive way to draw a conclusion on the publication trend analysis in SE field and its sub-areas over the last 13 years in accordance to the afore-mentioned time frames, respectively. Annual scientific production presents a mean to observe variation in scientific contributions over a specified time duration. The analysis represents the fact that the global publication trend in SE and its sub-areas touched its peak during the years 2015–2017. In Figure 3, the peak time duration is highlighted by red colour. The frequency of scientific publications in SE and its sub-areas depicts the intensity of research in this field in the past 13 years. Table 8 contains findings of RQ2 (see Table 6) in accordance to Table 3. Similarly, the sub-areas with research productivity in terms of published articles >40,000 and < 50,000 and published articles >50,000 for 13 years are highlighted with blue and red colours, respectively, in Table 8.

3.3.3 | Languages of research publications

Languages of scholarly articles were analysed to assess the usage of languages other than English for research publication [36]. Table 9 represents an analysis on the languages of scholarly articles for our research study during the four time periods (2007–2009), (2010–2013), (2014–2016) and (2017–2019). Scholarly articles written in English language constitute up to 98.75% of the entire pool of quality work included from

TABLE 6 Research questions with subpart of research questions/clarifications

Research Questions designed as per Research Query 5 (Section 3.1.1)	Subpart of Research Questions/clarification	Review time frames (as discussed in Table 2 or Table 3)
RQ1: **What are the types of documents being analysed in this research study?**	• Different types of documents included in our study are depicted. These categories are ISI WoS document types.	• Table 2 • Table 3
RQ2: **What is the annual scientific production in SE and its sub-areas?**	• What is the trend in annual scientific production for each of the four distinct time spans?	Table 2
	• How many research studies have been published in 12 notable sub-areas of SE?	
	• What is the yearly publication trend for each of the already mentioned 12 SE sub-areas? (RQ2 provides an insight into variation in productivity with respect to published documents).	Table 3
RQ3: **How many studies are published in languages other than English Language?**	• How many research studies are published in languages other than English for the specified time duration? (This parameter is included with a view of providing various language-based non-bias and non-discriminated research work).	Table 2
	• The languages of published research studies are analysed for each of the 12 distinguished sub–areas of SE	Table 3
RQ4: **Who are the most active researchers in SE?**	• Who are the top 10 most active authors in terms of publication count of the published documents?	Table 2
RQ5: **Which countries have strong collaborative research networks?**	• To exhibit research collaboration amongst countries we performed country-wise co-authorship analysis.	Table 2
RQ6: **How are the documents related with each other in terms of co-word?**	• Co-word analysis is used to demonstrate the relationship amongst studies based on the occurrence of keywords in research publications. Thus, it demonstrates how the studies are related to each other with respect to keywords.	Table 2

Abbreviations: RQ, reserach question; SE, software engineering; WoS, web of science.

ISI WoS for our research study for the above-mentioned four different time durations.

The results reflect that although major contributions are made by research scholars in English language. However, studies present in other languages also make a contribution. Therefore, to perform a detailed and comprehensive analysis, inclusion of research studies published in languages other than English becomes inevitable. A prominent decrease in research studies published in Chinese language over the 6 years (2014–2019) period is observed as compared to the time duration (2007–2013) in Table 9. This is probably due to the fact of publishing more scholarly articles in English language over the recent time period by the researchers in China.

Table 10 performs the analysis on each of the 12 sub-areas of SE discussed in Research Query 5 (Section 3.1.1).

3.3.4 | Author production

Analysis of top **10** authors for our research study is established on number of scientific contributions for the time duration (2007–2009), (2010–2013), (2014–2016) and (2017–2019) in accordance to the search string (Research Query 5 discussed in Section 3.1.1) is represented in Figure 5a–d. Top **10** authors are listed after careful evaluation and considerable amount of time. By using the WoS online interface, we came to know that multiple authors are grouped under same initials; therefore, vigilant assessment becomes necessary.

The evaluation of top 10 authors leads to the fact that multiple authors occupy a single rank in four non-identical time frames. The relationship between number of authors and corresponding rank position is clearly depicted in Figure 4.

Figure 4 is necessary to understand the reason behind having more than 10 authors in Figure 5a–d. In Figure 4., it can be seen that more than one authors are present at a particular rank position. This is due to the fact that authors having same number of publications occupy the same rank. Figure 4 depicts that in case of tie, how many authors occupy a single rank position. For example, for the year 2007–2009, at rank 10, there are four authors because all four authors have equal number of publications.

3.3.5 | Collaboration network amongst countries

Research collaboration at an international level is determined on account of having at least one author belonging to a different country [37].

TABLE 7 WoS document types over all four distinct time durations

Types of documents	2007–2009	2010–2013	2014–2016	2017–2019
Proceedings paper	26,225	25,515	28,661	22,237
Article	7,855	13,243	11,822	14,585
Book chapter	572	1248	597	447
Editorial material	231	368	292	260
Review	113	204	236	481
Book	22	62	29	17
Book review	4	9	2	1
Meeting abstract	1	9	9	5
Correction	1	8	5	7
Letter	12	8	5	7
News item	6	4		2
Biographical item	2	2		1
Retracted publication		1		2
Software review	5	1	4	1
Bibliography	1			
Reprint	1		1	1
Early access				30
Data paper				7
Retraction				1

Abbreviation: WoS, web of science.

We have used VOSviewer to perform collaboration network analysis. In VOSviewer, units of analysis are represented as nodes/item and links represent that two items have worked together on at least one scientific study. The distance between two nodes determines the strength of the link and the relation between two items [38]. The weight of a node/item is reflected by the size of the corresponding node/item. Greater weights are reflected by larger sizes of circles [38]. Different colour schemes represent clusters to which an item belongs [38]. Countries who have frequently co-authored research publications belong to the same cluster.

To examine the extent of collaboration, the evaluation of co-authored documents is conducted with countries as a unit of analysis established on a full counting method as per Research Query 5 (Section 3.1.1) for the four non-identical time frames mentioned in Table 2.

Top 20 countries in each figure with minimum number of five co-authored publications are shown in Network Visualisation (Figure 6a–d) for four distinct time duration. USA is identified as the leader in research collaboration on account of collected collaborative data. It is evident from Figure 6a–d that People's Republic of China leads the research activity of Asian countries as it has one of the strongest research association networks.

