



Highly-cited papers in software engineering: The top-100



Vahid Garousi^{a,*}, João M. Fernandes^b

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

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

ARTICLE INFO

Article history:

Received 27 May 2015

Revised 11 November 2015

Accepted 11 November 2015

Available online 23 November 2015

Keywords:

Software engineering

Highly-cited papers

Top cited

Most cited

Most frequently cited

Bibliometrics

ABSTRACT

Context: According to the search reported in this paper, as of this writing (May 2015), a very large number of papers (more than 70,000) have been published in the area of Software Engineering (SE) since its inception in 1968. Citations are crucial in any research area to position the work and to build on the work of others. Identification and characterization of highly-cited papers are common and are regularly reported in various disciplines.

Objective: The objective of this study is to identify the papers in the area of SE that have influenced others the most as measured by citation count. Studying highly-cited SE papers helps researchers to see the type of approaches and research methods presented and applied in such papers, so as to be able to learn from them to write higher quality papers which will likely receive high citations.

Method: To achieve the above objective, we conducted 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.

Results: 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.

Conclusion: It is concluded that it is important to identify the highly-cited SE papers and also to characterize the overall citation landscape in the SE field. We hope that this paper will encourage further discussions in the SE community towards further analysis and formal characterization of the highly-cited SE papers.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

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 [1]. Citations are usually used to quantify the impact of papers and journals, a practice not without controversy. Modern, formal use of citations in scientific literature dates back only to the nineteenth century as scholars and scientists started to give continuity to their body of ideas [1]. In 1955, Eugene Garfield published the Science Citation Index (SCI) [2], the first systematic effort to track citations in the scientific literature.

Under the rubric of bibliometrics, citation counts have been incorporated into metrics intended to measure the impact of researchers, papers, journals, universities and even countries. 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'. Such an excellence is usually measured by citation counts [3]. 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 in different disciplines [4].

A large number of papers have been published in the area of Software Engineering (SE) since its inception in 1968. In Dec. 2014, a search for papers published in the venues (e.g., journals and conferences) including the phrase "software", as indexed by the Scopus publication search engine, returned almost 70,000 records (with details reported in Section 3.3). Identification and classification of highly-cited papers in various areas of science, e.g., medicine, physics and social sciences, are quite common and regularly reported, e.g., [1,4–20]. However, only a few studies [21–26] have identified and analyzed highly-cited papers in SE in small-scale (e.g., only in a selected subset of venues), the last of which was published in 2010. Thus, there is a need for more recent and more comprehensive such studies in SE.

Identification and classification of highly-cited SE papers provide various benefits for researchers and practitioners, e.g., (1) the results help new researchers to see the type of contributions, approaches and

* Corresponding author. Tel.: +90 (312) 297-7500 x.128.

E-mail addresses: vahid.garousi@hacettepe.edu.tr (V. Garousi), jmf@di.uminho.pt (J.M. Fernandes).

research methods used and presented in highly-cited papers so as to be able to learn from them to write papers which will be of high quality and will likely receive high citations, (2) the classifications help both established and new researchers to spot the active and more impactful topics and thus they can carry one further incremental research on those areas, (3) using the results, researchers and practitioners can notice the most cited researchers and collaborate with them, get advice from them, etc., and (4) the results would help practitioners spot the highest quality work in specific areas of SE and aim at utilizing techniques, tools or findings reported in those studies in their real-world SE challenges.

To position the current study w.r.t. the related work, to highlight, up front, the reasons behind the different analyses presented and what the reader is supposed to take away from the paper, we note that this is the first most comprehensive study about top-cited papers in SE. As discussed in Section 2.3, since various research questions and objectives w.r.t. characterization of top-cited papers have been followed in other disciplines (e.g., neurosurgery [11] and ecology [7]), we narrow our focus in this work on a subset of those objectives as listed next: (1) identifying the top-cited papers which would yield the benefits discussed in the above paragraph, (2) characterizing the individual citation counts for top papers to see the trends in SE, to compare them to other areas (similar to what has been done in neurosurgery [11] and in all areas of science [18]), and to comparatively assess the scale of impact/popularity across disciplines, (3) mapping of the top-cited studies based on their SE fields of study (similar to what has been done in neurosurgery [11] and invasion ecology [7]) to assess the extent to which each SE sub-area is represented in the top list, (4) identifying the top venues for top papers as to inform researchers of the top venues which might behave as an ‘external factor’ [12] leading to increased popularity (citation) of their planned submissions, and (5) characterizing authorship in top papers so as we can assess the authorship team sizes in top papers in SE.

With the above motivations in mind, the objective of this paper is to systematically identify, rank, characterize, and classify the papers in the area of SE that have influenced others the most as measured by citation count. This paper also intends to encourage further discussions in the SE community towards formal characterization of the highly-cited SE papers, similar to what is regularly done in other disciplines, e.g., [6]. To achieve the above objective, we designed and conducted a systematic method to identify the top-100 highly-cited SE papers in terms of two metrics: total number of citations and average annual number of citations.

The remainder of the paper is organized as follows. Section 2 discusses the related work. Section 3 describes our research method, including the goal and research questions tackled in this study. Section 4 presents the results of the study. Section 5 summarizes the findings and implications, and discusses the potential threats to validity of our study. Finally, Section 6 concludes this study and states the future work directions.

2. Related work

The following three groups of work are related to this study and are discussed in the following sections.

- Studies reviewing highly-cited papers in SE
- Bibliometrics studies in SE
- Studies reviewing highly-cited papers in other disciplines

2.1. Studies reviewing highly-cited papers in SE

There are only a few studies [21–26] which analyze highly-cited papers in SE, that we were able to find in the academic search engines. Table 1 lists those papers and their notable findings.

The sequential series of four papers by Wohlin [21–24] analyzes the most cited papers in SE journals between 1999 and 2002. As discussed by Wohlin, the intention of the analysis in those four papers was twofold: (1) to identify the most cited papers, and (2) to invite the authors of the most cited papers to contribute to a special section of the Information and Software Technology journal.

The study reported in [25] conducted a classification of papers published in seven top journals and seven top international conferences in SE based on SE subjects (e.g., verification, testing and tools) and found that 73% of journal papers and 89% of conference papers have focused on 20% of the topics, an incarnation of the Pareto rule (a.k.a., the 80/20 rule).

The study reported in [26] focused on software metrics and selected the most cited papers published between 2000 and 2005 and then classified, using a systematic mapping approach, the papers based on their research approach. Among the findings is that most papers were journal papers with an empirical content, particularly papers published in the IEEE Transactions on Software Engineering (TSE).

2.2. Bibliometrics studies in SE

Another body of work related to this study is the bibliometrics studies in SE. Bibliometric rankings are quite common in SE. Table 2 lists a few representative studies [29–36] along with their notable findings.

The series of 12 papers by Glass and Chen, three of which are cited in Table 2 [29,30,33], was an ongoing, annual event that identified the top 15 SE scholars and institutions for a sliding five-year period in systems and software engineering 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 [31] presents a bibliometric assessment of Canadian SE scholars and institutions. Additional findings reported in [31] 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 [32] presents a bibliometric analysis of ten years of search-based SE.

Some recent systematic mapping studies, such as [34], report, as a part of their studies, bibliometric analyses of SE sub-areas, e.g., development of scientific software in [34]. Among the findings reported in [34] 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 British researchers.

The study reported in [35] 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 reports a bibliometric study [36] which focuses on authorship trends in SE. The researcher collected 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. The author indicates that the number of authors of articles in SE is increasing on average around 0.40 authors/decade. Also, the results indicate that until 1980, the majority of the articles have one author, while articles from the 90s until today with 3 or 4 co-authors 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.

Table 1
Studies reviewing highly-cited papers in SE (sorted by years of publications).

Ref.	Year	Topic	Notable findings
[21]	2005	An analysis of the most cited papers in software engineering journals-1999	<ul style="list-style-type: none"> An analysis of the 20 most cited SE journal papers in the 20 year period of 1979–1999 is presented. Most cited papers are ranked using two metrics: absolute numbers of citations and the average number of citations per year. The research topics and methods of the most cited papers in 1999 are compared with those from the most cited papers in 1994 to provide a picture of similarities and differences between the years. The top cited paper is “use case maps as architectural entities for complex systems” [27] with only 25 citations.
[22]	2007	An analysis of the most cited papers in software engineering journals-2000	<ul style="list-style-type: none"> The paper describing the SPIN model checker [28] by G.J. Holzmann published in 1997 is the first using both metrics.
[23]	2008	An analysis of the most cited papers in software engineering journals-2001	<ul style="list-style-type: none"> The most productive author in the 20-year period of 1981–2001 is Victor Basili.
[24]	2009	An analysis of the most cited papers in software engineering journals-2002	<ul style="list-style-type: none"> The top cited paper is “Preliminary guidelines for empirical research in software engineering” with 64 citations.
[25]	2008	An analysis of research topics in software engineering–2006	<ul style="list-style-type: none"> The paper examines all the 691 papers published in a selected list of venues in 2006. 73% of journal papers focus on 20% of subjects in SE, including testing and debugging, management, and software/program verification. 89% of conference papers focus on 20% of subjects in SE, including software/program verification, testing and debugging, and design tools and techniques. The average number of 7 top journals and 7 top international conferences in SE references cited by a journal paper is about 33, whereas this number becomes around 24 for a conference paper.
[26]	2010	What’s up with software metrics: A preliminary mapping study	<ul style="list-style-type: none"> The paper selects the most cited papers in the years 2000–2005 inclusive restricting the number of papers in the pool to 15 in each year. Papers are systematically mapped based on research and contribution facets. The study suggests that the metrics community is influenced primarily by journal papers with an empirical content, particularly papers published in TSE. Metrics research is not dominated by object-orientation but a reasonably large proportion of papers (approximately one third) are OO related. Compared with less cited papers, the most cited papers were more frequently journal papers, and empirical validation or data analysis studies.

2.3. Studies reviewing highly-cited papers in other disciplines

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 [4]. More recently, the cover story of the October 2014 issue of the prestigious Nature magazine was “The top 100 papers” [18]. That Nature issue includes several papers (e.g., [16]) on the issue of highly-cited papers in various scientific disciplines.

According to [4], as of 2010, the search in the Web of Science yielded 321 papers dealing with “highly-cited”, “most cited”, “top cited” and “most frequently cited”. The authors browsed the list of these papers and made a selection of the most relevant papers for the current paper as shown in Table 3. As we can see, most of the studies are neutral of scientific areas, while a few have analyzed the highly-cited papers in specific disciplines, e.g., neurosurgery [11] and medical physics [19].

As we can observe under “notable findings” in Table 3, various research questions and objectives w.r.t. characterization of top-cited papers have been followed in other disciplines, e.g., (1) literature aging of highly-cited papers based on their citation patterns [20], (2) are highly-cited research papers an appropriate frame of reference for identifying “world class” scientific excellence [5]?, (3) review papers are over-represented in the top list compared to the average [6], (4) the citation curves of highly-cited papers usually follow a typi-

cal pattern of rise and decline, and (5) there is a distinction between quality dynamics and visibility dynamics of top papers [6]. Overall, we can clearly see that papers characterizing top-cited papers are quite widespread in many disciplines and the need for them is arising as countries, funding agencies and universities are trying to gauge research performance and identify top researchers and research impact. As discussed in Section 1, we narrow our focus in this work on a subset of those objectives, as we further discuss in Section 3.

3. 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.

3.1. Goal and research questions

The research approach we have used in our study is the Goal, Question, Metric (GQM) methodology [38]. Using the GQM’s goal template [38], the goal of this study is to systematically identify the highly-cited papers in the field of SE, analyze the coverage of SE topics in top papers and to conduct initial comparison of the citations of the top-100 SE papers with other disciplines, from the point of view of the researchers in this area. Based on the above goal, we raised the following research questions (RQs). We should note that the goal and RQs of the study are *exploratory* and *descriptive* in nature [39].

Table 2
A few selected bibliometrics studies in SE (sorted by years of publications).

