sustainability

MDPI

Review

A Visualization Review of Cloud Computing Algorithms in the Last Decade

Junhu Ruan [1,2,3], Felix T. S. Chan [2], Fangwei Zhu [3], Xuping Wang [3] and Jing Yang [4,*]

1 College of Economics and Management, Northwest A & F University, Yangling 712100, China; rjh@nwafu.edu.cn or junhu.ruan@polyu.edu.hk
2 Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China; f.chan@polyu.edu.hk
3 Faculty of Management and Economics, Dalian University of Technology, Dalian 116023, China; zhufw@dlut.edu.cn (F.Z.); wxp@dlut.edu.cn (X.W.)
4 Department of Electronic Information Engineering, Handan Polytechnic College, Handan 056001, China
* Correspondence: yanguang@hebeu.edu.cn; Tel.: +86-310-316-4643

Academic Editor: Marc A. Rosen
Received: 30 July 2016; Accepted: 8 October 2016; Published: 11 October 2016

Abstract: Cloud computing has competitive advantages—such as on-demand self-service, rapid computing, cost reduction, and almost unlimited storage—that have attracted extensive attention from both academia and industry in recent years. Some review works have been reported to summarize extant studies related to cloud computing, but few analyze these studies based on the citations. Co-citation analysis can provide scholars a strong support to identify the intellectual bases and leading edges of a specific field. In addition, advanced algorithms, which can directly affect the availability, efficiency, and security of cloud computing, are the key to conducting computing across various clouds. Motivated by these observations, we conduct a specific visualization review of the studies related to cloud computing algorithms using one mainstream co-citation analysis tool—CiteSpace. The visualization results detect the most influential studies, journals, countries, institutions, and authors on cloud computing algorithms and reveal the intellectual bases and focuses of cloud computing algorithms in the literature, providing guidance for interested researchers to make further studies on cloud computing algorithms.

Keywords: cloud computing; algorithms; visualization review; CiteSpace

1. Introduction

In cloud computing environments, users can share computing resources and data located all over the world, as if obscured by a cloud. The cloud computing scheme has competitive advantages such as on-demand self-service, rapid computing and elasticity, cost reduction, almost unlimited storage, device and location independence, and environmental friendliness [1–4], so it has received extensive attention from both academia and industry in recent years.

Cloud computing has prevailed in the literature for about 10 years, and plenty of related works have been presented. To summarize the extant studies and find potential research directions, some review works related to cloud computing have been consecutively reported [5–15]. Readers can refer to the review works in Section 2.2 for details.

In the literature, various aspects on cloud computing have been reported, such as service availability [5,6,16], computing schemes and algorithms [17,18], data security [10–12,19], and energy management [9,20]. Among these aspects, advanced algorithms are the key to conducting the computing across various clouds, which have drawn special attention. Thus, our research goal in this work is to make a specific visualization review of the studies related to cloud computing algorithms.

The above-mentioned review works have summarized extant studies related to cloud computing and observed enlightening insights to help interested researchers, but few analyze these studies based on the citations. The times cited of one paper represent the influence of the paper, so co-citation analysis is an effective way to measure the impact of studies, journals, countries, institutions, and authors in the development of one specific research area [21]. Meanwhile, it can provide scholars a strong support to identify the intellectual bases and leading edges of a specific field. Motivated by these observations, we conduct a specific visualization review of the extant studies related to cloud computing algorithms using one mainstream co-citation analysis tool—CiteSpace. The main contributions of the work include: (i) The most influential studies and journals on cloud computing algorithms from 2006 to 2015 are identified, providing guidance for interested researchers to make further studies; (ii) From the country and institution visualization, we find the research development path of cloud computing algorithms and identify the current hot regions; (iii) From the author visualization, we detect the most influential authors focusing on cloud computing algorithms; (iv) Thirty keywords are identified to reveal the intellectual bases and focuses of cloud computing algorithms in the literature.

The work is organized as follows. In Section 2, we present a brief summary on extant definitions and review works on cloud computing. In Section 3, we introduce our review methodology, including the data collection process and visualization review tool. In Section 4, we present visualization results from seven aspects: cited references, cited journals, countries, institutions, authors, keywords, and categories. Section 5 concludes the review work.

2. Related Studies

This section introduces some definitions of cloud computing in the literature and presents a brief summarization on extant review works related to cloud computing.

