This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2014.2321168, IEEE Transactions on Cloud Computing

A Scientometric Analysis of Cloud Computing Literature

Leonard Heilig and Stefan Voß

Abstract—The popularity and rapid development of cloud computing in recent years has led to a huge amount of publications containing the achieved knowledge of this area of research. Due to the interdisciplinary nature and high relevance of cloud computing research, it becomes increasingly difficult or even impossible to understand the overall structure and development of this field without analytical approaches. While evaluating science has a long tradition in many fields, we identify a lack of a comprehensive scientometric study in the area of cloud computing. Based on a large bibliographic data base, this study applies scientometric means to empirically study the evolution and state of cloud computing research with a view from above the clouds. By this, we provide extensive insights into publication patterns, research impact and research productivity. Furthermore, we explore the interplay of related subtopics by analyzing keyword clusters. The results of this study provide a better understanding of patterns, trends and other important factors as a basis for directing research activities, sharing knowledge and collaborating in the area of cloud computing research.

Index Terms—cloud computing, cloud computing research, scientometric analysis, scientometrics, keyword cluster analysis.

1 INTRODUCTION

ALTHOUGH cloud computing is a relatively young field of research, the great interest in academia and practice has led to a considerable amount of publications in recent years. The interdisciplinary nature as well as technical and non-technical potentials and challenges of cloud computing (e.g., discussed in [1], [2], [3], [4]) are some of the main reasons for the rapid development. Given the significantly increasing number of publications, it becomes more and more important to investigate the current state and evolution of cloud computing research. Quantitative studies measuring and analyzing science activities form a type of research commonly known as scientometrics [5]. By providing a view on a research field from a meta-perspective [6], [7], scientometric studies facilitate the development and improvement of an academic discipline [5], [8] serving as a vital basis for defining and debating future research agendas [9]. Assuming that scientific activities are reflected through scientific publications, scientometric studies apply empirical measures to analyze scientific output of a specific field in order to better understand the dynamics and structure of its development. Thereby, it is possible to explore the body of publications extensively, for example, to observe citation patterns, number and types of citations, number and structure of authors and so forth. Going further, a scientometric study gives some indication of research activities in general, such as with respect to knowledge sharing, research quality, socio-organizational structures, influential countries/affiliations/authors, development of key topics, structural change, and economic impact of research. For further reading see, e.g., [6], [8], [10], [11], [12], [13]. Moreover, scientometrics, as an evaluation tool of science, increasingly impacts the resource distribution of research institutions [12].

Regarding these facts, it is surprising that not much work has been devoted to scientometric analysis of cloud computing research and even more so regarding a comprehensive scientometric study of the field. At this point, we are aware of three scientometric studies in the context of cloud computing. The authors of [14] investigate 510 publications related to cloud computing that are obtained from the Web of Science (WoS) database for the years 2001-2010. They look at the productivity of authors and contributing countries by analyzing the number of publications which is aggregated by WoS. In [15], 89 journal papers related to cloud computing research in China are investigated for the time period between 1993 and 2010. Based on the data of the Chinese Journal Full-text Database (CNKI), the authors examine the distribution of the number of journal papers, authors, subjects and funded papers. In [16], scientometric methods are applied to analyze the research progress of cloud security research from 2008 to 2011 in China. The authors investigate 103 journal articles of 76 journals provided by CNKI. They analyze types of contributing affiliations and identify the key topics exclusively focussing on cloud security. In general, these studies lack of important insights, such as given by an overview of current research topics and trends, citation patterns and top publications.

• *The authors are with the Institute of Information Systems, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany. E-mail: {leonard.heilig, stefan.voss}@uni-hamburg.de.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2014.2321168, IEEE Transactions on Cloud Computing

Due to the relatively small number of publications being analyzed and the specific area under consideration, the implications of those studies are limited. The existing studies especially apply straight count measures to analyze the respective literature. Without using specific scientometric techniques, for instance different methods to evaluate the productivity of authors or algorithms for a keyword cluster analysis, it is difficult to generate novel insights. Consequently, the main objective of this study is to provide a more comprehensive view on the overall cloud computing research area within a relevant time frame in order to present empirical and relevant findings. In this paper, we extend the observation period and apply a variety of methods including quantitative and computational algorithms to analyze key aspects of cloud computing research.

By considering the number of publications in the area of cloud computing, we observe a significant interest especially from 2008 onwards. Since then, the number of articles dramatically increased year after year. In Google Scholar, for instance, the number of search results nearly doubles each year since 2008. The significant increase of scientific literature can also be recognized in the Scopus and WoS database. Prior to 2008, Scopus indexed only three publications related to cloud computing, WoS only one publication. At the time of this study (as of January 12, 2014), Scopus covers 15,376 relevant publications while WoS covers 8,262 publications. To the best of our knowledge, published cloud computing articles in the time period from 2011 to 2013 are completely unexplored by means of scientometrics.

In this paper, we present a comprehensive scientometric study that empirically explores publications related to cloud computing covered by Elsevier's Scopus database from 2008 to 2013. In total, we investigate 15,376 publications. To the best of our knowledge, this is the first scientometric study that assesses such a large number of peer-reviewed publications. Thus, the results of this study stand on a broad empirical basis and may encourage scholars to conduct more comprehensive scientometric studies in other academic disciplines. We provide extensive insights into publishing patterns (e.g., contributing countries, distribution of outlets), analyze frequent keywords and keyword clusters to identify widely discussed topics and their relationships, provide insights into citation patterns of outlets, publications, affiliations and authors, and investigate the research productivity of affiliations and scientists in the area of cloud computing. Thereby, we present novel insights from a meta-perspective that may help to better understand the evolution, state and trends of cloud computing research. Due to limitations of space, this study does not intend to give an overview of cloud computing in general (for further comprehensive reading see, e.g., [17], [18], [19]).

TABLE 1
Number of publications per year

Database	2008	2009	2010	2011	2012	2013	$\sum$
Scopus	74	641	1926	4038	5146	3551	15376
ISI WoS	5	441	719	1543	3106	2448	8262

The remainder of this paper is organized as follows. Section 2 briefly describes the methodology and methods being applied in this scientometric study. In Section 3, publication patterns are investigated and further analyzed to understand the general composition of the field from different perspectives. The current focus of cloud computing research and dependencies between topics are observed by analyzing top keywords and keyword clusters in Section 4. Subsequently, the impact and productivity of cloud computing research is examined in Sections 5 and 6. Finally, a conclusion is presented in Section 7.

2 METHODOLOGY

The collection of relevant publications and citations establishes the foundation for a scientometric analysis of a specific research area [7]. As indicated, this study intends to cover a large part of peer-reviewed cloud computing articles published in the last six years. By this, we aim to obtain empirical evidence for supporting the metascientific findings of this scientometric study. In this section we describe our proceeding regarding data collection, data processing and proof-reading.

2.1 Data collection and cleansing

As manual processing of bibliographic data can be extremely cost- and labor-intensive [9], we use Elsevier's Scopus to collect and process structured data of articles. The Scopus database has decisive advantages over other bibliographic databases such as Thomson Reuters WoS. First, the amount of cloud computing articles being covered by Scopus is almost twice as high as in WoS (see Table 1); this includes article-in-press publications of a large number of journals. Second, Scopus provides advanced functionality to export structured data including citation and bibliographical information as well as abstracts, keywords, and references. A general limitation of using bibliographic databases is, however, that novel articles are often not covered which explains the decrease in the number of publications between the years 2012 and 2013.