Ref.	Year	Topic	Notable findings
[29]	2008	An assessment of systems and software engineering scholars and institutions (2001–2005)	<ul style="list-style-type: none"> 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.
[30]	2009	An assessment of systems and software engineering scholars and institutions (2002–2006)	<ul style="list-style-type: none"> 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.
[31]	2010	A bibliometric assessment of Canadian software engineering scholars and institutions (1996–2006)	<ul style="list-style-type: none"> 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).
[32]	2011	Ten years of search-based software engineering: a bibliometric analysis	<ul style="list-style-type: none"> 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.
[33]	2011	An assessment of systems and software engineering scholars and institutions (2003–2007 and 2004–2008)	<ul style="list-style-type: none"> 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.
[34]	2011	Development of scientific software: a systematic mapping, bibliometrics study and a paper repository	<ul style="list-style-type: none"> 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.
[35]	2013	A bibliometric/geographic assessment of 40 years of software engineering research (1969–2009)	<ul style="list-style-type: none"> 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. Over the 40-year period (1969–2009), in total about 60% of the SE literature has been contributed by only 7% of all countries. The US is the clear leader, followed by UK and China. The SE research output of different countries does not necessarily correlate with their GDPs. The share of contributions to the SE discipline by the American researchers has declined from 71.43% (in 1980) to 14.90% (in 2008). 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).
[36]	2014	Authorship trends in SE	<ul style="list-style-type: none"> Around 70,000 entries from the DBLP for 122 conferences and journals, for the period 1971–2012, were collected. The number of authors of articles in SE is increasing on average around 0.40 authors/decade. 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.

- RQ 1: Citation landscape of the SE literature: What is the distribution of citations for the SE papers? E.g., what ratio of SE papers has had no citations?
- RQ 2: The top papers:
 - RQ 2.1: What are the highly-cited papers in SE?
 - RQ 2.2: How do the citation counts of the top-100 SE papers compare with those of the top-100 papers in all areas of science (data from [18]) and other disciplines?
- RQ 3: Coverage of SE topics in top papers: Which areas of SE have been covered by the top-cited papers? For example, are most of the top papers focused on a few of the SE areas (e.g., requirements, design, and testing)? or are the distributions quite balanced? As discussed in Section 2.1, previous works have also studied such type of questions, e.g., a related study [25] found that 73% percent of SE journal papers focused on 20% of the SE topics.
- RQ 4: Top venues for top papers: What journals and conferences have been the publication venues for top papers? Answering this RQ will enable us to identify the venues in which top papers are published. It is often believed and also systematically analyzed

that venue reputation may have an impact on the high citation of papers [12].

- RQ 5: Authorship in top papers: What are the authorship trends among the top papers? E.g., are most of the papers single-authored or written by a high number of authors?

3.2. Metrics

The most important metric(s) that we had to select was regarding the identification of the highly-cited papers. Absolute numbers of citations is the most obvious metric. As the second complementary metric, we selected the average annual number of citations to a given paper. The latter metric is widely used in addition to the absolute numbers of citations, e.g., in medical physics [19], and in SE, e.g., [21–24], since bibliometricians “prefer to compare citation counts for papers of similar age” [18]. In other words, that metric normalizes the effect of publication year (age) on the total numbers of citations.

Table 3

A summary of studies reviewing highly-cited papers in other disciplines (sorted by year of publication).

Ref.	Year	Topic	Notable findings
[20]	1988	Citation patterns of highly-cited papers and their relationship to literature aging	<ul style="list-style-type: none"> • Citation patterns of 400 very highly-cited scientific papers, as of 1988, were identified and the relationship of citation patterns to literature aging rates was investigated. • Standardized citation counts for 1972 through 1980 were used as variables in a cluster analysis which groups papers with similar citation patterns. • Among highly-cited papers published in 1972, two basic citation patterns were identified: <ul style="list-style-type: none"> ◦ One group was highly-cited in the first years following publication and declines in terms of citations thereafter ◦ The second group reached its citation peak in the sixth year following publication and declines in terms of citations in the seventh, eighth, and ninth years of the series. ◦ (3) Both groups show general evidence of aging.
[5]	2002	Benchmarking international scientific excellence: Are highly-cited research papers an appropriate frame of reference?	<ul style="list-style-type: none"> • The findings indicate that these high performance papers provide a useful analytical framework - both in terms of transparency, cognitive and institutional differentiation, as well as its scope for domestic and international comparisons - providing new indicators for identifying "world class" scientific excellence
[6]	2003	Characteristics of highly-cited papers	<ul style="list-style-type: none"> • The majority of the papers represent regular journal papers (81%), although review papers (12%) are over-represented compared to the average. • The citation curves of highly-cited papers follow a typical pattern of rise and decline. • The paper introduces a conceptual distinction between quality dynamics and visibility dynamics of top papers.
[7]	2006	Who cites who in the invasion zoo: insights from an analysis of the most highly-cited papers in invasion ecology	<ul style="list-style-type: none"> • Papers were classified into broad research fields under ecology, similar to a systematic mapping [37] • The annual citation rate increased with time over the analyzed period (1981–2003), by 1.0 citations per year.
[8, 9]	2010	Do scientific advancements lean on the shoulders of giants?	<ul style="list-style-type: none"> • In all fields, the papers which went on to be most highly-cited were more likely to reference previous highly-cited work than were less popular papers. • "To be the best, cite the best"
[10]	2010	Are highly-cited papers more international?	<ul style="list-style-type: none"> • It is concluded that international papers are not well represented among high impact papers in research specialities, but dominate highly-cited papers from small countries, and from cities and institutions within them. • Domestic papers from the USA comprise about half of the highly-cited papers in the research specialities
[1]	2010	Assessing what distinguishes highly-cited from less-cited papers published in the <i>Interfaces</i> journal (in the area of operations research)	<ul style="list-style-type: none"> • The study found that competition papers, longer papers, tutorials, and papers with larger numbers of references to prior literature tend to have a higher number of citations.
[11]	2010	The 100 highly-cited works in neurosurgery	<ul style="list-style-type: none"> • The 100 most cited manuscripts in neurosurgical journals appeared in 3 of 13 journals dedicated to neurosurgery. • The individual citation counts for these papers ranged from 287 to 1,515. • A mapping of the studies based on their fields of study and research approaches was conducted. • The time of publication, field of study, nature of the work, and the journal in which the work appears are possible determinants of the likelihood of citation and impact.
[12]	2011	Mining typical features for highly-cited papers	<ul style="list-style-type: none"> • By integrating papers' external features and quality features, the feature space used to model highly-cited papers was established. • A list of prediction features for highly-cited papers were extracted on the basis of a multi-classifier system. • The findings show that both the papers' inner quality and external features, mainly represented as the reputation of the authors and journals, contribute to generation of highly-cited papers in future.
[13]	2012	Bibliometric characteristics of highly-cited papers from Taiwan, 2000–2009	<ul style="list-style-type: none"> • Taiwan's output of highly-cited papers was greatest in the categories of engineering, clinical medicine, and physics, while those in agricultural sciences and mathematics exceeded the expected output level in relative terms.
[17]	2014	Prediction of highly-cited papers	<ul style="list-style-type: none"> • The authors developed methods for early detection of candidate breakthroughs, based on dynamics of publication citations. • The study proposed two forecasting models that were validated using statistical methods to derive confidence levels. • These findings can be used to inform research portfolio management practices.

(continued on next page)

Table 3 (continued)

Ref.	Year	Topic	Notable findings
[14]	2014	Are the authors of highly-cited papers also the most productive ones?	<ul style="list-style-type: none"> The authors used bibliometric indicators, applied to the 2004–2008 publications authored by academics of Italian universities and indexed in the Web of Science. There is a moderate correlation between the phenomenon of being a top-productive scientist and the probability of having produced highly-cited papers.
[16]	2014	Is your most cited work your best?	<ul style="list-style-type: none"> The authors asked the 400 most highly-cited biomedical scientists to score their top-ten papers in six ways. On average, authors tended to give their blockbuster papers high scores for dimensions that reflect evolution: continuous progress, broader interest and greater synthesis. They gave their blockbuster papers lower scores on average for dimensions that reflect revolution: disruptive innovativeness and surprise.
[15]	2014	What makes papers highly-cited?	<ul style="list-style-type: none"> This paper examined drivers of paper citations using 776 papers that were published from 1990 to 2012 in a high-impact social sciences journal, <i>The Leadership Quarterly</i>. The regression models developed in this paper showed that quantitative, review, method, and theory papers were significantly more cited than were qualitative papers or agent-based simulations. Regarding statistical conclusion validity of quantitative papers, papers having endogeneity threats received significantly fewer citations than did those using a more robust design or an estimation procedure that ensured correct causal estimation. The study makes several general recommendations on how to improve research practice and paper citations.
[4]	2014	How are excellent (highly-cited) papers defined in bibliometrics	<ul style="list-style-type: none"> The analysis was carried out in order to acquire an overview of the methods used to identify excellent papers and an indication of an "average" or "most frequent" bibliometric practice. The top 1% is used most frequently in the papers, followed by the top 10%.
[18]	2014	The top 100 papers	<ul style="list-style-type: none"> Many of the world's most famous papers are not in the top 100. Vast majority of the papers in the top 100 list describe experimental methods or software that have become essential in their fields. The most cited work in history is a 1951 paper describing an assay to determine the amount of protein in a solution. It has gathered, as of 2014, more than 305,000 citations. The colossal size of the scholarly literature means that the top-100 papers are extreme outliers. Thomson Reuter's Web of Science holds some 58 million items. Only 14,499 papers have more than 1,000 citations. Many methods papers have "become a standard reference that one cites in order to make clear to other scientists what kind of work one is doing". Bibliometricians prefer to compare citation counts for papers of similar age and in comparable fields. The list reveals just how powerfully research has been affected by computation and the analysis of large data sets. Position of any particular methods paper or search engine at the top of the citation charts is also down to luck and circumstance.
[19]	2014	Highly-cited papers in medical physics	<ul style="list-style-type: none"> The top ten papers published in the <i>Medical Physics</i> journal, were ranked using two metrics: total citations and citations per year. Compared to a similar survey ten years ago, some "old classics" have remained, but half of the entries in the top-10 are new. For this particular journal, the mean number of references per paper has risen from 23.2 in 2002 to 31.7 in 2012. What might be the reasons for this increase? It is known that online search engines greatly increase access to published work, which may well lead to more extensive referencing. Social media has also been shown to increase dissemination. The paper recommends authors to take time to distill the key papers relating to their manuscript, and not just include large strings of references to demonstrate that they are widely read on the subject.

3.3. Data extraction

3.3.1. Selection of the publication search engine

To identify the highly-cited SE papers, we had to select a suitable publication search engine. For systematic selection of such a search engine, by reviewing the related review studies [1,4–20], we devised three important selection criteria:

1. The publication search engine should provide the highest quality and reliability in terms of coverage of the SE literature, i.e., including all the SE papers,
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 highly-cited SE papers that we were looking for.

Table 4
Rating of the three candidate publication search engines in terms of the three selection criteria.

Criteria	Publication search engines		
	Scopus	Web of science	Google Scholar
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.
2-Including citation data	Yes	Yes	Yes
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.

Table 5
The search query that was developed to extract of all SE papers from Scopus.

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

To find the candidate publication search engines, we reviewed a large number of bibliometrics studies, in SE (e.g., [29–35]), and fields other than SE (e.g., [40–43]). 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 search engines that researchers regularly use in various bibliometrics studies. DBLP was not further considered, since it does not include citation data. In Table 4, we discuss how the remaining three candidate publication search engines rate in terms of the selection criteria discussed above.

Regarding criterion #1, as shown in Table 4, Scopus scores better than Web of Science, since Scopus has the feature to search by "Source name" (venue names). 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. We included in the search query 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 bibliometrics study [35] in which a bibliometric/geographic assessment of 40 years of SE research (1969–2009) was reported. All the major SE venues including the top SE conferences and journals, e.g., the top 25 venues as listed by the Google Scholar listing¹ 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 publication 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 4, Google Scholar became ineligible for our selection, since exporting the list of extracted papers to files is not automatically possible in a convenient manner (except that one has to write 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 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 results 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.

In conclusion, by summarizing the outcomes with respect to our three selection criteria, the Scopus publication search engine was chosen as the publication search engine from which the highly-cited papers would be identified. A recent paper published in the Nature magazine, titled "The top 100 papers" [18], which was discussed in Section 2, also used Scopus. There have been empirical studies, e.g., [40–43], which have compared the performance and coverage of Web of Science versus Scopus in several fields, e.g., social sciences. Some studies, e.g., [42], have found empirically that Scopus is better than Web of Science in certain aspects, e.g., "larger coverage of titles" [42].

3.3.2. Extraction of all SE papers from Scopus

Having selected Scopus as the publication search engine to conduct the search for the highly-cited SE papers, the next step was to actually conduct the search for those papers. To rank the list of the highly-cited SE papers, we needed to find all the SE papers indexed by Scopus along with their citation counts.

