

Quantity versus impact of software engineering papers: a quantitative study

Vahid Garousi^{1,2}  · João M. Fernandes³ 

Received: 6 February 2017 / Published online: 12 June 2017
© Akadémiai Kiadó, Budapest, Hungary 2017

Abstract According to the data from the Scopus publication database, as analyzed in several recent studies, more than 70,000 papers have been published in the area of Software Engineering (SE) since late 1960's. According to our recent work, 43% of those papers have received no citations at all. Since citations are the most commonly used metric for measuring research (academic) impact, these figures raise questions (doubts) about the (non-existing) impact of such a large set of papers. It is a reality that typical academic reward systems encourage researchers to publish more papers and do not place a major emphasis on research impact. To shed light on the issue of volume (quantity) versus citation-based impact of SE research papers, we conduct and report in this paper a quantitative bibliometrics assessment in four aspects: (1) quantity versus impact of different paper types (e.g., conference versus journal papers), (2) ratios of uncited (non-impactful) papers, (3) quantity versus impact of papers originating from different countries, and (4) quantity versus impact of papers by each of the top-10 authors (in terms of number of papers). To achieve the above objective, we conducted a quantitative exploratory bibliometrics assessment, comprised of four research questions, to assess quantity versus impact of SE papers with respect to the aspects discussed above. We extracted the data through a systematic, automated and repeatable process from the Scopus paper database, which we also used in two previous papers. Our results show that the distribution of SE publications has a major inequality in terms of impact overall, and also when categorized in terms of the above four aspects. The situation in the SE literature is similar to the other

✉ Vahid Garousi
vgarousi@gmail.com

João M. Fernandes
jmf@di.uminho.pt

¹ SnT center, University of Luxembourg, Luxembourg, Luxembourg

² Software Engineering Research Group, Department of Computer Engineering, Hacettepe University, Ankara, Turkey

³ Department of Informatics/Centro ALGORITMI, School of Engineering, University of Minho, Braga, Portugal

areas of science as studied by previous bibliometrics studies. Also, among our results is the fact that journal articles and conference papers have been cited 12.6 and 3.6 times on average, confirming the expectation that journal articles have more impact, in general, than conference papers. Also, papers originated from English-speaking countries have in general more visibility and impact (and consequently citations) when compared to papers originated from non-English-speaking countries. Our results have implications for improvement of academic reward systems, which nowadays mainly encourage researchers to publish more papers and usually neglect research impact. Also, our results can help researchers in non-English-speaking countries to consider improvements to increase their research impact of their upcoming papers.

Keywords Bibliometrics · Software engineering · Research impact · Countries · Authors · Exploratory study

Introduction

According to the data from the Scopus publication database, as analyzed in two recent studies (Garousi and Fernandes 2016; Garousi and Mäntylä 2016), more than 70,000 papers have been published in the area of Software Engineering (SE) since its inception in late 1960's. The SE research literature has grown tremendously, so there is a need for bibliometrics studies in this area. Bibliometrics is a set of methods to quantitatively analyze research literature. Citation analysis is one of the most widely used bibliometric methods to assess research impact, productivity and quality (Hamrick et al. 2010; King 2004; Moed 2006).

Citations are used to document sources of information, to acknowledge prior relevant research, and to substantiate claims. As such, citations play a key role in the evolution of knowledge and research impact (Hamrick et al. 2010). Citation analysis is widely used to quantify the impact of papers, scholars, journals, and even nations (King 2004; Moed 2006). Modern, formal use of citations in scientific literature dates back to the nineteenth century as scholars and scientists started to give continuity to their body of ideas (Hamrick et al. 2010). In 1955, Eugene Garfield published the Science Citation Index (SCI) (Garfield 1955), the first systematic effort to track citations in the scientific literature.

Under the rubric of bibliometrics, citation counts have been incorporated into various metrics intended to measure the research impact. Many countries are moving towards research policies that emphasize excellence; consequently, they develop evaluation systems to identify universities, research groups, and individual researchers that can be said to be 'excellent', which is usually measured by citation counts (Danell 2011). As the subject of research excellence has received increasing attention (in science policy) over the last few decades, an increasing number of bibliometric studies have been published dealing with and characterizing impact of papers in different disciplines (Bornmann 2014). Although a number of bibliometric studies have been published in SE, e.g., Wong et al. (2008, 2009) and Eric et al. (2011), there is a need for more recent bibliometrics studies and also studies covering larger pools of SE papers. Motivated by those needs, the goal of this study is to provide an overview of the publication and citation landscapes of the SE research literature, and to conduct an exploratory bibliometric assessment on quantity (number of) versus impact of the papers (as measured by citations), in the SE research

literature, from the point of view of the researchers in this area. Our study aims at answering a number of questions in this context, e.g.:

1. How do citations compare for different paper types (e.g., conference and journal papers) in SE?, e.g., do conference papers receive, on average, less citations compared to journal papers?
2. How does the quantity versus impact of SE papers relate for different countries?, i.e., does having a higher number of papers from a given country necessarily mean higher impact (citations) on the SE literature from that country?

The significance of our study and its results is that it touches on the issues of quantity (i.e., number of) versus impact of publications, focusing in the area of SE. With controversial discussions in all research communities about “least publishable unit (LPU)” (Lawrence 2003) and phrases such as “getting more for less” (Broad 1981) since early 1980’s and “stop the numbers game” (Parnas 2007), there is a need, more than ever, to raise and analyze the issue of quantity versus impact of publications in SE. As Sarewitz mentions: “The pressure to publish pushes down quality” (Sarewitz 2016). We are observing that some researchers opt for more publications with marginal impact or quality, while some researchers prefer to publish less, but higher impactful or higher quality papers [the so-called ‘scientometric bubble’ phenomenon (Génova et al. 2016)]. We want to quantitatively assess to what extent these issues are the case in the SE community.

In the bibliometrics and scientometrics literature, there have been many studies on assessing quantity versus impact of papers. However, no similar study has been conducted in the SE literature yet. While a very large number of SE papers have been published, it is natural to question the differences in the impact of different papers as measured by the number of citations they have received. Also, we raise and answer this question in the context of different paper types (conference versus journal papers), different countries and also distinct authors.

The current paper is a follow-up to two recent works by the authors and their colleagues (Garousi and Fernandes 2016; Garousi and Mäntylä 2016). The study in Garousi and Fernandes (2016) identified the top-100 highly-cited papers in SE, while the study in Garousi and Mäntylä (2016) characterized the citations and research topics landscape of SE. The current study differs from Garousi and Fernandes (2016) and Garousi and Mäntylä (2016) in that this study takes a different and novel goal by assessing the issue of quantity versus impact of SE papers.

The remainder of the paper is organized as follows. “[Background and related work](#)” section discusses the background and related work. “[Research method](#)” section describes our research method, including the goal and research questions tackled in this study. “[Goal and research questions](#)” Section presents the results of the study. “[Results](#)” section summarizes the findings and implications, and discusses the potential threats to validity of our study. Finally, “[Discussions](#)” section concludes this study and points out the future work directions.

Background and related work

We review two categories of related work next: (1) bibliometrics studies on quantity versus impact of papers, and (2) bibliometrics studies in SE. Since our study partially builds on top of our recent works (Garousi and Fernandes 2016; Garousi and Mäntylä 2016;

Fernandes 2014), we then briefly review them afterwards. We then review the different views in the community on using citations as an indicator of research impact.

Bibliometrics studies in SE

To the best of our knowledge, there is no bibliometrics study analyzing quantity versus impact of SE papers. Thus, we only review the bibliometrics studies in SE, which are quite common in SE. Table 1 lists a few representative studies along with their notable findings.

The series of 12 papers by Glass and Chen, three of which (Wong et al. 2008, 2009; Eric et al. 2011) are cited in Table 1, was an ongoing, annual initiative that aimed at identifying the top 15 scholars and institutions in systems and software engineering, for a sliding 5-year period between 1995 and 2006. The rankings were based on the number of papers published in a selected set of leading SE journals.

The study reported in Garousi and Varma (2010) presents a bibliometric assessment of Canadian SE scholars and institutions. Additional findings reported in Garousi and Varma (2010) include a correlation analysis of the SE research productivity (output in terms of number of papers) of Canadian provinces versus their national research grant amounts. Focusing on specific sub-areas under SE, the study reported in de Freitas and de Souza (2011) presents a bibliometric analysis of ten years of search-based SE.

Some recent systematic mapping studies report, as a part of their studies, bibliometric analyses of SE sub-areas, e.g., development of scientific software in Farhoodi et al. (2013). Among the findings reported in Farhoodi et al. (2013) is that the most active authors in the area of development of scientific software were mostly located in the US (approximately 50%), followed by the Canadian and the British researchers.

The study reported in Garousi and Ruhe (2013) is a bibliometric/geographic assessment of 40 years of SE research (1969–2009) in which the entire set of 26,624 SE papers indexed by the ISI Web of Knowledge were studied to find the most active countries.

Fernandes (the second author of the current paper) reports in Fernandes (2014) a bibliometric study which focuses on authorship trends in SE. Around 70,000 entries from the DBLP (a well-known online computer science bibliography website) for 122 conferences and journals, for the period 1971–2012, were collected. The results indicate that the number of authors of articles in SE is increasing on average around 0.40 authors per decade. Also, the results indicate that until 1980, the majority of the articles have one author, while articles with 3 or 4 co-authors, published from the 1990s until today, represent almost half of the total number of papers. Since the average number of authors of scientific articles is increasing, it was the opinion of the researcher that the system of authorship is becoming inappropriate, in the sense that it is more difficult to credit all the authors for the specific contributions they made to each article. Therefore, Fernandes suggests that the SE community must establish an agreed publishing standard to define how to assign the academic contribution to all collaborators of a research project.

Garousi (the first author of the current paper) recently conducted and published a bibliometric assessment of Turkish software engineering scholars and institutions covering years 1992–2014 (Garousi 2015). Among the results are that: (1) Turkey produces only about 0.5% of the world-wide SE knowledge, as measured by the number of papers in Scopus, which is very negligible; (2) there is a lack of diversity in the general SE spectrum in Turkey, e.g., the author noticed very little focus on requirements engineering, software maintenance and evolution, and architecture. This denotes the need to further diversification in SE research topics in Turkey; and (3) in total, 89 papers in the pool (30.8% of the total) are internationally-authored SE papers. Having a good level of international

Table 1 A few selected bibliometrics studies in SE (sorted by years of publications)

Ref.	Year	Topic	Notable findings
Wong et al. (2008)	2008	An assessment of systems and software engineering scholars and institutions (2001–2005)	The rankings are calculated based on the number of papers published in journals: IEEE TSE, TOSEM, JSS, SPE, EMSE, IST, and IEEE Software The top scholar is Magne Jørgensen of Simula Research Laboratory, Norway The top institution is Korea Advanced Institute of Science and Technology, Korea
Wong et al. (2009)	2009	An assessment of systems and software engineering scholars and institutions (2002–2006)	The top-ranked scholar is Magne Jørgensen of Simula Research Laboratory, Norway The top-ranked institution is Korea Advanced Institute of Science and Technology, Korea
Garousi and Varma (2010)	2010	A bibliometric assessment of Canadian software engineering scholars and institutions (1996–2006)	The study uses two metrics: impact factors, and h-index, based on papers published in top 12 selected software engineering journals and conferences The top-ranked institution is Carleton University The top-ranked scholars (by each of the two metrics) are Lionel Briand (formerly with Carleton University) and Gail Murphy (from UBC)
de Freitas and de Souza (2011)	2011	Ten years of search-based software engineering: a bibliometric analysis	The study covers 740 publications of the SBSE community from 2001 through 2010 The performed bibliometric analysis concerned mainly in four categories: publication, sources, authorship, and collaboration. The study also analyzed the applicability of bibliometric laws in SBSE, such as Bradfords and Lotka
Eric et al. (2011)	2011	An assessment of systems and software engineering scholars and institutions (2003–2007 and 2004–2008)	The top-ranked institution is Korea Advanced Institute of Science and Technology, Korea for 2003–2007, and Simula Research Laboratory, Norway for 2004–2008 Magne Jørgensen is the top-ranked scholar for both periods
Farhoodi et al. (2013)	2011	Development of scientific software: a systematic mapping, bibliometrics study and a paper repository	17 out of 130 publications in the pool were cited more than 25 times The most active author in the field is Diane Kelly, with Royal Military College of Canada, with a total of ten (co-authored) publications The authors' most frequent affiliations are located in the US (approximately 50%), followed with a large distance by Canada and the UK

Table 1 continued

Ref.	Year	Topic	Notable findings
Garousi and Ruhe (2013)	2013	A bibliometric/geographic assessment of 40 years of software engineering research (1969–2009)	<p>The first bibliometric quantitative analysis of publications in SE, including relative and absolute growth in the number of all SE publications as well as an analysis among countries</p> <p>Over the 40-year period (1969–2009), in total about 60% of the SE literature has been contributed by only 7% of all countries</p> <p>The US is the clear leader, followed by UK and China</p> <p>The SE research output of different countries does not necessarily correlate with their GDPs</p> <p>The share of contributions to the SE discipline by the American researchers has declined from 71.43% (in 1980) to 14.90% (in 2008)</p> <p>China is the country with the biggest share growth in the number of SE publications (from 0.82% of the entire SE publications in 1991 to 13.82% in 2009)</p>
Fernandes (2014)	2014	Authorship trends in SE	<p>Around 70.000 entries from the DBLP for 122 conferences and journals, for the period 1971–2012, were collected</p> <p>The number of authors of articles in SE is increasing on average around 0.40 authors/decade</p> <p>Until 1980, the majority of the articles have one author, while articles from 90 s until today with 3 or 4 authors represent almost half of the total number of papers</p>
Garousi (2015)	2015	Bibliometric assessment of Turkish software engineering scholars and institutions (1992–2014)	<p>Turkey produces only about 0.5% of the world-wide SE knowledge, as measured by the number of papers in Scopus, which is very negligible unfortunately</p> <p>There is a lack of diversity in the general SE spectrum in Turkey, e.g., we noticed very little focus on requirements engineering, software maintenance and evolution, and architecture. This denotes the need to further diversification in SE research topics in Turkey.</p> <p>In total, 89 papers in the pool (30%) are internationally-authored SE papers. Having a good level of international collaborations is a good sign for the Turkish SE community</p>

Table 1 continued

Ref.	Year	Topic	Notable findings
Garousi and Fernandes (2016)	2016	Highly-cited papers in software engineering: The top-100	<p>A study, comprised of five research questions, to identify and classify the top-100 highly-cited SE papers in terms of two metrics: total number of citations and average annual number of citations</p> <p>By total number of citations, the top paper is “A metrics suite for object-oriented design”, cited 1817 times and published in 1994. By average annual number of citations, the top paper is “QoS-aware middleware for Web services composition”, cited 154.2 times on average annually and published in 2004</p> <p>It was concluded that it is important to identify the highly-cited SE papers and also to characterize the overall citation landscape in the SE field. It was hope that this paper would encourage further discussions in the SE community towards further analysis and formal characterization of the highly-cited SE papers, as it has been done in other fields</p>
Garousi and Mäntylä (2016)	2016	Citations, research topics and active countries in SE	<p>The number of SE papers published per year has grown tremendously and, as of 2015, about 6000 to 7000 papers are published every year</p> <p>Out of all the 71,668 SE papers in the pool indexed in the Scopus publication database, 30,958 papers (~ 43% of the pool) had no citations at all</p> <p>Using text mining of articles titles, we found that currently the hot research topics in SE are: (1) web services, (2) mobile and cloud computing, (3) industrial (case) studies, (4) source code and (5) test generation</p> <p>A small share of large countries produce the majority of the papers in SE while small European countries are proportionally the most active in the area of SE, based on the number of papers</p>

collaborations is a good sign for the Turkish SE community. The current article follows the same bibliometric approach as was conducted in Garousi (2015) to extract all the SE papers (details are discussed in “[Research method](#)” section).

Garousi and Fernandes (2016) conducted and reported in 2016 a bibliometric assessment to identify the top-100 highly-cited papers in SE in terms of two metrics: total number of citations and average annual number of citations. These two researchers argue that, as the subject of research excellence has received increasing attention (in science

policy) over the last few decades, increasing numbers of bibliometric studies have been published dealing with characterizing and ranking highly-cited papers (Bornmann 2014). For example, the cover story of the October 2014 issue of the prestigious *Nature* magazine was “The top 100 papers” (Noorden et al. 2014). That *Nature* issue includes several papers (e.g., Ioannidis et al. 2014) on the issue of highly-cited papers in various scientific disciplines. Garousi and Fernandes (2016) report, among other things, that: by total number of citations, the top paper is “A metrics suite for object-oriented design” (Chidamber and Kemerer 1994), cited 1817 times and published in 1994. By average annual number of citations, the top paper is “QoS-aware middleware for Web services composition” (Zeng et al. 2004), cited 154.2 times on average annually and published in 2004. The researchers also identify works pointing out possible determinants of the likelihood of high citations, e.g., based on a paper entitled “Highly-cited works in neurosurgery” (Ponce and Lozano 2010), the determinants are: the time of publication, field of study, nature of the work, and the journal in which the work appears. One would wonder if those determinants are also applicable in the SE domain. For example, it seems that publishing in *IEEE Transactions on Software Engineering* increases the chances of the highest impact, as 47% of the most cited papers in SE were published in that journal.