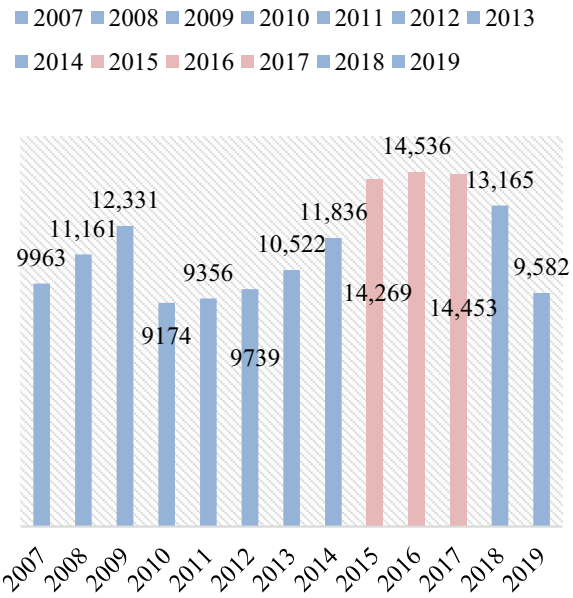

FIGURE 3 Annual publication trend for the time span (2007–2019)

3.3.6 | Co-word analysis

Co-word analysis is used to establish a relationship amongst various articles/topics on account of the occurrence of keywords [39]. Similarity in the type of articles is predicted by the frequent occurrence of keywords. Co-word analysis has been previously used as a metric to perform bibliometric analysis in the research studies [40–43]. Figure 7a–d exhibit the relatedness of keywords as per Research Query 5 (Section 3.1.1) in the form of Network Visualisation where top 20 keywords are selected based on a full counting method provided that the minimum number of the occurrence of keyword is 100.

4 | DISCUSSION

The discussion section examines the findings of six research questions (RQ1–RQ6) formulated in Table 6 (Section 3.2).

4.1 | Detailed discussion on findings of Research Questions

4.1.1 | Inclusion of multi-disciplinary fields and non-identical time spans

The inclusion of 85 multi-disciplinary research areas (Table 4) categorised on account of Inclusion Criterion 1 (Figure 1a.) ensures maximum coverage of related data to perform an in-depth evaluation of 6 RQs (Table 6, Section 3.2). Inclusion Criterion 2 (Figure 1b.) is used to present a clear picture of variation in research trend across multiple time frames based on refinement of categories. These 85 multiple fields also

TABLE 8 Annual publication count of research studies in SE sub-areas (Research Query 5 in Section 3.1.1) for the time duration of 13 years ((2007- 2013) and (2014 – 2019))

Sub-areas of SE	Time Frame (2007 – 2013)							Time Frame (2014 – 2019)					
	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Automation	186	218	264	202	198	221	265	296	388	360	415	382	273
Mining	198	251	328	228	231	284	361	362	450	474	553	467	367
Maintenance	373	406	407	321	328	381	401	431	508	499	556	480	334
Evolution	403	457	477	381	421	409	457	559	598	615	674	576	462
Requirements	1400	1588	1779	1218	1235	1278	1483	1624	1840	1965	2055	1774	1255
Quality	1116	1377	1591	1099	1172	1218	1466	1652	2059	2138	2133	1995	1520
Architecture	2013	2288	2320	1656	1668	1646	1885	2110	2475	2586	2534	2340	1712
Testing	1683	1731	2125	1495	1629	1813	2071	2142	2745	2802	2732	2619	2035
Development	2996	3427	3690	2766	2716	2803	2949	3358	3849	3913	4035	3617	2447
Design	3900	4175	4724	3579	3611	3811	3950	4403	5264	5342	5258	4704	3502
Modeling or Modelling	4110	4696	5205	3909	3976	4110	4352	4771	5804	5868	5719	5321	4219
System	5982	6667	7268	5386	5274	5692	6081	6767	7867	7933	7838	7112	5048

Abbreviation: SE, software engineering.

TABLE 9 Publication languages of scholarly articles over four distinct time frames

Year 2007–2009		Year 2014–2016	
Top four languages	**Corresponding number of research studies**	**Top four languages**	**Corresponding number of research studies**
English	33,313	English	40,191
Chinese	40	Portuguese	209
German	39	Spanish	182
Spanish	29	Russian	21
Others	34	Others	38
Year 2010–2013		**Year 2017–2019**	
Top four languages	**Corresponding number of research studies**	**Top four languages**	**Corresponding number of research studies**
English	37,991	English	36,723
Chinese	476	Spanish	165
Portuguese	138	Portuguese	164
Spanish	136	Turkish	59
Others	50	Others	89

reflect the fact that the importance and application of SE is not only limited to the SE sub-areas but is also extended across multiple entirely different domains. Table 5 further gives an insight into top 10 research areas for each of the four distinct time frames. Engineering, Telecommunications, Mathematics, Automation Control Systems and Education Educational Research are the areas that are a part of top 10 Research Areas for a continuous period of 13 years.

4.1.2 | An evaluation of annual publication pattern

The publications pattern in Figure 3 represents a clear increment in the publication trend over the 3-year time period where, for three consecutive years (2015, 2016 and 2017), the number of scholarly articles published are well above 14,000 articles/year. It is evident that the research publication trend of publishing more than 13,000 scholarly articles/year also continues in the year 2018. The annual publication pattern in SE and its related sub-areas is a clear indication of how well SE has established itself over the last few years. The growth in the publication trend is predicted to evolve in the coming years with the technological advancements and enhanced research facilities. The observations from Table 8 further affirm the fact that Software Development, Software Modeling or Modelling, Software Systems and Software Designs are the most well-established sub-areas of SE. However, the emergence and evolution of some other areas of SE is also visible from the results of Table 8. Especially, Mining, Evolution, Maintenance and Automation are the emerging fields in SE.