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 discovered by the first author during an informal search for the SE papers authored by the Turkish SE community which later resulted in a publication [44]. By further experimentation, we found that this approach is

¹ https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_softwaresystems

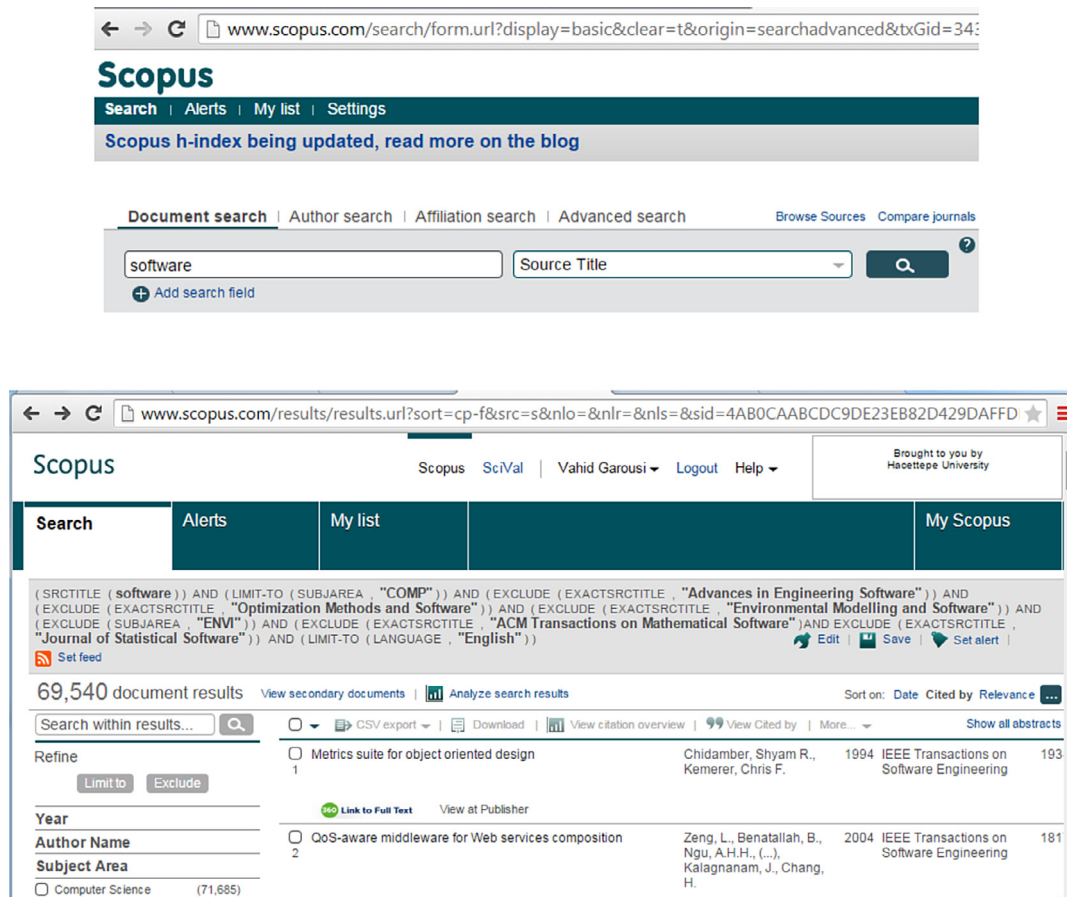


Fig. 1. Two screenshots showing the method used to identify the top papers in the Scopus publication search engine.

indeed quite reliable in terms of coverage of the SE literature and has been used in other disciplines as well [1,4–20].

In the Scopus search interface, we included the phrase “software” under “source title” as shown in Fig. 1. The exact search query that was developed to extract all SE papers from Scopus is shown in Table 5 along with explanations for each phrase in the query. We conducted several rounds of iterative review and excluded unrelated venues (such as, Journal of Optimization Methods and Software) and also non-English papers.

We should also note that the data extraction phase of this study 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 (it seems that there is some sort of a batch processing scheme in place). Thus, the data for 2014 were partial. Furthermore, the citations for papers in 2014 were relatively very low since they were either “In Press” or recently published. For example, our analysis showed that the 2,443 papers (partial count as per the Scopus approach discussed above) published in 2014 had 203 citations, while for 6403 papers published in 2013, there were 3365 citations. Due to the partial situation of the 2014 data set, we decided to not include 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 listing² in the area of Software Systems,

were included in the results returned by Scopus since all the names include the word ‘software’.

Furthermore, we were also aware that a number of 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 (WCORE)
- 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 the programming language community 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 include 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 (in the first week of May 2015), and as a result, 3240 additional papers

² https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_softwaresystems

www.scopus.com/results/results.url?sort=plf-f&src=s&st1=Object-Oriented+Programming%2c+Systems%2c+Lang

The screenshot shows the Scopus search interface. At the top, there are navigation tabs for 'Search', 'Alerts', and 'My list'. Below these, the search query is displayed as 'SRCTITLE (object-oriented programming, systems, languages)'. The search results show '1,768 document results'. There are options to 'View secondary documents' and 'Analyze search results'. A search bar is present with the text 'Search within results...'. Below the search bar, there are options to 'Refine' results, including 'Limit to' and 'Exclude'. A table shows the number of results for each year from 2010 to 2014. The results list includes two entries: 'EventBreak: Analyzing the responsiveness of user interfaces through performance-guided tes' and 'Welcome from the SPLASH 2014 general chair'. Each entry has a 'Link to Full Text' button and a 'View at Publisher' link.

Year	Count
2014	(54)
2013	(53)
2012	(62)
2011	(63)
2010	(154)

Fig. 2. Screenshot showing the query used to identify papers published in the proceedings of the Conference on Object-Oriented Programming, Systems, Languages and Applications (OOPSLA).

were found and added to the pool. As an example, Fig. 2 shows the query used to extract the list of papers published in the proceedings of the Conference on Object-Oriented Programming, Systems, Languages and Applications (OOPSLA).

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 include scientific papers, thus we included records of the following types only: articles, articles in press, book chapters, conference papers and review papers (e.g., survey and systematic review papers), and excluded the rest.

Once we had the pool of papers, we reviewed the records to ensure their integrity, e.g., not having duplicate records of a given paper. It was somewhat surprising that data exported from Scopus had some duplicates. We cleaned up the data set and after applying all the above steps, the final paper pool size to 71,668 papers. To ensure transparency and replicability of our analysis, the entire raw and ranking data for all the papers is available as an Excel file which can be downloaded online [45].

3.4. Initial observations in the dataset

3.4.1. Growth of the SE literature

In terms of the growth of the SE literature, Fig. 3 shows the number of SE papers included in Scopus by their publication year. The earliest publication year was 1972 from which 29 papers were included in Scopus. The annual numbers of papers have grown in the latest years (starting around 2005) and have reached to about 7000 papers each year since 2008.

3.4.2. Authorship trends

Once we had the dataset, we calculated the average number of authors for each paper. We found that the dataset is consistent with the ones used in other SE bibliometric studies, in terms of authorship trends, namely in what concerns the average number of authors for paper, per year, in SE. The line shown in Fig. 4, which is related to the dataset used for the study reported in this article, is quite similar to the one derived in [36] (Fig. 7). As we can see, the trend for the average number of authors in SE is around 1.5 author per paper in the

1970's and it has reached about three authors for a typical SE paper in the years after 2010.

4. Results

4.1. RQ 1: citation landscape of the SE literature

RQ 1 raises the following two sub-questions: What is the distribution of citations for all the SE papers? What ratio of SE papers has had no citations?

The pool of 71,668 papers, along with their citation counts, was automatically extracted from Scopus into a CSV file. To address RQ 1, Fig. 5 shows a scatter plot of all the papers' citation counts versus publication years, along with the corresponding box plots. Note that there are 71,668 points on this graph. As we can see, the data in the X-axis (publication years) are somewhat skewed, while the Y-axis data (citations) are extremely skewed. As a consequence, the two box plots in the top and right of Fig. 5 have a very large number of 'outliers' shown as '*'.

Out of all the 71,668 SE papers in the pool indexed in the Scopus publication search engine, 30,958 papers (~43% of the pool) had no citations at all, 10,095 papers (~14% of the pool) had only one citation. In total, 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 Section 4.2). Fig. 6 shows the histogram of the citation data for all the SE papers.

4.2. RQ 2: identifying the highly-cited SE papers

We report in the following the top-100 highly-cited papers by absolute number of citations, and then by average total number of citations. We then compare the two rankings versus one another.

4.2.1. The top papers by total number of citations

The top-100 papers by total number of citations are shown in Table 6. When there is a draw in the ranking (number of citations), we put the most recent paper first, e.g., the papers in positions 34 and 35 in Table 6.

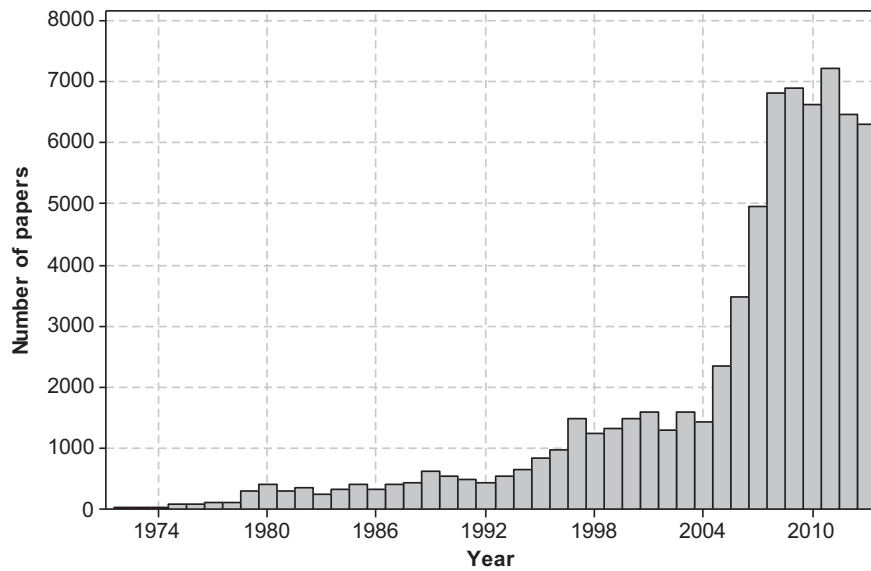


Fig. 3. Number of SE papers included in Scopus by their publication year.

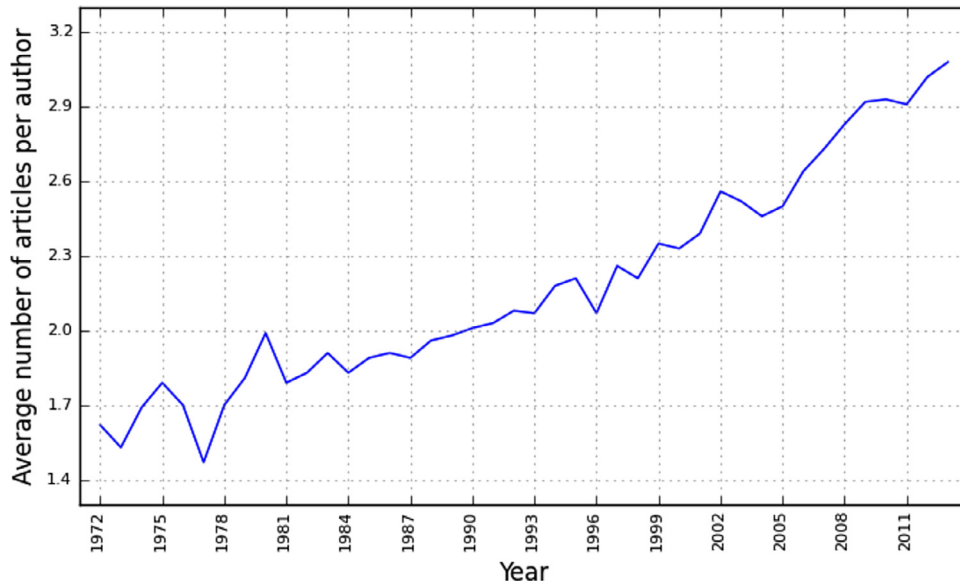


Fig. 4. Average number of authors for articles, per year, in software engineering.