Garousi and Mäntylä (2016) report another bibliometrics study of citations, research topics and active countries in SE. They found that the number of SE papers published per year has grown tremendously and, as of 2015, about 6000–7000 papers are published every year. Furthermore, they found that nearly half of the SE papers (43%) are not cited at all.

Our recent work in citation analysis of the software engineering literature

Since our study partially builds on top of two of our recent works (Garousi and Fernandes 2016; Garousi and Mäntylä 2016) (as also discussed in Table 1), we briefly review some findings from them next.

We extracted in Garousi and Fernandes (2016) and Garousi and Mäntylä (2016) the entire pool of SE papers from Scopus. In terms of the growth of the SE literature, Fig. 1 shows the number of SE papers included in Scopus by their publication year. The earliest publication year was 1972 from which 29 papers are included in Scopus. The annual numbers of papers have grown in the latest years (starting around 2005) and have reached around 6500 papers each year since 2008.

Observing such increasing trends have made the authors curious about the potential root causes and associated implications. Justifying some of these trends (e.g., major increase in number of papers from 2005 until 2008 in Fig. 1) is by no means trivial and would be speculative and hypothetical since no precise data/insights are available at this point. Many possible root causes may be playing roles in this case, e.g., (1) increase in the number of SE-specific venues and also the number of papers in venues starting 2005, (2) changes in the choice of venues included in the Scopus database, and (3) a spike in the number of researchers and research students trained in SE. Follow-up studies can aim at justifying those trends in more depth, e.g., counting the number of SE-specific venues held annually.

In Garousi and Fernandes (2016) and Garousi and Mäntylä (2016), we automatically extracted the pool of 71,668 papers, along with their citation counts, from Scopus into a CSV file. Figure 2 shows a scatter plot of all the papers’ citation counts versus publication years, along with the corresponding box plots. Note that on this graph there are theoretically 71,668 points, as many as the papers in the pool. The data in the X-axis (publication years) are somewhat skewed, while the Y-axis data (citations) are extremely skewed. As a

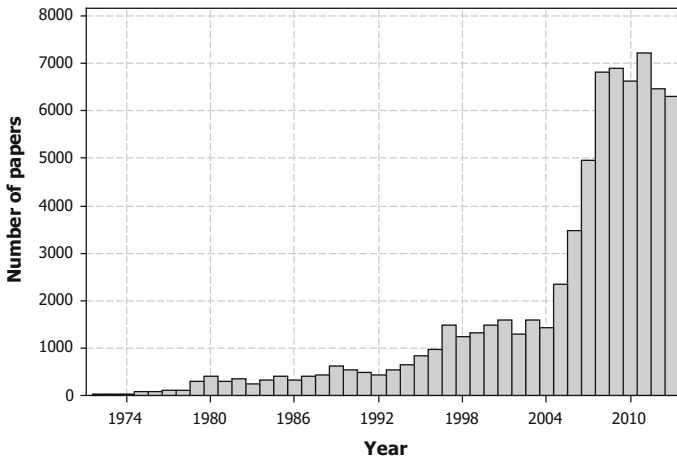


Fig. 1 Number of SE papers included in Scopus by their publication year. Adopted from Garousi and Fernandes (2016) and Garousi and Mäntylä (2016)

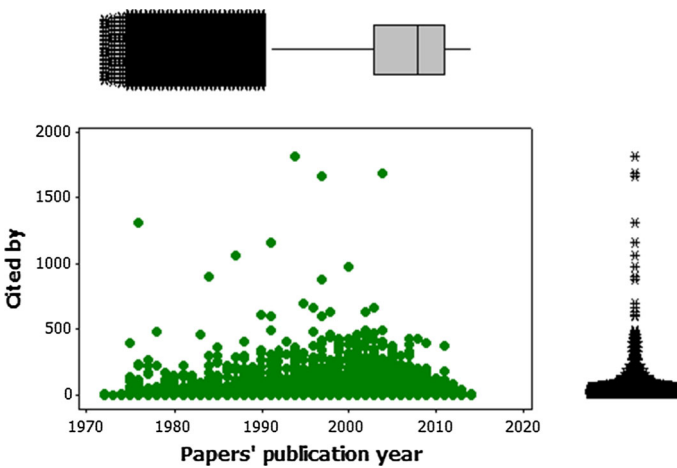


Fig. 2 Scatter plot of all the 71,668 SE papers' publication years and citations, along with their box plots. Adopted from Garousi and Fernandes (2016) and Garousi and Mäntylä (2016)

consequence, the two box plots in the top and right of Fig. 2 have a very large number of ‘outliers’ shown as ‘*’.

Out of all the 71,668 SE papers in the pool indexed in Scopus, 30,958 papers (~43% of the pool) had no citations at all. 10,095 papers (~14% of the pool) had only one citation. On the other hand, 30,615 papers (~43% of the pool) had received more than one citation. The sum of all the citation numbers is 448,050. Thus, the average citation value is 6.82 per paper. The highest cited paper was cited 1817 times (to be discussed in further detail in “RQ 2 Quantity versus impact: by venues” section). Figure 3 shows the histogram of the citation data for all the SE papers (the outlier higher citation values, more than 400, have been cropped for the brevity of the graph).

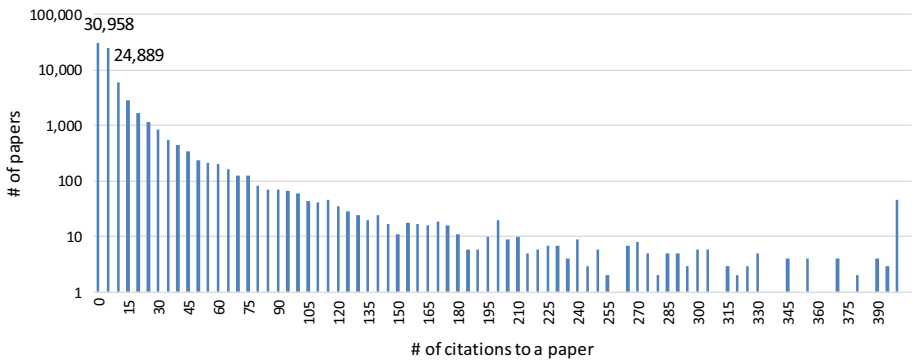


Fig. 3 Histogram of citation data for all the SE papers included in Scopus. Adopted from Garousi and Fernandes (2016) and Garousi and Mäntylä (2016)

Bibliometrics studies on quantity versus impact of papers

In the bibliometrics and scientometrics literature, there is a body of studies which assess quantity versus impact of papers. We randomly selected a few of those works (Sandström and van den Besselaar 2016; Allison and Stewart 1974; Merton 1968; Cole et al. 1978; Halfman and Leydesdorff 2010; Ghosh et al. 2014; Rahm 2008; Freyne et al. 2010) and review them next.

Entitled “Quantity and/or quality? The importance of publishing many papers”, a recent study (Sandström and van den Besselaar 2016) assessed the following question: Do highly productive researchers (those publishing high quantity of papers) have significantly higher probability to produce top-cited papers? Or do high productive researchers mainly produce a sea of irrelevant papers? The authors of Sandström and van den Besselaar (2016) argue that the answer on the above questions is important, as it may help to answer the question of whether the increased competition and increased use of indicators for research evaluation and accountability focus has perverse effects or not. The study used a Swedish author disambiguated dataset consisting of 48.000 researchers and their publications in the Web of Science (WoS) database during the period of 2008–2011 with citations until 2014 to investigate the relation between productivity and production of highly cited papers. As the analysis showed, quantity does make a difference, i.e., researchers publishing high quantity of papers indeed have a higher probability to produce top-cited papers.

Allison and Stewart (1974) demonstrated that citations to papers are more unequally distributed than counts of publications themselves. While such extreme inequality is of interest in itself, several researchers have examined variations in inequality in order to test theories about social processes in science. For example, the closely related notions of the Matthew effect¹ and cumulative advantage suggest that inequality of both publications and citations ought to increase as a cohort of scientists grows older (Merton 1968). On the other hand, (Cole et al. 1978) used the inequality of citations to papers and to persons as an indicator of the level of consensus within disciplines and research areas.

¹ In sociology, the Matthew effect (or accumulated advantage) is the phenomenon where “the rich get richer and the poor get poorer”.

A 2010 study (Halffman and Leydesdorff 2010) used the Gini coefficient² to assess whether the inequality (in terms of number of publications and citations) among universities are increasing or not. The results of the study did not support the thesis that universities are becoming more unequal.

More recently, in 2014, two new metrics (based on Gini and k-indices) were proposed (Ghosh et al. 2014) and were used to analyze the citation inequality in disciplines, academic institutions and science journals. Both of these measures suggested a universal nature of academic inequalities in terms of citations.

A 2008 study (Rahm 2008) assessed the number of papers and citations for the top-100 venues (journal and conference) in Computer Science (CS) based on data from Google Scholar and Web of Science, and found that the average citations for journal and conference papers are similar (7.5 and 7.3, respectively).

Another 2010 bibliometrics study (Freyne et al. 2010) reported that citations represent a trustworthy measure of CS research quality, whether in articles in conference proceedings or in CS journals. The study confirmed, by quantitative means, the belief among computer scientists that conference publications enjoy greater status in CS than in other disciplines. The study did not use the citation counts of individual papers but the impact factor of the venues to derive the conclusions. Another interesting finding of the study was the following: “The view that conference rejection rates are a good proxy for conference quality did not hold up to scrutiny, reflecting a low coefficient of correlation between the rejection rate of conferences and their Google Scholar scores”.

More recently, in the context of altmetrics, there have also been efforts assessing quantity versus impact of papers, e.g., Buttlere and Buder (2017). Altmetrics are non-traditional metrics proposed as an alternative to more traditional citation metrics. Altmetrics also cover other aspects of the impact of a work, such as how many data and knowledge bases refer to it, article views, downloads, or mentions in social media and news media. The paper in Buttlere and Buder (2017) compares papers ‘quality’/‘impact’ to persons’ ‘intelligence’/‘personality’ in the context of altmetrics.

On using citations as an indicator of research impact

Citations are the most widely used means to assess research impact. While the issue of using citations as a metric for research impact is seen as controversial by some researchers, many other researchers accept this metric albeit its limitations. There is a debate in the community on this controversial subject, e.g., Parnas (2007), Nieminen et al. (2006), Reuters (2016) and Saha et al. (2003). We discuss some of the views on this subject based on the literature.

A 2006 paper entitled ‘*The relationship between quality of research and citation frequency*’ (Nieminen et al. 2006), published in the *Medical Research Methodology* journal, assessed whether statistical reporting and statistical errors in the analysis of the primary outcome are associated with the number of citations received. The authors evaluated all original research articles published in 1996 in four psychiatric journals. The impact of each paper was assessed and the number of citations received until 2005 was obtained from the Web of Science database. The authors then examined whether the number of citations was associated with the quality of the statistical analysis and reporting. The study found that

² The Gini coefficient (sometimes expressed as a Gini ratio or a normalized Gini index) is a measure of statistical dispersion intended to represent the income distribution of a nation’s residents, and is the most commonly used measure of inequality.

extended description of statistical procedures had a positive effect on the number of citations received. In the considered cohort of published research, measures of reporting quality and appropriate statistical analysis were not associated with the number of citations. The journal in which a study is published appears to be as important as the statistical reporting quality in increasing the number of citations.

The Highly-Cited Researchers (HCR) project (Reuters 2016) by Thomson Reuters is another indicator of popularity and acceptance of citations as an important metric for ranking and assessing the success of researchers in the community.

In a paper entitled '*Stop the Numbers Game*' (Parnas 2007), Parnas warns that counting papers slows down the rate of scientific progress. He also criticizes using citation values solely as indicators of research impact. He mentioned that: "Some citations are negative. Others are included only to show that the topic is of interest to someone else or to prove that the author knows the literature. Sometimes authors cite papers they have not studied; we occasionally see irrelevant citations to papers with titles that sound relevant but are not. Finally, the importance of some papers is not recognized for many years. A low citation count may indicate a paper that is so innovative it was not initially understood". There are also others who use objective discussions to disagree with Parnas' opinion. For example, Grigore (2007) argues with Parnas' proposal for an alternative strategy for research impact evaluations as he suggests: "When serving on recruiting, promotion, or grant-award committees, read the candidate's papers and evaluate the contents carefully". In other words, Parnas suggests the evaluation should be a thorough review. Here is what Parnas says about reviews in Parnas (2007): "Anyone with experience as an editor knows there is tremendous variation in the seriousness, objectivity, and care with which referees perform their task". In conclusion, Grigore (2007) objects Parnas' opinion and suggestions and believes that there is a contradiction: "The same observation applies to Parnas' proposed solution!".

There are also various forum-like online sources in which various groups of researchers have argued in favor and against judging a paper by citation count, e.g., Various authors (2016). In this particular sources, scientific studies on the subject were also discussed, e.g., Siler et al. (2015). In the latter study, arguments in favor of using paper citations were summarized as follows. Scientists cite work for a myriad of reasons. However, the vast majority of citations are either positive or neutral in nature. The authors worked with the assumption that scientists prefer to build upon other quality research with their own work. As Latour and Woolgar (1979) suggest, citation is an act of deference, as well as the means by which intellectual credit and content flows in science. Relatedly, the authors also assume that most scientists want to produce quality work and will seldom attempt to garner credit and attention by blatantly doing bad work. Thus, on the whole, "the attention and impact associated with citations provides a reasonable measure of quality. Citations provide an objective and quantitative measure of credit and attention flows in science". Many researchers agree that citations constitute a more solid indicator to measure quality (or alternatively impact, relevance, or popularity) of a given paper than, for instance, the impact factor of the journal in which that paper is published (Latour and Woolgar 1979). In particular, an excellent, self-contained, and clearly-written paper is more likely to be more cited than papers that just publish a minimum part of a research result. In fact, today we are witnessing an article-inflation phenomenon, a scientometric bubble as indicated in Génova et al. (2016), where many authors prefer to have more papers than impactful or higher-quality papers.

In Siler et al. (2015), arguments against using paper citations for assessing paper quality are summarized as follows. Because citations are often distributed exponentially, with a few articles garnering disproportionate attention (Lotka 1926), some researchers also used the logarithm of citation counts as a dependent variable to diminish the potential influence

of a few highly-cited outlier articles. The other factor is the ‘social status’ effect. Scientists often rely on heuristics to judge quality; status of scholars, institutions, and journals are common means of doing so. Unsurprisingly, citations received by manuscripts were positively correlated with the impact factor of the journal in which it was published.

To further show that citation and other metrics based on it are widely used to assess the quality of work by researchers and journals, Fig. 4 shows screenshots from credible online sources which use citation and other metrics based on it (e.g., *h*-index and *i10*-index) to highlight the portfolio and impact of researchers.

Paper citations are also used in aggregate forms, e.g., impact factor of academic journals and researchers’ *h*- and *g*-indexes. The impact factor of an academic journal is a measure reflecting the annual average number of citations to recent articles published in that journal. The *h*- and *g*-indexes are author-level metrics that attempt to measure both the productivity and citation impact of the publications of a scientist or scholar. These indexes are based on the ordered list of the scientist’s most cited papers and the number of citations that they have received in other publications.