In order to cover a large part of publications in cloud computing, a generic search query '*cloud computing*' is used in the fields *title*, *abstract*, and *keywords*. The asterisk acts as wildcard character. The search query finds 16,042 articles for the observation period 2008-2013 (as of January 12, 2014).

Data cleansing actions are carried out to detect and remove inaccurate data records (e.g., authors/title not specified, double entries, etc.). Finally, we retain 15,376 publications containing 273,477 references and 32,620 unique keywords. Only 94.57 % of those publications have a non-empty bibliography resulting in an average of 18.81 references per article. The majority of publications, at an average 97.38 %, is written in English.

2.2 Data processing

Based on an extensive bibliographic data basis, several meta-data attributes (e.g., authors, keywords, document type, etc.) of the collected publications are utilized to analyze certain aspects. In the following, we briefly describe the methods being applied to investigate research productivity, research impact, and keyword clusters.

2.2.1 Research productivity

The review of literature reveals four basic approaches to measure research productivity: *straight count*, *author position*, *equal credit* and *normalized page size* [9], [20]. The straight count method assigns a score of one to each of the co-authors of an article. This approach, however, undervalues the productivity of single-author papers and favors individual co-authors of multi-author papers. The author position approach, in contrast, assigns a score based on the original position of authorship and favors first authors. To calculate the scores based on the position of authors, the formula proposed by [21] is used [9]. Although the method considers that the first author often is the main contributor of an article, this approach is error-prone to multi-author papers where the names of co-authors are arranged in alphabetical order. The equal credit method aims to compensate these errors; it calculates a per-author score based on the reciprocal of the number of authors so that the score of each co-author is reduced by every additional co-author. The last approach, normalized page size, is not considered in this study for the following two reasons. First, it assigns a score to each of the authors based on the number of pages and authors per publication and thus favors quantity rather than quality. Second, outlets often limit the number of pages per publication which would lead to an error-prone analysis of research productivity. In this study, we primarily focus on the equal credit method.

2.2.2 Research impact

In order to measure the research impact, indices based on individual citations are applied. This includes individual citations of journals, conferences, and authors as well as the normalized citation impact index (NCII) to consider the longevity of publications [9]. Regarding research productivity and research impact, we also discuss the Matthew effect which describes the phenomenon that highly recognized scientists get most of the credit for contributions that are also presented by many other, relatively unknown scientists [22]. As the credit given by the scientist's peers again influences recognition, the effect leads to accumulated advantages for those authors. Author and/or publication visibility is furthermore influenced by positive network membership effects, such as given by influential outlets, research institutes or research collaborations [23].

2.2.3 Keyword analysis and other relevant aspects

To further explore key topics and aspects in cloud computing from a meta-perspective, additional methods are implemented based on the given data basis. This includes algorithms for analyzing keyword clusters as well as for classifying and aggregating bibliographic data.

2.3 Proofreading

As indicated, one purpose of the scientometric analysis is to reduce the effort of analyzing a great amount of peer-reviewed papers. Although the indexing of scientific publications is highly standardized, some inconsistencies can be detected, such as for the name of research affiliations. To ensure the correctness of results, generated outputs are validated by manual proof-reading activities to identify inconsistencies. The resulting semi-automatic process guarantees the data quality and quality of results of this study.

3 Analysis of Publishing Patterns

We begin by analyzing the basic structure of cloud computing research from different perspectives. First of all, the distribution of involved research disciplines is investigated in order to evaluate major disciplines involved in the development of the field. Subsequently, the distribution of contributing countries is analyzed for all publications and for well-recognized publications. On the level of publications, we further analyze the number of authors, distribution of document types and the number of publications per outlet.

3.1 Academic disciplines

To obtain an understanding of the general structure and development of the cloud computing research area, we start with a scientometric analysis of academic disciplines (see Table 2). Based on the chosen publication outlet, which may be attributed to more than one subject area, each publication is categorized by Scopus [24].

The figures of the distribution of publications show some interesting patterns. While the contribution in the subject area *Computer Science* is nearly constant, the number of contributions in other disciplines

TABLE 2
Subject areas (Avg. ≥ 1 %)

Subject area	2008 (%)	2009 (%)	2010 (%)	2011 (%)	2012 (%)	2013 (%)	Avg. (%)
Computer Science	54.55	59.57	58.87	60.74	59.46	49.29	57.1
Engineering	20.66	11.65	11.24	13.82	14.90	23.48	16.0
Mathematics	4.13	11.55	14.85	13.82	10.43	11.03	11.0
Social Sciences	2.48	5.07	4.36	3.35	3.45	4.02	3.8
Business, Management and Accounting	7.44	3.04	2.08	2.52	2.22	1.61	3.2
Decision Sciences	1.65	2.03	2.11	1.33	1.65	1.33	1.7
Economics, Econometrics and Finance	4.13	-	0.85	0.89	0.34	0.22	1.1
Materials Science	0.83	0.91	1.19	1.14	0.78	1.11	1.0

slightly indicates the effects of the hype surrounding cloud computing in an early phase of development. In particular for the business-related subject areas, a peak of popularity and the subsequent effect of *disillusionment* can be identified. According to Gartner's *Hype Cycle for Emerging Technologies* [25], cloud computing reached the *peak of inflated expectations* between 2008 and 2009 which is reflected by the number of contributions in certain disciplines. This implies that research activities in cloud computing are partially affected by ongoing expectations.

The figures further indicate that the majority of contributions in the area of cloud computing is related to *Computer Science*. This demonstrates that a lot of research is primarily concerned with the technology itself. This observation is also made in other publications, such as in [4]. According to [4], it is important to equally consider business-related issues associated with cloud computing. The numbers further reveal a slight trend towards the application of cloud technologies to support research and/or business-related activities, such as in engineering and mathematics, but also that general initial thoughts regarding potential applications are underpinned with more theoretical considerations (e.g., regarding mathematical models in cloud pricing). We expect that the trend towards a more productive use of cloud computing in practice will continue in the future. Social and business-related disciplines as well as the interaction between them will become increasingly important in order to understand the implications of cloud computing from different perspectives.

3.2 Contributing countries and authorship

To obtain a deeper insight into contribution patterns, we further investigate the distribution of publications per country and authorship patterns. As shown in Table 3 (R.: Rank), the majority of research on cloud computing is carried out by researchers from China (22.50 %) and the United States (19.16 %).

As these numbers alone do not provide enough insight about the relevance of contributions per country, we also generate a ranking of contributing countries for publications that are cited by at least 50 other publications. The numbers in the right area of Table 3 show that mostly authors from the United States (36.52 %) and Australia (10.43 %) have contributed widely recognized publications.

TABLE 3
Contributing countries (left: all publications; right: publications cited at least 50 times

R.	Country	f (%)	R.	Country	f (%)
1	China	22.50	1	USA	36.84
2	USA	19.16	2	Australia	10.53
3	Germany	5.07	3	Germany	8.77
4	India	4.86	4	China	7.02
5	UK	4.47	5	UK	6.14
6	Taiwan	3.60	6	Canada	5.26
7	South Korea	3.29	7	Spain	2.63
8	Australia	3.17	7	India	2.63
9	Japan	3.15	7	New Zealand	2.63
10	Italy	2.85	7	South Korea	2.63
11	Canada	2.63	7	Austria	2.63
12	Spain	2.51	12	Greece	1.75
13	France	2.39	12	Israel	1.75
14	Austria	1.24	12	Switzerland	1.75
15	Brazil	1.20	12	Singapore	1.75
16	Greece	1.14	16	Italy	0.88
17	Singapore	1.01	16	Cyprus	0.88
Total		84.24	Total		96.49

Regarding the authorship of publications the average number of authors per publication over the last six years is depicted in Figure 1. The distribution of authorship shows that for more than half of the publications n the number of co-authors is between two and four. In conjunction with the relatively high percentage of publications with five or more authors, the distribution demonstrates that collaboration may have some advantages over research by individual researchers. The interdisciplinary nature of cloud computing research may be one of the main reasons for the dominance of joint works.