2.1. The Origin and Definition of Cloud Computing

As stated by Prof. Regalado [22], the origin of cloud computing is not clear. In the literature, the term of cloud computing was early used in a 1996 Compaq business plan [22]. The concept suddenly became a hot term after the usage of Amazon and Google in around 2006, and drew extensive attention in the following years, as Figure 1 shows. Obviously, the popularity of cloud computing will last a long time, because both Big Data and Internet of Things will need the support of a matched computing scheme.

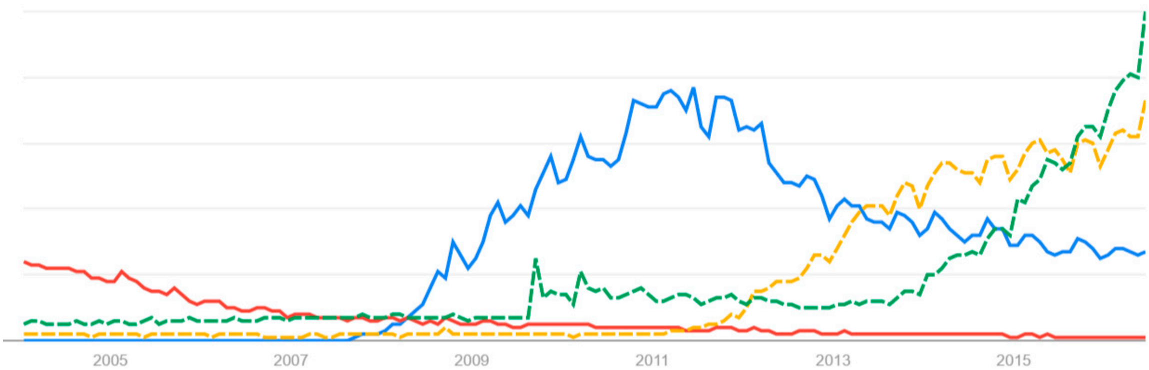

Figure 1. Cloud computing related terms in Google trends (**Red**: Grid Computing; **Blue**: Cloud Computing; **Yellow**: Big Data; **Green**: Internet of Things).

Although the term of cloud computing is recognized just for about 10 years, the literature has reported plenty of versions on its definition. Geelan [23] summarized the definitions of cloud computing from 21 experts. In the summarization, some original definitions of cloud computing were discussed and identified. For example, Cohen defined the cloud computing as "one of those catch all buzz words that tries to encompass a variety of aspects ranging from deployment, load balancing, provisioning, business model and architecture" [23]. Readers can refer to the report for other statements on cloud computing. However, none of them presents a comprehensive and well recognized definition.

Then, Vaquero et al. [24] presented a cloud definition: "Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services)", and analyzed the differences of the cloud approaches from the grid approaches. Wang et al. [25] gave a definition of computing cloud as "a set of network enabled services, providing scalable, QoS guaranteed, normally personalized, inexpensive computing infrastructures on demand, which could be accessed in a simple and pervasive way". Armbrust et al. [8] stated that "cloud computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services". These studies not only formulated definitions of cloud or cloud computing, but also identified the specific characteristics and functionalities of cloud computing, such as user-centric interfaces, on-demand service provisioning, and virtualization.

Later on, Mell and Grance, from the National Institute of Standards and Technology (NIST), formulated a definition of cloud computing [26]: "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction". They also specified the definition using five characteristics, three service models, and four deployment models. This definition has attained wide recognition from both academia and industry, and can be taken as a standard definition of cloud computing.

In Section 2.2, we also summarize other definitions of cloud computing.

2.2. Review Works Related to Cloud Computing

In the literature, plenty of review works on cloud computing have been reported, as Tables 1 and 2 show. These review works can be divided into two categories: general reviews and specific reviews.

In Table 1, we summarize five general review works on cloud computing and highlight the issues needed to be solved in the development of cloud computing. Besides, other meaningful analyses are also presented, as stated below. Vouk [5] discussed the concept of cloud computing from the aspects of service-oriented architecture, components, workflows, virtualization, and users. Armbrust et al. [6,8] analyzed the in-depth drives of the coming of cloud computing, and proposed the top 10 obstacles and opportunities for cloud computing. Kim [7] listed several concerns in the adoption of cloud computing such as availability, security and privacy, interoperability, and compliance. Zhang et al. [9] formulated a layered model of cloud computing and analyzed its business model.