In a paper entitled ‘Impact factor: a valid measure of journal quality?’ (Saha et al. 2003), the authors assess the validity of the impact factor as a measure of quality for

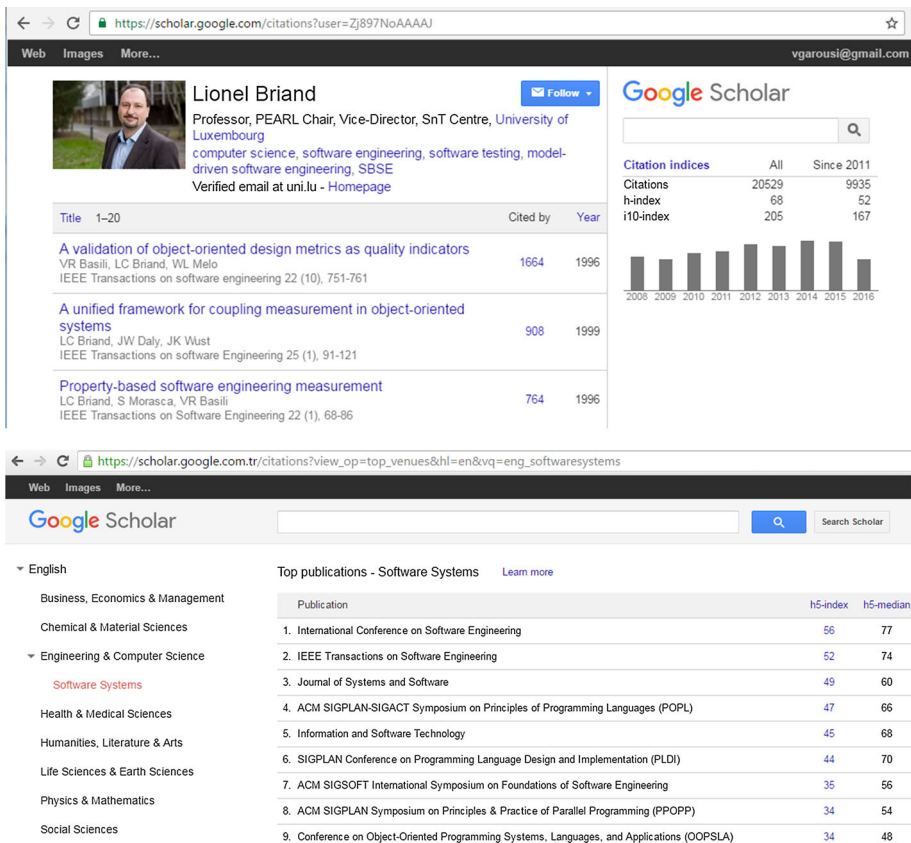


Fig. 4 Screenshots from credible online sources which use citation and metrics based on it to assess the quality of work by researchers and journals

general medical journals by testing its association with journal quality as rated by clinical practitioners and researchers. The authors conclude that “Impact factors may be a reasonable indicator of quality for general medical journals”, thus supporting the notion of using citations as an indicator of research quality.

As discussed above, the majority of the research community, at large, still thinks the main metric for research impact is the number of citations a paper receives, e.g., the following quote from a recent bibliometric paper (Mingers and Lipitakis 2010): “Assessing the quality of the knowledge produced by business and management academics is increasingly being metricated. Moreover, emphasis is being placed on the impact of the research rather than simply where it is published. The main metric for impact is the number of citations a paper receives”. Also, major reports and studies are regularly prepared and circulated among the funding agencies in both national and international levels on measuring research performance by citations, e.g., a report prepared for the Netherlands organization for scientific research on bibliometric indicators of research performance in computer science (Moed and Visser 2007).

Since “Citation measures neglect impact outside the academy” (Priem et al. 2016) and to address that limitation, more recently, newer more modern metrics have been proposed to assess impact of research works beyond just academia. ‘Altmetrics’ are the most popular set of metrics in this domain which are non-traditional metrics proposed as an alternative to more traditional citation-based impact metrics (Priem et al. 2016; Piwowar 2013). The term altmetrics was proposed in 2010, as a generalization of article level metrics, and has its roots in the *#altmetrics* hashtag. In addition to citations, Altmetrics cover other aspects of the impact of a work, such as how many data and knowledge bases refer to it, article views, downloads, or mentions in social media and news media.

Although altmetrics are often thought of as metrics about articles, they can be applied to people, journals, books, data sets, presentations, videos, source code repositories, web pages, etc. They are related to Webometrics, which had similar goals but evolved before the social web.

Sources such as LibGuides at Duke University Medical Center (2016) believe that “Like citations, altmetrics are measures of attention, not quality. (But altmetrics come much closer to indicating quality than citations currently do).” However, altmetrics is still a relatively young field, and research is still needed into the motivations that cause others to bookmark, share, blog about, and otherwise discuss scholarship online. That said, much more research is needed before any accurate measures of quality can be confidently used. Also, “Altmetrics are meant to supplement citations, not replace them” (2016). But a major limitation of altmetrics is that they are only measurable for “recent papers” (Altmetric LLP 2016), i.e., old papers are not usually actively ‘discussed’ on the web, e.g., in Twitter or news articles. Research works are usually discussed in online sources when they are published (released).

As an example output of the Altmetric web service (www.altmetric.com), Fig. 5 shows an example Altmetric report of a SE paper (Garousi et al. 2016) which reports that the paper has been ‘mentioned by’ 7 tweets and has had 8 readers on the Mendeley online reference manager (www.mendeley.com). The report includes both quantitative and qualitative pieces, e.g., “In the top 25% of all research outputs scored by Altmetric”, “Among the highest-scoring outputs from this source (#17 of 324)”, and “Good Attention Score compared to outputs of the same age (79th percentile)”.

As another output of the Altmetric web service, Fig. 6 depicts a screenshot of the list of top 100 articles in 2015 under the ‘Information and Computer Sciences’ subject area. The

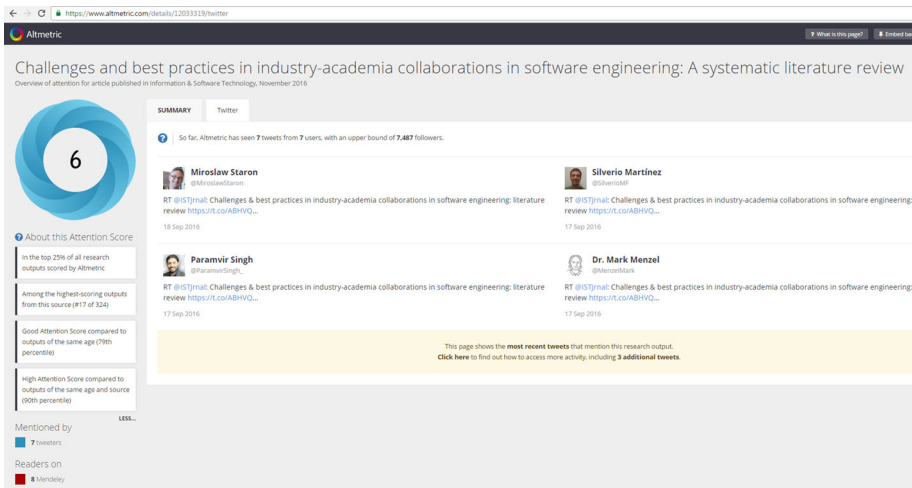


Fig. 5 Screenshots from the Altmetric report for a SE paper (Garousi et al. 2016)

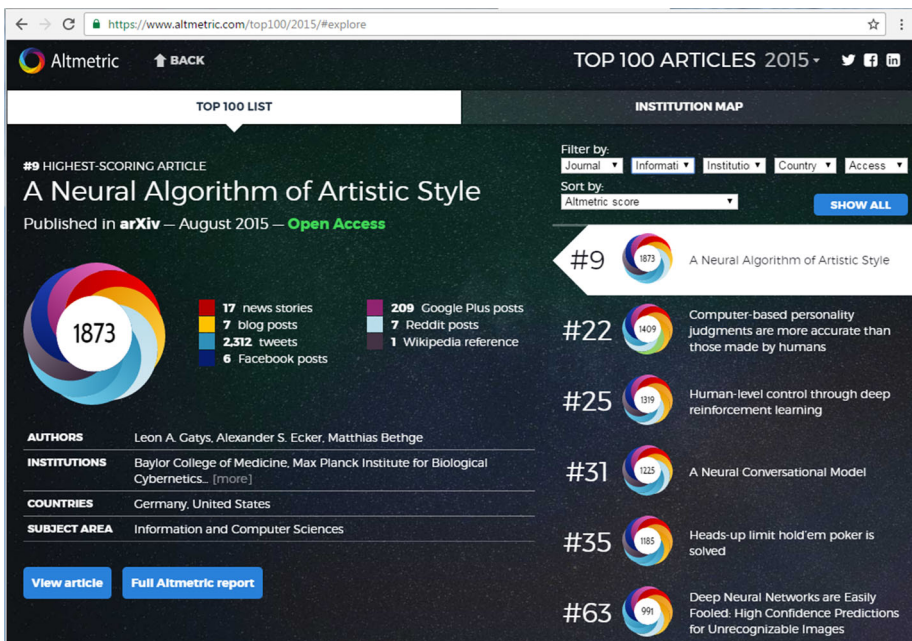


Fig. 6 Screenshots from the Altmetrics web service, top 100 articles of 2015 in the ‘Information and Computer Sciences’ subject area

description of the webpage in the Altmetric web service states is quite interesting: “What academic research caught the public imagination in 2015?”.

In summary, we discussed both the traditional citation-based and also more modern Altmetrics-based approaches to assess research impact, and we saw that each approach has its own advantages and disadvantages.

While we agree that citation is not the best and most objective measure of research impact and researchers productivity, similar to many other researchers, we see no other way of measuring research impact without having to form comprehensive committees and then conducting effort-intensive initiatives to examine papers' impact by which one still would not be able to guarantee objectivity, as per (Grigore 2007). Thus, similar to many other contexts (e.g., promotion and hiring committees, and research agencies), we use the citation measure in this work as an only one indicator (not the only one) for research impact, while keeping its limitations in mind.

Research method

In the following, the goal, research questions of our study and the metrics we have used are presented. We then present the data extraction phase of our study.

Goal and research questions

The research approach we have used in our study is the Goal, Question, Metric (GQM) methodology (Basili 1992). Using the GQM's goal template (Basili 1992), the goal of this study is to conduct an exploratory bibliometric assessment on quantity (the number of papers that were published) versus research impact (as measured by the number of citations) of the papers in the SE research literature. Based on the above goal, we raised the following five research questions (RQs):

- *RQ 1: Quantity versus impact of papers by paper types* How do citations compare for different paper types (e.g., conference and journal papers)? For example, do conference papers receive, on average, less or more citations than journal papers?
- *RQ 2: Quantity versus impact of papers by venues* How do citations compare for different venues?
- *RQ 3: Ratios of uncited (non-impactful) papers* What is the ratio of uncited (non-impactful) papers in SE? How does SE compare to other areas of science w.r.t. this metric? How has the ratio changed over the years (for the papers published in different years)?
- *RQ 4: Quantity versus impact of papers for top countries* How does the quantity versus impact of papers relate for different countries?
- *RQ 5: Quantity versus impact of papers for top authors* How does the quantity versus impact of papers relate for the different most productive authors (i.e., those that publish the largest numbers of papers)?

In terms of the SE research method perspective, we should note that the goal and RQs of the study are *exploratory* and *descriptive* in nature (Easterbrook et al. 2008).

Data extraction and pool of papers

The data extraction approach was adopted and the pool of papers was imported from our recent two papers (Garousi and Fernandes 2016; Garousi and Mäntylä 2016). To ensure that this paper is self-contained, we briefly explain again those aspects next, but full details about extraction approach and the pool of papers can be found in Garousi and Fernandes (2016) and Garousi and Mäntylä (2016).

To extract the set of all SE papers, we had to select a suitable publication search engine. For a systematic selection of such a search engine, by reviewing the related review studies (Hamrick et al. 2010; Bornmann 2014; Noorden et al. 2014; Ioannidis et al. 2014; Ponce and Lozano 2010; Tijssen et al. 2002; Aksnes 2003; Pyšek et al. 2006; Bornmann et al. 2010; Corby 2010; Persson 2010; Wang et al. 2011; Miyairi and Chang 2012; Abramo et al. 2014; Antonakis et al. 2014; Newman 2014; Eaton 2014; Aversa 1985), we devised three important criteria for proper selection of the suitable publication search engine:

1. The publication search engine should provide the highest quality and reliability in terms of coverage of the SE literature, i.e., including the SE papers published in all relevant and credible venues,
2. The publication search engine should include the citation data for papers,
3. The publication search engine should provide a convenient/usable interface to search and extract the SE papers that for further processing and analysis.

To find the candidate publication search engines, we reviewed a large number of bibliometrics studies, in SE (e.g., Wong et al. 2008, 2009; Eric et al. 2011; Garousi and Varma 2010; de Freitas and de Souza 2011; Farhoodi et al. 2013; Garousi and Ruhe 2013) and other fields (e.g., Archambault et al. 2009; Falagas et al. 2008; Abrizah et al. 2013; Chadegani et al. 2013). We short-listed the candidate publication search engines as follows: DBLP (www.dblp.org), Scopus (www.scopus.com), Web of Science (www.webofknowledge.com) and Google Scholar (scholar.google.com). These search engines are among the most popular ones that researchers regularly use in various bibliometrics studies. DBLP was not further considered, since it does not include citation data. In Table 2, we discuss how the remaining three search engines rate in terms of the selection criteria discussed above.

Regarding criterion #1, as shown in Table 2, Scopus scores better than Web of Science, since Scopus has the feature to search by “Source name” (venue name). Thus, when using Scopus, quality and reliability of the search results in terms of complete coverage of the SE domain can be achieved to a great extent, as we discuss in the following. The search query includes the phrase “software” in venue names which we found to be a suitable approach to ensure including almost all major SE journals and conferences in the search approach. Given the nature of SE papers, quality and reliability of search results in terms of complete coverage cannot be guaranteed using Web of Science, since searching by paper title having the phrase “software engineering” does not guarantee including all the SE papers, as many SE papers do not explicitly include that phrase in their title, nor in the abstract, nor in the keywords. The first author actually experienced this challenge in a recent study (Garousi and Ruhe 2013), in which a bibliometric/geographic assessment of 40 years of SE research (1969–2009) is reported. All the major SE venues including the top SE conferences and journals, e.g., the top 25 venues as listed by the Google Scholars listing³ in the area of Software Systems, were included in the results returned by Scopus when the search via source name including ‘software’ was conducted.

Regarding criterion #2, all three candidate search engines include citation data (i.e., the number of times a given paper has been cited).

Regarding criterion #3, as we discuss in Table 2, Google Scholar became ineligible for our selection, since exporting the list of extracted papers to files is not automatically possible in a convenient manner (unless one writes complex scripts), and we were not able to find any API for it. One can easily imagine that manual analysis of huge number of SE

³ https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_softwaresystems.

Table 2 Rating of the candidate publication search engines in terms of the selection criteria

Criteria	Publication search engines			
	Scopus	Web of Science	Google Scholar	DBLP
1. Quality and reliability in terms of coverage of the SE literature	Since Scopus has the feature to search by “Source name” (venue names), quality and reliability of search results in terms of complete coverage can be achieved to a great extent	Given the nature of SE papers, quality and reliability of search results in terms of complete coverage cannot be guaranteed	Given the nature of SE papers, quality and reliability of search results in terms of complete coverage cannot be guaranteed	Given the nature of SE papers, quality and reliability of search results in terms of complete coverage cannot be guaranteed
2. Inclusion of citation data?	Yes	Yes	Yes	No
3. Convenient/usable interface for searching and data extraction	Allows saving the list of all extracted papers into CSV files	Only allows saving the list of extracted papers into CSV files on a page by page basis	Exporting the list of extracted papers to files is not automatically possible We were not able to find any API for it	No automated interface

papers using Google Scholar would be very time consuming. Web of Science only allows saving the list of extracted papers into CSV files on a page by page basis, e.g., if the paper search returns 100 pages of papers, exporting the data would be very tedious. Only Scopus allows saving the list of all extracted papers into CSV files. Thus, this is an advantage of Scopus over Web of Science.

A recent paper published in the Nature magazine, titled “The top 100 papers” (Noorden et al. 2014), which was discussed in “[Background and related work](#)” section, also used Scopus. There have been empirical studies, e.g., Mingers and Lipitakis (2010), Archambault et al. (2009), Falagas et al. (2008), Abrizah et al. (2013), Chadegani et al. (2013), Harzing and Alakangas (2016), which have compared the performance and coverage of these search engines in other fields, e.g., social sciences. Some studies, e.g., Abrizah et al. (2013), have found empirically that Scopus is better than Web of Science in certain aspects, e.g., “larger coverage of titles” (Abrizah et al. 2013). A longitudinal and cross-disciplinary comparison of citation data among Google Scholar, Scopus and the Web of Science is reported in Harzing and Alakangas (2016). The citation data of Web of Science and Google Scholar in the field of business and management is compared in Mingers and Lipitakis (2010). The analysis reported in Harzing and Alakangas (2016) shows “a consistent and reasonably stable quarterly growth for both publications and citations across the three databases. This suggests that all three databases provide sufficient stability of coverage to be used for more detailed cross-disciplinary comparisons”. Thus, given its sufficient quality, as reported in the previous studies, e.g., Mingers and Lipitakis (2010), Archambault et al. (2009), Falagas et al. (2008), Abrizah et al. (2013), Chadegani et al. (2013) and Harzing and Alakangas (2016), we choose Scopus as it also meets the three selection criteria in Table 2.

Having selected Scopus as the publication search engine to conduct the search for all the SE papers, the next step was to actually conduct the search for those papers. We found that, when conducting searches in Scopus, including the phrase “software” in “source title” (a term used in Scopus interface meaning the conference or journal where a paper has been published) is a suitable approach to ensure targeting the entire SE literature with a high precision (coverage). This finding was made by the first author during an informal search for the SE papers authored by the Turkish SE community which later resulted in publication (Garousi 2015). By further experimentation, we found that this approach for extracting the entire SE paper dataset is indeed reliable in terms of coverage of the SE literature and has been used in other disciplines for extracting large paper datasets as well, e.g., (Bornmann 2014; Noorden et al. 2014; Ioannidis et al. 2014; Abramo et al. 2014; Antonakis et al. 2014; Newman 2014; Eaton 2014; Aversa 1985).