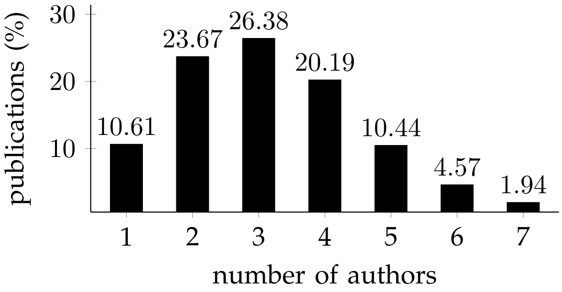

Fig. 1. Co-authorship distribution ($n = 15376$)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2014.2321168, IEEE Transactions on Cloud Computing

TABLE 4
Referencing patterns

Min. citations	n	f
0	14541	18.81
1	5183	22.00
5	1415	24.44
10	661	24.93
25	213	28.06
50	86	31.10
100	31	30.68
150	17	31.86
200	9	29.44
300	4	36.00
350	3	37.33

3.3 Referencing

We next look at referencing patterns of publications with a non-empty bibliography ($n = 14541$, see Table 4). A table row describes the average number of references f depending on the number of citations a publication receives. The number of considered publications is expressed by n. For instance, a publication which is cited by 100 or more publications contains on average 30.68 references. By comparing those numbers we find that frequently cited publications (cited by at least 25 publications) contain on average about ten references (f) more than other publications. The coverage of significant literature is recognized as a main criterion for high quality research [26]. Although these findings stand on a broad empirical basis, we have to consider that outlets often limit the maximum number of pages per publication which also affects the number of references.

3.4 Form of publication

The selection of an appropriate outlet often has an influence on the visibility and impact of an article. Consequently, it is interesting to analyze which type of publication venue researchers prefer for conveying their ideas and insights to the research community. As the document type is specified for all observed cloud computing articles, it is possible to analyze the distribution of document types for a respective research field. Table 5 indicates that most articles on cloud computing, at an average 73.88 %, are preferably published through conference proceedings. This can be explained by the fact that most of the research activities are carried out by the computer science research community, as shown in Table 2, where conference publications have always had a dominant presence and are legitimized as the primary means of publication [27], [28]. One reason is that the review and publishing of journal papers take a long time whereas conference papers are usually published much faster. In particular in a rapidly growing field of research, as in the case with cloud computing, a timely presentation is important; otherwise, it may happen that an idea gets obsolete or is presented, in a similar form, by other researchers.

To extend this analysis, we further investigate the distribution of publications for conferences and journals. The number of publications per conference only provides limited insights as it is usually dependent on the size of a conference and is limited to a particular year. The number of publications per journal, in contrast, reveals common journals for publishing articles in the area of cloud computing. Table 6 shows a list of journals and the related number of publications (f). The numbers indicate that journal articles are published by a wide variety of scientific journals emphasizing the various theoretic roots, such as in distributed computing [1]. Journals with a specific focus on cloud computing, such as *Journal of Cloud Computing* and *IEEE Transactions on Cloud Computing*, are not listed in the ranking mainly due to their novelty.

4 Frequent keywords and keyword clusters

Keywords are an effective tool to abstractly represent and classify the content of a scientific article. From a meta-perspective, keywords provide the foundation for analyzing the key topics and aspects representing a respective research area. The emergence of new popular topics can be quickly identified by looking at the occurrence of keywords within a specific timeframe. By analyzing co-occurrences of keywords, it is also possible to identify topics or aspects that are strongly related to each other.

The scientometric study extracts 32,620 unique keywords from the *author keywords* and *index keywords* field of the collected source articles. Note that keywords are not always specified by the authors. In case of missing keywords, index keywords are manually assigned by professional indexers based on several thesauri [24]. By this, we observe that keywords are specified for 97.18 % of the analyzed publications, while for 15.57 % of those publications, index keywords have been assigned by Scopus. In case that no author keywords are specified, we use, if available, the indexed keywords of a publication for the analysis of frequent keywords and keyword clusters.

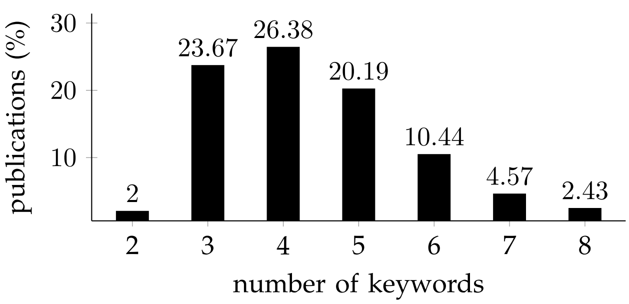

Fig. 2. Number of keywords per pub. ($n = 14942$)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2014.2321168, IEEE Transactions on Cloud Computing

TABLE 5
Number of publications by document type

Outlet	2008 (%)	2009 (%)	2010 (%)	2011 (%)	2012 (%)	2013 (%)	Overall (%)
Conference Paper	71.62	75.04	81.52	79.49	77.36	58.24	73.88
Journal Article	24.32	21.06	15.42	17.31	18.15	33.17	21.57
Journal Article in Press	-	0.16	0.31	0.07	0.82	5.74	1.42
Review	-	0.47	0.88	1.09	1.28	1.55	1.05
Other	4.05	3.28	1.87	2.03	2.39	1.30	2.07
n	74	641	1926	4038	5146	3551	15376

TABLE 6
Number of publications per journal ($f \geq 30$)

R.	Journal	f
1	*Future Generation Computer Systems*	100
2	*Journal of Supercomputing*	56
3	*IEEE Transactions on Parallel and Distributed Systems*	50
4	*Fujitsu Scientific and Technical Journal*	49
5	*Concurrency Computation Practice and Experience*	47
5	*Computer*	47
7	*Journal of Huazhong University of Science and Technology*	44
7	*IEEE Internet Computing*	44
9	*Tongxin Xuebao/Journal on Communications*	43
10	*Journal of Grid Computing*	40
11	*Journal of Theoretical and Applied Information Technology*	35
11	*IT Professional*	35
13	*Journal of Convergence Information Technology*	34
14	*Journal of Computational Information Systems*	33
15	*Cluster Computing*	33
16	*Jisuanji Xuebao/Chinese Journal of Computers*	31
17	*NEC Technical Journal*	30

Regarding the average distribution of the number of keywords per publication, we observe that commonly between three and six keywords are used to capture the core topic of a publication (see Figure 2). This observation is not specific for cloud computing; publishers often specify the minimum and/or maximum number of keywords per publication. In order to reduce the variability of keyword terms, publishers usually provide a predefined list of standardized keywords being relevant for a specific journal or conference.