From the presented issues in Table 1, we can get the following observations:

(1) The computing across different clouds is one basic open issue in the development of cloud computing, which is reflected in the terms such as service construction among clouds, cloud metadata management, image portability and formation, collaboration applications, application and data integration across clouds, server consolidation, multimedia transmission and data mining, and virtual machine migration.

(2) Advanced service and software platforms are in urgently needed in order to deal with issues such as cloud computing software platforms, software frameworks, novel cloud architectures, service availability, automated service provisioning, service management, scalable storage, storage technologies, and data management.

(3) Security is one of the key concerns, which is reflected in the presented issues such as security, data lock-in, data confidentiality and auditability, data transfer bottlenecks, bugs in large-scale distributed systems, and data security.

Table 1. General reviews on cloud computing.

	Concepts of Cloud Computing	Issues Needed to Be Solved
Vouk [5]	... embraces cyberinfrastructure, and builds upon... research in virtualization, distributed computing, "grid computing", utility computing, and, more recently, networking, Web and software services.	(1) Service construction among clouds (2) Cloud metadata management (3) Optimization of loading times (4) Image portability and formation (5) Security
Armbrust et al. [6,8]	... refers to both the applications delivered as services over the Internet and the hardware and systems software in the datalefts...	(1) Service availability (2) Data lock-in (3) Data confidentiality and auditability (4) Data transfer bottlenecks (5) Performance unpredictability (6) Scalable storage (7) Bugs in large-scale distributed systems (8) Quick scaling (9) Reputation fate sharing (10) Software licensing
Kim [7]	... being able to access files, data, programs and third party services from a Web browser via the Internet that are hosted by a third party provider.	(1) Cloud computing software platform (2) Collaboration applications (3) Application and data integration across clouds (4) Multimedia transmission and data mining (5) Service management
Zhang et al. [9]	... a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.	(1) Automated service provisioning (2) Virtual machine migration (3) Server consolidation (4) Energy management (5) Traffic management and analysis (6) Data security (7) Software frameworks (8) Storage technologies and data management (9) Novel cloud architectures

Table 2. Specific reviews on the security of cloud computing.

	Concept of Cloud Computing	Security Issues
Hu et al. [10]	... a conglomerate of several different computing technologies and concepts like grid computing, virtualization, autonomic computing, service oriented architecture, peer-to-peer computing, and ubiquitous computing.	(1) Encryption-on-demand (2) Trusted virtual data left implementation (3) Trusted cloud computing (4) Privacy model (5) Intrusion detection strategy (6) Dirichlet reputation model (7) Anonymous bonus point system (8) Network slicing (9) Privacy
Subashini and Kavitha [11]	... a style of computing where massively scalable IT enabled capabilities are delivered "as a service" to external customers using Internet technologies.	(1) Security related to third party resources (2) Application security (3) Data transmission security (4) Data storage security
Khorshed et al. [12]	... is a system, where the resources of a data left are shared using virtualization technology, which also provide elastic, on demand and instant services to its customers and charges customer usage as utility bill.	(1) Abuse and nefarious use of cloud computing (2) Insecure application programming interfaces (3) Malicious insiders (4) Shared technology vulnerabilities (5) Data loss/leakage (6) Account, service and traffic hijacking (7) Unknown risk profile

Extensive related studies on the security of cloud computing have been reported [27]. Several specific review works on the security are summarized in Table 2. In addition, there are some other specific reviews related to cloud computing. Dinh et al. [13] argued that mobile cloud computing (MCC) could bring new types of services and facilities for mobile users to take full advantage of cloud computing, and they also analyzed the architectures, advantages, and applications of MCC. Jula et al. [14] thought that service composition was the key to conduct cloud computing over clouds, and presented a specific review on cloud computing service composition. Hashem et al. [15] reviewed the rise of big data in cloud computing, and discussed the relationship between big data and cloud computing, big data storage systems, and so on.

3. Methodology

3.1. Data Collection

For a review work, it is important to collect and analyze related studies in a proper way. In this work, we collected related studies on cloud computing algorithms in Web of Science (WoS). WoS includes extensive citation databases such as Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Conference Proceedings Citation Index—Science, Conference Proceedings Citation Index—Social Science & Humanities, and recent Emerging Sources Citation Index. Thus, it is a good choice to use WoS to collect related studies and their citation relations.