In the Scopus search interface, we included the phrase “software” under “source title” as shown in Fig. 7. The exact search query that was developed to extract all SE papers from Scopus is shown in Table 3 along with explanations for each phrase in the query. We conducted several rounds of iterative review and excluded venues unrelated to SE (such as, Journal of Optimization Methods and Software) and also non-English papers.

Let us repeat from Garousi and Fernandes (2016) and Garousi and Mäntylä (2016) that, to get the pool of SE papers to feed our two recent studies (Garousi and Fernandes 2016;

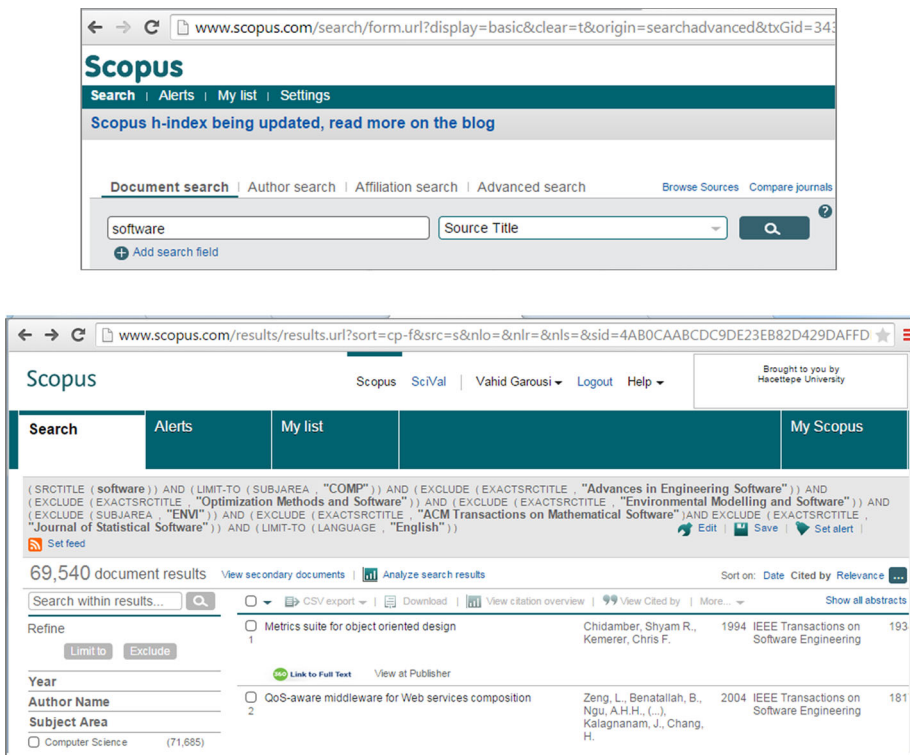


Fig. 7 Two screenshots showing the method used to extract all SE papers from Scopus

Table 3 The search query that was developed to extract all SE papers from Scopus

Search query	Explanations
(SRCTITLE(software)) AND	Only venues with the “software” phrase
(LIMIT-TO(SUBJAREA, “COMP”)) AND	Only the sub-area of computer science
(EXCLUDE(SUBJAREA, “ENVI”)) AND	Excluding the environmental science sub-area
(EXCLUDE(EXACTSRCTITLE, “Advances in Engineering Software”)) AND	Excluding this particular journal
(EXCLUDE(EXACTSRCTITLE, “Optimization Methods and Software”)) AND	Excluding this particular journal
(EXCLUDE(EXACTSRCTITLE, “Environmental Modelling and Software”)) AND	Excluding this particular journal
(EXCLUDE(EXACTSRCTITLE, “ACM Transactions on Mathematical Software”)) AND	Excluding this particular journal
(EXCLUDE(EXACTSRCTITLE, “Journal of Statistical Software”)) AND	Excluding this particular journal
(LIMIT-TO(LANGUAGE, “English”))	Only including papers written in English

Garousi and Mäntylä 2016), the data extraction phase was conducted on Dec. 25, 2014. Even if the analysis was done at the end of 2014, as per our analysis, we found that it takes a while for the Scopus search engine to record/import all the data from other sources. Thus, the data for 2014 were partial. Furthermore, the number of citations for papers in 2014 was relatively very low, since those papers were either “In Press” or recently published. For example, our analysis showed that the 2443 papers (partial count as per the Scopus approach discussed above) published in 2014 had 203 citations (0.08 citations per paper), while for 6403 papers published in 2013, there were 3365 citations (0.53 citations per paper). Due to the partial situation of the 2014 data set, we decided to exclude the 2014 papers altogether in our dataset and used 2013 as the last publication year.

As a result of applying the above approach, we had an initial dataset of 69,540 papers. Obviously, all the major SE venues including the top SE conferences and journals, such as the top 25 venues as listed by the Google Scholars listing⁴ in the area of Software Systems, are included in the results returned by Scopus since all their names include the word ‘software’.

Furthermore, we were also aware that some SE-related venues do not have the phrase “software” in their titles, like the following ones:

- Venues on requirements engineering: Springer Journal on Requirements Engineering and the International Requirements Engineering Conference (RE).
- Venues including the “Formal Methods” phrase: Formal Methods in System Design (journal), and the International Symposium on Formal Methods (FM).
- International Conference on Program Comprehension (ICPC).
- Working Conference on Reverse Engineering (WCRE).

⁴ https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_softwaresystems.

- International Conference on Model-Driven Engineering Languages and Systems (MODELS).
- International Conference Technology of Object-Oriented Languages and Systems (TOOLS).
- European Conference on Object-Oriented Programming (ECOOP).
- Object-Oriented Programming, Systems, Languages & Applications (OOPSLA).

We should mention that, at some point, the line between SE and other related disciplines, such as programming languages, becomes “gray”. Thus, for the purpose of this study, we had to draw the border somewhere. As we have listed in the above additional list of venues not including the phrase “software”, we included those that have a focus on object-oriented concepts and thus related to the design phase of SE.

Thus, we conducted searches for the above venues separately, and as a result, 3240 additional papers were found and added to the pool. As an example, Fig. 8 shows the query used to extract the list of papers published in the proceedings of the OOPSLA conferences. Thus, in summary, the importance of the choices regarding search strings and venues were considered with extreme care. Given that both authors have had long track record and expertise in conducting bibliometrics studies in the past, e.g., Garousi and Fernandes (2016), Garousi and Mäntylä (2016), Garousi and Varma (2010), Farhoodi et al. (2013), Garousi and Ruhe (2013), Fernandes (2014) and Garousi (2015), they discussed the choices regarding the search strings and venues very carefully in several iterations and data (paper set) extractions were conducted/improved in iterations among the two authors to ensure rigor and high quality in the data set and to minimize threats to vitality with respect to choices regarding search strings and venues. To prevent duplicated rows (papers) in the dataset, we used the automated duplicate detection of the Excel tool on the paper titles’ column.

We should add that Scopus stores the following 12 document (resource) types: article, article in press, book, book chapter, conference paper, conference review, editorial, erratum, letter, note, review, and short survey. We only wanted to consider scientific papers,

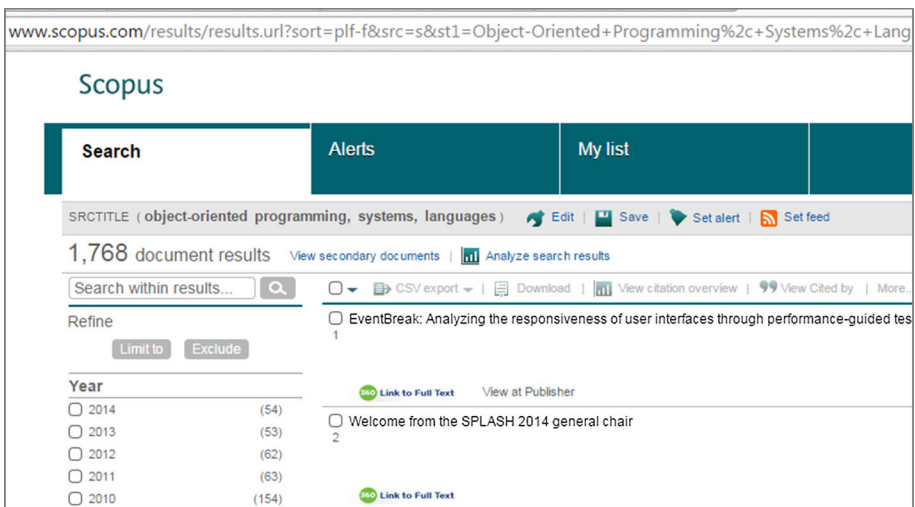


Fig. 8 Screenshot showing the query used to identify papers published in the proceedings of the Conference on Object-Oriented Programming, Systems, Languages and Applications (OOPSLA)

thus we only included records of the following types: article, article in press, book chapter, conference paper, and review paper (e.g., survey or systematic review paper), and excluded the rest.

To address RQs 2-5, we needed to filter the master paper dataset by country and authors. For this purpose, we conducted new searches in Scopus on Dec. 28, 2015, using the method discussed next. Firstly, we extracted the set of SE papers using the same search string as in Table 3. Then, we filtered the master pool of papers by country and authors, as shown in Figs. 9 and 10.

Note that in the 2014 search (Fig. 7), “document types” were excluded manually after we got the data from Scopus, but in the 2015 searches (Figs. 9, 10), we did this right inside Scopus itself. Also, since RQs 2-5 are independent from RQ 1, the authors decided to include papers from 2014 for RQs 2-5 while, to ensure reusing the same master dataset from Garousi and Fernandes (2016) and Garousi and Mäntylä (2016) and comparability of results with trends of Garousi and Fernandes (2016) and Garousi and Mäntylä (2016), RQ 1 was based on the exact dataset as used in Garousi and Fernandes (2016) and Garousi and Mäntylä (2016).

Once we had the pool of papers, we reviewed the records to ensure their integrity, e.g., not having duplicate records for a given paper. It was somewhat surprising that data exported from Scopus had many duplicates. We cleaned up the data set and, after applying all the above steps, the final paper pool was finalized with 71,668 papers. To ensure transparency and replicability of our analysis, the entire raw data for all the papers is available as an Excel file which can be downloaded online (Garousi and Mäntylä 2015). Furthermore, the source files for all the analyses reported in this paper can be found in another online folder (Garousi and Fernandes 2016).

Results

We present the results for the five RQs in this section.

RQ 1-Quantity versus impact: for different paper types

Many researchers in the CS and SE communities argue that conference papers have comparable (or even higher) impact compared to journal articles. Nuseibeh (2011) indicates that there are no significant difference between conference articles and journal articles. Patterson (2004) discusses that, in computing, it is common to prefer conferences instead of journals. This reality contrasts with the prevailing academic tradition where the primary means of publishing is in journals (Vrettas and Sanderson 2015; Vardi 2009).

Freyne et al. (2010) present quantitative evidence that articles in leading computing conferences match the impact (citations) of articles in mid-ranking journals and surpass the impact of articles in journals in the bottom half of the Thompson Reuters rankings. Contrarily, Garousi and Fernandes (2016) provided statistics that show that, amongst the most cited papers in software engineering, 85% are published in journals or magazines (such as the IEEE Software), while only 15% are published in conference and workshop proceedings. These numbers seem to show that journal papers in SE may have more visibility and impact than conference papers.

We aimed at providing additional evidence regarding this issue (conference versus journal papers). Scopus supports grouping of a pool of papers by their types (the so called

The screenshot shows the Scopus search results page for a query. The search bar contains a complex query. The results are sorted by 'Date Cited by Relevance'. The 'Country/Territory' filter is set to 'United States' (16,352). The results list includes:

Refine	Document Title	Author Name	Year	Source Title	Count
<input type="checkbox"/> Restoring security of evolving software models using graph	Birger, J., Jijens, J., Wenzel, S.	2014	International Journal on Software Tools for Technology Transfer	0	
<input type="checkbox"/> Analysis of distance functions for similarity-based test suite reduction in the context of model-based testing	Coutinho, A.E.V.B., Cartaxo, E.G., Machado, P.D.L.	2014	Software Quality Journal	0	
<input type="checkbox"/> Differential precondition checking: a language-independent, reusable analysis for refactoring engines	Overbey, J.L., Johnson, R.E., Hafz, M.	2014	Automated Software Engineering	0	
<input type="checkbox"/> An overview of model checking practices on verification of PLC software	Ovattman, T., Aral, A., Polat, D., Ünver, A.O.	2014	Software and Systems Modeling	0	
<input type="checkbox"/> Integrating deductive verification and symbolic execution for abstract object creation in dynamic logic	De Gouw, S., De Boer, F., Ahrendt, W., Babel, R.	2014	Software and Systems Modeling	0	
<input type="checkbox"/> Does the level of detail of UML diagrams affect the maintainability of source code?: a family of experiments	Fernández-Sáez, A.M., Genero, M., Caviano, D., Chaudron, M.R.V.	2014	Empirical Software Engineering	0	

The screenshot shows the Scopus search results page for the same query, but with the 'Country/Territory' filter set to 'United States' (16,352). The results are sorted by 'Date Cited by Relevance'. The 'CSV export' button is highlighted. The results list includes:

Refine	Document Title	Author Name	Year	Source Title	Count
<input checked="" type="checkbox"/> Differential precondition checking: reusable analysis for refactoring engines	Overbey, J.L., Johnson, R.E., Hafz, M.	2014	Automated Software Engineering	0	
<input checked="" type="checkbox"/> Supporting and accelerating reproducible empirical research in software evolution and maintenance using TraceLab Component Library	Dit, B., Moritz, E., Linares-Vasquez, M., Poshvanyk, O., Cleland-Huang, J.	2014	Empirical Software Engineering	0	
<input checked="" type="checkbox"/> Charting the API minefield using software telemetry data	Kechagia, M., Mitropoulos, D., Spinellis, D.	2014	Empirical Software Engineering	0	
<input checked="" type="checkbox"/> Exploring the costs of technical debt management – a case study	Guo, Y., Spinola, R.O., Seaman, C.	2014	Empirical Software Engineering	0	
<input checked="" type="checkbox"/> Certifying a java type resolution function using program transformation, annotation, and reflection	Winter, V., Reinke, C., Guerrero, J.	2014	Software Quality Journal	0	
<input checked="" type="checkbox"/> Normalizing variations in feature vector structure in keystroke dynamics authentication systems	Syed, Z., Banerjee, S., Cukic, B.	2014	Software Quality Journal	0	
<input checked="" type="checkbox"/> Testing variability-intensive systems using automated analysis: an application to Android	Galinido, J.A., Turner, H., Benavides, D., White, J.	2014	Software Quality Journal	0	

Fig. 9 Two screenshots showing the method used to filter the pool of papers by country

Scopus

The screenshot shows the Scopus search interface. At the top, there are tabs for 'Search', 'Alerts', and 'Lists'. Below the search bar, a message states: 'Your Search has been saved in My Scopus > Saved searches.' The search results are displayed as '70,712 document results'. A search filter is visible: '(SRCTITLE (software)) AND (LIMIT-TO (SUBJAREA, "COMP")) AND (EXCLUDE (EXACTSRCTITLE, "Advances in Engineering Software")) AND (EXCLUDE (EXACTSRCTITLE, "Environmental Modelling and Software")) AND (EXCLUDE (SUBJAREA, "ENVI")) AND (EXCLUDE (EXACTSRCTITLE, "ACM Transactions on Software Engineering")) AND (EXCLUDE (DOCTYPE, "ed")) AND (LIMIT-TO (LANGUAGE, "English")) AND (EXCLUDE (PUBYEAR, 2016)) OR EXCLUDE (PUBYEAR, 2015) TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "ip")'. Below the search bar, there are options to 'View secondary documents' and 'Analyze search results'. A search filter is visible: 'Search within results...'. The 'Refine' section shows 'Limit to' and 'Exclude' buttons. The 'Year' section is empty. The 'Author Name' section is highlighted with a red box and contains the following list:

Author Name	Citation Count
<input checked="" type="checkbox"/> Plattini, M.	(160)
<input type="checkbox"/> Harman, M.	(123)
<input type="checkbox"/> De Lucia, A.	(117)
<input type="checkbox"/> Xie, T.	(115)
<input type="checkbox"/> Antoniol, G.	(107)
<input type="checkbox"/> Di Penta, M.	(104)
<input type="checkbox"/> Harrold, M.J.	(104)
<input type="checkbox"/> Wohlin, C.	(104)
<input type="checkbox"/> Khoshgofaar, T.M.	(102)
<input type="checkbox"/> Tonello, P.	(101)

The 'Subject Area' and 'Document Type' sections are empty. The search results list shows the following items:

- Restoring security of evolving software models using graph transformation (1)
 - [Link to Full Text](#) [View at Publisher](#)
- Analysis of distance functions for similarity-based test suite reduction in the context of model-based testing (2)
 - [Link to Full Text](#) [View at Publisher](#)
- An overview of model checking practices on verification of PLC software (3)
 - [Link to Full Text](#) [View at Publisher](#)
- Differential precondition checking: a language-independent, reusable analysis for refactoring engines (4)
 - [Link to Full Text](#) [View at Publisher](#)
- Integrating deductive verification and symbolic execution for abstract object creation in dynamic logic (5)

Fig. 10 A screenshot showing the method used to extract the pool of papers by each top author

“documents types” in Scopus), as shown in Table 4. As we can see, journal and conferences papers, by covering respectively about 30 and 62% of the pool, are in the majority. As another analysis, we calculated the average number of citations per document type, as also shown in Table 4. Review papers (such as survey papers and systematic literature reviews) and journal articles (with 18.5 and 12.6 citations on average) are the top two types in terms of this metric. Thus, it seems that, as one would expect, review papers are quite popular and receive a relatively high number of citations. In terms of median values, only journal and review articles have non-zero values, denoting that for the other types, the data is highly skewed towards zero (no citations).