The ranking of the top keywords with a high frequency (f greater than or equal to 100) is provided in Table 8. The results indicate that recent research activities are mainly focussed on the technology itself (e.g., *virtualization*, *scheduling*, *energy efficiency*, *load balancing*), current challenges (e.g., *security*, *privacy*, *interoperability*, *quality of service*, *monitoring*) and the utilization of scalable cloud resources (e.g., *scalability*, *optimization*, *MapReduce*). The literature reveals that these topics are also discussed, e.g. with regard to research challenges, in widely recognized publications, such as in [2], [3], [4], [17], [19]. This shows that those publications have a huge impact on the direction of cloud computing research. Popular links to other fields of research are also revealed, for instance with the keywords *internet of things* and *grid computing*. Again, the dominance of computer science related research in cloud computing is evident. By analyzing frequent keywords per year, the emergence and growing popularity of specific topics can be documented. For example, the growing importance of cloud computing to efficiently process large and complex masses of structured and unstructured data is depicted by keywords shown in Table 7. Based on foundational research concerned with methods (e.g., data mining techniques, MapReduce algorithm) and technologies (e.g., Hadoop) we recognize the emergence of new research topics in the field of cloud computing, such as *Big Data* and the *Internet of Things*. Consequently, the keyword analysis can be used as a tool for identifying current research trends.

TABLE 7
Yearly occurrence of keywords

f	2008	2009	2010	2011	2012
MapReduce	2	4	45	120	133
Hadoop	1	7	29	68	98
Data Mining	1	13	14	28	52
Big Data	0	0	0	3	27
Internet of Things	0	0	0	1	9

As an area of interest is usually characterized by more than one keyword, we further analyze the co-occurrences of keywords within the keyword list of publications, also referred to as keyword cluster. Frequent keyword clusters unveil interconnections between different aspects and topics. While the analysis of top keywords is straight forward, the analysis of keyword clusters is computationally complex since every possible combination of relevant keywords

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2014.2321168, IEEE Transactions on Cloud Computing

TABLE 8
Top keywords ($f \geq 100$)

R.	Keyword	f	R.	Keyword	f
1	cloud computing	9354	28	scalability	158
2	virtualization	680	30	middleware	156
3	security	510	31	performance	153
4	cloud	439	32	energy efficiency	152
5	MapReduce	398	33	resource management	150
6	grid computing	370	34	cloud security	144
7	computer systems	365	35	PaaS	140
8	web services	309	36	load balancing	138
9	virtual machines	270	36	distributed computer systems	138
10	Hadoop	267	38	data mining	133
10	internet	267	39	computer simulation	132
12	cloud services	266	40	SOA	131
13	SaaS	259	41	algorithms	130
14	information technology	253	41	architecture	130
15	resource allocation	229	43	computing resource	126
16	privacy	227	43	interoperability	126
17	scheduling	225	45	QoS	120
18	data centers	222	46	data center	119
18	quality of service	222	46	distributed systems	119
20	IaaS	200	48	computer science	115
21	virtualizations	188	49	fault tolerance	110
22	virtual machine	185	49	clouds	110
23	cloud storage	175	51	internet of things	109
24	access control	173	52	information management	108
24	optimization	173	53	virtual reality	107
26	distributed computing	165	54	service provider	102
27	software as a service	164	55	monitoring	101
28	mobile cloud computing	158			

TABLE 9
Top keyword clusters of length 2 ($f \geq 35$)

R.	Keyword cluster		f
1	Hadoop	MapReduce	124
2	privacy	security	88
3	SaaS	PaaS	81
4	PaaS	IaaS	77
5	SaaS	IaaS	75
6	virtual machines	computer simulation	71
7	computing environments	computer systems	63
8	virtual machines	virtualizations	61
9	cloud services	distributed database systems	58
10	virtualizations	virtual reality	55
11	cloud services	web services	44
12	virtualization	security	40
12	computing system	computer systems	40
14	virtualization	cloud	39
15	data centers	virtual machines	38
16	cloud computing environments	computer systems	37
16	service oriented architecture (soa)	information services	37

TABLE 10
Top keyword clusters of length 3 ($f \geq 15$)

R.	Keyword cluster			f
1	PaaS	IaaS	SaaS	70
2	cloud services	distributed database systems	Web services	30
3	computer program	internet	software	22
4	virtualizations	virtual reality	virtual machines	20
5	computer simulation	virtualizations	virtual machines	18
5	information retrieval	internet	information storage and retrieval	18
7	methodology	computer program	software	17
8	computational biology	biology	methodology	16
8	software as a service	platform as a service	infrastructure as a service	16
10	PaaS	IaaS	virtualization	15
10	computer program	computational biology	software	15

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2014.2321168, IEEE Transactions on Cloud Computing

TABLE 11
General citation patterns

Year	2008	2009	2010	2011	2012	2013
Number of publications	74	641	1926	4038	5146	3551
Number of citations	1581	8035	11089	8609	3889	585
Longevity (in years)	6	5	4	3	2	1
Overall NCII score	263.50	1607.00	2772.25	2869.67	1944.50	585.00
Avg. NCII score / pub.	3.56	2.51	1.44	0.71	0.38	0.16

TABLE 12
Number of citations per outlet ($n = 15376$)

Outlet	2008 (%)	2009 (%)	2010 (%)	2011 (%)	2012 (%)	2013 (%)	Overall (%)
Conference Paper	84.19	47.69	52.60	50.32	48.65	15.04	49.75
Journal Paper	15.81	46.45	31.12	40.66	45.18	70.94	41.69
Review	-	2.14	13.45	7.13	3.81	11.45	7.60
Other	-	3.72	2.82	1.89	2.37	2.56	2.67
Number of citations	1581	8035	11089	8609	3889	585	33788

needs to be compared. The implemented algorithm incrementally increases the length of keyword clusters in order to identify all relevant co-occurrences. In the following, we perform a keyword cluster analysis for keyword clusters with two elements (Table 9) and three elements (Table 10). To obtain meaningful keyword clusters, we removed the keyword *cloud computing* since it mainly demonstrates the relation to cloud computing research (used in 60.84 % of all observed publications). The findings reveal a strong interrelation between certain keywords. For the two element keyword clusters, for instance, we see that the terms *Hadoop* and *MapReduce* as well as *privacy* and *security* are often used together. For three element keyword clusters we obtain, for instance, that the combination of the terms *IaaS*, *PaaS* and *SaaS* is common. By studying the generated keyword clusters, it is possible to identify dependencies between certain keywords. For example, the term *cloud service* is related to *web services* that rely on *distributed database systems* or that *virtualization* has an influence on the *security* of cloud computing. Again, the technical focus of research is obvious. Consequently, our findings help to better understand the relevant topics in cloud computing research and how they are related to each other. Furthermore, the keyword cluster analysis provides the foundation for generating a topic network consisting of nodes (topics) and edges (relationship between topics).

5 CITATION PATTERNS

While the numbers given in the previous subsections provide insights into publishing patterns and key disciplines of cloud computing, a primary concern of a scientometric study is to evaluate the impact of contributions. A measure for analyzing the impact of contributions is the aggregated number of citations a publication receives. The aggregated citations of a publication express how often this publication is referenced in other publications. For the overall number of publications we receive 33,788 citations. The number of citations varies between 0 and 1,002 with an average of 2.197 references per article. In the following, we evaluate the impact of contributions from different perspectives.

5.1 Overall citation patterns

First, we analyze the distribution and impact of citations in general. As time has a significant influence on the number of citations a publication receives, we use the NCII in order to make the citation numbers of several years comparable. The NCII takes into account the longevity of a publication which refers to the number of years the publication has been in print [9].

$$NCII = \frac{\text{number of citations per publication}}{\text{publication longevity (in years)}} \quad (1)$$

The figures presented in Table 11 reveal some interesting patterns. The most noticeable observation is that the average NCII per publication is declining which demonstrates the significant impact of early and fundamental publications on other publications. We also recognize a significant decline of citations from 2011 to 2012. The main reason for this is that many references used in publications of 2013 are still not covered by Scopus (see Table 1).