Specific steps of collecting studies related to cloud computing algorithms in the work are as follows: (1) Open the WoS search website: https://apps.webofknowledge.com/; (2) Select the Web of Science Core Collection and conduct the advanced search with the setting: TS = (Cloud* and Comput* and algorithm*) NOT WC = (Astronomy & Astrophysics or Meteorology & Atmospheric Sciences or Optics or Geoscience or Geography), where TS denotes the topic of the studies (TS consists of Title, Abstract, Keywords, and Keywords Plus which is the result of Thomson Reuters editorial expertise in Science [28]), WC denotes Web of Science Category [28] (Here we use WC to exclude the records dealing with the real cloud using some algorithms), and * denotes any other texts; (3) The language, document type, and timespan are set as "English", "article", and "2006–2015", respectively. Using the above setting, we retrieved 1865 records and then manually deleted 32 records which are not related to cloud computing, and downloaded the full records and cited references of the remaining 1833 studies. The full records and cited references were saved in the text format for the requirement of the visualization tool—CiteSpace.

3.2. Visualization Tool

Author co-citation analysis (ACA)—originally introduced by White and Griffith [29]—is an effective method to identify and visualize the intellectual structure of one specific research field [30]. Citation frequency is the main basis of existing ACA tools. By analyzing the co-citations among related studies, the intellectual bases, research fronts, and development trends in the field can be identified. Thus, the visualization of co-citations is helpful for researchers to conduct further studies.

CiteSpace, one of the mainstream ACA tools, was initially developed by Chaomei Chen in 2004 [31,32], and is lastly updated by Chen's research team. The latest version is CiteSpace 5.0.R1 SE, available at http://cluster.ischool.drexel.edu/~cchen/citespace/download/ [33]. The principle of CiteSpace is based on co-citation analysis theory and pathfinder network scaling algorithms, so CiteSpace can detect the development paths and trends of one specific subject. Intellectual turning points play key roles in the development and formulation of scientific domains. By identifying these turning points, CiteSpace can find the development path of one subject, which is beneficial for researchers to understand the subject and catch research fronts.

In this work, we use the latest version of CiteSpace, that is, 5.0.R1 SE. Before installing CiteSpace, we equipped our computer with Java Runtime (JRE) 8.0. With 1833 full records on cloud computing algorithms, we present the visualization results in the following section.

4. Visualization Results

CiteSpace includes 11 visualization functions by analyzing different terms such as author, institution, country, term, keyword, category, cited reference, cited author, and cited journal. In this work, we make co-citation visualization, cooperation visualization, and co-occurrence visualization by analyzing cited reference, cited journal, country, institution, author, keyword, and category.

4.1. Cited Reference Visualization

Cited reference is one key aspect of co-citation visualization, which can identify the most influential studies of one specific research domain. The time slicing is from 2006 to 2015, and the year per slice is set as 1. The node type is selected as cited reference, and the strength among links is measured by the Cosine metric:

$$\text{Cosine}(x,y) = \frac{C_x C_y}{||C_x||\,||C_y||}$$

where $C_x C_y$ represents the co-citation counts between paper x and paper y, and $||C_x||$ and $||C_y||$ represent the times cited of paper x and paper y, respectively.

Filtering strategy is another key setting for the co-citation analysis. In this work, we use the most common strategy, that is, Top N strategy. This filtering strategy selects out top N records in terms of citation frequencies during the slice. Considering the number of our records, we take N as 100, that is, 100 records are selected in each slice. After 2104 iterations, the cited reference visualization is obtained, as Figure 2 shows. In Figure 2, the seven most influential studies on cloud computing or their algorithms from 2006 to 2015 are identified (Here the threshold of times cited is set as 30).

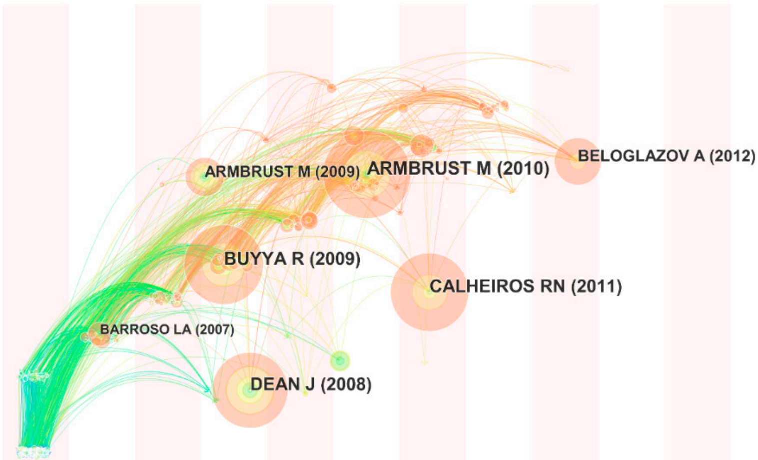

Figure 2. Cited reference visualization.