As discussed in “Bibliometrics studies in SE” section, we were aware of a 2008 study (Rahm 2008) which assessed the number of papers and citations for the top-100 venues (journal and conference) in computer science. That study reports that the average citations for journal and conference papers in computer science are similar (7.5 and 7.3, respectively). Our results are not in alignment with the results of that study and, thus, the data show that the SE literature is quite different in this aspect from the general CS literature as the SE journal papers have received more citations in average (12.6) than conference papers (3.6). We wanted to study the data of the 2008 study (Rahm 2008) in detail to be able to analyze and justify the differences between the results of our study and theirs in terms of average number of citations to journal versus conference papers. However, unlike our study in which we have open-sourced the raw study we have used for analysis, the 2008 study did not provide the raw data. Thus, such an analysis was unfortunately not possible.

To visually compare papers quantity and their impact, Fig. 11 shows the Average Citations per Paper (ACPP) metric for the five different paper types. Furthermore, Fig. 12

Table 4 Statistics by document (paper) types (ACPP stands for Average Citations per Paper)

	Document (paper) types				
	Article	Article in press	Review article	Book chapter	Conference paper
Total # in the pool	21,274	452	683	945	44,726
% Of the pool (%)	29.7	0.6	1.0	1.3	62.4
Total times cited	268,153	145	12,610	2357	163,042
Times cited (average), defined as ACPP	12.6	0.3	18.5	2.5	3.6
Times cited (median)	2	0	4	0	0
# With no citations	6238	367	200	566	23,367
# with 1 citation	2648	57	49	129	6563
# With > 1 citation	12,391	30	437	253	14,799
% With no citations (%)	29.3	81.2	29.3	59.9	52.2
% With 1 citation (%)	12.4	12.6	7.2	13.7	14.7
% With > 1 citation (%)	58.2	6.2	63.5	26.5	33.1

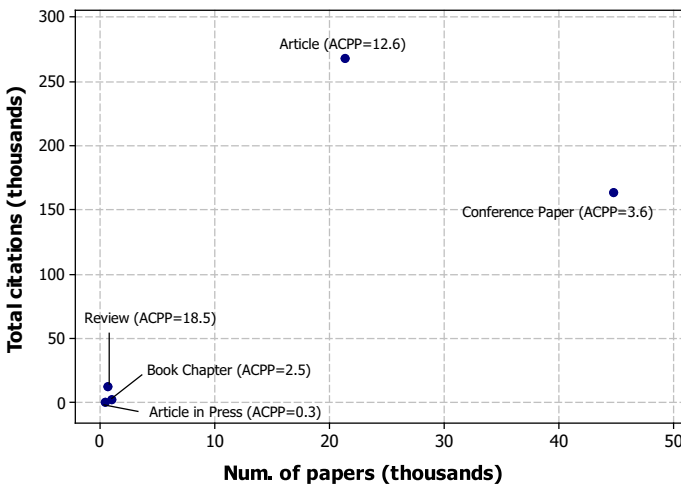


Fig. 11 Number of papers versus total citations for different paper types (both axes are in thousands)

depicts some of the Table 4 data by grouping citations for different documents types (we show both regular and percentage stack charts). As we can see, ACPP for the review articles are the highest (ACPP = 18.5), succeeded by regular journal papers (ACPP = 12.6), and then conference papers (ACPP = 3.6).

RQ 2-Quantity versus impact: by venues

For RQ 2, we wanted to analyze quantity versus impact for different venues. For this purpose, we randomly selected from the dataset four representative SE journals and one representative SE conference as follows:

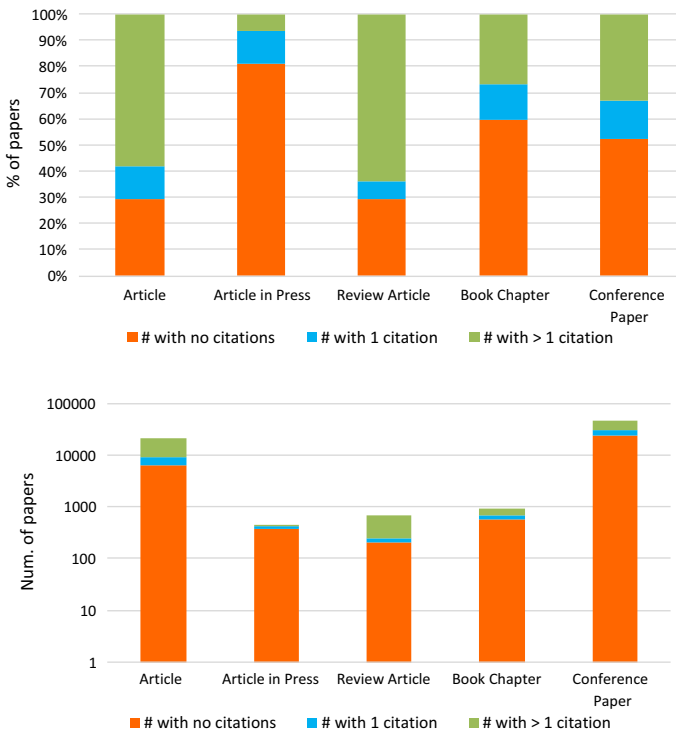


Fig. 12 Grouping of citations for different documents types (both regular and percentage stack charts)

- IEEE Transactions on Software Engineering (TSE).
- ACM Transactions on Software Engineering and Methodology (TOSEM).
- Elsevier Empirical Software Engineering journal (ESE).
- WorldScientific International Journal of Software Engineering and Knowledge Engineering (IJSEKE).
- ACM/IEEE International Conference on Software Engineering (ICSE).

From the master dataset of 71,668 papers, we carefully filtered the subset of data for each of the venues above, saved them as a different sub-dataset, and then conducted citation analysis on each of the datasets. Figure 13 shows the boxplots of citation data of papers published in the above four SE journals and the one conference (ICSE). We have calculated both the Average Citations per Paper (ACPP) and also the Average normalized Citations per Paper (AnCPP), which is the absolute number of citations divided by the number of years passed after the publication of each paper. Figure 13 also shows the total number of papers published in each venue from the beginning of its history until 2013 (by the labels “ n ”). For example, $n = 2491$, $ACPP = 36.8$, and $AnCPP = 2.5$ for TSE. For better understanding, Fig. 13 also shows the boxplots without “outliers”. We should note that we have used the Minitab software for most of the charts in this papers.

As depicted in Fig. 13, there is a major difference in the citation landscape among the five venues. Among the four journals, TOSEM leads in terms of both metrics (ACPP and AnCPP), and IJSEKE is the lowest among the four journals. In terms of both metrics, the well-known ICSE conference is ranked lower than TSE, TOSEM and EE but slightly higher than IJSEKE.

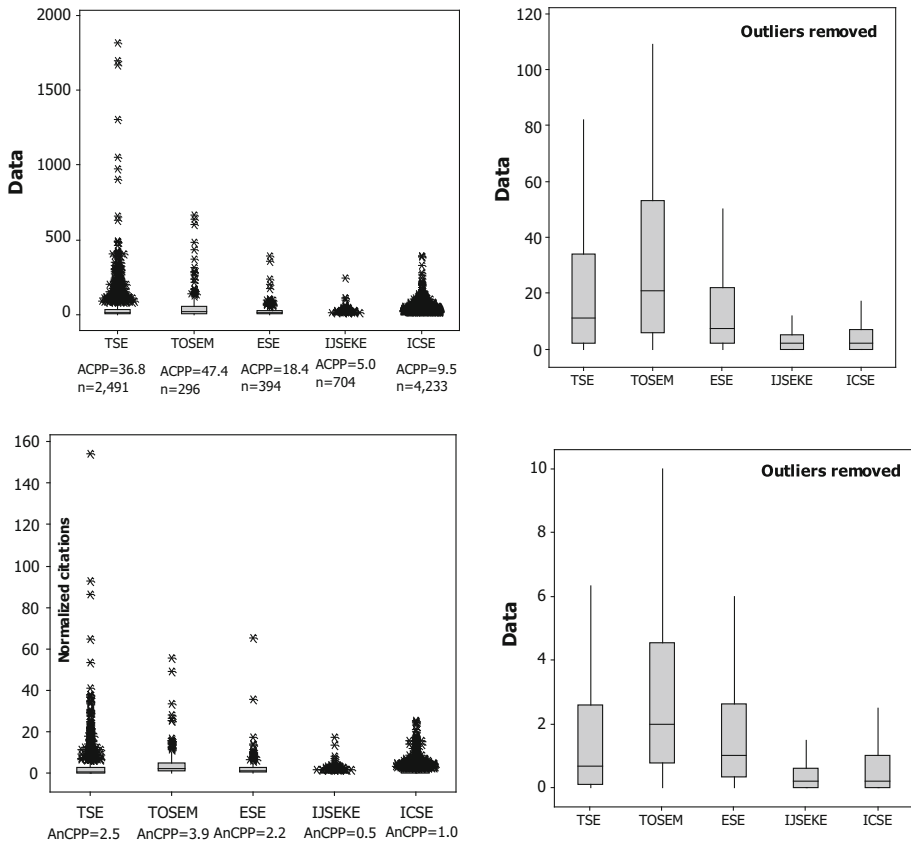


Fig. 13 Boxplots of citation data of papers published in four SE journals and one SE conference

To assess quantity (number of papers) versus impact of papers published in the above five venues, Fig. 14 a scatterplot of the two metrics (ACPP and AnCPP) for the venues. TOSEM seems to be the leading venue in this set since with the lowest quantity (number of papers), it has received the highest ACPP and AnCPP values.

We also wondered about ratios of uncited papers to all papers in these venues and calculated that metrics, as shown in Table 5. Recall from “Bibliometrics studies on quantity versus impact of papers” section that in the set of all 71,668 SE papers, 43.0% of the papers were uncited [had received no citations as per our analysis in Garousi and Fernandes (2016) and Garousi and Mäntylä (2016)]. With an uncited paper ratio of only 8.1%, TOSEM also ranks the highest in this aspect. With uncited paper ratios of 36.5 and 33.1%, respectively, IJSEKE and ICSE are the lowest ranked venues among the set of five, with values close to the ratio corresponding to the master set of all 71,668 SE papers (43.0%).

RQ 3-Quantity versus impact: ratios of uncited (non-impactful) papers

As discussed in the related work section, our recent works (Garousi and Fernandes 2016; Garousi and Mäntylä 2016) measure the ratio of uncited (non-impactful) papers in SE.

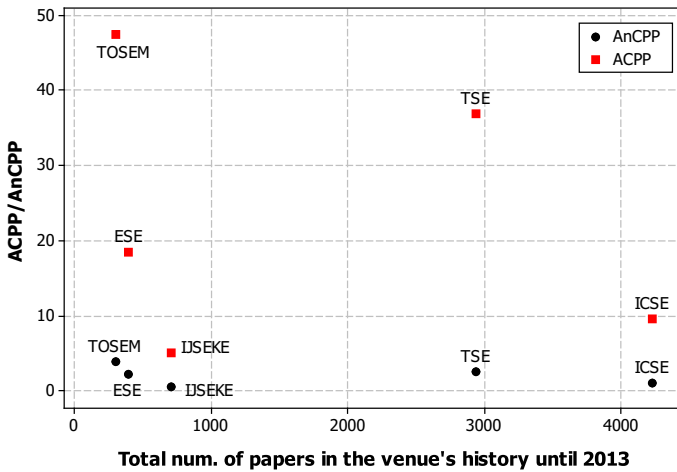


Fig. 14 Scatterplot of quantity (number of papers) versus impact of papers published in five SE venues

Table 5 Ratios of uncited papers in the entire pool and for several different representative venues

Subset of papers	Ratio of uncited to all papers (%)
All the 71,668 papers	43.0
IEEE TSE	13.2
ACM TOSEM	8.1
Elsevier ESE	10.1
WorldScientific IJSEKE	36.5
IEEE/ACM ICSE	33.1

30,958 papers (43% of the pool) had no citations at all. 10,095 papers (14% of the pool) had only one citation.

We also wanted to compare the ratio of uncited papers in SE to other disciplines, for which we found a data source (Biswas and Kirchherr 2016). According to Biswas and Kirchherr (2016), “82% of articles published in humanities are not even cited once. No one ever refers to 32% of the peer-reviewed articles in the social and 27% in the natural sciences”. Based on our analysis in Garousi and Mäntylä (2016), out of all the 71,668 SE papers in the pool indexed in the Scopus publication database, 30,958 papers (43% of the pool) had no citations at all. We visually compare the ratio of uncited papers in SE to the situation in humanities, social and natural sciences in Fig. 15. The ratio of uncited papers in SE (43%) is slightly higher than those ratios in natural and social sciences (27 and 32%, respectively) and much lower than (about half of) the case in humanities (a staggering 82%).

To assess the citation trends and ratios of uncited papers grouped by publication years, we show in Fig. 16 the individual-value plot and box plot (excluding outliers) of paper citations versus years of publications ($n = 71,668$ papers). Figure 17 shows the ratios of uncited papers by years of publications. As we can see in Fig. 17, after excluding the first 3 years (1972–74), the ratio of uncited papers for a given publication year ranges from 30 to 80%. As expected, it is slightly the case that in the pool of more recent papers (e.g., published after 2010), higher ratios are uncited (between 50 and 70%).

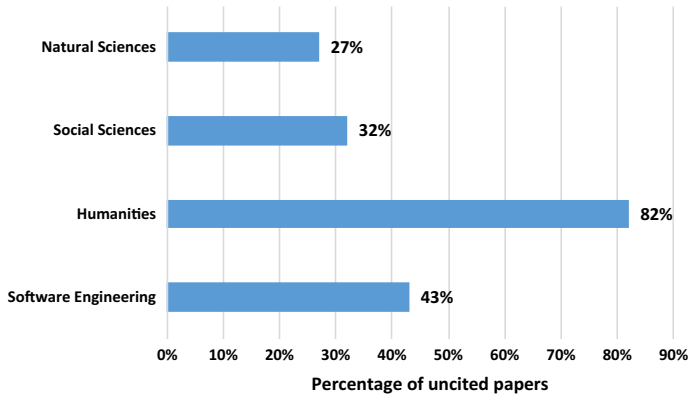


Fig. 15 Comparing the ratio of uncited papers in SE to the situation in humanities, social and natural sciences [based on data from Biswas and Kirchherr (2016) and Garousi and Mäntylä (2016)]

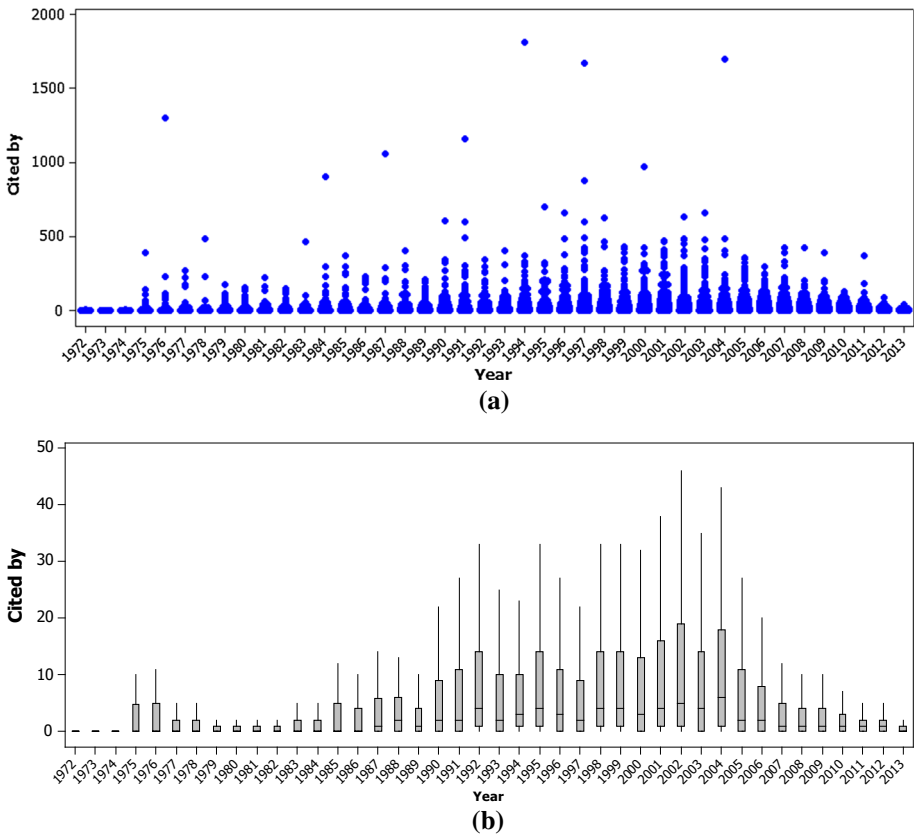


Fig. 16 Individual-value plot and box plot (excluding outliers) of paper citations versus years of publications ($n = 71,668$ papers)

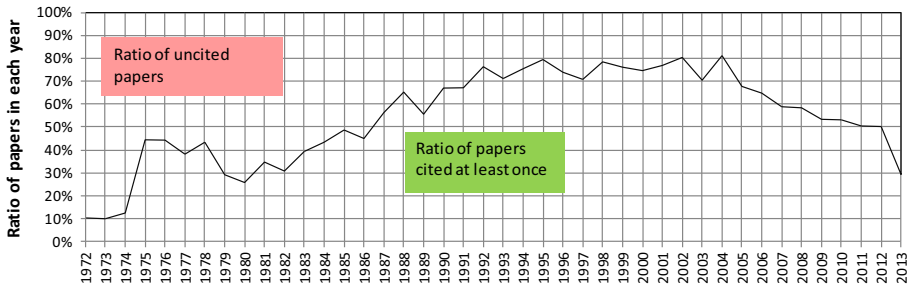


Fig. 17 Ratios of cited/uncited papers by years of publications

RQ 4-Quantity versus impact: for the top-10 countries

For each search query, Scopus provides statistics of countries based on author affiliations. Thus, our data pool supports that kind of information and let us conduct country-level analyses.