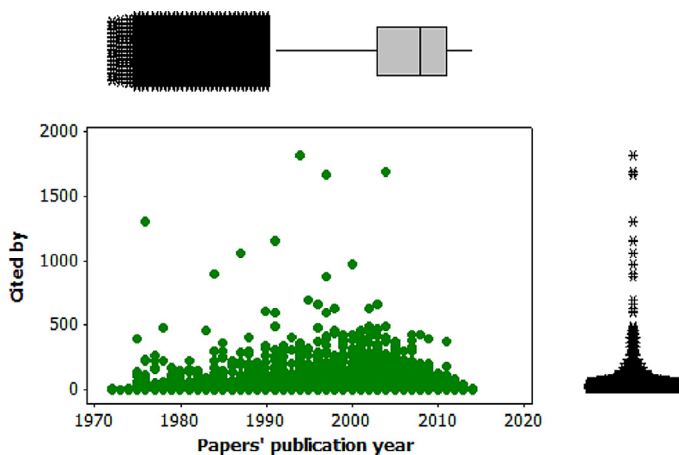


Fig. 5. Scatter plot of all the 71,668 SE papers' publication years and citations, along with their box plots.

According to the papers' titles, we can see that various topics are represented in the top papers list, and we can notice both old and recent papers. Also, the top-100 papers is a mixture of actual SE-papers and methodology papers targeted towards SE, although not being necessarily SE-papers as such, e.g., items #21 and #48 which are guideline papers versus the others in Table 6. There are primary studies (technical non-survey papers) as well as secondary studies (e.g., surveys) in the list, e.g., comparing items #20 and #34 which are secondary studies with the others in the list.

The column “# in the other ranking” in Table 6 tells whether each paper also appears in the top-100 ranking by average annual number of citations, reported in Section 4.2.2.

It would be interesting to conduct cross-discipline comparison of citation values for the top papers. From the collection of papers discussed in Section 2.3, one study in the area of neurosurgery [11] reported that the individual citation counts for the top-100 papers in that particular discipline ranged from 287 to 1,515. We can see that the top SE papers have similar ranges as the citation counts in Table 6

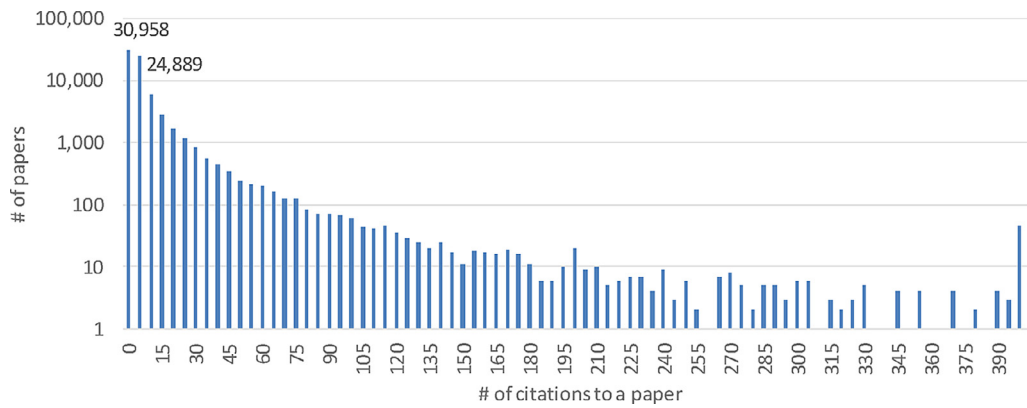


Fig. 6. Histogram of citation data for all the SE papers included in Scopus.

range from 290 to 1,817. However, we should mention that the number of papers and citations between communities are hard to compare and shall be done with care, since different communities have various differences, e.g., different sizes and distinct reference styles.

As another external comparison, Fig. 7 shows the citation counts of each of the top-100 SE papers (listed in Table 6) versus top-100 papers in all areas of science (data from [18]). For example, in the rank #1, the Y value is 1817 (the number of citation to paper #1 in Table 6). For brevity, the y-axis is in logarithmic scale. We can observe that the top SE papers have about one to two orders of magnitude less citation than the top papers in all areas of science.

Fig. 8 shows the scatter plot of publication year and citations for the top-100 papers ranked by total number of citations. A quadratic regression fit is also shown. Both old and recent papers have received high citations, while the latter group seems to have slightly less citations on average which is as one would expect, i.e., recent top papers have had less time to get exposure and thus citations. The top three papers have been cited more than 1600 times, and the top six papers have received more than 1000 citations each.

4.2.2. The top papers by average annual number of citations

The top-100 papers by average annual number of citations are shown in Table 7, alongside with a column showing the total number of citations, for cross comparison purposes. As one would expect, there are overlaps between the two top-100 rankings, which is discussed in Section 4.2.3.

Fig. 9 shows the scatter plot of publication year and citations for the top-100 papers ranked by average annual number of citations. A quadratic regression fit is also shown.

At this point, we would like to comment on the issue of older papers as related to the average number of citations per year. As the SE community has grown over the years, newer papers have a slight advantage when it comes to the number of citations per year. The community was very small about five decades ago, i.e., recall from Section 3.4 that the earliest publication year of SE papers in Scopus is 1972 from which 29 papers were included in the pool. Elapse of time may also affect the total number of citations, since technology has changed and hence some very good older papers may not be relevant anymore, and given that it was a smaller community when the papers were relevant, such papers have received fewer citations in total. This relates to the issue of papers' exposure, i.e. fewer years since the publication of a paper means less exposure, but at the same time, newer papers are exposed to a larger community than what new papers were 50 years ago. To have better indicators, we may want to take the size of the community into account when looking at the actual impact of a paper.

4.2.3. The two rankings versus one another

By a close examination of the two rankings, we found out that 65 papers appear in both rankings. So, each ranking has 35 papers which do not appear in the other ranking, as shown by the columns "in the other ranking" in Tables 6 and 7. As a result, the union of the two top-100 rankings results in a set of 135 papers.

Fig. 10 shows the scatter plot of data for both rankings, in which a quadratic regression fit is also shown. The chart has 135 points, which correspond to the papers in the union of the two rankings. The correlation of the two series is 0.67 denoting a quite strong positive correlation. This denotes that, if a top paper receives high citations, it is quite likely that it will also receive high average annual citations, and vice versa. Let us stress that a few number of data points in the top corner of Fig. 10 (mainly three points) have high impact on the slope of the quadratic regression fit.

Follow-up future studies are needed to analyse citation patterns in detail to assess whether top SE papers are cited less as years go by, i.e., do top-cited papers get less popular by time? Previous works such as [6,20] have studied such a phenomenon, e.g., the study reported in [6] shows that the citation curves of highly-cited papers, in the data pool considered in that study, follow a typical pattern of rise and decline. Of course, it is expected that there will be differences in citation patterns among different top papers, e.g., papers presenting guidelines for empirical SE are always considered useful by researchers and will likely continue to get cited. However, trendy topics, such as cloud computing, might become popular (highly cited) during a time interval and then become forgotten (less cited) after the trend has passed by.

4.3. RQ 3: coverage of SE topics in top papers

For RQ 3, our goal was to classify the top papers by their coverage of SE topics. For this purpose, we adopted the classification of SE areas presented in the 2014 version of the well-known Guide to the Software Engineering Body of Knowledge (SWEBOK) [46]. SWEBOK divides the SE body of knowledge into 12 core knowledge areas and three foundational knowledge areas (computing, mathematical and engineering foundations) as listed in Table 8.

The classification was conducted by the first author and reviewed by the second one to ensure better results. A paper could be classified under more than one SWEBOK knowledge area, e.g., the paper entitled "Colored Petri Nets and CPN Tools for modeling and validation of concurrent systems" [47], published in 2007, covers three knowledge areas of SWEBOK: requirements, models and methods, and quality. To ensure transparency, replication and further analysis, the online spreadsheet which includes the detailed classification of top papers is publicly available in an online Google spreadsheet [48].

Table 6
Top-100 papers by total number of citations.

#	Title	Year	Cited by	Annual average	# In the other ranking
1	A metrics suite for object oriented design	1994	1817	86.5	4
2	QoS-aware middleware for Web services composition	2004	1696	154.2	1
3	The model checker SPIN	1997	1669	92.7	3
4	Complexity measure	1976	1304	33.4	33
5	Graph drawing by force-directed placement	1991	1162	48.4	12
6	An intrusion-detection model	1987	1055	37.7	19
7	A classification and comparison framework for software architecture description languages	2000	973	64.9	6
8	Program slicing	1984	903	29.1	46
9	Uppaal in a nutshell	1997	875	48.6	11
10	4 + 1 view model of architecture	1995	698	34.9	26
11	Developing multi-agent systems: The Gaia methodology	2003	663	55.3	8
12	A validation of object-oriented design metrics as quality indicators	1996	661	34.8	27
13	Two case studies of open source software development: Apache and Mozilla	2002	635	48.8	10
14	Understanding code mobility	1998	627	36.9	21
15	Reverse engineering and design recovery: A taxonomy	1990	605	24.2	78
16	A formal basis for architectural connection	1997	600	33.3	34
17	Software risk management: Principles and practices	1991	598	24.9	71
18	Towards modelling and reasoning support for early-phase requirements engineering	1997	494	27.4	55
19	Modeling and verification of time dependent systems using time Petri nets	1991	490	20.4	-
20	Search-based software test data generation: A survey	2004	488	44.4	15
21	Preliminary guidelines for empirical research in software engineering	2002	487	37.5	20
22	Testing software design modeled by finite-state machines	1978	486	13.1	-
23	The STATEMATE semantics of Statecharts	1996	482	25.4	69
24	CCFinder: A multilingual token-based code clone detection system for large scale source code	2002	479	36.8	22
25	The pragmatics of model-driven development	2003	475	39.6	17
26	Goal-oriented requirements engineering: A guided tour	2001	470	33.6	32
27	DiamondTouch: A multi-user touch technology	2001	463	33.1	35
28	Software function, source lines of code, and development effort prediction: a software science validation	1983	463	14.5	-
29	FORM: A feature-oriented reuse method with domain-specific reference architectures	1998	462	27.2	58
30	A taxonomy and survey of grid resource management systems for distributed computing	2002	449	34.5	28
31	Discovering models of software processes from event-based data	1998	434	25.5	67
32	A critical success factors model for ERP implementation	1999	432	27.0	59
33	Adaptive service composition in flexible processes	2007	427	53.4	9
34	Empirical studies of agile software development: A systematic review	2008	423	60.4	7
35	Open graph visualization system and its applications to software engineering	2000	423	28.2	52
36	The AETG system: An approach to testing based on combinatorial design	1997	421	23.4	84
37	A critique of software defect prediction models	1999	416	26.0	63
38	An empirical study of speed and communication in globally distributed software development	2003	414	34.5	29
39	Toward reference models for requirements traceability	2001	408	29.1	45
40	Object-oriented metrics that predict maintainability	1993	406	18.5	-
41	Dynamically discovering likely program invariants to support program evolution	2001	405	28.9	48
42	HyTech: A model checker for hybrid systems	1997	405	22.5	88
43	A survey of software refactoring	2004	401	36.5	23
44	Recovering traceability links between code and documentation	2002	401	30.8	39
45	The Tame project: Towards improvement-oriented software environments	1988	401	14.9	-
46	Patterns in property specifications for finite-state verification	1999	393	24.6	74
47	System structure for software fault tolerance	1975	393	9.8	-
48	Guidelines for conducting and reporting case study research in software engineering	2009	392	65.3	5
49	Prioritizing test cases for regression testing	2001	390	27.9	53
50	Estimating software project effort using analogies	1997	390	21.7	-
51	N degrees of separation: Multi-dimensional separation of concerns	1999	388	24.3	77
52	Coloured Petri nets and CPN Tools for modelling and validation of concurrent systems	2007	387	48.4	13
53	Bandera: extracting finite-state models from Java source code	2000	382	25.5	68
54	Model-based performance prediction in software development: A survey	2004	379	34.5	30
55	Method engineering: Engineering of information systems development methods and tools	1996	376	19.8	-
56	CloudSim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms	2011	371	92.8	2
57	LSCs: Breathing life into message sequence charts	2001	368	26.3	62
58	Alloy: A lightweight object modelling notation	2002	367	28.2	51
59	Understanding quality in conceptual modeling	1994	367	17.5	-
60	Selecting software test data using data flow information	1985	367	12.2	-
61	Interface automata	2001	363	25.9	64
62	Supporting controlled experimentation with testing techniques: An infrastructure and its potential impact	2005	354	35.4	24

(continued on next page)

Table 6 (continued)