5.2 Outlet citations

As a next step we analyze the distribution of citations with respect to different publication outlets. The results shown in Table 12 outline that the main sources of references are conference papers (49.75 %) and journal papers (41.69 %). Regarding the number of contributions per outlet, as analyzed in Section 3.4, we observe that journal contributions are more efficient with respect to received citations. In addition, the figures indicate that books and/or book chapters

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2014.2321168, IEEE Transactions on Cloud Computing

TABLE 13
Conference citations ($f \geq 120$)

R.	Conference	f
1	*2nd IEEE International Conference on Cloud Computing Technology and Science, CloudCom 2010*	536
2	*3rd IEEE International Conference on Cloud Computing, CLOUD 2010*	516
3	*Grid Computing Environments Workshop, GCE 2008*	485
4	*9th IEEE/ACM International Symposium on Cluster Computing and the Grid, CCGRID 2009*	406
5	*4th IEEE International Conference on Cloud Computing, CLOUD 2011*	369
6	*1st ACM Symposium on Cloud Computing, SoCC 2010*	315
7	*10th IEEE International Conference on High Performance Computing and Communications, HPCC 2008*	307
8	*10th IEEE/ACM International Conference on Cluster, Cloud, and Grid Computing, CCGRID 2010*	296
9	*16th ACM Computer and Communications Security Conference, CCS 2009*	232
10	*3rd IEEE International Conference on Cloud Computing Technology and Science, CloudCom 2011*	202
11	*2nd IEEE International Conference on Cloud Computing, CLOUD 2009*	200
12	*19th ACM International Symposium on High Performance Distributed Computing, HPDC 2010*	196
13	*ICSE Workshop on Software Engineering Challenges of Cloud Computing, ICSE 2009*	179
14	*11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGrid 2011*	160
15	*5th IEEE International Conference on Cloud Computing, CLOUD 2012*	155
16	*4th IEEE International Conference on eScience, eScience 2008*	147
17	*24th IEEE International Conference on Advanced Information Networking and Applications, AINA 2010*	122

TABLE 14
Journal citations ($f \geq 100$)

R.	Journal	f	n	f (rel)	IF	5-Year IF
1	*Future Generation Computer Systems*	1569	100	15.69	1.864	2.033
2	*IEEE Internet Computing*	755	44	17.16	2.039	2.498
3	*Computer*	703	47	14.96	1.675	2.403
4	*IEEE Security and Privacy*	415	23	18.04	0.962	1.019
5	*IEEE Transactions on Parallel and Distributed Systems*	374	50	7.48	1.796	2.031
6	*Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems*	352	23	15.30	-	-
7	*IT Professional*	340	35	9.71	0.482	-
8	*Communications of the ACM*	338	11	30.73	2.511	2.564
9	*Journal of Internet Services and Applications*	276	11	25.09	-	-
10	*IBM Journal of Research and Development*	231	8	28.88	0.688	1.684
11	*IEEE Pervasive Computing*	206	2	103.00	2.055	2.566
12	*Proceedings of the VLDB Endowment*	202	12	16.83	-	-
13	*Journal of Parallel and Distributed Computing*	198	26	7.62	-	-
14	*BMC Bioinformatics*	193	10	19.30	3.024	3.510
15	*Ruan Jian Xue Bao/Journal of Software*	178	14	12.71	-	-
16	*Computer Journal*	161	11	14.64	0.755	0.954
17	*Journal of Grid Computing*	160	40	4.00	1.603	-
18	*Decision Support Systems*	158	4	39.50	2.201	3.037
19	*Journal of Convergence Information Technology*	136	34	4.00	-	-
20	*Nature Photonics*	130	2	65.00	27.254	31.567
21	*IEEE Transactions on Intelligent Transportation Systems*	123	2	61.50	3.064	3.263
22	*International Journal of Information Management*	118	11	10.73	1.843	1.898
23	*Computing in Science and Engineering*	115	10	11.50	1.729	1.676
24	*Genome Biology*	106	1	106.00	10.288	8.959
25	*Advances in Information Sciences and Service Sciences*	102	29	3.52	-	-
26	*Synthesis Lectures on Computer Architecture*	101	3	33.67	-	-

currently play a minor role in the area of cloud computing. This is contradictory to other fields [12] where books and book chapters are cited more frequently (see, e.g., scientometric studies in the journal *Scientometrics* [6], [10]). Although scientific review articles play a minor role, they receive a large part of the remaining citations. This can be explained by the simple fact that a review article can be seen as a valuable measure to handle the rapidly growing body of knowledge in cloud computing.

5.3 Conference and journal citation patterns

The figures of the previous section have shown that most of the citations are received through conference and journal papers. To further investigate how the citations are distributed among different conferences and journals, we generate both a ranking for conferences (see Table 13) and journals (see Table 14). Again, we apply a straight count method to analyze the citation patterns in order to clearly differentiate between research impact and productivity. By considering the conference citations, we observe that widely cited publications are mainly published by cloud computing-specific symposia. We also see that influential publications are published by three main conferences: *IEEE International Conference on Cloud Computing* (Google Scholar h5-index: 30), *IEEE International Conference on Cloud Computing Technology and Science* (Google Scholar h5-index: 29) and *IEEE/ACM International Symposium on Cluster,*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2014.2321168, IEEE Transactions on Cloud Computing

TABLE 15
Top cited publications (NCII score $\geq$ 30.0)