The ranking of the countries with more than 500 papers in the pool is shown in Fig. 18. We adopt an ‘inflated’ metric, since for papers with multiple country affiliations, all the involved countries are considered with equal weights in the Scopus data, i.e., in those cases, the paper is fully credited to each country of origin. We would have liked to experiment and utilize a fractional credit metric for authorship but the Scopus features in this regard are limited and did not provide such data. Future works can conduct further in-depth analyzes in this regard and may assess whether utilizing a fractional credit metric might reveal significantly different results. As a result of using the above metric and dataset, the sum of the values in Fig. 18 is obviously larger than the pool size (total number of SE papers).

Fig. 18 Ranking of the countries with more than 500 papers (contributions)

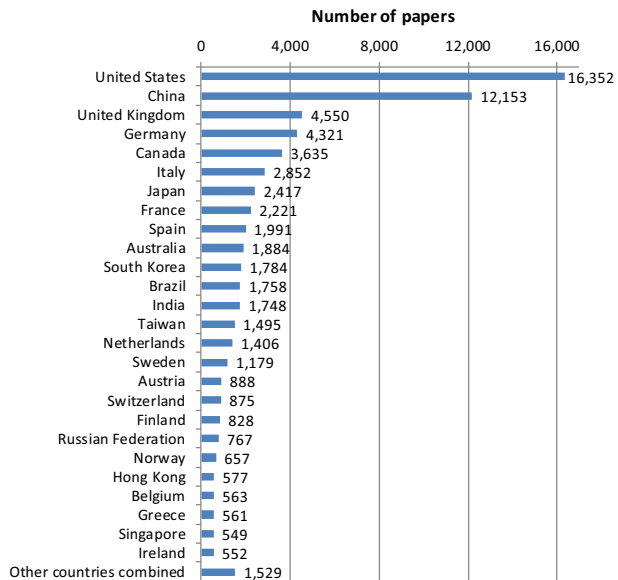


Figure 19 shows the number of papers versus total citations for top-15 countries (taken from Fig. 18). To assess the issue of citations of papers from English-speaking countries versus non-English speaking countries, we have also included the data for Ireland and New Zealand in Fig. 19 as well. Table 6 shows the detailed numerical data for Fig. 19. In terms of the ACPP metric, the US (15.2) is the first and Canada (12.4) is the second. UK (11.9), the Netherlands (11.5), Australia (11.3), and Italy (11.3) with very close ACPP values are the next ones.

Furthermore, Fig. 20 depicts the grouping of citations for the top 15 countries plus Ireland and New Zealand (we show both regular and percentage stack charts). Furthermore, Fig. 21 shows the interval plot of the countries’ citation data showing 95% confidence intervals which make it easier to compare the distributions. The results in Figs. 20 and 21 allow us to speculate that papers originated from English-speaking countries tend to have more visibility and impact (citations) than papers originated from non-English-speaking countries. Although, as we can see in Fig. 21, non-English-speaking countries such as Germany, Italy and Netherlands have also performed well in terms of citation data. As the confidence intervals in Fig. 21 depict, there are statistically-significant differences among the distributions of some of the countries’ citation data, e.g., between that of the US and China, but no statistically-significant differences among the distributions of some other pairs, e.g., between Canada and the UK. For brevity, we do not report the exhaustive list of all pair-wise statistical test results of each pair of country datasets using the Mann–Whitney test in this paper, since the number of combinations is simply too high (selection of 2 out of 17 countries in this list would yield 136 pairs). Given the fact that we have provided the dataset publicly online (Garousi and Fernandes 2016), the reader can conduct more detailed statistical analysis and comparisons.

There may be many reasons for the above observation, but we believe that the skills in mastering the English language and consequently in better conveying the research contributions are very important to make the paper citable. This observation has been studied and is well documented in other research areas, e.g., Belcher (2007), Uzuner (2008), Flowerdew (1999), Vasconcelos et al. (2007), West et al. (2009) and Schliesser (2016). We should still explicitly highlight that a non-trivial ratios of papers from even English-speaking countries are uncited papers, e.g., for the case of six even English-speaking

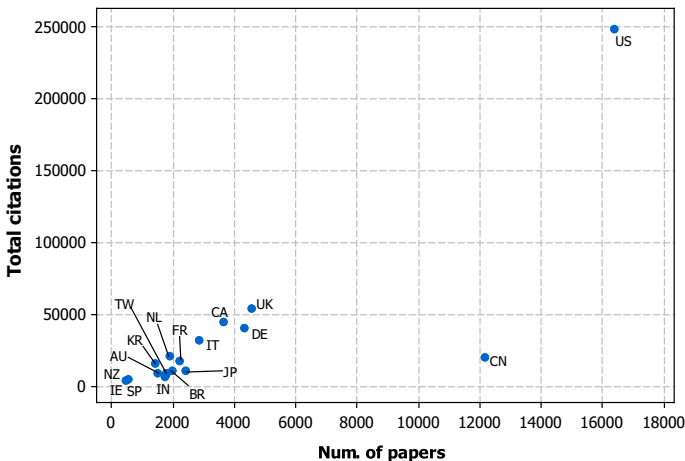


Fig. 19 Number of papers versus total citations for top 15 countries, plus Ireland and New Zealand

Table 6 Numerical data for Fig. 19

Rank (by num. of papers)	Country name	Num. papers	Num. citations	ACPP	Rank (by ACPP)
1	United States	16,352	248,523	15.2	1
2	China	12,153	20,073	1.7	17
3	United Kingdom	4550	54,009	11.9	3
4	Germany	4321	40,558	9.4	7
5	Canada	3635	44,908	12.4	2
6	Italy	2852	32,100	11.3	6
7	Japan	2417	10,448	4.3	15
8	France	2221	17,753	8.0	10
9	Spain	1991	11,144	5.6	12
10	Australia	1884	21,326	11.3	5
11	South Korea	1784	8664	4.9	13
12	Brazil	1758	8369	4.8	14
13	India	1748	6564	3.8	16
14	Taiwan	1495	8947	6.0	11
15	Netherlands	1406	16,173	11.5	4
26	Ireland	552	4516	8.2	9
32	New Zealand	460	4149	9.0	8

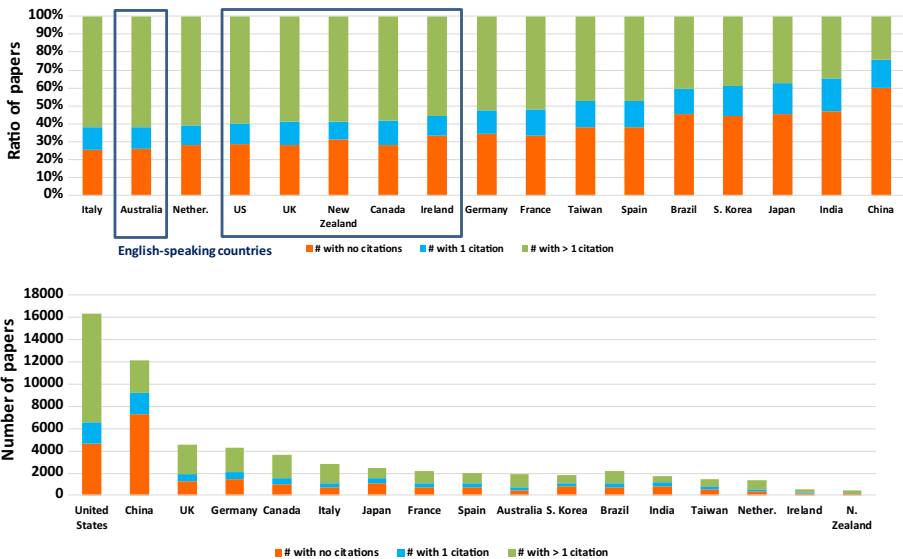


Fig. 20 Grouping of citations for top countries (both regular and percentage stack charts)

countries highlight inside a box in Fig. 20, the ratios of uncited papers have ranged between 25.9 and 33.3%, thus denoting that only by being authored by authors from a English-speaking country does not save a paper from bring uncited.

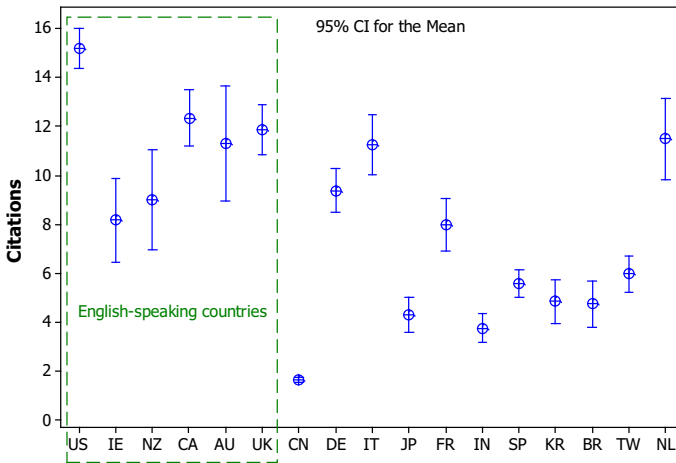


Fig. 21 Interval plot of the countries’ citation data showing 95% confidence intervals (CIs)

Another possible reason is related to the observation that researchers prefer in some cases to cite the works of colleagues geographically close to them, while ignoring those from other parts of the world (Wong and Kokko 2005). Obviously, not all papers affiliated to universities/organizations located in countries where English is the/an official language (such as USA, UK, Canada, Australia, Ireland, and New Zealand) are written by native-English speakers, since many researchers in those countries are foreigners (immigrants). Anyway, in general, those researchers are more exposed to English than researchers affiliated with, for example, an Asian country (such as China, Japan, and India).

Another observation is related to the involvement timeline of different countries in the SE research literature. Figure 22 shows the scatter plots of all publications and their citations for the top-3 countries (USA, China and UK). Note that the Y-axis is in a logarithmic scale and all X and Y axes are the same to ease comparisons. As we can observe, American researchers have been actively involved in the SE research arena since 1975. Chinese SE researchers did not become active until later 1990’s. Finally, British SE researchers started to be actively involved since the late 1980’s. As we can see, the overall ACPP of Chinese papers is almost an order of magnitude lower than American SE papers (ACPP = 1.7 versus 15.2).

RQ 5-Quantity versus impact: for the top-10 authors

To assess quantity versus impact of papers for the top-10 authors, we identified the top-10 authors in terms of number of papers in Scopus (see Fig. 10) and then downloaded the full list of their papers from Scopus (as discussed in Data extraction and pool of paper “Data extraction and pool of papers” section). Figure 23 shows the number of papers versus total citations for the top-10 authors. Values in parentheses are ACPP. With ACPP of 51.5, the late Mary J. Harrold (who passed away on September 2013) has the highest average impact for each of her papers. Mark Harman with ACPP = 26.2 is the second and Claus Wohlin with ACPP = 22.0 is the third in the ranking.

A word of caution related to the list of top-10 authors is needed. Ranking top scholars in SE (or in any other domain) is a controversial topic, e.g., refer to the series of 12 papers by Glass and Chen (Wong et al. 2008, 2009; Eric et al. 2011) that ranked the top-15 SE

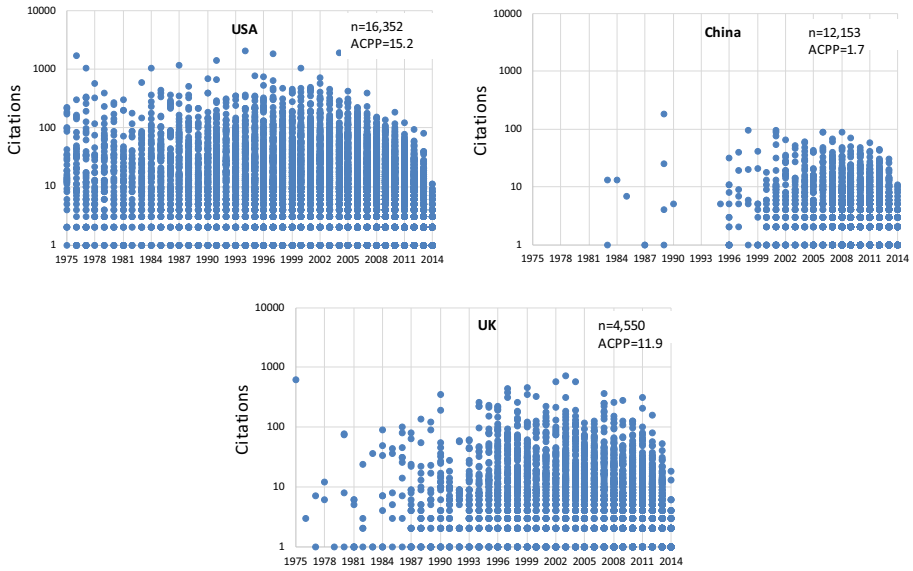


Fig. 22 Scatter plots of all publications and their citations for top-3 countries (Y-axis is in logarithmic scale)

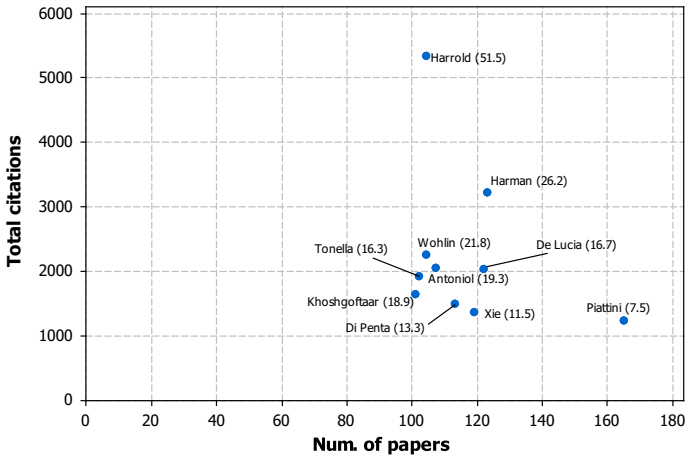


Fig. 23 Number of papers versus total citations for top 10 authors (in terms of # of papers); Values in parentheses are ACP

scholars and institutions between 1995 and 2006. It may be the case that the names of other very influential SE scholars, such as the past recipients of the ACM SIGSOFT Outstanding Research Award, do not appear in this list. The underlying reasons is that we are using the citation data as provided by Scopus and impact in our study is being measured by paper citations. Other dimensions of impact, e.g., impact of a SE scholar on the field and also on practice, without by necessarily getting high number of citations, should be studied by other follow-up studies.