The study entitled "A View of Cloud Computing" by M. Armbrust et al. in 2010 [8] has the most citation frequency. This study was published on a top journal in the computer area, that is, Communications of the ACM. In this review, authors argued that confusion existed in the popularity of cloud computing, and proposed 10 obstacles and opportunities for cloud computing. As we can see, another most influential work is also from Armbrust's team at University of California, Berkeley, that is, "Above the Clouds: A Berkeley View of Cloud Computing" [6]. Besides listing the obstacles and opportunities, this report also analyzed the reasons of the emergence of cloud computing and presented observations about cloud computing economic models. As we can see, these two works belong to qualitative analysis studies with no models or algorithms, but really present enlightening insights for further studies. Although these qualitative works did not present specific algorithms, they indeed have influenced studies related to cloud computing greatly, at least observed by citation relationships among the 1833 records.

Among the most influential studies, two specific algorithms or tools on cloud computing were observed, that is, MapReduce and CloudSim. Dean and Ghemawat [17], from Google Inc. (Mountain View, CA, USA), presented the work "MapReduce: Simplified Data Processing on Large Clusters" which introduced the programming model, basic implementation modules, and extensive functions of MapReduce. Calheiros et al. [18], from the Cloud Computing and Distributed Systems (CLOUDS) Laboratory at the University of Melbourne, reported a toolkit for the modeling and simulation of cloud computing environments—CloudSim, from the architecture, design, and implementation. Actually, the same research team chaired by Prof. Buyya had presented a toolkit for modeling and simulating of grid computing environments, namely, GridSim [34]. Another two identified influential works are also from Prof. Buyya's team, that is, "Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the Fifth Utility" [35] and "Energy-Aware Resource Allocation Heuristics for Efficient Management of Data Centers for Cloud Computing" [36]. Both studies were published in the journal—Future Generation Computer Systems.

As we can see, the above most influential works include two overview papers besides four specific contributions to cloud computing algorithms. In order to find out why the overview papers have a profound influence, we check citied articles of the overview paper "A View of Cloud Computing", and find out two main underlying reasons. One is that scholars need a strong and significant motivation to develop their cloud computing algorithms. For example, Calheiros et al. [18] cited the overview paper in say that "the importance of these services was highlighted in a recent report from the University of Berkeley as: cloud computing, the long-held dream of computing as a utility has the potential to transform a large part of the IT industry, making software even more attractive as a service". The other is that interested researchers want to find out what on earth "cloud computing" is before presenting suitable algorithms. Armbrust et al. [8] presented a definition of cloud computing with clear descriptions of the differences between cloud computing and other similar terms such as software as a service (SaaS) and grid computing. In addition, another important reason why so many researchers cite the Armbrust et al.' s overview paper is that the cited work is published in one top journal (i.e., Communications of the ACM) by one famous scholar from one prestigious institution (i.e., the University of Berkeley).

The last identified influential work is reported by Barroso and Hölzle from Google Inc. [37]. Barroso and Hölzle argued that energy management should be a key issue for servers and data center operations, focusing on the reduction of both economic costs and environmental impacts. Interestingly, no cloud or grid related terms were mentioned in the work. After checking some studies citing the work, the term of energy proportionality proposed in the work takes the attention of the popularity, because energy saving is recognized as one important performance indicator of cloud computing algorithms. The cloud computing scheme deals with large amounts of data, management and switching of communications, producing incomprehensibly huge energy consumption. Thus, in order to implement the energy proportionality, wide attention has been paid to develop energy-saving algorithms to improve the energy usage profile of every component in cloud computing systems.

4.2. Cited Journal Visualization

Cited journal is another aspect of co-citation visualization. Using the same setting with those in Section 4.1, we can get the cited journal visualization by selecting the node type as cited journal. After 736 iterations, the cited journal visualization was obtained, as Figure 3 and Table 3 show (Here the threshold of citation frequency is set as 150).