TABLE 10 Publication languages of scholarly articles over two distinct time frames for 12 distinct SE sub-areas

	Top four languages for 2007 – 2013		Top four languages for 2014 – 2019	
SE sub-areas	**Languages**	**Corresponding number of research studies**	**Languages**	**Corresponding number of research studies**
Mining	English	1865	English	2651
	Chinese	11	Spanish	8
	Spanish	3	Turkish	7
	Portuguese	2	Portuguese	6
			Chinese	1
Evolution	English	2985	English	3447
	Portuguese	10	Portuguese	20
	Chinese	7	Spanish	12
	Spanish	3	Others	5
Quality	English	8924	English	11,301
	Chinese	46	Spanish	80
	Portuguese	27	Portuguese	76
	Spanish	26	Ukrainian	13
	Others	16	Others	27
Development	English	21,065	English	20,824
	Chinese	133	Spanish	168
	Portuguese	69	Portuguese	160
	Spanish	55	Turkish	17
	Others	25	Others	50
Architecture	English	13,368	English	13,646
	Chinese	57	Portuguese	46
	Spanish	21	Spanish	42
	Portuguese	17	Turkish	8
	Others	13	Others	15
Maintenance	English	2596	English	2779
	Chinese	13	Portuguese	18
	Portuguese	3	Spanish	5
	Spanish	3	Russian	3
	Others	2	Others	3
Testing	English	12,418	English	14,913
	Chinese	75	Portuguese	72
	Portuguese	16	Spanish	52
	Spanish	16	Turkish	21
	Others	22	Others	17
Requirements	English	9869	English	10,403
	Chinese	54	Portuguese	40
	Spanish	22	Spanish	38
	Portuguese	21	Turkish	10
	Others	15	Others	22

TABLE 10 (Continued)

SE sub-areas	Top four languages for 2007 – 2013		Top four languages for 2014 – 2019	
	Languages	**Corresponding number of research studies**	**Languages**	**Corresponding number of research studies**
Modeling or modelling	English	30,011	English	31,329
	Chinese	188	Portuguese	146
	Spanish	60	Spanish	134
	Portuguese	57	Russian	35
	Others	42	Others	58
Automation	English	1538	English	2083
	Chinese	9	Spanish	11
	Spanish	3	Portuguese	6
	Portuguese	2	Ukrainian	5
	Others	2	Others	9
Design	English	27,363	English	28,245
	Chinese	288	Spanish	99
	Spanish	49	Portuguese	69
	Portuguese	24	Turkish	20
	Others	26	Others	40
System	English	41,796	English	42,102
	Chinese	324	Portuguese	191
	Spanish	84	Spanish	151
	Portuguese	79	Turkish	36
	Others	67	Others	85

Abbreviation: SE, software engineering.

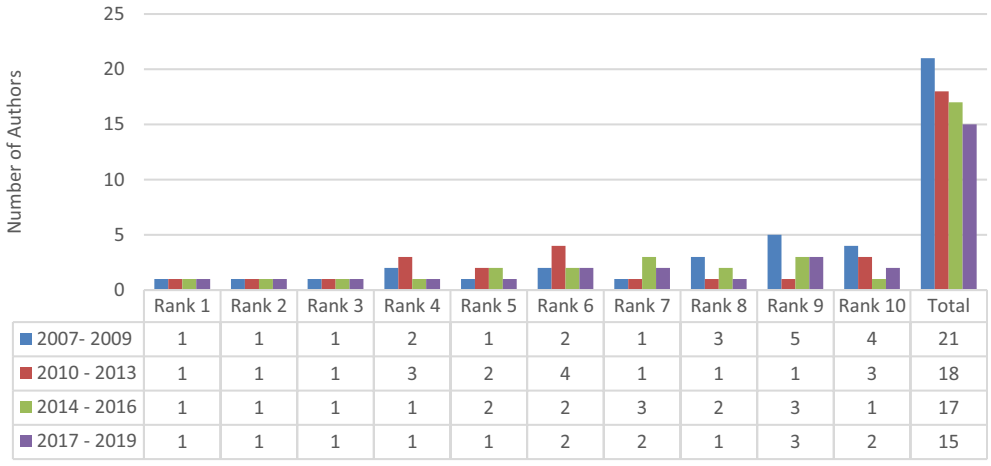

FIGURE 4 Number of authors at a given rank position

4.1.3 | An assessment on languages of scholarly articles:

English continued to be the dominating language of research articles published so far but the contributions made by other languages in this aspect is also significant. Hence, inclusion of articles published in languages other than English becomes mandatory to avoid researcher bias in the findings.

Another important finding from Tables 9 and 10 depicts that the contributions made in Chinese, Spanish, Portuguese and Turkish languages cannot be ignored. It can be safely stated that English is a preferred publication language of