#	Title	Year	Cited by	Annual average	# In the other ranking
63	Model checking programs	2003	352	29.3	44
64	X10: An object-oriented approach to Non-Uniform Cluster Computing	2005	351	35.1	25
65	Building knowledge through families of experiments	1999	351	21.9	96
66	Qualitative methods in empirical studies of software engineering	1999	345	21.6	–
67	Automated software test data generation	1990	345	13.8	–
68	A unified framework for coupling measurement in object-oriented systems	1999	342	21.4	–
69	Representing and using nonfunctional requirements: A process-oriented approach	1992	341	14.8	–
70	Agent-based software engineering	1997	338	18.8	–
71	Data mining static code attributes to learn defect predictors	2007	329	41.1	16
72	Conflicts in policy-based distributed systems management	1999	329	20.6	–
73	Experiments on the effectiveness of dataflow- and controlflow-based test adequacy criteria	1994	329	15.7	–
74	Model-checking algorithms for continuous-time Markov chains	2003	327	27.3	57
75	Tactical approaches for alleviating distance in global software development	2001	326	23.3	85
76	Abstractions for software architecture and tools to support them	1995	325	16.3	–
77	STATEMATE: A working environment for the development of complex reactive systems	1990	324	13.0	–
78	On the unification power of models	2005	322	32.2	38
79	The evolving philosophers problem: Dynamic change management	1990	320	12.8	–
80	Model checking JAVA programs using JAVA PathFinder	2000	318	21.2	–
81	Specification and analysis of system architecture using Rapide	1995	315	15.8	–
82	Analysis of the requirements traceability problem	1994	312	14.9	–
83	Methodology for controlling the size of a test suite	1993	312	14.2	–
84	State transition analysis: a rule-based intrusion detection approach	1995	306	15.3	–
85	The JEDI event-based infrastructure and its application to the development of the OPSS WFMS	2001	305	21.8	100
86	Constraint-based automatic test data generation	1991	305	12.7	–
87	A systematic review of software development cost estimation studies	2007	303	37.9	18
88	Soft computing and fuzzy logic	1994	303	14.4	–
89	Kronos: A verification tool for real-time systems	1997	302	16.8	–
90	A taxonomy of scheduling in general-purpose distributed computing systems	1988	302	11.2	–
91	Quantitative analysis of faults and failures in a complex software system	2000	300	20.0	–
92	Empirical validation of object-oriented metrics on open source software for fault prediction	2005	299	29.9	41
93	Predicting fault incidence using software change history	2000	299	19.9	–
94	Spawn: A distributed computational economy	1992	298	13.0	–
95	N-version approach to fault-tolerant software	1985	297	9.9	–
96	A methodology for collecting valid software engineering data	1984	297	9.6	–
97	The FRACTAL component model and its support in Java	2006	295	32.8	36
98	CARISMA: Context-Aware Reflective middleware System for Mobile Applications	2003	294	24.5	75
99	Developing multi-agent systems with a FIPA-compliant agent framework	2001	292	20.9	–
100	Eliciting security requirements with misuse cases	2005	290	29.0	47

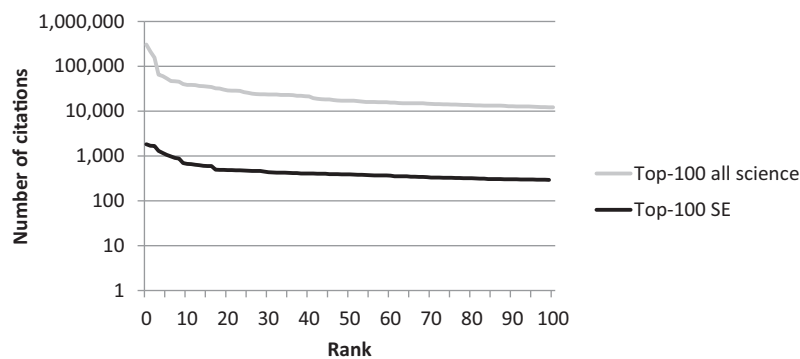


Fig. 7. Comparing the citation counts of top-100 SE papers versus top-100 papers in all areas of science (data from [18]).

Fig. 11 shows the coverage frequency of different SWEBOK knowledge areas by top SE papers. As we can see, “models and methods” and “software design” knowledge areas are the most popular topics in the top papers (in 50 and 34 of the top 135 papers, respectively). Note that such models and methods are not just any typical models and methods. As SWEBOK [46] defines: “software engineering models and methods impose structure on software engineering with the goal of making that activity systematic, repeatable, and ultimately more success-oriented”. For example, the following methods have been mentioned in the SWEBOK: heuristic methods, formal methods and prototyping.

We carefully followed the SWEBOK’s definitions and terminology in our classifications.

Three of the SWEBOK knowledge areas, namely professional practice, SE economics and configuration management, have been represented quite sparsely in the top papers (in 1, 2 and 4 of the top 135 papers, respectively). Three foundational knowledge areas (computing, engineering and mathematical foundations) have been covered by several top papers. For example, relating to computing foundations, an approach for management of computer networks is discussed in [49].

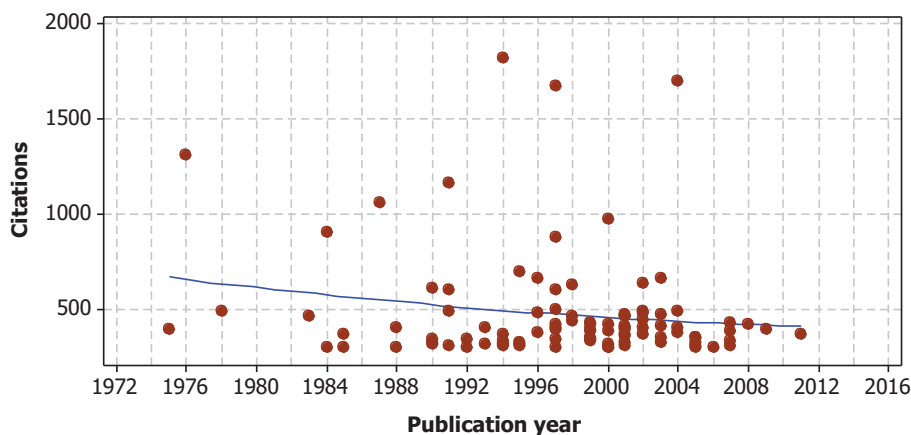


Fig. 8. Scatter plot of publication year and citations for the top-100 papers ranked by total number of citations.

Relating to engineering foundations, simulation was discussed in [50,51]. Empirical methods and experimental techniques were discussed in [52]. Note that SWEBOK categorizes empiricism under the engineering foundations knowledge areas. Relating to mathematical foundations, graphs and graph drawing (in the context of SE) are the focus of [53]. Markov models and logic are the focus of [54].

One paper [55] was classified under the SE professional practice knowledge area. That paper investigated team and group communication in globally distributed software development.

Our results are quite contrary to the finding of Cai and Card [25] in which they found that 73% of SE journal papers focused on 20% of the SE topics. However, let us note that that study used a different classification for SE topics compared to ours, and also that they mapped a collection of SE journal papers, not necessarily the highly-cited ones.

We also wondered about the number of SWEBOK knowledge areas covered in each single paper, and whether a typical top paper covers more than one SWEBOK knowledge area. Fig. 12 shows the histogram of number of SWEBOK knowledge areas covered in each paper. As we can see, almost half of the pool (69 of the 135 papers, 51%) only covered one knowledge area.

As another way to assess coverage of SE topics by top papers, we decided to use the notion of word clouds, which is a popular information visualization model. A word cloud is a visual representation for text data, typically used to depict keywords in a given context. Fig. 13 shows the word cloud of top papers' titles. An online tool named Wordle (www.wordle.net) was used to generate this word cloud. For brevity, the word "software" has been removed. As we can see, keywords such as "engineering", "development", "model", "code" and "requirements" are among the most common words.

4.4. RQ 4: venues for top papers

Table 9 lists the venues for top papers, which include a total of 25 venues: 15 journals, 1 magazine, 5 conferences and 4 symposia. The IEEE Transactions on Software Engineering (TSE) is the clear top venue, including nearly half of the top papers (64 of 135 papers). IEEE Software magazine with 10 papers is the second. The ACM Transactions on Software Engineering and Methodology is the third with 8 papers. The first conference is the International Conference on Software Engineering (ICSE) that holds the fifth place with 6 papers. In Table 9, the values inside parenthesis under the column "Venue type and rank" denote the rank of each venue among the venues of that type, i.e., journals, magazines, conferences, and symposia.

Entitled "The 100 highly-cited works in neurosurgery", the paper reported in [11] reported that the 100 most cited papers in the neurosurgical journals appeared in 3 of 13 journals dedicated to neurosurgery. It would be interesting to perform such an analysis in the

SE. First, one shall find out how many journals are dedicated to SE. Different bibliometrics studies in SE, e.g., [29–35], have considered different sets of top SE journals, which are often categorized in different tiers. To get an estimate on the number of SE journals, we used the list of ISI-listed SE journals prepared by an active SE researcher and posted online [56], which contains the names of 18 SE journals.

Using the list of 18 SE journals posted online in [56] and the data from Table 9, we find that the venue concentration of top SE papers is not as dense as the case in neurosurgery [11], as the top 135 papers have been published in 12 of the 18 ISI-listed SE journals (shown in Table 9). But almost half of the top papers appeared in only one venue (IEEE TSE). Although it is entirely speculative in this stage, it seems that similar to what is often believed and also systematically analyzed in the related work in other disciplines, e.g., [12], venue reputation may have some impact on the high citation of SE papers.

4.5. RQ 5: authorship in top papers

In terms of authorship, 371 different researchers are the authors of the 135 papers that appear in both top-100 lists. Only 30 researchers (8.1%) have more than one paper among those 135 papers. Basili has four papers. Briand, Harel, Harman, Harrold, Henzinger, Kitchenham, and Rothermel each have three papers, and 22 other researchers have two papers in the tops-100. The other 341 authors have authored each one just one paper. Additionally, as shown in Fig. 14, the majority (77%) of those 135 papers have at most 3 co-authors. Only two papers [57,58] have more than 10 authors. 25 top papers (18.5%) have been authored by single authors, meaning that it is perfectly possible for a single researcher to produce (write) papers with high impact.

5. Discussions

5.1. Summary of findings, trends, and implications

We present in the following the summary of findings, trends, and implications for the RQs.

- RQ 1: Out of all the 71,668 papers in the pool, 30,958 papers (43.2% of the pool) had no citations at all. 10,095 papers (14.1% of the pool) had only one citation and 30,615 papers (42.7% of the pool) had received more than one citation. The average citation value was 6.82 per paper. It would be interesting to compare the citation landscape of SE with other disciplines in future studies. Do we have a higher or lower ratio of papers without citations in SE compared to other disciplines? Results of such upcoming studies would highlight the need for critical discussions in the community on the issue of quantity versus quality of publications,

Table 7

Top-100 papers by average annual number of citations.