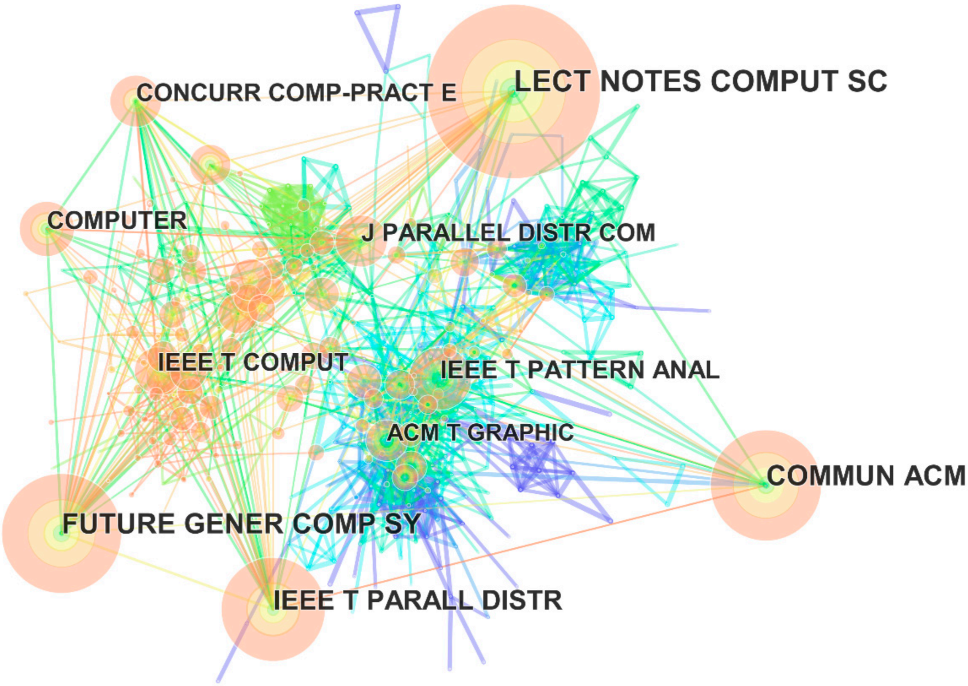

Figure 3. Cited journal visualization.

Table 3. The identified journals.

Full Journal Names	Abbreviated Journal Names	Citation Frequency	Publishers
Lecture Notes in Computer Science	LECT NOTES COMPUT SC	573	Springer
Future Generation Computer Systems	FUTURE GENER COMP SY	391	Elsevier
Communications of the ACM	COMMUN ACM	377	Association for Computing Machinery
IEEE Transactions on Parallel and Distributed Systems	IEEE T PARALL DISTR	337	IEEE
IEEE Transactions on Pattern Analysis and Machine Intelligence	IEEE T PATTERN ANAL	228	IEEE
Computer	COMPUTER	182	IEEE
Journal of Parallel and Distributed Computing	J PARALLEL DISTR COM	177	Elsevier
Concurrency and Computation: Practice and Experience	CONCURR COMP-PRACT E	175	John Wiley & Sons
IEEE Transactions on Computers	IEEE T COMPUT	169	IEEE
ACM Transactions on Graphics	ACM T GRAPHIC	156	Association for Computing Machinery

As Table 3 shows, the top five journals are Lecture Notes in Computer Science, Future Generation Computer Systems, Communications of the ACM, IEEE Transactions on Parallel and Distributed Systems, and IEEE Transactions on Pattern Analysis and Machine Intelligence. Lecture Notes in Computer Science ranks top on the list, mostly because it is a series book focusing on frequently publishing new developments in computer science and information technology research. Two of the above five journals, that is, Future Generation Computer Systems and Communications of the ACM, reported the most influential studies identified in Section 4.1. This reveals the important role of key studies in improving the position of academic journals. Note that several meteorological journals are identified mainly because some studies on dealing with real clouds are included in the records.