R.	Authors	NCII	f	f^G
1	Armbrust M., Fox A., Griffith R., Joseph A.D., Katz R., Konwinski A., Lee G., Patterson D., Rabkin A., Stoica I., Zaharia M. (2010) A view of cloud computing. Communications of the ACM 53(4):50-58.	250.50	1002	5876
2	Buyya R., Yeo C.S., Venugopal S., Broberg J., Brandic I. (2009) Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation Computer Systems 25(6):599-616.	190.80	954	2018
3	Foster I., Zhao Y., Raicu I., Lu S. (2008) Cloud Computing and Grid Computing 360-degree compared. In: Grid Computing Environments Workshop, pp. 1-10.	63.00	378	1582
4	Nurmi D., Wolski R., Grzegorczyk C., Obertelli G., Soman S., Youseff L., Zagorodnov D. (2009) The eucalyptus open-source cloud-computing system. In: 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, CCGRID 2009, pp. 124-131.	62.40	312	1300
5	Zhang Q., Cheng L., Boutaba R. (2010) Cloud computing: State-of-the-art and research challenges. Journal of Internet Services and Applications 1(1):7-18.	60.75	243	631
6	Subashini S., Kavitha V. (2011) A survey on security issues in service delivery models of cloud computing. Journal of Network and Computer Applications 34(1):1-11.	52.00	156	502
7	Sotomayor B., Montero R.S., Llorente I.M., Foster I. (2009) Virtual infrastructure management in private and hybrid clouds. IEEE Internet Computing 13(5):14-22.	49.60	248	554
8	Marston S., Li Z., Bandyopadhyay S., Zhang J., Ghalsasi A. (2011) Cloud computing - The business perspective. Decision Support Systems 51(1):176-189.	47.33	142	477
9	Li H., Homer N. (2010) A survey of sequence alignment algorithms for next-generation sequencing. Briefings in Bioinformatics 11 (5):473-483.	45.50	182	284
10	Wang H., Ma Y., Pratx G., Xing L. (2011) Toward real-time Monte Carlo simulation using a commercial cloud computing infrastructure. Physics in Medicine and Biology 56(17):175-181.	44.67	134	142
11	Ristenpart T., Tromer E., Shacham H., Savage S. (2009) Hey, you, get off of my cloud: Exploring information leakage in third-party compute clouds. In: Proceedings of the ACM Conference on Computer and Communications Security, pp. 199-212.	42.80	214	720
12	Li B.-H., Zhang L., Wang S.-L., Tao F., Cao J.-W., Jiang X.-D., Song X., Chai X.-D. (2010) Cloud manufacturing: A new service-oriented networked manufacturing model. Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems 16(1):1-7.	42.25	169	198
13	Satyanarayanan M., Bahl P., Caceres R., Davies N. (2009) The case for VM-based cloudlets in mobile computing. IEEE Pervasive Computing 8(4):14-23.	41.00	205	422
13	Hillerkuss et al. (2011) 26 Tbit s-1 line-rate super-channel transmission utilizing all-optical fast Fourier transform processing. Nature Photonics 5(6):364-371.	41.00	123	149
14	Iosup A., Ostermann S., Yigitbasi N., Prodan R., Fahringer T., Epema D. (2011) Performance analysis of cloud computing services for many-tasks scientific computing. IEEE Transactions on Parallel and Distributed Systems 22(6):931-945.	40.67	122	261
15	Xu X. (2012) From cloud computing to cloud manufacturing. Robotics and Computer-Integrated Manufacturing 28(1):75-86.	36.50	73	-
15	Buyya R., Yeo C.S., Venugopal S. (2008) Market-oriented cloud computing: Vision, hype, and reality for delivering IT services as computing utilities. In: Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications, HPCC 2008, pp. 5-13.	36.50	219	1193
17	Rochwerger B., Breitgand D., Levy E., Galis A., Nagin K., Llorente I.M., Montero R., Wolfsthal Y., Elmroth E., Caceres J., Ben-Yehuda M., Emmerich W., Galan F. (2009) The Reservoir model and architecture for open federated cloud computing. IBM Journal of Research and Development 53(4):1-11.	35.40	177	371
18	Zissis D., Lekkas D. (2012) Addressing cloud computing security issues. Future Generation Computer Systems 28(3):583-596.	32.50	65	226
19	Chen K., Zheng W.-M. (2012) Cloud computing: System instances and current research. Ruan Jian Xue Bao/Journal of Software 20(5):1337-1348.	31.40	157	209
20	Wang F.-Y. (2010) Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications. IEEE Transactions on Intelligent Transportation Systems 11(3):630-638.	30.75	123	142
21	Dikaiakos M.D., Katsaros D., Mehra P., Pallis G., Vakali A. (2009) Cloud computing: Distributed internet computing for IT and scientific research. IEEE Internet Computing 13(5):10-13.	30.00	150	329

Cloud and Grid Computing. The reason is mainly that these venues feature high quality and high impact research and serve a large academic audience. Our results correspond to the Google Scholar h5-metrics (http://scholar.google.de, date: 01/12/14). By adding the journal impact factor (IF) from the 2012 Journal Citation Reports (JCR), which calculates the average number of citations per publications based on the two preceding years, we observe that the average number of citations per journal article and the IF are only slightly correlated ($r = 0.47$). An even smaller correlation is observed by using the impact factor based on the last 5 years ($r = 0.41$). This demonstrates that the impact of journal articles published within the emerging cloud computing research area is not greatly influenced by the IF of journals. It should be noted that the IF, aiming to reflect the importance of journals, is not without any controversial discussion. A problem in the emerging area of cloud computing research could be, for instance, that relatively new, but influential journals are not indexed in the JCRs.

TABLE 16
Top cited authors (NCII score ≥ 60.0)

R.	Name	Affiliation	Country	n	NCII score	f
1	Buyya, Rajkumar	U. of Melbourne	Australia	15	407.50	2211
2	Patterson, David	U. of California, Berkeley	United States	1	250.50	1002
2	Griffith, Rean	U. of California, Berkeley	United States	1	250.50	1002
2	Konwinski, Andy	U. of California, Berkeley	United States	1	250.50	1002
2	Fox, Armando	U. of California, Berkeley	United States	1	250.50	1002
2	Katz, Randy	U. of California, Berkeley	United States	1	250.50	1002
2	Joseph, Anthony D.	U. of California, Berkeley	United States	1	250.50	1002
2	Zaharia, Matei	U. of California, Berkeley	United States	1	250.50	1002
2	Armbrust, Michael	U. of California, Berkeley	United States	1	250.50	1002
2	Rabkin, Ariel	U. of California, Berkeley	United States	1	250.50	1002
2	Lee, Gunho	U. of California, Berkeley	United States	1	250.50	1002
2	Stoica, Ion	U. of California, Berkeley	United States	1	250.50	1002
13	Yeo, Cheeshin	U. of Melbourne	Australia	3	244.63	1243
14	Venugopal, Srikumar	U. of Melbourne	Australia	2	227.30	1173
15	Brandic, Ivona	Vienna U. of Technology	Austria	2	198.30	1092
16	Broberg, James	U. of Melbourne	Australia	1	190.80	954
17	Foster, Ian T.	U. of Chicago	United States	3	121.93	656
18	Zhang, Lin	Beihang U.	China	6	102.92	408
18	Tao, Fei	Beihang U.	China	6	102.92	407
20	Youseff, Lamia	U. of California, Santa Barbara	United States	2	80.23	444
21	Li, Bohu	Beihang U.	China	3	67.92	291
21	Beloglazov, Anton	U. of Melbourne	Australia	4	67.42	232
23	Yong, Zhao	U. of Chicago	United States	1	63.00	378
23	Lu, Shiyong	U. of Chicago	United States	1	63.00	378
23	Raicu, Ioan	U. of Chicago	United States	1	63.00	378
26	Montero, Rubén Santiago	U. Complutense de Madrid	Spain	2	62.93	289
26	Llórente, Ignacio Martn	U. Complutense de Madrid	Spain	2	62.93	310
27	Wolski, Rich	U. of California, Santa Barbara	United States	1	62.40	312
27	Obertelli, Graziano	U. of California, Santa Barbara	United States	1	62.40	312
27	Nurmi, Daniel	U. of California, Santa Barbara	United States	1	62.40	312
27	Soman, Sunil	U. of California, Santa Barbara	United States	1	62.40	312
27	Grzegorczyk, Chris	U. of California, Santa Barbara	United States	1	62.40	312
27	Zagorodnov, Dmitrii	U. of California, Santa Barbara	United States	1	62.40	312
33	Cheng, Lu	U. of Waterloo	Canada	1	60.75	243
33	Zhang, Qi	U. of Waterloo	Canada	1	60.75	243
33	Boutaba, Raouf	U. of Waterloo	Canada	1	60.75	243

5.4 Publication citation patterns

Next we focus on the top publications in cloud computing. For this purpose, we calculated the total count of citations as well as the NCII for each publication (note that not all articles in the ranking are referenced in the bibliography), as depicted in Table 15. To compare the results with another citation measure, a column f^G is added containing the citation count of Google Scholar (http://scholar.google.de, date: 01/12/14).