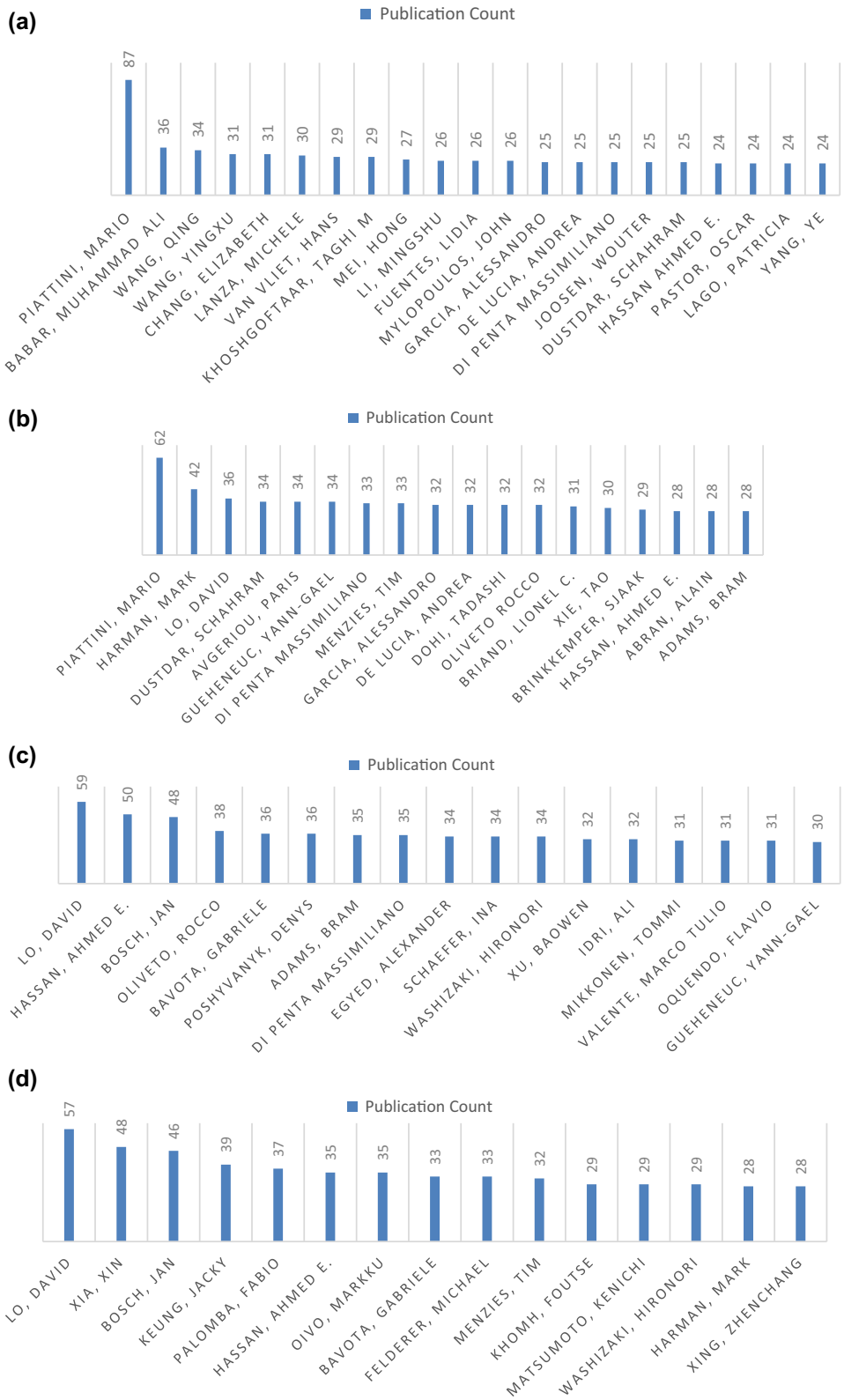

FIGURE 5 (a) Top 10 authors in terms of frequency of publications for (2007–2009). (b) Top 10 authors in terms of frequency of publications for (2010–2013). (c) Top 10 authors in terms of frequency of publications for (2014–2016). (d) Top 10 authors in terms of frequency of publications for (2017–2019)

research scholars all over the world and perhaps this fact also influenced the Chinese researchers; thus, a significant decrement in Chinese as language of published research articles in recent times is observed from Tables 9 and 10. These results further consolidate our decision of including languages other than English to perform a comprehensive bibliometric analysis.

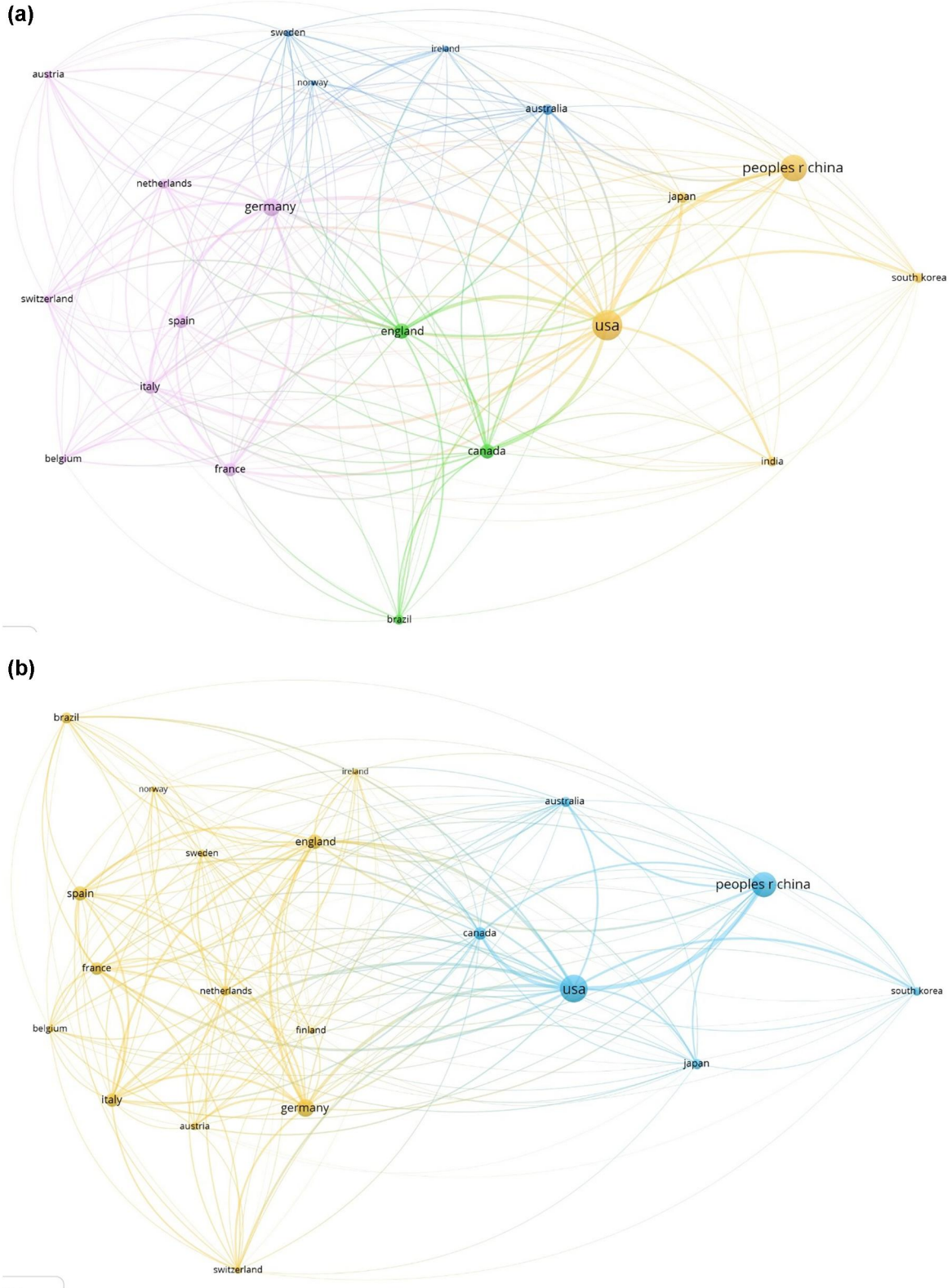

FIGURE 6 (a) Collaboration network amongst top 20 countries for the time duration (2007–2009). (b) Collaboration network amongst top 20 countries for the time duration (2010–2013). (c) Collaboration network amongst top 20 countries for the time duration (2014–2016). (d) Collaboration network amongst top 20 countries for the time duration (2017–2019)