Referring to Fig. 23 again, we can notice the large variance in the ACPP values, ranging from the minimum value of 7.5 for Mario Piattini up to the maximum of 51.5 for Mary J. Harrold. This once again denotes the spectrum of impact versus volume of papers and resonates the message of previous papers such as “getting more for less” (Broad 1981) in the area of SE, i.e., with 104 papers, Harrold’s ACPP impact is 51.5, while with 165 papers, Piattini’s ACPP impact is only 7.5.

Figure 24 shows the grouping of citations for top authors. Harrold again performs well in this chart as only seven of her 104 papers in Scopus have zero or one citations, while the rest (the other 97 papers) have at least two citations. Figure 25 shows boxplots of citation data for top authors.

Four of the top-10 authors (Antoniol, De Lucia, Di Penta, and Tonella, all from Italy) regularly collaborate with each other. In total, our pool has 79 papers co-authored by two of them (as shown by the bold numerical values in Table 6). Additionally, there are more papers co-authored by at least two of the top-10 authors. Table 7 shows the distribution of the 100 papers co-authored by seven of the top-10 authors. We can see that Antoniol and Di Penta are co-authors in 33 papers. Only Piattini, Wohlin, and Khoshgofaar have no papers co-authored with other top-10 authors.

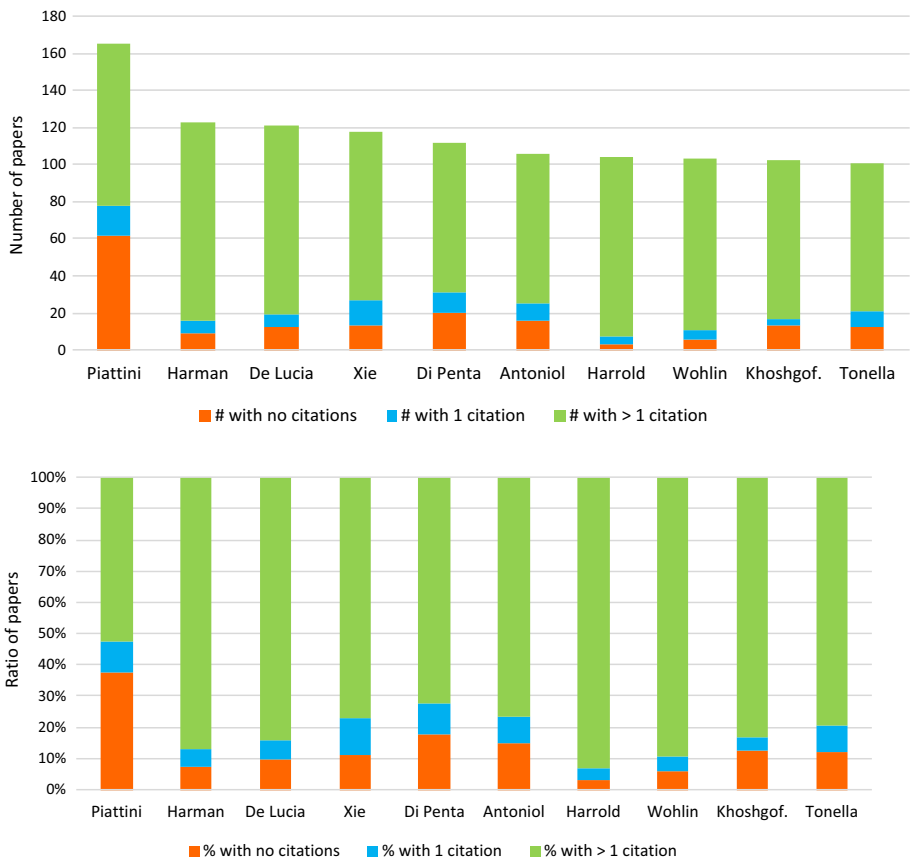


Fig. 24 Grouping of citations for top authors

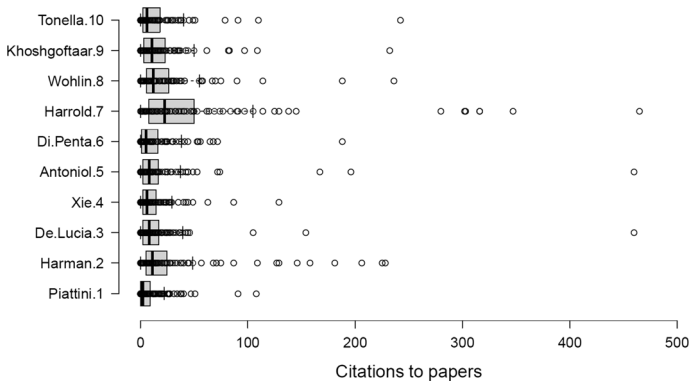


Fig. 25 Boxplots of citation data for top authors

Table 7 Joint authorship (number of joint papers) among seven top authors: Antoniol, De Lucia, Di Penta, Harman, Harrold, Tonella and Xie

	Antoniol	De Lucia	Di Penta	Harman	Harrold	Tonella
De Lucia	9					
Di Penta	33	15				
Harman	–	1	–			
Harrold	–		–	1		
Tonella	15	1	6	10	–	
Xie	–	–	1	1	–	–

Table 8 *h*- and *g*-indexes, number of citations and ACP for the top-10 authors (ordered by number of papers)

Authors	Num. of papers	<i>h</i> -index	<i>g</i> -index	Total citations	ACPP
Piattini	165	<i>19</i>	<i>29</i>	1238	7.5
Harman	123	27	53	3221	26.2
De Lucia	121	22	40	2043	16.7
Xie	118	21	31	1368	11.5
Antoniol	106	23	43	2068	19.3
Di Penta	112	22	34	1498	13.3
Harrold	104	36	72	5354	51.5
Wohlin	103	26	43	2264	21.8
Khoshgoftaar	102	24	39	1924	18.9
Tonella	101	23	38	1648	16.3

Table 8 shows the *h*- and *g*-index for the top-10 authors, calculated based on their SE papers. Maximum numbers in each column are in bold and minimum numbers are in italics. Maximum numbers in all the metrics (except number of papers) belong to Mary J. Harrold. Minimum numbers in all the metrics (except number of papers) belong to Mario Piattini. Harrold is by far the researcher with the highest impact, since she has the highest values for those two indexes (36 and 72), in spite of being among the top-10 authors with

the smallest number of published papers (104). The four Italian top-10 authors (Antoniol, De Lucia, Di Penta, and Tonella) also show similar values for the indexes; in particular each one has an *h*-index equal to 22 or 23.

Discussions

Summary of the findings

We present the summary of the findings for each of the RQs in the following:

RQ 1: Quantity versus impact: for different paper types

Among the five types of scientific papers included in Scopus (articles, articles in press, book chapters, conference papers, and review papers), journal and conferences papers, by covering respectively about 30 and 62% of the pool, are majority, in terms of numbers (quantity). Thus, as expected, we see that conferences papers are more, in quantity, compared to journal papers.

To assess impact versus quantity of paper types in a “macro” scale, we defined and used the ACP metric. We found that review papers and journal articles (with 18.5 and 12.6 citations on average) are the top two types in terms of this metric. Interpretation of the above results reveals that, as expected, review papers (such as survey papers and systematic literature reviews) are more popular and cited higher than other paper types in SE. Furthermore, again as expected, journal articles are cited more than conferences papers.

We also conducted cross comparisons between our results with other similar studies, e.g., a 2008 study (Rahm 2008) which assesses the number of papers and citations for the top-100 venues (journal and conference) in computer science (CS). That study reported that the average citations for journal and conference papers in CS are quite similar (7.5 and 7.3, respectively). Our results are not that aligned with the results of the study (Rahm 2008). Thus, the comparison of results reveals that the SE literature is quite different in this aspect from the general CS literature, as the SE journal papers have received more citations in average (12.6) than conference papers (3.6). Thus, it seems that in the SE community, researchers cite journal papers more compared to conferences papers.

In terms of ratios of papers with no citations in each venue type, 29.3% of journal papers versus 81.2% articles in press were uncited. Only 29.3% of review articles versus 52.2% of conference papers were uncited, which confirmed the general expectations for those types of papers.

RQ 2: Quantity versus impact: by venues

We populated the datasets for papers published in four representative SE journals (TSE, TOSEM, ESE and IJSEKE) and one representative SE conference (ICSE). We found that there is a major difference in the citation landscape among those five venues. Among the four journals, we found that TOSEM leads in terms of both citation metrics (ACPP and AnCPP), and IJSEKE is the lowest among the four journals. In terms of both metrics, the well-known ICSE conference is ranked lower than TSE, TOSEM and EE but slightly higher than IJSEKE.

When assessing quantity (number of papers) versus impact of papers published in the above five venues, TOSEM was the leading venue among the above five, denoting that with a small number of papers, it has had the highest impact in terms of citations. This could possibly be due to strict rules of this journal in accepting papers and reminding us of the famous quote: “less is more!”

RQ 3: Quantity versus impact: ratios of uncited (non-impactful) papers

30,958 papers (43%) in the pool had no citations at all. 10,095 papers (14% of the pool) had only one citation. We also compared the ratio of uncited papers in SE to the situation in humanities, social and natural sciences. The ratio of uncited papers in SE (43%) is slightly higher than those ratios in natural and social sciences (27 and 32%, respectively) and much lower than (about half of) the case in humanities (a staggering 82%). But still it is surprising that about half of the papers in the SE community stay uncited. The above data can be interpreted in different ways, e.g.: (1) many researchers publish papers in topics which no one later cites (uses or reads); (2) some SE venues have very low quality thresholds for acceptance of papers which result in having many accepted papers with low qualities.

We furthermore assessed the citation trends and ratios of uncited papers grouped by publication years. After excluding the first three years (1972–74), the ratio of uncited papers for a given publication year ranges from 30 to 80%. As expected, it is slightly the case that in the pool of more recent papers (e.g., published after 2010), higher ratios are uncited (between 50 and 70%). These trends confirmed the general expectation that newer papers would receive less or no citations until they get known in the community.

RQ 4: Quantity versus impact: for the top-10 countries

The top-3 countries in terms of number of SE papers are US, China, and UK. In terms of the ACP metric, the US (ACPP = 15.2) is the first and Canada (12.4) is the second. UK (11.9), the Netherlands (11.5), Australia (11.3), and Italy (11.3) have very close ACP values and are the next ones. The results allow us to speculate that papers originated from English-speaking countries have more visibility and impact (and consequently citations) than papers originated from non-English-speaking countries. There may be many reasons for this fact, but we believe that the skills in mastering the English language and consequently in better conveying the research contributions are very important to make a given paper more citable. This observation has also been documented in several other research areas (Vasconcelos et al. 2007; West et al. 2009; Schliesser 2016). Another possible reason is related to the observation that researchers prefer in some cases to cite the works of colleagues geographically close to them, while ignoring those from other parts of the world (Wong and Kokko 2005). Interpretation of the above results also means that higher-education policy makers in non-English-speaking countries should pay close attention to this issue and ensure proper training of their researchers in English writing to ensure producing papers with higher citations.

Another observation was regarding the involvement timeline of different countries in the SE research literature. American researchers have been actively involved in the SE research arena since 1975. Chinese SE researchers' involvement did not become active until later 1990's. Finally, British SE researchers started to be actively involved since the late 1980's. Also, the average (expected) ACP of Chinese papers is almost an order of magnitude lower than American SE papers (1.7 versus 15.2). This last issue raises

questions on why papers from certain countries have received low citations. Also, two possible root causes may be the “bias in location and selection of studies” (Egger and Smith 1998) to cite by researchers in English-speaking countries, and the “region-based citation bias” (Paris et al. 1998), in which researchers prefer to cite papers of those researchers who are regionally close to themselves.

RQ 5: Quantity versus impact: for the top-10 authors

We identified the top-10 authors in terms of number of papers in Scopus and then downloaded the full list of their papers from Scopus. With an ACPP value of 51.5, the late Mary J. Harrold has the highest average impact for each of her papers. Mark Harman with ACPP = 26.2 is the second and Claus Wohlin with ACPP = 22.0 is the third in the ranking.

We noticed the large variance in the ACPP values among the top-10 authors, ranging from the minimum value of 7.5 for Mario Piattini up to the maximum of 51.5 for Mary Harrold. This once again denotes the spectrum of impact versus volume of papers and resonates the message of previous papers such as “getting more for less” (Broad 1981) in the area of SE, i.e., with 104 papers, Harrold’s ACPP impact is 51.5, while with 165 papers, Piattini’s ACPP impact is 7.5. By analyzing the list of authors in by the top authors, we also found that four of the top-10 authors (Antoniol, De Lucia, Di Penta, and Tonella, all from Italy) regularly collaborate with each other.

Implications

It is the opinion of the authors that the SE community should pay more attention to the impact of papers versus their quantity, which was put nicely by David Parnas as “Stop the numbers game” (Parnas 2007). In fact, among the top-10 SE authors, Harrold is by far the one with more impact (i.e., with more citations to her papers), even if she is not the author with the largest number of papers. Typical academic reward systems encourage researchers to publish more papers and usually neglect research impact (Parnas 2007; Tongai 2013; Foster et al. 2015). Parnas recommends against this by expressing recommendations such as: “If you get a letter of recommendation that counts numbers of publications, rather than commenting substantively on a candidate’s contributions, ignore it”.

Our results also have implications for researchers in non-English-speaking countries to help them plan improvement strategies to increase their papers’ impact.

As a major issue, we should note that we only measured research (academic) impact (measured by citations) of papers in this study. Another important notion of impact for academic research is, for example, its impact on industrial practice which is not really easy to measure (Osterweil et al. 2008). In other words, the research impact of academic research is a type of ‘internal’ impact while its industry impact denotes a form of ‘external’ impact. One form of addressing this issue consists, for example, in counting papers that are authored simultaneously by both academic and industrial collaborators.

Limitations and potential threats to validity

In this section, the potential threats to the validity of the study are discussed in the context of the four types of threats to validity based on a standard checklist presented in Wohlin et al. (2000). We also discuss the steps that we have taken to minimize or mitigate those potential threats.

Internal validity

Internal validity reflects the extent to which a causal conclusion based on a study is warranted (Wohlin et al. 2000). We follow a systematic approach for the selection of publication database as described in “[Background and related work](#)” section. In order to make sure that this study and its rankings are repeatable, search engines, search terms were carefully defined and reported. Also, to ensure transparency and replicability of our analysis, the entire raw dataset is publicly available as an Excel file and can be downloaded online (Garousi and Mäntylä 2015; Garousi and Fernandes 2016). We are aware that limitation of search terms and search engines could lead to incomplete set of papers in the pool. We empirically found that, when conducting searches in Scopus, including the phrase “software” in venue names is an effective way to ensure targeting the entire SE literature with a high precision. The same approach was used in our previous studies (Garousi and Fernandes 2016; Garousi and Mäntylä 2016) and showed to be reliable. Additional searches were conducted to include in our study papers from SE venues that do not have the word “software” in their titles.

We are aware that the final set of venues considered in this study may not include some other SE conferences and journals that the reader was expecting to see incorporated. In fact, the process of selecting the venues is subjective, since there is no unique way of classifying the scientific fields addressed by a given venue. However, we believe our approach is appropriate, since we have included all the major SE venues.

Construct validity

Construct validities are concerned with issues that measure to what extent the object of study truly represents theory behind the study (Wohlin et al. 2000). Threats related to this type of validity in this study were suitability of RQs and the metrics that we analyzed (e.g., citation count). In “[On using citations as an indicator of research impact](#)” section, we discuss in detail the issue of using citations as an indicator of research impact and their associated risks and limitations and report various opinions from the literature either in support or against using that metric. As discussed, the majority of the community still considers that the main metric for research impact is the number of citations a paper receives. To limit potential construct threats in this study, the GQM approach was used to preserve the tractability between research goal, questions and measurements. RQs were designed to cover our goal.

Conclusion validity

Conclusion validity of a study deals with whether correct conclusions are reached through rigorous and repeatable treatments (Wohlin et al. 2000). In this study, all the discussions and conclusions are directly based on the data and their relevant statistics.

External validity

External validity is concerned with to what extent the results of this study can be generalized (Wohlin et al. 2000). Generalizability is not applicable in this study, since we did not assess a particular case (as in “case studies”). The results of this study are not meant to be generalized to fields outside SE.

Conclusions and future work

This paper presents an exploratory bibliometrics assessment of the SE research literature in term of quantity versus impact of papers (as measured by citation counts). As the trends throughout this paper depicted, the SE literature is very active and the number of papers in this area is increasing each year. However, about 43% of the papers in this area have received no citations at all. This raises the following questions: why is there such a large ratio of uncited papers? How does this trend compare to other research areas? Is it because we have too many less-known venues that publish papers not seen or read by others? Does this have anything to do with papers quality or venues quality?

In terms of impact, we found that review papers and journal articles are the top two types of publications in SE in terms of the ACPP metric, with 18.5 and 12.6. These values contrast with papers published in conferences (3.6), yet very popular venues to publish papers in SE in particular and CS in general. These values seem to indicate that SE researchers should target journals (instead of conferences), as the impact is in general higher.