5.5 Author and affiliation citation patterns

In order to get an overview of the most influential authors, we aggregate the NCII for each author. Each co-author receives the full count of citations of a corresponding article. A list of top authors is given in Table 16. A column for the number of publications being considered (n) and the frequency of citations (f) are added. We see that *Rajkumar Buyya* of the *University of Melbourne*, a co-author of two top publications (see Table 15), is currently the most influential author in the area of cloud computing in terms of citations. This observation was already made for the time period from 2001-2010 in [14]. Regarding the number of publications, we observe that, for many authors, only one publication is relevant for the citation index. This underlines the large impact of a handful of publications in the field; therefore, the ranking of top authors is strongly related to the ranking of top publications. Furthermore, the figures indicate that most of the top authors are currently from the United States and Australia, which corresponds to the scientometric analysis of top contributing countries, depicted in Table 3. The individual impact of authors significantly influences the impact of research affiliations (see Table 17). Note that *Manjrasoft*, a company founded by *Rajkumar Buyya*, strongly collaborates with the *University of Melbourne* explaining its high impact.

6 Research Productivity

The evaluation of research productivity is an important measure to identify the most active research institutes and scholars in the field. The insights may help, for instance, to build fruitful research collaborations and reflect the global distribution of research.

First we analyze the institutional research productivity. Table 18 shows a ranking of research institutes ordered by the number of contributed publications (f). The numbers demonstrate the dominance

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2014.2321168, IEEE Transactions on Cloud Computing

TABLE 17
Top 15 cited affiliations

R.	Affiliation	Country	NCII	*f*
1	U. of Melbourne	Australia	555.78	2382
2	Vienna U. of Technology	Austria	263.17	1172
3	U. of California, Berkeley	United States	250.50	1002
4	Manjrasoft	Australia	193.60	968
5	Tsinghua U.	China	180.62	856
6	U. of Chicago	United States	179.23	1517
7	Beihang U.	China	176.02	713
8	U. of California, Santa Barbara	United States	144.28	652
9	Carnegie Mellon U.	United States	129.82	513
10	U. of Maryland	United States	124.90	659
11	Arizona State U.	United States	110.97	394
12	Wayne State U.	United States	100.82	509
13	U. of Waterloo	Canada	100.22	385
14	Purdue U.	United States	96.63	371
15	U. of Innsbruck	Austria	94.90	323

TABLE 18
Top 15 contributing affiliations

R.	Name	Country	*f*
1	Beijing U. of Posts and Telecommunications	China	369
2	Tsinghua U.	Taiwan	253
3	Beihang U.	China	172
4	Wuhan U.	China	130
5	Northeastern U.	USA	117
6	Huazhong U. of Science and Technology	China	106
7	Arizona State U.	USA	101
8	U. of Melbourne	Australia	97
9	National U. of Defense Technology	China	96
10	National Taiwan U.	Taiwan	95
11	Purdue U.	USA	90
12	Peking U.	China	84
13	Shanghai Jiao Tong U.	China	81
14	Zhejiang U.	China	80
15	Vienna U. of Technology	Austria	78
15	Nanjing U.	China	78

of publications coming from Chinese institutions (as already demonstrated in Table 3). Most of these institutions have an excellent overall reputation (e.g., *University of Melbourne*, *Peking University*, *Tsinghua University*) and dispose of sufficient monetary and personell resources. In sum this provides the basis for promoting new generations of highly qualified scientists and enables to employ several scientists working on particular topics, such as in the area of cloud computing. The concentration of cloud computing research attracts new generations of scientists as they may benefit from a broad range of knowledge and expertise. This effect is consistent with the Matthew effect as researchers benefit from the intrinsic recognition of their affiliation and the associated network benefits.

Finally, we perform a scientometric analysis to evaluate the productivity of authors in the area of cloud computing. As discussed in Section 2, several methods for measuring the productivity of individual authors exist. In this paper, we focus on the equal credit method as it seems to be the best compromise among the discussed methods. Based on the equal credit method, each individual author receives a score based on the reciprocal of the number of authors per article, depicted in Table 19. We observe that most of the top productive individual contributors are from China (7), United States (6), Austria (5) and Australia (4). By comparing the results with the results of the author position method and straight count method, we observe that *Rajkumar Buyya* is not only the most influential researcher, but also the most productive one in the area of cloud computing. This demonstrates that a high impact of authors can have a positive effect on the individual productivity as it attracts interest from other researchers to collaborate in order to benefit from the author's recognition. Combined with the top cited publications, the numbers indicate the existence of the Matthew effect in the area of cloud computing. Authors who gained high recognition in an early stage of research development by contributing ideas and discoveries through using appropriate outlets are repeatedly rewarded by other scientists.

7 Conclusions

Cloud computing attracts a lot of interdisciplinary attention and is a rapidly developing field of research.

TABLE 19
Individual productivity (equal credit method, score ≥ 5.5)

R.	Author	Affiliation	Country	Score
1	Buyya, Rajkumar	U. of Melbourne	Australia	29.360
2	Brandic, Ivona	Technische U. Wien	Austria	11.426
3	Chen, Jinjun	U. of Technology Sydney	Australia	10.936
4	Puliafito, Antonio	U. degli Studi di Messina	Italy	10.525
5	Dustdar, Schahram	Technische U. Wien	Austria	10.327
6	Jin, Hai	Huazhong U. of Science and Technology	China	9.931
7	Rong, Chunming	U. of Stavanger	Norway	9.461
8	Petcu, Dana	U. de Vest din Timisoara	Timisoara	8.900
9	Huh, Euinam	Kyung Hee U.	South Korea	8.567
10	Srirama, Satish Narayana	U. of Tartu	Estonia	8.250
11	Prodan, Radu	U. of Innsbruck	Austria	8.108
12	Yang, Chaotung	Tunghai U.	Taiwan	7.400
13	Leymann, Frank	U. Stuttgart	Germany	6.947
14	Lee, Sungyoung	Kyung Hee U.	South Korea	6.777
15	Fox, Geoffrey Charles	Indiana U.	United States	6.659
16	Tsai, Weitek	Arizona State U. (Downtown Phoenix)	United States	6.410
17	Wang, Xingwei	Northeastern U. China	China	6.283
18	Zhang, Yaoxue	Tsinghua U.	China	6.175
19	Li, Baochun	U. of Toronto	Canada	6.033
19	Li, Keqin	State U. of New York at New Paltz	United States	6.033
21	Villari, Massimo	U. degli Studi di Messina	Italy	5.871
22	Deters, Ralph	U. of Saskatchewan	Canada	5.833
22	Zhang, Xuyun	Huazhong U. of Science and Technology	China	5.833
24	Truong, Honglinh	Technische U. Wien	Austria	5.758
25	Schikuta, Erich	U. Wien	Austria	5.700
26	Fu, Song	U. of North Texas	United States	5.667
27	Wang, Yeqiao	Beijing U. of Posts and Telecommunications	China	5.617

In this paper, we conduct a scientometric analysis to comprehensively investigate the development and current state of cloud computing related publications based on a large bibliographic data basis provided by Scopus.

The results of this study reveal that past and current research is dominated by computer science research conveyed especially through conference proceedings. The focus of research activities is predominantly influenced by fundamental and highly recognized scientists and publications. In this regard, we demonstrate the Matthew effect in the area of cloud computing. Given the results of the keyword analysis it is obvious that the past and current focus of cloud computing research lies mainly on the technology itself rather than on socioeconomic issues. Current trends, such as depicted by keywords related to data analysis and *Big Data*, demonstrate the increasing importance of shifting the focus of research to socioeconomic issues to solve understudied problems and better utilize the potentials of cloud computing. This may help to increase the overall value of cloud computing and may facilitate further adoption in both an academic and practical context. Thus, the results of the scientometric analysis may help the relatively new field of cloud computing to (re-) define itself in order to provide a clear direction and objectives for research. The analysis of main contributing affiliations and highly influential authors and publications may help, especially new generations of scholars, to get an overview of important publications, topics, outlets and to identify main contributors and driving forces in the area of cloud computing. Thus, the results of this study can be used to better understand patterns, trends and other important factors for directing individual research activities, efficiently extending research networks and selecting appropriate publication outlets for sharing individual knowledge.