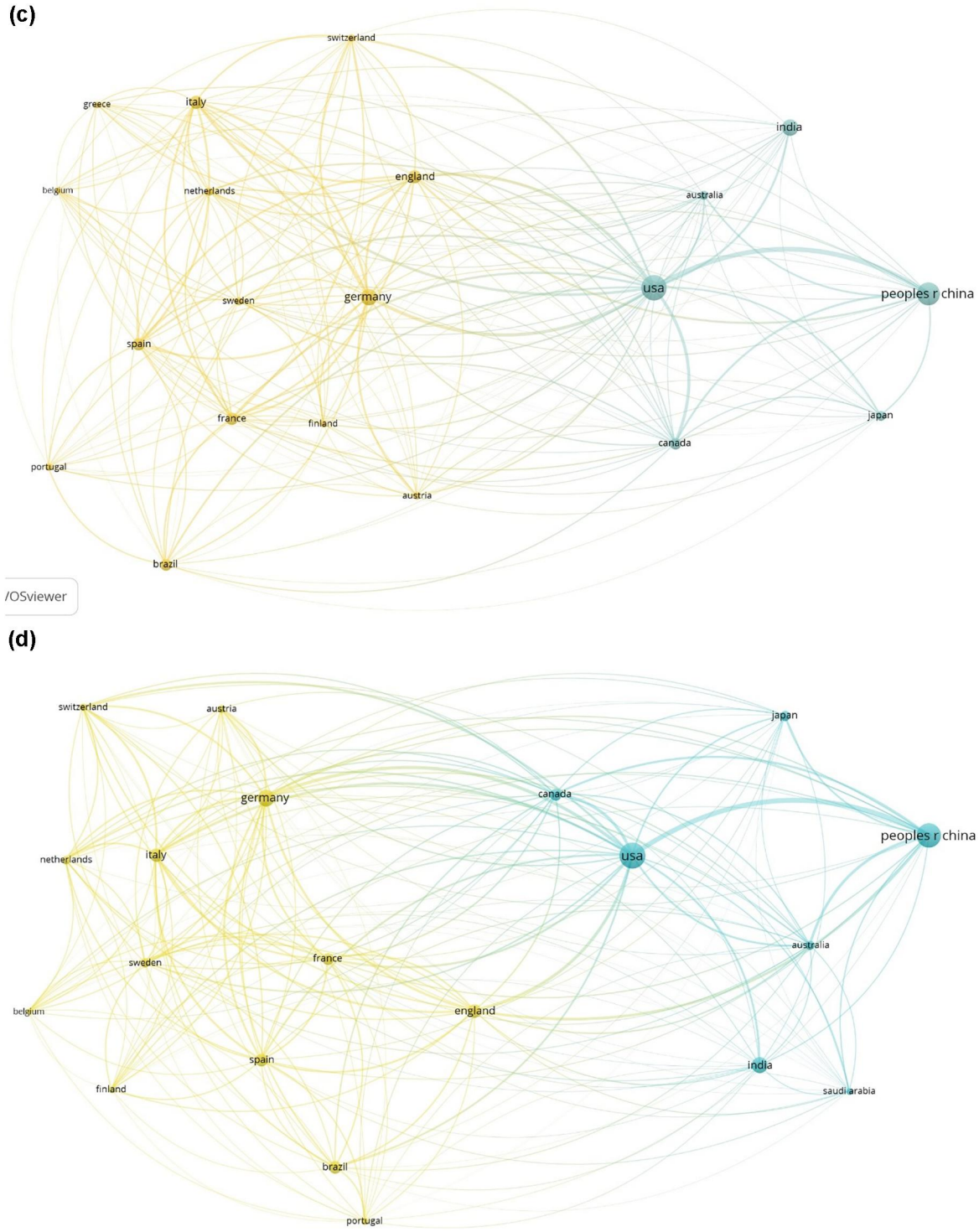

FIGURE 6 (Continued)

4.1.4 | Analysis on top 10 authors' scientific production

In Figure 5a–d, Bavota Gabriele is identified as one of the most active early stage researchers for the last 5 years whose earliest publication dates back to the year 2010.

Major contributions are made by some of the most prominent research scholars over the time duration of 13 years. The aforementioned figures also suggest that maintaining more or less same frequency of publication over longer periods of time is the key to be acknowledged as one of the leading researchers.