The pool that we have made publicly available (Garousi and Mäntylä 2015; Garousi and Fernandes 2016) can be used to conduct other thematic and demographic analysis in SE and its sub-domains. Also, this bibliometric approach can be repeated periodically to analyze the growth and trends in the field in upcoming years and compare the future trends to the findings of this study.

We pointed out a number of future research directions in our recent work (Garousi and Fernandes 2016), which are also applicable in the context of this work, i.e., related to what makes SE papers highly-cited (-impactful). In a paper entitled “Highly-cited works in neurosurgery” (Ponce and Lozano 2010), possible determinants of the likelihood of high citations were listed as: the time of publication, field of study, nature of the work, and the journal in which the work appears. It would be interesting to investigate whether those determinants are also applicable in the SE domain. We think that other factors should also be added to this list, e.g., writing style and proper usage of English. Our other future work directions are the followings: (1) to replicate this study after several years to see the differences (if any), and (2) to adopt interesting ideas and approaches from other disciplines to assess quantity versus impact of SE papers. Similar to existing empirical studies, e.g., Mingers and Lipitakis (2010), Archambault et al. (2009), Falagas et al. (2008), Abrizah et al. (2013), Chadegani et al. (2013) and Harzing and Alakangas (2016), which have compared the performance and coverage of paper search engines (e.g., Google Scholar, Scopus and Web of Science) in other fields, e.g., social sciences, an interesting future work will be to conduct experimental investigation on data stability, consistency and coverage of citation data among various paper search engines focusing on software engineering.

Acknowledgements Vahid Garousi was partially supported by several internal grants provided by the Hacettepe University and the Scientific and Technological Research Council of Turkey (TÜBİTAK). João M. Fernandes was supported by FCT - Fundação para a Ciência e Tecnologia within the Project Scope UID/CEC/00319/2013.

References

- Abramo, G., Cicero, T., & D’Angelo, C. A. (2014). Are the authors of highly cited articles also the most productive ones? *Journal of Informetrics*, 8, 89–97.
- Abrizah, A., Zainab, A. N., Kiran, K., & Raj, R. G. (2013). LIS journals scientific impact and subject categorization: a comparison between Web of Science and Scopus. *Scientometrics*, 94, 721–740.

- Aksnes, D. W. (2003). Characteristics of highly cited papers. *Research Evaluation*, 12, 159–170.
- Allison, P. D., & Stewart, J. A. (1974). Productivity differences among scientists: Evidence for accumulative advantage. *American Sociological Review*, 39, 596–606.
- Altmetric LLP, “Bookmarklet for Researchers,” in <https://www.altmetric.com/products/free-tools/bookmarklet/>. Last Accessed May 2017.
- Antonakis, J., Bastardo, N., Liu, Y. H., & Schriesheim, C. A. (2014). What makes articles highly cited? *Leadership Quarterly*, 25, 152–179.
- Archambault, É., Campbell, D., Gingras, Y., & Larivière, V. (2009). Comparing bibliometric statistics obtained from the web of science and scopus. *Journal of American Society for Information Science*, 60, 1320–1326.
- Aversa, E. (1985). Citation patterns of highly cited papers and their relationship to literature aging: A study of the working literature. *Scientometrics*, 7, 383–389.
- Basili, V. R. (1992) Software modeling and measurement: The Goal/Question/Metric paradigm. In: Technical Report, University of Maryland at College Park.
- Belcher, D. D. (2007). Seeking acceptance in an English-only research world. *Journal of Second Language Writing*, 16, 1–22.
- Biswas, A. K. and Kirchherr, J. (2016) Prof, no one is reading you. <http://www.straitstimes.com/opinion/prof-no-one-is-reading-you>. Last Accessed May 2017.
- Bornmann, L. (2014). How are excellent (highly cited) papers defined in bibliometrics? A quantitative analysis of the literature. *Research Evaluation*, 23, 166–173.
- Bornmann, L., de Moya Anegón, F., & Leydesdorff, L. (2010). Do scientific advancements lean on the shoulders of giants? A bibliometric investigation of the ortega hypothesis. *PLoS ONE*, 5, e13327.
- Broad, W. (1981). The publishing game: Getting more for less. *Science*, 211, 1137–1139.
- Buttliere, B., & Buder, J. (2017). Personalizing papers using Altmetrics: Comparing paper ‘Quality’ or ‘Impact’ to person ‘Intelligence’ or ‘Personality’. *Scientometrics*, 111, 219–239.
- Chadegani, A. A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., et al. (2013). A comparison between two main academic literature collections: Web of science and scopus databases. *Asian Social Science*, 9, 18–26.
- Chidamber, S. R., & Kemerer, C. F. (1994). A metrics suite for object oriented design. *IEEE Transactions on Software Engineering*, 20, 476–493.
- Cole, S., Cole, J. R., & Dietrich, L. (1978). Measuring the cognitive state of scientific disciplines. In Y. Elkana, J. Lederberg, R. K. Merton, A. Thackray, & H. Zuckerman (Eds.), *Toward a Metric of Science: The Advent of Science Indicators* (pp. 209–251). Hoboken: Wiley.
- Corby, Z. (2010). To be the best, cite the best. *Nature News*. doi:10.1038/news.2010.539.
- Danell, R. (2011). Can the quality of scientific work be predicted using information on the author’s track record? *Journal of the American Society for Information Science and Technology*, 62, 50–60.
- de Freitas, F., & de Souza, J. (2011). Ten years of search based software engineering: A bibliometric analysis. In M. Cohen & M. O. Cinnéide (Eds.), *Search Based Software Engineering* (Vol. 6956, pp. 18–32). Berlin Heidelberg: Springer.
- Easterbrook, S., Singer, J., Storey, M.-A., & Damian, D. (2008). Selecting empirical methods for software engineering research. In F. Shull, J. Singer, & D. K. Sjøberg (Eds.), *Guide to advanced empirical software engineering* (pp. 285–311). London: Springer.
- Eaton, D. J. (2014). Highly cited papers in medical physics. *Medical Physics*, 41, 43–44.
- Egger, M., & Smith, G. D. (1998). Bias in location and selection of studies. *BMJ British Medical Journal*, 316, 61–66.
- Eric, W. W., Tse, T. H., Glass, R. L., Basili, V. R., & Chen, T. Y. (2011). An assessment of systems and software engineering scholars and institutions (2003–2007 and 2004–2008). *Journal of Systems and Software*, 84, 162–168.
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PUBMED, scopus, web of science, and google scholar: Strengths and weaknesses. *The FASEB Journal*, 22, 338–342.
- Farhoodi, R., Garousi, V., Pfahl, D., & Sillito, J. P. (2013). Development of scientific software: A systematic mapping, bibliometrics study and a paper repository. *International Journal of Software Engineering and Knowledge Engineering*, 23, 463–506.
- Fernandes, J. M. (2014). Authorship trends in software engineering. *Scientometrics*, 101, 257–271.
- Flowerdew, J. (1999). Problems in writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing*, 8, 243–264.
- Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and innovation in scientists’ research strategies. *American Sociological Review*, 80, 875–908.
- Freyne, J., Coyle, L., Smyth, B., & Cunningham, P. (2010). Relative status of journal and conference publications in computer science. *Communications of the ACM*, 53, 124–132.

- Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, *122*, 108–111.
- Garousi, V. (2015). A bibliometric analysis of the Turkish software engineering research community. *Springer Journal on Scientometrics*, *105*, 23–49.
- Garousi, V., & Fernandes, J. M. (2016a). Highly-cited papers in software engineering: The top-100. *Information and Software Technology*, *71*, 108–128.
- Garousi, V. and Fernandes, J. M. (2016) All source data for quantity versus quality: A bibliometric assessment of the number versus impact of software engineering research papers. <https://goo.gl/JYgRQB>. Last Accessed May 2017.
- Garousi, V., and Mäntylä, M. V. (2015) All source data for bibliometrics study of the software engineering community in <https://goo.gl/G8f0M0>. Last Accessed May 2017.
- Garousi, V., & Mäntylä, M. V. (2016). Citations, research topics and active countries in software engineering: A bibliometrics study. *Elsevier Computer Science Review*, *19*, 56–77.
- Garousi, V., Petersen, K., & Özkan, B. (2016). Challenges and best practices in industry-academia collaborations in software engineering: A systematic literature review. *Information and Software Technology*, *79*, 106–127.
- Garousi, V., & Ruhe, G. (2013). A bibliometric/geographic assessment of 40 years of software engineering research (1969–2009). *International Journal of Software Engineering and Knowledge Engineering*, *23*, 1343–1366.
- Garousi, V., & Varma, T. (2010). A bibliometric assessment of canadian software engineering scholars and institutions (1996–2006). *Canadian Journal on Computer and Information Science*, *3*, 19–29.
- Génova, G., Astudillo, H., & Fraga, A. (2016). The scientometric bubble considered harmful. *Science and Engineering Ethics*, *22*, 227–235.
- Ghosh, A., Chattopadhyay, N., & Chakrabarti, B. K. (2014). Inequality in societies, academic institutions and science journals: Gini and k-indices. *Physica A: Statistical Mechanics and its Applications*, *410*, 30–34.
- Grigore, R. (2007) David Parnas is wrong!. <http://rgrig.blogspot.com.tr/2007/10/david-parnas-is-wrong.html>. Last Accessed May 2017.
- Halfman, W., & Leydesdorff, L. (2010). Is inequality among universities increasing? Gini coefficients and the elusive rise of elite universities. *Minerva*, *48*, 55–72.
- Hamrick, T. A., Fricker, R. D., & Brown, G. G. (2010). Assessing what distinguishes highly cited from less-cited papers published in interfaces. *Interfaces*, *40*, 454–464.
- Harzing, A.-W., & Alakangas, S. (2016). Google scholar, scopus and the web of science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, *106*, 787–804.
- Ioannidis, J. P. A., Boyack, K. W., Small, H., Sorensen, A. A., & Klavan, R. (2014). Is your most cited work your best? *Nature*, *514*, 561–562.
- King, D. A. (2004). The scientific impact of nations. *Nature*, *430*, 311–316.
- Latour, B., & Woolgar, S. (1979). *Laboratory Life: The Construction of Scientific Facts*. Los Angeles: Sage.
- Lawrence, P. A. (2003). The politics of publication. *Nature*, *422*, 259–261.
- LibGuides at Duke University Medical Center, “Altmetrics: a primer,” in <http://guides.mclibrary.duke.edu/altmetrics/home>. Last Accessed May 2017.
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, *16*, 317–323.
- Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science*, *159*, 56–63.
- Mingers, J., & Lipitakis, E. A. E. C. G. (2010). Counting the citations: a comparison of Web of Science and Google Scholar in the field of business and management. *Scientometrics*, *85*, 613–625.
- Miyairi, N., & Chang, H.-W. (2012). Bibliometric characteristics of highly cited papers from Taiwan, 2000–2009. *Scientometrics*, *92*, 197–205.
- Moed, H. F. (2006). *Citation analysis in research evaluation*. Berlin: Springer.
- Moed, H. F. and Visser, M. S. (2007) Developing bibliometric indicators of research performance in computer science: An exploratory study. In: *Research Report, Centre for Science and Technology Studies (CWTS), Leiden University, the Netherlands*.
- Newman, M. E. J. (2014). Prediction of highly cited papers. *Europhysics letters (EPL)*, *105*, 6.
- Nieminen, P., Carpenter, J., Rucker, G., & Schumacher, M. (2006). The relationship between quality of research and citation frequency. *BMC Medical Research Methodology*, *6*, 1–8.
- Noorden, R. V., Maher, B., & Nuzzo, R. (2014). The top 100 papers. *Nature*, *514*, 550–553.
- Nuseibeh, B. (2011). Editorial: What makes a publication archival? *Software Engineering IEEE Transactions on*, *37*, 145.

- Osterweil, L. J., Ghezzi, C., Kramer, J., & Wolf, A. L. (2008). Determining the impact of software engineering research on practice. *Computer*, *41*, 39–49.
- Paris, G., Leo, G. D., Menozzi, P., & Gatto, M. (1998). Region-based citation bias in science. *Nature*, *396*, 210.
- Parnas, D. L. (2007). Stop the numbers game. *Communications of the ACM*, *50*, 19–21.
- Patterson, D. A. (2004). The health of research conferences and the dearth of big idea papers. *Communication of the ACM*, *47*, 23–24.
- Persson, O. (2010). Are highly cited papers more international? *Scientometrics*, *83*, 397–401.
- Piwowar, H. (2013). Altmetrics: Value all research products. *Nature*, *493*, 159.
- Ponce, F. A., & Lozano, A. M. (2010). Highly cited works in neurosurgery. Part I: The 100 top-cited papers in neurosurgical journals. *Journal of Neurosurgery*, *112*, 223–232.
- Priem, J., Taraborelli, D., Groth, P., Neylon, C. (2016) Altmetrics: A manifesto, in <http://altmetrics.org/manifesto>, 2010. Last Accessed May 2017.
- Pyšek, P., Richardson, D. M., & Jarošík, V. (2006). Who cites who in the invasion zoo: insights from an analysis of the most highly cited papers in invasion ecology. *Preslia*, *78*, 437–468.
- Rahm, E. (2008). Comparing the scientific impact of conference and journal publications in computer science. *Inf. Serv. Use*, *28*, 127–128.
- Reuters, T. (2016) Highly cited researchers. <http://hcr.stateofinnovation.thomsonreuters.com/>. Last Accessed May 2017.
- Saha, S., Saint, S., & Christakis, D. A. (2003). Impact factor: A valid measure of journal quality? *Journal of the Medical Library Association*, *91*, 42–61.
- Sandström, U., & van den Besselaar, P. (2016). Quantity and/or quality? The importance of publishing many papers. *PLoS ONE*, *11*, e0166149.
- Sarewitz, D. (2016). The pressure to publish pushes down quality. *Nature*, *533*, 147.
- Schliesser, E. (2016) On 'Me 2 research' and English-as-Second-language citation rates. <http://digressionsnimpresions.typepad.com/digressionsnimpresions/2014/10/on-me-2-research-and-non-english-citation-rates.html>. Last Accessed May 2017.
- Siler, K., Lee, K., & Bero, L. (2015). Measuring the effectiveness of scientific gatekeeping. *Proceedings of the National Academy of Sciences*, *112*, 360–365.
- Tijssen, R., Visser, M., & van Leeuwen, T. (2002). Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics*, *54*, 381–397.
- Tongai, I. (2013) Incentives for researchers drive up publication output. <http://www.universityworldnews.com/article.php?story=20130712145949477>. Last Accessed May 2017.
- Uzuner, S. (2008). Multilingual scholars' participation in core/global academic communities: A literature review. *Journal of English for Academic Purposes*, *7*, 250–263.
- Vardi, M. Y. (2009). Conferences vs. journals in computing research. *Communication of the ACM*, *52*, 5.
- Various authors. (2016). Why is it said that judging a paper by citation count is a bad idea?. <http://academia.stackexchange.com/questions/37021/why-is-it-said-that-judging-a-paper-by-citation-count-is-a-bad-idea>. Last Accessed May 2017.
- Vasconcelos, S. M. R., Sorenson, M. M., & Leta, J. (2007). Scientist-friendly policies for non-native English-speaking authors: timely and welcome. *Brazilian Journal of Medical and Biological Research*, *40*, 743–747.
- Vrettas, G., & Sanderson, M. (2015). Conferences versus journals in computer science. *Journal of the Association for Information Science and Technology*, *66*, 2674–2684.
- Wang, M., Yu, G., & Yu, D. (2011). Mining typical features for highly cited papers. *Scientometrics*, *87*, 695–706.
- West, R., & Stenius, K. (2009). Use and abuse of citations. In T. F. Babor, K. Stenius, S. Savva, & J. O'Reill (Eds.), *Publishing addiction science: A guide for the perplexed*. Multi-Science Publishing Co.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2000). *Experimentation in Software Engineering: An Introduction*. Berlin: Kluwer Academic Publishers.
- Wong, B. B. M., & Kokko, H. (2005). Is science as global as we think? *Trends in Ecology & Evolution*, *20*, 475–476.
- Wong, W. E., Tse, T. H., Glass, R. L., Basili, V. R., & Chen, T. Y. (2008). An assessment of systems and software engineering scholars and institutions (2001–2005). *Journal of Systems and Software*, *81*, 1059–1062.
- Wong, W. E., Tse, T. H., Glass, R. L., Basili, V. R., & Chen, T. Y. (2009). An assessment of systems and software engineering scholars and institutions (2002–2006). *Journal of Systems and Software*, *82*, 1370–1373.
- Zeng, L., Benatallah, B., Ngu, A. H. H., Dumas, M., Kalagnanam, J., & Chang, H. (2004). QoS-aware middleware for Web services composition. *IEEE Transactions on Software Engineering*, *30*, 311–327.