#	Title	Year	Cited by	Annual average	# in the other ranking
1	QoS-aware middleware for Web services composition	2004	1696	154.2	2
2	CloudSim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms	2011	371	92.8	56
3	The model checker SPIN	1997	1669	92.7	3
4	A metrics suite for object oriented design	1994	1817	86.5	1
5	Guidelines for conducting and reporting case study research in software engineering	2009	392	65.3	48
6	A classification and comparison framework for software architecture description languages	2000	973	64.9	7
7	Empirical studies of agile software development: A systematic review	2008	423	60.4	34
8	Developing multi-agent systems: The Gaia methodology	2003	663	55.3	11
9	Adaptive service composition in flexible processes	2007	427	53.4	33
10	Two case studies of open source software development: Apache and Mozilla	2002	635	48.8	13
11	Uppaal in a nutshell	1997	875	48.6	9
12	Graph drawing by force-directed placement	1991	1162	48.4	5
13	Coloured Petri nets and CPN Tools for modelling and validation of concurrent systems	2007	387	48.4	52
14	KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera	2011	181	45.3	-
15	Search-based software test data generation: A survey	2004	488	44.4	20
16	Data mining static code attributes to learn defect predictors	2007	329	41.1	71
17	The pragmatics of model-driven development	2003	475	39.6	25
18	A systematic review of software development cost estimation studies	2007	303	37.9	87
19	An intrusion-detection model	1987	1055	37.7	6
20	Preliminary guidelines for empirical research in software engineering	2002	487	37.5	21
21	Understanding code mobility	1998	627	36.9	14
22	CCFinder: A multilinguistic token-based code clone detection system for large scale source code	2002	479	36.8	24
23	A survey of software refactoring	2004	401	36.5	43
24	Supporting controlled experimentation with testing techniques: An infrastructure and its potential impact	2005	354	35.4	62
25	X10: An object-oriented approach to Non-Uniform Cluster Computing	2005	351	35.1	64
26	4 + 1 view model of architecture	1995	698	34.9	10
27	A validation of object-oriented design metrics as quality indicators	1996	661	34.8	12
28	A taxonomy and survey of grid resource management systems for distributed computing	2002	449	34.5	30
29	An empirical study of speed and communication in globally distributed software development	2003	414	34.5	38
30	Model-based performance prediction in software development: A survey	2004	379	34.5	54
31	The physics of notations: Toward a scientific basis for constructing visual notations in software engineering	2009	204	34.0	-
32	Goal-oriented requirements engineering: A guided tour	2001	470	33.6	26
33	A complexity measure	1976	1304	33.4	4
34	A formal basis for architectural connection	1997	600	33.3	16
35	DiamondTouch: A multi-user touch technology	2001	463	33.1	27
36	The FRACTAL component model and its support in Java	2006	295	32.8	97
37	The Palladio component model for model-driven performance prediction	2009	196	32.7	-
38	On the unification power of models	2005	322	32.2	78
39	Recovering traceability links between code and documentation	2002	401	30.8	44
40	Systematic literature reviews in software engineering - A systematic literature review	2009	185	30.8	-
41	Empirical validation of object-oriented metrics on open source software for fault prediction	2005	299	29.9	92
42	An analysis and survey of the development of mutation testing	2011	119	29.8	-
43	Regression testing minimization, selection and prioritization: A survey	2012	89	29.7	-
44	Model checking programs	2003	352	29.3	63
45	Toward reference models for requirements traceability	2001	408	29.1	39
46	Program slicing	1984	903	29.1	8
47	Eliciting security requirements with misuse cases	2005	290	29.0	100
48	Dynamically discovering likely program invariants to support program evolution	2001	405	28.9	41
49	Benchmarking classification models for software defect prediction: A proposed framework and novel findings	2008	202	28.9	-
50	Empirical evaluation of the tarantula automatic fault-localization technique	2005	284	28.4	-
51	Alloy: A lightweight object modelling notation	2002	367	28.2	58
52	Open graph visualization system and its applications to software engineering	2000	423	28.2	35
53	Prioritizing test cases for regression testing	2001	390	27.9	49
54	The DaCapo benchmarks: Java benchmarking development and analysis	2006	248	27.6	-
55	Towards modelling and reasoning support for early-phase requirements engineering	1997	494	27.4	18
56	Self-managed systems: An architectural challenge	2007	219	27.4	-
57	Model-checking algorithms for continuous-time Markov chains	2003	327	27.3	74
58	FORM: A feature-oriented reuse method with domain-specific reference architectures	1998	462	27.2	29
59	A critical success factors model for ERP implementation	1999	432	27.0	32
60	Model driven security: From UML models to access control infrastructures	2006	239	26.6	-

(continued on next page)

Table 7 (continued)

#	Title	Year	Cited by	Annual average	# in the other ranking
61	An overview of JML tools and applications	2005	263	26.3	–
62	LSCs: Breathing life into message sequence charts	2001	368	26.3	57
63	A critique of software defect prediction models	1999	416	26.0	37
64	Interface automata	2001	363	25.9	61
65	Model-driven development of complex software: A research roadmap	2007	207	25.9	–
66	Seven process modeling guidelines (7PMG)	2010	129	25.8	–
67	Discovering models of software processes from event-based data	1998	434	25.5	31
68	Bandera: extracting finite-state models from Java source code	2000	382	25.5	53
69	The STATEMATE semantics of Statecharts	1996	482	25.4	23
70	Semantics and analysis of business process models in BPMN	2008	176	25.1	–
71	Software risk management: Principles and practices	1991	598	24.9	17
72	Is mutation an appropriate tool for testing experiments?	2005	248	24.8	–
73	Predicting the location and number of faults in large software systems	2005	246	24.6	–
74	Patterns in property specifications for finite-state verification	1999	393	24.6	46
75	CARISMA: Context-aware reflective middleware system for mobile applications	2003	294	24.5	98
76	The software model checker Blast: Applications to software engineering	2007	194	24.3	–
77	N degrees of separation: Multi-dimensional separation of concerns	1999	388	24.3	51
78	Reverse engineering and design recovery: A taxonomy	1990	605	24.2	15
79	Locating features in source code	2003	289	24.1	–
80	Accurate online power estimation and automatic battery behavior based power model generation for smartphones	2010	120	24.0	–
81	Analyzing CUDA workloads using a detailed GPU simulator	2009	144	24.0	–
82	Search algorithms for regression test case prioritization	2007	191	23.9	–
83	MuJava: An automated class mutation system	2005	238	23.8	–
84	The AETG system: An approach to testing based on combinatorial design	1997	421	23.4	36
85	Tactical approaches for alleviating distance in global software development	2001	326	23.3	75
86	Model-driven development: A metamodeling foundation	2003	275	22.9	–
87	Who should fix this bug?	2006	206	22.9	–
88	HyTech: A model checker for hybrid systems	1997	405	22.5	42
89	Model transformation: The heart and soul of model-driven software development	2003	269	22.4	–
90	OmniTouch: Wearable multitouch interaction everywhere	2011	89	22.3	–
91	Formalizing cardinality-based feature models and their specialization	2005	222	22.2	–
92	Software fault interactions and implications for software testing	2004	244	22.2	–
93	Key establishment in large dynamic groups using one-way function trees	2003	265	22.1	–
94	Watermarking, tamper-proofing, and obfuscation - Tools for software protection	2002	287	22.1	–
95	Lessons from applying the systematic literature review process within the software engineering domain	2007	176	22.0	–
96	Building knowledge through families of experiments	1999	351	21.9	65
97	Test case prioritization: A family of empirical studies	2002	285	21.9	–
98	A framework for QoS-aware binding and re-binding of composite web services	2008	153	21.9	–
99	Advancing candidate link generation for requirements tracing: The study of methods	2006	196	21.8	–
100	The JEDI event-based infrastructure and its application to the development of the OPSS WFMS	2001	305	21.8	85

Table 8

Knowledge areas of the version 3.0 of the SWEBOK [46].

1.	Software requirements
2.	Software design
3.	Software construction
4.	Software testing
5.	Software maintenance
6.	Software configuration management
7.	Software engineering management
8.	Software engineering process
9.	Software engineering models and methods
10.	Software quality
11.	Software engineering professional practice (professionalism)
12.	Software engineering economics
13.	Computing foundations
14.	Mathematical foundations
15.	Engineering foundations

which was put nicely by David Parnas as “Stop the numbers game” [59].

- RQ 2: 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. The individual citation counts for the top-100 papers range from 290 to 1,817, which is favorably comparable to the citation data set of

highly-cited papers from another discipline (neurosurgery [11]). For the case of ranking by average annual number of citations, the values ranged from 21.8 to 154.2. The union of the two top-100 rankings based on the two metrics results in 135 papers. Inspired by the quote: “stand on the shoulders of giants”, we believe that reading the top papers relevant to one’s chosen research area would be useful for young SE researchers, as it permits the reader to understand some key aspects for the popularity of a paper, such as clarity, structure, types of results, style, and evaluation (validation) approach. It would enable young SE researchers to improve on the process of technical writing, by having access to papers that are popular in terms of citations. It is the belief of the authors that if established researchers train their young researchers to read and study highly-cited papers in detail, using the top list as identified in this paper and also by offering courses such as “Research methods in SE” [60] (offered by the first author), high-quality young researchers are effectively and efficiently trained.

- RQ 3: While papers presenting professional practice, SE economics and configuration management are in a minority among the top list, the other contribution types (e.g., models and methods) are quite well represented. Two papers discussing simulation have made it to the top list. Three foundational knowledge areas (computing, engineering and mathematical foundations) have been covered by several top papers. Our results are quite contrary to the finding of Cai and Card [25] in which they report that 73%

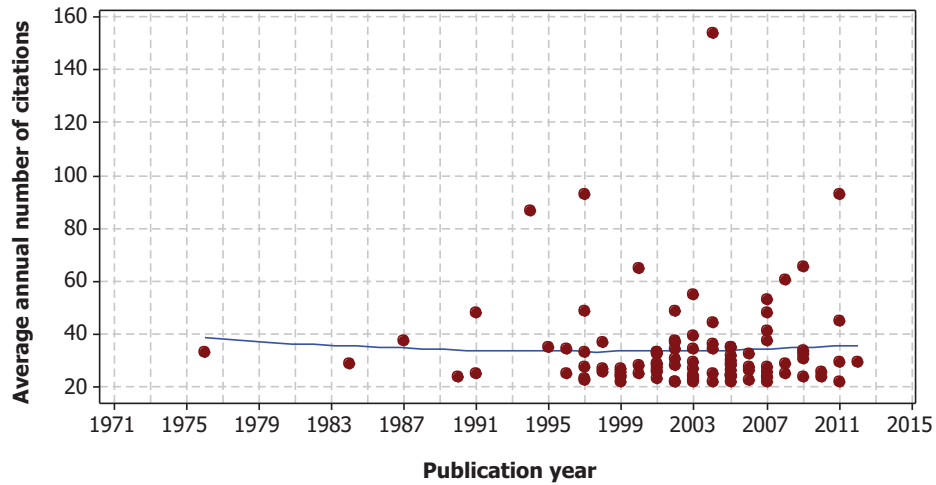


Fig. 9. Scatter plot of publication year and citations for the top-100 papers ranked by average annual number of citations.

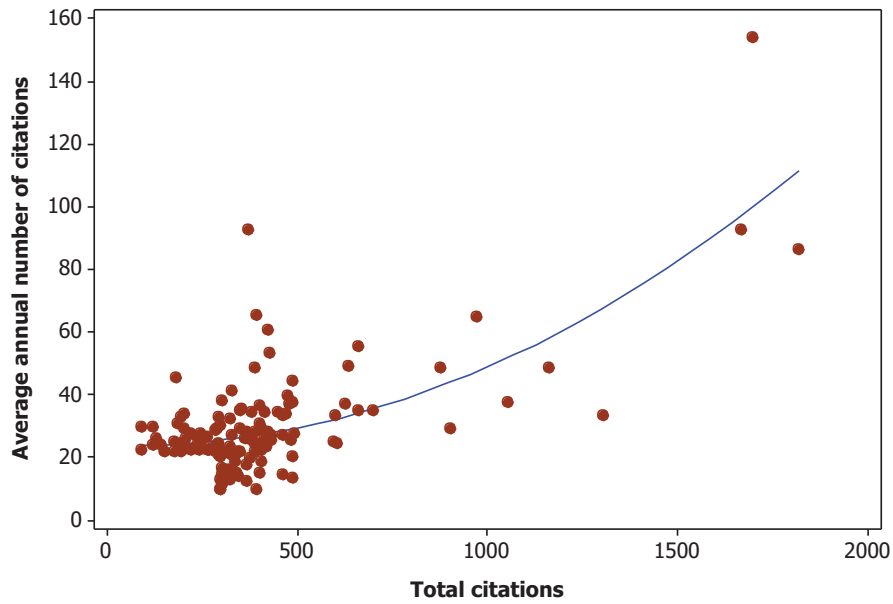


Fig. 10. Scatter plot of total citations versus average annual number of citations for papers in both rankings.



Fig. 11. Coverage frequency of SWEBOK knowledge areas by top SE papers.

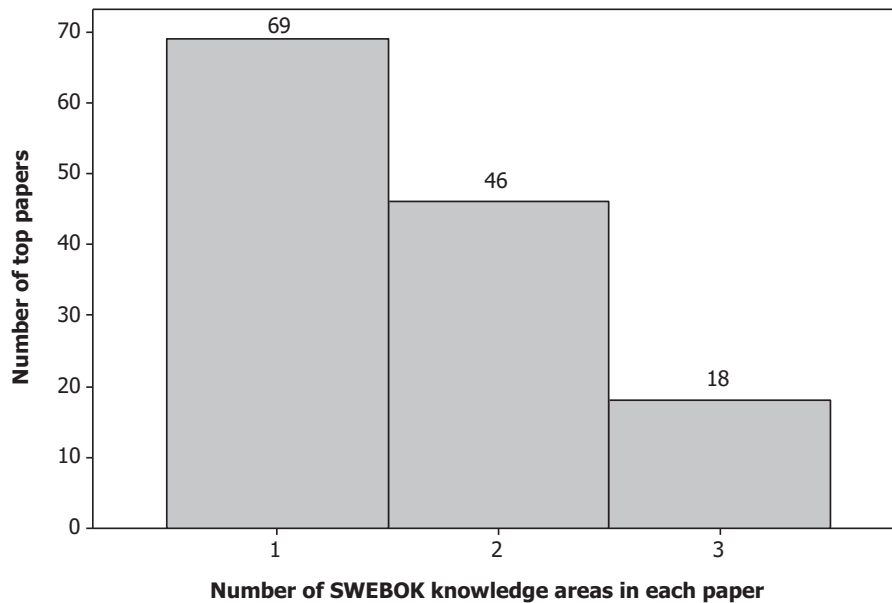


Fig. 12. Number of SWEBOK knowledge areas in each paper.