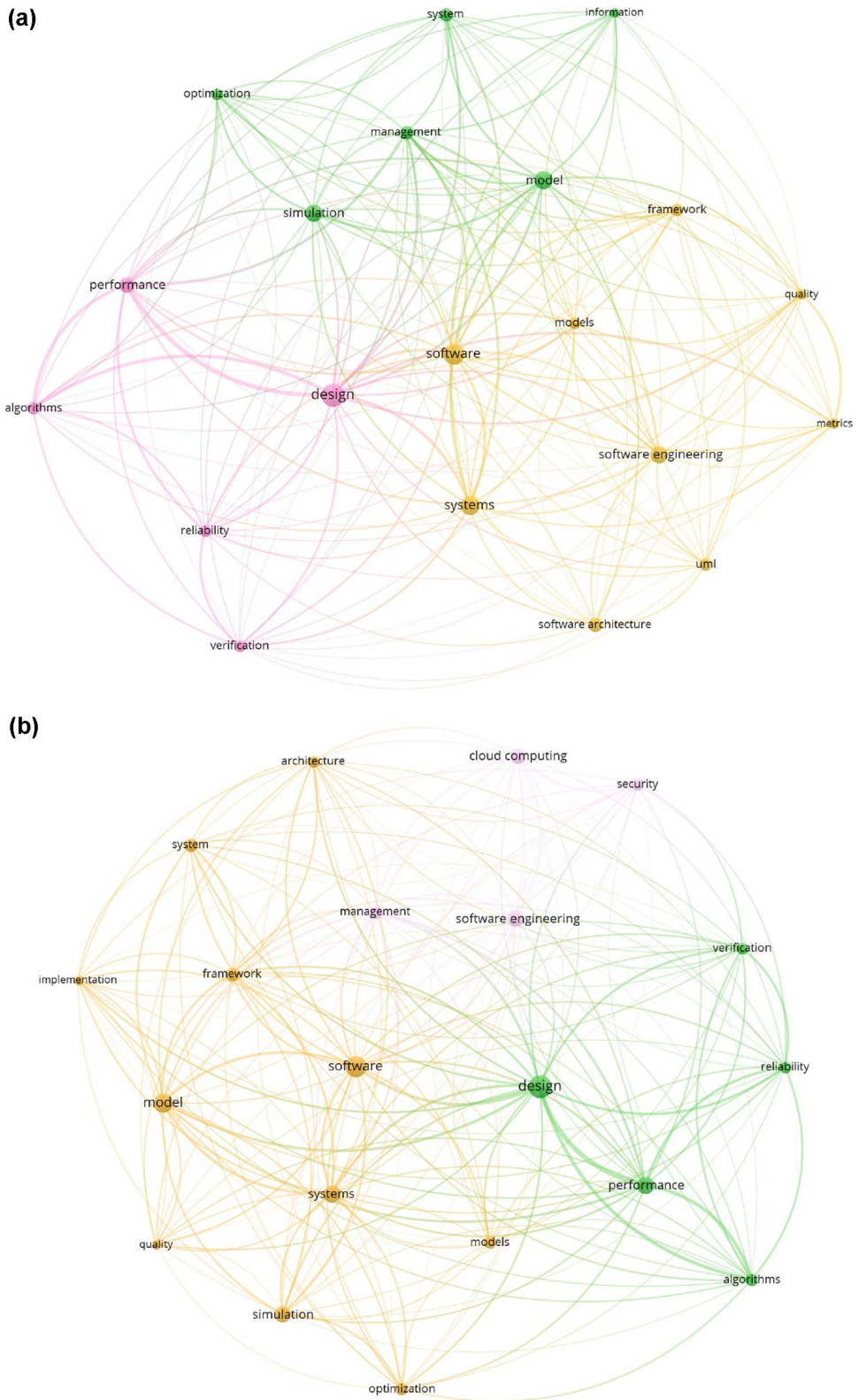

FIGURE 7 (a) Co-word analysis of top 20 words for the time duration (2007–2009). (b) Co-word analysis of top 20 words for the time duration (2010–2013). (c) Co-word analysis of top 20 words for the time duration (2014–2016). (d) Co-word analysis of top 20 words for the time duration (2017–2019)

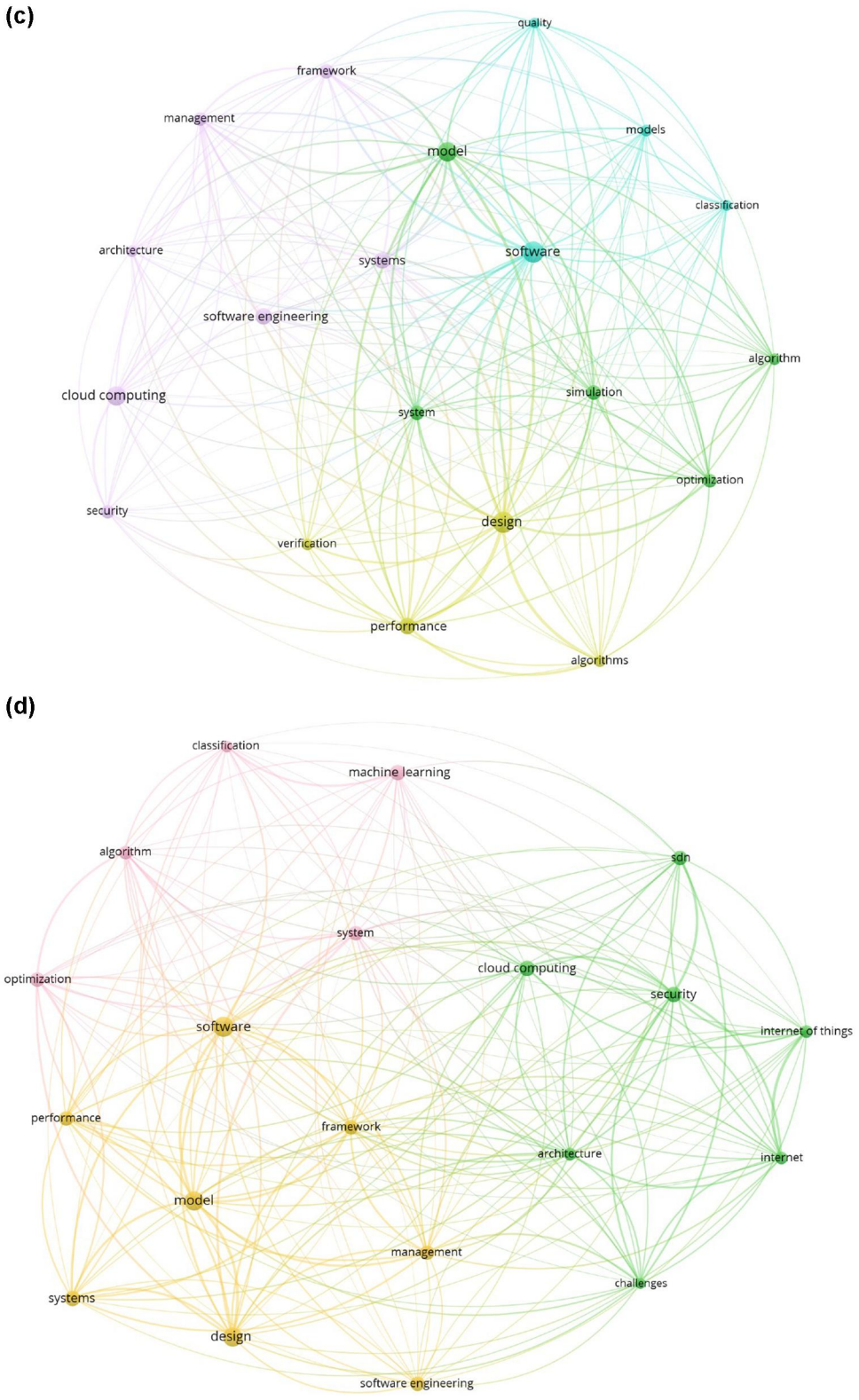

FIGURE 7 (Continued)

The analysis of Figure 5a–d represents the fact that the male research community dominates the research world with respect to the number of publications. For a time span (2007–2009), 78% of published articles belong to male research community as far as the list of top 10 scholars is concerned. Moreover, for two different time frames (2010–2013) and (2017–2019), the list of the top 10 scholars exhibits the absolute supremacy of male scholars with respect to number of scholarly publications.