The empirical findings of the scientometric analysis are partially reflected in widely recognized publications (e.g., relevant topics, research challenges). In general, this demonstrates the strength of a scientometric analysis to extensively investigate a field of interest. As demonstrated, the results of the scientometric study are not only valuable for discussing and defining future research agendas in the area of cloud computing. Moreover, the semi-automated process of assessing a large amount of publications makes it possible to easily obtain a general overview of a particular research area. This, in contrast, is not possible with structured literature reviews. Therefore, the study represents a good starting point for academics and practitioners to identify the sources and concentration of the existing knowledge base. In addition, research trends and important topics can be observed over a specific period of time by means of keyword analysis.

For further research, we intend to investigate and visualize collaboration structures among authors as well as the relationship between topics and authors in order to better understand the dynamics and network structure of this field. In this regard, we aim to analyze the patterns of trends for specific

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2014.2321168, IEEE Transactions on Cloud Computing

topics. In addition, we plan to compare the results with other metrics in order to further evaluate our results. Technically, we indend to further improve the applied data processing algorithms in order to further reduce manual proof-reading activities by means of data mining and machine learning.

REFERENCES

[1] M. A. Vouk, "Cloud computing - issues, research and implementations," *Journal of Computing and Information Technology*, vol. 16, no. 4, pp. 235–246, 2004.

[2] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-of-the-art and research challenges," *Journal of Internet Services and Applications*, vol. 1, no. 1, pp. 7–18, 2010.

[3] T. Dillon, C. Wu, and E. Chang, "Cloud computing: Issues and challenges," in *24th International Conference on Advanced Information Networking (AINA)*. IEEE, 2010, pp. 27–33.

[4] S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang, and A. Ghalsasi, "Cloud computing - the business perspective," *Decision Support Systems*, vol. 51, no. 1, pp. 176–189, 2011.

[5] B. R. Lewis, G. F. Templeton, and X. Luo, "A scientometric investigation into the validity of IS journal quality measures," *Journal of the Association for Information Systems*, vol. 8, no. 12, pp. 619–633, 2007.

[6] A. van Raan, "Scientometrics: State-of-the-art," *Scientometrics*, vol. 38, no. 1, pp. 208–218, 1996.

[7] S. Schwarze, S. Voß, G. Zhou, and G. Zhou, "Scientometric analysis of container terminals and ports literature and interaction with publications on distribution networks," in *Computational Logistics, Lecture Notes in Computer Science, LNCS 7555*. Berlin: Springer, 2012, pp. 33–52.

[8] D. Straub, "The value of scientometric studies: An introduction to a debate on IS as a reference discipline," *Journal of the Association for Information Systems*, vol. 7, no. 5, pp. 241–246, 2006.

[9] A. Serenko and N. Bontis, "Meta-review of knowledge management and intellectual capital literature: citation impact and research productivity rankings," *Knowledge and Process Management*, vol. 11, no. 3, pp. 185–198, 2004.

[10] W. Hood and C. Wilson, "The literature of bibliometrics, scientometrics, and informetrics," *Scientometrics*, vol. 52, no. 2, pp. 291–314, 2001.

[11] L. Leydesdorff, "Indicators of structural change in the dynamics of science: Entropy statistics of the SCI journal citation reports," *Scientometrics*, vol. 53, no. 1, pp. 131–159, 2002.

[12] S. Voß and X. Zhao, "Some steps towards a scientometric analysis of publications in machine translation," in *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications*, M. H. Hamza, Ed. Acta Press, 2005, pp. 651–655.

[13] L. Leydesdorff and T. Schank, "Dynamic animations of journal maps: Indicators of structural changes and interdisciplinary developments," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 11, pp. 1810–1818, 2008.

[14] K. Sivakumaren, S. Swaminathan, and G. Karthikeyan, "Growth and development of publications on cloud computing: a scientometric study," *International Journal of Information Library and Society*, vol. 1, no. 1, 2012.

[15] Q. Bai and W.-h. Dong, "Scientometric analysis on the papers of cloud computing," *Sci-Tech Information Development & Economy*, vol. 5, no. 1, pp. 6–8, 2011.

[16] T. Wang and G. Huang, "Research Progress of Cloud Security From 2008 to 2011 in China," *Information Science*, no. 1, pp. 153–160, 2013.

[17] I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud computing and grid computing 360-degree compared," in *Grid Computing Environments Workshop GCE'08*, 2008, pp. 1–10.

[18] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599–616, 2009.

[19] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and I. Stoica, "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[20] C. W. Holsapple, L. E. Johnson, H. Manakyan, and J. Tanner, "Business computing research journals: A normalized citation analysis," *Journal of Management Information Systems*, vol. 11, no. 1, pp. 131–140, 1994.

[21] G. S. Howard, D. A. Cole, and S. E. Maxwell, "Research productivity in psychology based on publication in the journals of the american psychological association." *American Psychologist*, vol. 42, no. 11, pp. 975–986, 1987.

[22] R. K. Merton, "The Matthew effect in science," *Science*, vol. 159, no. 3810, pp. 56–63, 1968.

[23] J. R. Faria and R. K. Goel, "Returns to networking in academia," *Netnomics*, vol. 11, no. 2, pp. 103–117, 2010.

[24] Scopus, "Content coverage guide," 2012. [Online]. Available: http://www.info.sciverse.com/UserFiles/sciverse_scopus_content_coverage_0.pdf

[25] Gartner, "Hype cycles," 2014. [Online]. Available: http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

[26] D. W. Straub, S. Ang, and R. Evaristo, "Normative standards for IS research," *SIGMIS Database*, vol. 25, no. 1, pp. 21–34, 1994.

[27] M. Y. Vardi, "Conferences vs. journals in computing research," *Communications of the ACM*, vol. 52, no. 5, pp. 5–5, 2009.

[28] Computing Research Association (CRA), "Best practices memo evaluating computer scientists and engineers for promotion and tenure," *Computing Research News*, 1999.

Leonard Heilig holds a B.Sc. (University of Münster, Germany) and a M.Sc. (University of Hamburg, Germany) in Information Systems. Currently he holds a position at the Institute of Information Systems at the University of Hamburg. He spent some time at the University of St Andrews (Scotland, UK) focusing on security management, web technologies and software engineering. Practical experiences include work at companies like Adobe Systems, Airbus Group Innovations and Beiersdorf Shared Services. His current interest focuses on cloud computing. Related applications incorporate mobile workforce management systems.

Stefan Voß is professor and director of the Institute of Information Systems at the University of Hamburg. Previous positions include (full) professor and head of the department of Business Administration, Information Systems and Information Management at the University of Technology Braunschweig (Germany) from 1995 up to 2002. He holds degrees in Mathematics (diploma) and Economics from the University of Hamburg and a Ph.D. and the habilitation from the University of Technology Darmstadt. His current research interests are in quantitative / information systems approaches to supply chain management and logistics including applications in maritime shipping, public mass transit and telecommunications. He is author and co-author of several books and numerous papers in various journals. Stefan Voß serves on the editorial board of some journals including being Editor of Netnomics and Editor of Public Transport. He is frequently organizing workshops and conferences. Furthermore, he is consulting with several companies.