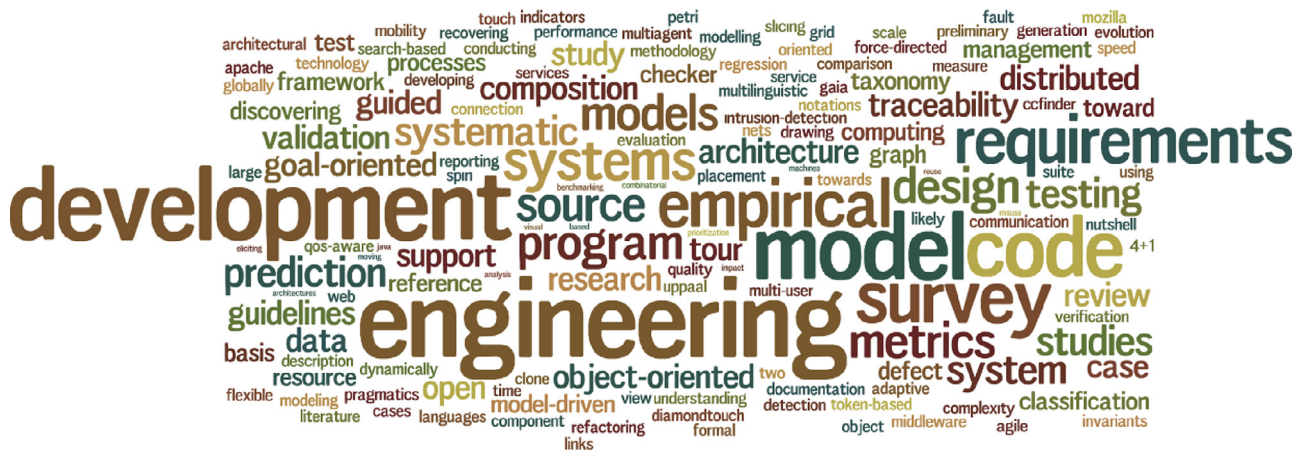


Fig. 13. Word cloud of top papers' titles (tool used: www.wordle.net).

of SE journal papers focused on 20% of the SE topics. However, when compared to mapping of top-cited studies in neurosurgery (as a study conducted in another discipline) [11], we find reasonable similarities, i.e., the top papers quite homogeneously cover all sub-areas of the field.

- RQ 4: The venue concentration of top SE papers is not as dense as the case in neurosurgery [11], as the top 135 papers have been published in 12 of the 18 ISI-listed SE journals. But almost half of the top papers appeared in only one venue (IEEE TSE). Our results once again confirm the list of the commonly-accepted list of top SE venues and that venue rank may behave as an 'external factor' [12] leading to increased popularity (citation) of papers.
- RQ 5: In terms of authorship, 371 different researchers are the authors of the 135 papers that appear in both tops-100. Only 30 researchers (8.1%) have more than one paper among those 135 papers. 25 top papers (18.5%) have been authored by single authors, meaning that it is perfectly possible for a single researcher to produce papers with high impact.

In general, we believe that the results reported in this paper are useful for both young and senior SE researchers. They show some patterns (like venue or topic) that are likely to affect the citation success of a publication.

5.2. Limitation 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 [61]. We also discuss the steps that we have taken to minimize or mitigate those potential threats.

5.2.1. Internal validity

Internal validity reflects the extent to which a causal conclusion based on a study is warranted [61]. The systematic approach that has been utilized for the selection publication search engine and SE papers is described in Section 2. In order to make sure that this study and the ranking are repeatable, search terms are carefully defined and reported. Also, to ensure transparency and replicability of our analysis, the entire raw and ranking data for all the 71,668 papers is available as an Excel file which can be downloaded online [45]. The online spreadsheet which includes the detailed classification of top papers is also publicly available [48].

5.2.2. Construct validity

Construct validities are concerned with issues that are related to what extent the object of study truly represents theory behind the study [61]. Threats related to this type of validity in this study were

Table 9
Venues for top papers.

Venue name	Acronym	Venue type and rank	Publisher	Number of top papers
IEEE Transactions on Software Engineering	TSE	Journal (1)	IEEE	63
IEEE Software	IEEE Software	Magazine (1)	IEEE	10
ACM Transactions on Software Engineering and Methodology	TOSEM	Journal (2)	ACM	8
International Journal on Software Tools for Technology Transfer	STTT	Journal (3)	Springer	7
Software: Practice and Experience	SPE	Journal (4)	Wiley	6
International Conference on Software Engineering	ICSE	Conference (1)	–	6
Information and Software Technology	IST	Journal (5)	Elsevier	5
Journal of Systems and Software	JSS	Journal (6)	Elsevier	4
Software Testing, Verification and Reliability	STVR	Journal (7)	Wiley	3
International Conference on Requirements Engineering	RE	Conference (2)	–	3
ACM Symposium User Interface Software and Technology	UIST	Symposium (1)	–	3
Empirical Software Engineering	EMSE	Journal (8)	Springer	2
Conference on Object-Oriented Programming Systems, Languages, and Applications	OOPSLA	Conference (3)	–	2
Future of Software Engineering Symposium	FOSE	Symposium (2)	–	2
Annals of Software Engineering	ASE	Journal (9)	Springer	1
Automated Software Engineering	ASE	Journal (9)	Springer	1
Formal Methods in System Design	FMSD	Journal (9)	Springer	1
IET Software (formerly IEE Proceedings - Software)	IET Software	Journal (9)	IET	1
Requirements Engineering	REJ	Journal (9)	Springer	1
Software and System Modeling	SoSyM	Journal (9)	Springer	1
Software Process Improvement and Practice	SPIP	Journal (9)	Wiley	1
IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis	CODES+ISSS	Conference (4)	–	1
IEEE/ACM International Conference on Automated Software Engineering	ASE	Conference (4)	–	1
International Symposium on Performance Analysis of Systems and Software	ISPASS	Symposium (3)	–	1
ACM SIGSOFT Symposium on the Foundations of Software Engineering	FSE	Symposium (3)	ACM	1

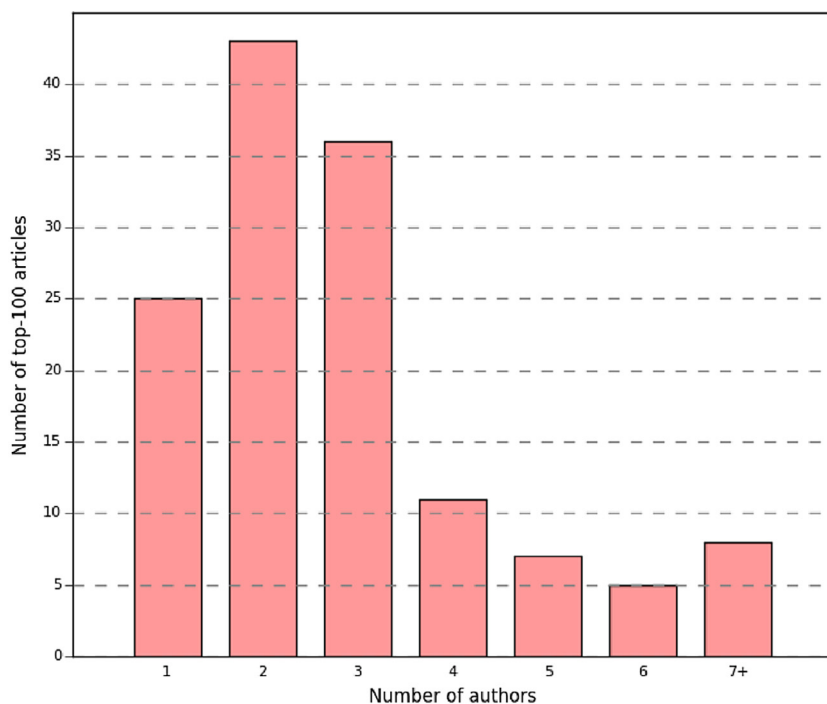


Fig. 14. Number of papers from both top-100 lists distributed according to number of authors.

suitability of RQs and categorization scheme used for the data extraction.

To limit potential construct threats in this study, the GQM approach was used to preserve the traceability between research goal, questions and measurements. RQs were designed to cover our goal and different aspects of the top papers. For designing a good categorization scheme for the systematic mapping, we adapted standard classifications from [37] and also have finalized the used schema through an iterative improvement process.

The citations reported by Scopus have two major limitations. Firstly, Scopus includes self-citations, i.e., those where authors cite

their own work. Additionally, the citations only include the ones made in Scopus-indexed papers, so many papers may have more citations than the ones provided by Scopus.

5.2.3. Conclusion validity

Conclusion validity of a study deals with whether correct conclusions are reached through rigorous and repeatable treatments [61].

Conclusions that are discussed throughout the paper are based on actual quantitative measures and statistics on the data extracted from the top papers. The systematic approach that we used to identify and map the top papers assures that, if the study is conducted by other

researchers, it is expected that results will not have major deviations from our results.

5.2.4. External validity

External validity is concerned with to what extent the results of this secondary study can be generalized [61]. The results of this study are not meant to be generalized to fields outside SE. However, we believe that given the rigour of the systematic approach that we used to identify and map the top papers, the results highlight the citation landscape of the SE and the highly-cited papers in this area.

6. Conclusions and future work

This study systematically identifies and classifies the highly-cited papers in the area of SE. The results of our study provide various benefits for researchers and practitioners in SE, e.g., (1) the results help new researchers to see the type of contributions, approaches and research methods applied in highly-cited papers, so as to learn from them in writing higher quality papers which will likely receive high citations, (2) the classifications help established and new researchers to spot the active and more impactful topics and thus they can carry on further incremental research on those areas, (3) researchers and practitioners can notice the most cited researchers and collaborate with them, get advice from, etc., and (4) the results help practitioners spot the highest quality work in specific areas of SE and aim at utilizing techniques, tools or findings reported in those studies.

While a few small-scale studies [21–26] to identify highly-cited SE papers were previously reported, they have not been as comprehensive as the current study. Inspired by the large number of studies reviewing highly-cited papers in other disciplines, e.g., [1,4–20], and the need for such studies in the SE, this study was the first step in accomplishing a comprehensive analysis of highly-cited papers in SE. We hope that this paper encourages further discussions in the SE community towards further analysis and formal characterization of the highly-cited SE papers.

The following are among our future work directions:

- Research directions related to what makes SE papers highly-cited:
 - In a paper entitled “What makes papers highly-cited?” [15], the authors found that regarding statistical conclusion validity of quantitative papers, papers having endogeneity threats received significantly fewer citations than did those using a more robust design or an estimation procedure that ensured correct causal estimation. It would be interesting to conduct such studies in the SE literature and findings could be used to improve research practice and paper citations.
 - In a paper entitled “Highly-cited works in neurosurgery” [11], 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. Are those determinants also applicable in the SE domain?
 - A 2004 study entitled “Why authors think their papers are highly-cited” [62] surveyed authors of highly-cited papers in 22 fields, in order to discover their opinions on why their papers are highly-cited. The responses of the authors were classified as follows: strong interest in a given topic, the novelty, the utility and the high significance of the research reported in the papers. Peter Moore, a chemist at Yale University, as quoted in [18], mentioned that: “If citations are what you want, devising a method that makes it possible for people to do the experiments they want at all, or more easily, will get you a lot further than, say, discovering the secret of the Universe”. It will be again interesting to conduct similar studies in the SE literature. This issue relates to the discussion in Section 4.2 in that the top-100 papers is a mixture of actual SE-papers and methodology papers targeted towards SE, although not being

necessarily SE-papers as such, e.g., items #21 and #48 which are guideline papers versus the others in Table 6.