4.1.5 | Analysis on collaboration network amongst countries

Collaboration is pivotal for a successful research in any field. Countries closely associated in research activities resulting in multiple co-authored scholarly documents belong to the same cluster where they are represented in the form of nodes in each cluster. However, since the world has become a global village, the research association is not only limited to one cluster, but also inter-clusters collaboration has become a reality and is the key for making research advancements in limited time. Generally, the countries possessing large collaborative networks tend to contribute more effectively as compared to others.

The research success of USA and China is apparent from the fact that both have one of the largest collaborative networks as depicted in Figure 6a–d.

4.1.6 | Co-word analysis

The degree of relatedness amongst articles is depicted by the occurrence of keywords and this relatedness factor forms the basis of grouping of keywords amongst different clusters (Figure 7a–d).

4.2 | Limitation of our research study and directions for future

4.2.1 | Limitations of data set and inclusion of all 19 types of documents

Our research data set incorporates only those research articles that are found in one of the most convenient and reliable databases, ISI WoS Core Collection. The possibility exists that ISI WoS database might not include some research studies. The combination of database is suggested to be evaluated in order to examine the excluded research articles. Furthermore, other inclusion criteria can be defined to enhance the data sets. We have included all 19 types of documents as listed in Table 7. Some researchers might not prefer to include all the listed documents (Table 7).

4.2.2 | Limited number of sub-areas

Our research work provides an insight into 12 sub-areas of SE with respect to some of the research questions articulated in Section 3.2. Inclusion of other SE domains is mandatory to perform a more detailed and comprehensive survey. Moreover, every sub-domain can be examined in accordance to the entire set of 6 research questions.

4.2.3 | Researchers' bias

Our study includes 85 multi-disciplinary fields refined on the grounds of the inclusion criterion discussed in Figure 1a,b. More critical analysis is required for the excluded research studies in accordance to the inclusion criterion (Figure 1a,b) to avoid slightest chance of researcher's bias.

Another important factor is the formulation of 6 RQs. Although, the RQs are designed after careful analysis to provide a comprehensive coverage of multiple parameters, the inclusion of other variables is possible.

4.2.4 | Future direction

We plan to extend our current work to perform analysis on top 10 institutions and top 10 countries with respect to entire 15 knowledge areas identified by SWEBOK [33].

5 | CONCLUSION

Our research study provides a structured and detailed bibliometric assessment of the SE domain over a period of 13 years. Despite the existence of various literature reviews and bibliometric analysis in the SE field, no research work provides such comprehensive and recent evaluation based on the key parameters (discussed in Table 6) in the SE field along with its prominent 12 sub-areas (as per Research Query 5 Section 3.1.1) including most of the knowledge areas of SWEBOK [33].

The uniqueness of our research study is evident from the fact that our research study incorporates a large dataset of *150,087* articles distributed across the non-identical time domains represented in Tables 2 and 3. The inevitable influence of the SE field on multiple domains leads to the inclusion of *85* multi-disciplinary areas based on the inclusion criteria (Figure 1a,b) specified in our research work. Due emphasis is being given to ensure that there is no researchers' bias in terms of languages of published documents by including research articles in language other than English. Our assessment further affirms the fact that a complete research study is not possible without the inclusion of published research contributions in languages other than English. A major shift in publication trend in SE and its sub-areas over a 13-year time duration is also observed in our work. Our research work not only presents an in-depth traditional bibliometric analysis but also includes evaluation on co-word and research collaboration networks of countries, which play a pivotal part in investigating the social aspects of research areas. Co-word assessment helps in analysing the degree of relatedness among multiple published scholarly articles. Social co-operation amongst various countries is evident by analysing collaboration networks. Analysis on the collaboration networks identified USA and People's Republic of China as two leading countries with largest research collaborative networks. Moreover, another important analysis of top 10

scholars' leads to the fact that majority of articles are published by male researchers. Maintaining consistency in terms of research papers for longer periods of times is the base to be spotted as one of the top scholars across the globe in terms of research publication count. Our research, to best of our knowledge, is novel in providing sound bibliometric assessment by including articles related to knowledge areas (10 out of 15) identified by SWEBOK V3.0. [33]. The statistical analysis used in our research study allowed us to obtain reliable indicators related to the quality of research. Trends in annual publication rate will help and encourage new researchers to publish in the field of SE in general and in various knowledge areas by SWEBOK in particular. Co-word assessment provided an insight into related terminologies related to the published research articles. Collaboration networks amongst countries provided insights into the collaborative networks of authors and their demographics.

ACKNOWLEDGEMENTS

We are extremely thankful to NED University of Engineering and Technology for providing us access to ISI Web of Science database, which is the pillar of our research study. Grant has been received by NED University of Engineering and Technology for our research work.

CONFLICT OF INTEREST

Dr. Muneera Bano is an associate editor of IET Software.

PERMISSION TO REPRODUCE MATERIALS FROM OTHER SOURCES

None.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on Web of Science. These data were derived from the following resources available in the public domain: Web of Science. Data available on request from the authors.

ORCID

Shehnila Zardari https://orcid.org/0000-0002-0525-3336

REFERENCES

1. Garousi, V.: A bibliometric analysis of the Turkish software engineering research community. Scientometrics. 105(1), 23–49 (2015)
2. Sommerville, I.: Software Engineering, ISBN-10. vol. 137035152, 9th ed, pp. 18. (2011)
3. Johnson, P., Ekstedt, M., Jacobson, I.: Where's the theory for software engineering? IEEE softw. 29(5), 96–96 (2012)
4. Roger, S.: Pressman. Software Engineering: A Practitioner's Approach, 7th ed, pp. 666–81. McGraw-Hill Higher Education, Vol. US (2010)
5. Wasserman, A.I.: Software engineering issues for mobile application development. In: Proceedings of the FSE/SDP Workshop on Future of Software Engineering Research. vol. 7, pp. 397–400. (2010)
6. Kelly, D.F.: A software chasm: software engineering and scientific computing. IEEE softw. 24(6), 120–119 (2007)
7. Garousi, V., Mäntylä, M.V.: Citations, research topics and active countries in software engineering: a bibliometrics study. Comput. Sci. Rev. 19, 56–77 (2016)
8. Ding, Y.: Scientific collaboration and endorsement: network analysis of coauthorship and citation networks. J.inf. 5(1), 187–203 (2011)
9. Garousi, V., Fernandes, J.M.: Highly-cited papers in software engineering: the top-100. Inf Software Technol. 71. 108-28 (2016)
10. Garousi V., Fernandes J.M.: Quantity versus impact of software engineering papers: a quantitative study. Scientometrics. 112(2), 963–1006 (2017)
11. Wohlin, C.: An analysis of the most cited articles in software engineering journals—1999. Inf Software Technol. 47(15), 957–964 (2005)
12. Wohlin, C.: An analysis of the most cited articles in software engineering journals-2000. Inf Software Technol. 49(1), 2–11 (2007)
13. Wohlin, C.: An analysis of the most cited articles in software engineering journals–2001. Inf Software Technol. 50(1-2), 3–9 (2008)
14. Garousi, V., Varma, T.: A bibliometric assessment of canadian software engineering scholars and institutions (1996–2006). Comput Inf Sci. 3(2), 19 (2010)
15. Glass, R.L.: An assessment of systems and software engineering scholars and institutions. J Syst Software. 27(1), 63–67 (1994)
16. Glass, R.L.: An assessment of systems and software engineering scholars and institutions (1995–1999). J Syst Software. 54(1), 77–82 (2000)
17. Glass, R.L., Chen, T.Y.: An assessment of systems and software engineering scholars and institutions (1996–2000). J Syst Software. 59(1), 107–113 (2001)
18. Glass, R.L., Chen, T.Y.: An assessment of systems and software engineering scholars and institutions (1997–2001). J Syst Software. 64(1), 79–86 (2002)
19. Glass, R.L., Chen, T.Y.: An assessment of systems and software engineering scholars and institutions (1998–2002). J Syst Software. 68(1), 77–84 (2003)
20. Glass, R.L., Chen, T.Y.: An assessment of systems and software engineering scholars and institutions (1999–2003). J Syst Software. 76(1), 91–97 (2005)
21. Tse, T.H., Chen, T.Y., Glass, R.L.: An assessment of systems and software engineering scholars and institutions (2000–2004). J Syst Software. 79(6), 816–819 (2006)
22. Ren, J., Taylor, R.N.: Automatic and versatile publications ranking for research institutions and scholars. Commun ACM. 50(6), 81–85 (2007)
23. Wong, W.E., et al.: An assessment of systems and software engineering scholars and institutions (2001–2005). J Syst Software. 81(6), 1059–1062 (2008)
24. Wong, W.E., et al.: An assessment of systems and software engineering scholars and institutions (2002–2006). J Syst Software. 82(8), 1370–1373 (2009)
25. Wong, W.E., et al.: An assessment of systems and software engineering scholars and institutions (2003–2007 and 2004–2008). J Syst Software. 84(1), 162–168 (2011)
26. Karanatsiou, D., et al.: A bibliometric assessment of software engineering scholars and institutions (2010–2017). J Syst Software. 147, 246–261 (2019)
27. Geist, R., et al.: Computing research programs in the US. Commun ACM. 39(12), 96–99. (1996)
28. Hamadicharef, B.: Scientometric study of the IEEE transactions on software engineering 1980–2010. In: Proceedings of the 2011 2nd International Congress on Computer Applications and Computational Science, pp. 101–106. Springer, Berlin (2012)
29. Fernandes, J.M.: Authorship trends in software engineering. Scientometrics. 101(1), 257–271 (2014)
30. Petersen, K., et al.: Systematic mapping studies in software engineering. In: 12th International Conference on Evaluation and Assessment in Software Engineering (EASE), pp. 1–10. (12 Jun 2008)
31. Garousi, V., Ruhe, G.: A bibliometric/geographic assessment of 40 years of software engineering research (1969–2009). Int J Software Eng Knowl Eng. 23(09), 1343–1366 (2013)
32. Mikki, S.: Comparing Google scholar and ISI Web of science for earth sciences. Scientometrics. 82(2), 321–331 (2010)
33. Bourque, P., (Dick) Fairley, R.E.: Guide to the Software Engineering Body of Knowledge (SWEBoK), v. 3.0. Accessed [05/07/2021]. www.swebok.org

34. Aria, M., Cuccurullo, C.: Bibliometrix: an R-tool for comprehensive science mapping analysis. Journal of informetrics. 11(4), 959–975 (2017)
35. Van Eck, N.J., Waltman, L.: Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 84(2), 523–538 (2010)
36. Davarpanah, M., Aslekia, S.: A scientometric analysis of international LIS journals: productivity and characteristics. Scientometrics. 77(1), 21–39 (2008)
37. Jung, J.: International research collaboration among academics in China and South Korea. Int J Chin Educ. 1(2), 235–254 (2012)
38. Van Eck, N.J., Waltman, L.: VOSviewer Manual, vol. 1, pp. 1–53. Universiteit Leiden, Leiden (2013 Dec 5)
39. Ravikumar, S., Agrahari, A., Singh, S.N.: Mapping the intellectual structure of scientometrics: a co-word analysis of the journal Scientometrics (2005–2010). Scientometrics. 102(1), 929–55 (2015)
40. Cambrosio, A., et al.: Historical scientometrics? Mapping over 70 years of biological safety research with coword analysis. Scientometrics. 27(2), 119–143 (1993)
41. Leung, X.Y., Sun, J., Bai, B.: Bibliometrics of social media research: a co-citation and co-word analysis. Int J Hospit Manag. 66, 35–45 (2017)
42. De la Hoz-Correa, A., Muñoz-Leiva, F., Bakucz, M.: Past themes and future trends in medical tourism research: a co-word analysis. Tourism Manag. 65, 200–211 (2018)
43. Corrales-Garay, D., Ortiz-de-Urbina-Criado, M., Mora-Valentín, E.M.: Knowledge areas, themes and future research on open data: a co-word analysis. Govern Inf Q. 36(1), 77–87 (2019)

How to cite this article: Alam, S., Zardari, S., Bano, M.: Software engineering and 12 prominent sub-areas: comprehensive bibliometric assessment on 13 years (2007–2019). IET Soft. 16(2), 125–145 (2022). https://doi.org/10.1049/sfw2.12046