- Perhaps, similar to software quality attributes, we need to think of and formalize quality attributes for top-cited papers. Similar to related work in other areas, SE researchers may also be able to formally characterize the highly-cited SE papers, e.g., using the conceptual distinction between quality dynamics and visibility dynamics, as reported in [6].
- As discussed in Section 3.1, the goal and RQs of the study were exploratory and descriptive in nature [39]. In the future, when aiming to answer the question of what makes SE papers highly-cited, we would need to raise causality and causality-comparative types [39] of RQs.
- It will be interesting to compare types and trends of top SE papers versus top papers in other disciplines. What types of papers (e.g., in terms of contribution types) get high citations in SE versus other disciplines?
- What are the citation patterns of the top papers? Are the top papers cited less as years go by? Do top-cited papers get less popular by time? As discussed in Section 2.1, previous works, such as [6,20], have also studied such a question, e.g., the study in [6] reports that the citation curves of highly-cited papers follow a typical pattern of rise and decline.
- Similar to the work of Ioannidis et al. [16], one could ask the most highly-cited SE scientists to score their top-ten papers in various ways. That study answered these questions: Are the most highly-cited papers the most important ones? Does science make progress mostly through evolution or through revolution? Are these two processes mutually exclusive or complementary, and which do high citations most reflect? Are surprising findings difficult to publish? Dimensions that reflected evolution in Ioannidis et al. work [16] were continuous progress, broader interest and greater synthesis. Dimensions that reflected revolution were disruptive innovativeness and surprise. It would be interesting to conduct such studies in the area of SE.
- Similar to [12], we plan to mine typical features for highly-cited papers and to assess the extent to which papers' inner quality and external features, as suggested by [12], and mainly represented as the reputation of the authors and journals, contribute to generation of highly-cited papers in the future.

Acknowledgments

Vahid Garousi was partially supported by several internal grants provided by the Hacettepe University. The authors would like to thank the anonymous reviewers for their insightful comments.

References

- [1] T.A. Hamrick, R.D. Fricker, G.G. Brown, Assessing what distinguishes highly cited from less-cited papers published in interfaces, *Interfaces* 40 (2010) 454–464.
- [2] E. Garfield, Citation indexes for science: a new dimension in documentation through association of ideas, *Science* 122 (1955) 108–111.
- [3] R. Danell, Can the quality of scientific work be predicted using information on the author's track record? *J. Am. Soc. Inf. Sci. Technol.* 62 (2011) 50–60.
- [4] L. Bornmann, How are excellent (highly cited) papers defined in bibliometrics? A quantitative analysis of the literature, *Res. Eval.* 23 (2014) 166–173.
- [5] R. Tijssen, M. Visser, T. van Leeuwen, Benchmarking international scientific excellence: are highly cited research papers an appropriate frame of reference? *Scientometrics* 54 (2002) 381–397.
- [6] D.W. Aksnes, Characteristics of highly cited papers, *Res. Eval.* 12 (2003) 159–170.
- [7] P. Pyšek, D.M. Richardson, J. V. Who cites who in the invasion zoo: insights from an analysis of the most highly cited papers in invasion ecology, *Preslia* 78 (2006) 437–468.
- [8] L. Bornmann, F. de Moya Anegón, L. Leydesdorff, Do scientific advancements lean on the shoulders of giants? a bibliometric investigation of the ortega hypothesis, *PLoS ONE* 5 (2010) 1–6.
- [9] Z. Corby, To be the best, cite the best, *Nat. News* (2010), doi:10.1038/news.2010.539.
- [10] O. Persson, Are highly cited papers more international? *Scientometrics* 83 (2010) 397–401.

- [11] F.A. Ponce, A.M. Lozano, Highly cited works in neurosurgery. Part I: the 100 top-cited papers in neurosurgical journals, *J. Neurosurg.* 112 (2010) 223–232.
- [12] M. Wang, G. Yu, D. Yu, Mining typical features for highly cited papers, *Scientometrics* 87 (2011) 695–706.
- [13] N. Miyairi, H.-W. Chang, Bibliometric characteristics of highly cited papers from Taiwan, 2000–2009, *Scientometrics* 92 (2012) 197–205.
- [14] G. Abramo, T. Cicero, C.A. D'Angelo, Are the authors of highly cited articles also the most productive ones, *J. Inf.* 8 (2014) 89–97.
- [15] J. Antonakis, N. Bastardoz, Y.H. Liu, C.A. Schriesheim, What makes articles highly cited? *Leadersh. Q.* 25 (2014) 152–179.
- [16] J.P.A. Ioannidis, K.W. Boyack, H. Small, A.A. Sorensen, R. Klavan, Is your most cited work your best? *Nature* 514 (2014) 561–562.
- [17] M.E.J. Newman, Prediction of highly cited papers, *Europhys. Lett. (EPL)* 105 (2014) 6.
- [18] R.V. Noorden, B. Maher, R. Nuzzo, The top 100 papers, *Nature* 514 (2014) 550–553.
- [19] D.J. Eaton, Highly cited papers in Medical Physics, *Med. Phys.* 41 (2014) 43–44.
- [20] E. Aversa, Citation patterns of highly cited papers and their relationship to literature aging: a study of the working literature, *Scientometrics* 7 (1985) 383–389.
- [21] C. Wohlin, An analysis of the most cited articles in software engineering journals—1999, *Inf. Softw. Technol.* 47 (2005) 957–964.
- [22] C. Wohlin, An analysis of the most cited articles in software engineering journals – 2000, *Inf. Softw. Technol.* 49 (2007) 2–11.
- [23] C. Wohlin, An analysis of the most cited articles in software engineering journals – 2001, *Inf. Softw. Technol.* 50 (2008) 3–9.
- [24] C. Wohlin, An analysis of the most cited articles in software engineering journals – 2002, *Inf. Softw. Technol.* 51 (2009) 2–6.
- [25] K.-Y. Cai, D. Card, An analysis of research topics in software engineering – 2006, *J. Syst. Softw.* 81 (2008) 1051–1058.
- [26] B. Kitchenham, What's up with software metrics? – A preliminary mapping study, *J. Syst. Softw.* 83 (2010) 37–51.
- [27] R.J.A. Buhr, Use case maps as architectural entities for complex systems, *IEEE Trans. Softw. Eng.* 24 (1998) 1131–1155.
- [28] G.J. Holzmann, The model checker SPIN, *IEEE Trans. Softw. Eng.* 23 (1997) 279–295.
- [29] W.E. Wong, T.H. Tse, R.L. Glass, V.R. Basili, T.Y. Chen, An assessment of systems and software engineering scholars and institutions (2001–2005), *J. Syst. Softw.* 81 (2008) 1059–1062.
- [30] W.E. Wong, T.H. Tse, R.L. Glass, V.R. Basili, T.Y. Chen, An assessment of systems and software engineering scholars and institutions (2002–2006), *J. Syst. Softw.* 82 (2009) 1370–1373.
- [31] V. Garousi, T. Varma, A bibliometric assessment of canadian software engineering scholars and institutions (1996–2006), *Can. J. Comput. Inf. Sci.* 3 (2010) 19–29.
- [32] F. de Freitas, J. de Souza, Ten years of search based software engineering: a bibliometric analysis, in: M. Cohen, M.Ó. Cinnéide (Eds.), *Search Based Software Engineering*, vol. 6956, Springer, Berlin Heidelberg, 2011, pp. 18–32.
- [33] W.W. Eric, T.H. Tse, Robert L. G., Victor R. B., T.Y. C., An assessment of systems and software engineering scholars and institutions (2003–2007 and 2004–2008), *J. Syst. Softw.* 84 (2011) 162–168.
- [34] R. Farhoodi, V. Garousi, D. Pfahl, J.P. Sillito, Development of scientific software: a systematic mapping, bibliometrics study and a paper repository, *Int. J. Softw. Eng. Knowl. Eng.* 23 (2013) 463–506.
- [35] V. Garousi, G. Ruhe, A bibliometric/geographic assessment of 40 years of software engineering research (1969–2009), *Int. J. Softw. Eng. Knowl. Eng.* 23 (2013) 1343–1366.
- [36] J. Fernandes, Authorship trends in software engineering, *Scientometrics* 101 (2014) 257–271.
- [37] K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson, Systematic mapping studies in software engineering, in: 12th International Conference on Evaluation and Assessment in Software Engineering (EASE), 2008.
- [38] V.R. Basili, Software modeling and measurement: the Goal/Question/Metric paradigm, University of Maryland at College Park, 1992 Technical Report.
- [39] S. Easterbrook, J. Singer, M.-A. Storey, D. Damian, Selecting empirical methods for software engineering research, in: F. Shull, J. Singer, D.K. Sjøberg (Eds.), *Guide to Advanced Empirical Software Engineering*, Springer, London, 2008, pp. 285–311.
- [40] É. Archambault, D. Campbell, Y. Gingras, V. Larivière, Comparing bibliometric statistics obtained from the web of science and Scopus, *J. Am. Soc. Inf. Sci.* 60 (2009) 1320–1326.
- [41] M.E. Falagas, E.I. Pitsouni, G.A. Malietzis, G. Pappas, Comparison of PubMed, Scopus, web of science, and google scholar: strengths and weaknesses, *FASEB J.* 22 (2008) 338–342.
- [42] A. Abrizah, A.N. Zainab, K. Kiran, R.G. Raj, LIS journals scientific impact and subject categorization: a comparison between Web of Science and Scopus, *Scientometrics* 94 (2013) 721–740.
- [43] A.A. Chadegani, H. Salehi, M.M. Yunus, H. Farhadi, M. Fooladi, M. Farhadi, et al., A comparison between two main academic literature collections: web of science and scopus search engines, *Asian Soc. Sci.* 9 (2013) 18–26.
- [44] V. Garousi, A bibliometric analysis of the Turkish software engineering research community, *Springer J. Scientometrics* 105 (1) (2015) 23–49.
- [45] V. Garousi and J.M. Fernandes, "All source data for highly-cited papers in software engineering," in <https://goo.gl/RQJTru>, Last accessed: May 2015.
- [46] P. Bourque, A. Abran, J. Garbajosa, G. Keeni, B. Shen, B. Shen, Guide to the Software Engineering Body of Knowledge (SWEBOK) version 3.0, IEEE Press, 2014 <http://www.computer.org/portal/web/swebok>.
- [47] K. Jensen, L.M. Kristensen, L. Wells, Coloured Petri Nets and CPN Tools for modelling and validation of concurrent systems, *Int. J. Softw. Tools Technol. Transf.* 9 (2007) 213–254.
- [48] V. Garousi and J.M. Fernandes, "Raw data for classification of highly-cited papers in software engineering," in <https://goo.gl/WkNFUa>, Last accessed: May 2015.
- [49] A. Fuggetta, G.P. Picco, G. Vigna, Understanding code mobility, *IEEE Trans. Softw. Eng.* 24 (1998) 342–361.
- [50] R.N. Calheiros, R. Ranjan, A. Beloglazov, C.A.F. De Rose, R. Buyya, CloudSim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms, *Softw. Pract. Exp.* 41 (2011) 23–50.
- [51] K.G. Larsen, P. Pettersson, W. Yi, Uppaal in a nutshell, *Int. J. Softw. Tools Technol. Transf.* 1 (1997) 134–152.
- [52] P. Runeson, M. Höst, Guidelines for conducting and reporting case study research in software engineering, *Empir. Softw. Eng.* 14 (2009) 131–164.
- [53] T.M.J. Fruchterman, E.M. Reingold, Graph drawing by force-directed placement, *Softw. Pract. Exp.* 21 (1991) 1129–1164.
- [54] C. Baier, B. Haverkort, H. Hermanns, J.P. Katoen, Model-checking algorithms for continuous-time Markov chains, *IEEE Trans. Softw. Eng.* 29 (2003) 524–541.
- [55] J.D. Herbsleb, A. Mockus, An empirical study of speed and communication in globally distributed software development, *IEEE Trans. Softw. Eng.* 29 (2003) 481–494.
- [56] R. Feldt, "ISI Listed SE Journals," in http://www.cse.chalmers.se/~feldt/advice/isi-listed_se_journals.html, Last accessed: May 2015.
- [57] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, et al., KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera, in: *Proceedings of the 24th annual ACM symposium on User interface software and technology*, Santa Barbara, California, USA, 2011.
- [58] S.M. Blackburn, R. Garner, C. Hoffmann, A.M. Khang, K.S. McKinley, R. Bentzur, et al., The DaCapo benchmarks: java benchmarking development and analysis, *SIGPLAN Not* 41 (2006) 169–190.
- [59] D.L. Parnas, Stop the numbers game, *Commun. ACM* 50 (2007) 19–21.
- [60] V. Garousi, "Graduate course: Research Methods in Software Engineering," <https://docs.google.com/spreadsheets/d/1DFLiSRsyKRLHfhz6OJgmdEwjAWEn-Rep0WzW1Ums3ao/edit?usp=sharing>, Last Accessed: Oct. 2015.
- [61] C. Wohlin, P. Runeson, M. Höst, M.C. Ohlsson, B. Regnell, A. Wesslén, *Experimentation in Software Engineering: An Introduction*, Kluwer Academic Publishers, 2000.
- [62] H. Small, Why authors think their papers are highly cited, *Scientometrics* 60 (2004) 305–316.