In the aspect of publishers, IEEE contributes the highest citations, with a total of 916 times recorded in Table 3, which shows the dominant position of IEEE in the area of cloud computing. Springer, Elsevier, and Association for Computing Machinery follow the list, with comparative impacts. From the journal visualization, publishers can recognize the contribution of their journals to the field of cloud computing, and make proper publishing policies to enhance their positions in the development of cloud computing research. For interested researchers, the results in Figure 3 and Table 3 are helpful for them to catch mainstream academic journals on cloud computing algorithms. To sum up, the visualization of cited journals provides supports for both publishers and researchers to make their research decisions, which is finally favorable to the academic development and practical application of cloud computing algorithms.

4.3. Country and Institution Visualization

Country is one important aspect of cooperation visualization. By choosing the node type as country, we got the country visualization of the 1833 studies on cloud computing algorithms after 41 iterations, as Figure 4 shows. In Figure 4, the threshold of citation frequency is 30, that is, all the countries with more than 30 citations are labeled out. We also label out the hot periods of these countries using different colors, as the top bar in Figure 4 shows.

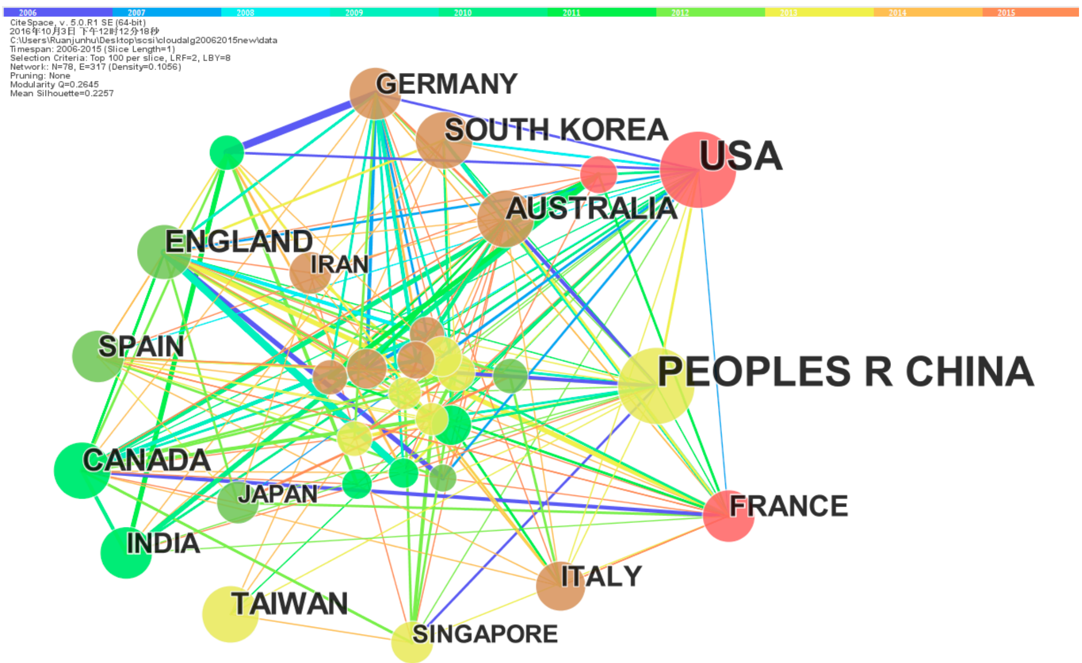

Figure 4. Country visualization.

Based on the citations, China, USA, Taiwan, Australia, Canada, South Korea, England, France and then other countries or regions make decreasing contributions to the intellectual bases of cloud computing algorithms. Meanwhile, these countries have different hot periods. Canada, Japan, England, India, and Spain focused on the cloud computing algorithms at an earlier stage, mainly before 2012; China, Taiwan, and Singapore were the next followers, mainly during from 2013; Then, the research attention was extended to more countries including Germany, South Korea, Australia, Italy, France, and USA. The contribution of one country on the research has a positive relation with the supporting funds. As Table 4 shows, the correlation coefficient between identified citation frequency and gross expenditure on R & D is about 0.93. For simplicity, we just use the 2014 GERD summarized by the Industrial Research Institute [38].

Table 4. The correlation of identified citation frequency and gross expenditure on R & D (GERD).

Countries or Regions	Citation Frequency	2014 GERD (Billion US $)
China	591	485.39
USA	503	343.78
Taiwan	107	24.02
Australia	103	24.75
Canada	102	30
South Korea	99	64.3
England	91	44.07
France	79	58.21
Spain	75	19.18
Germany	72	103.2
India	71	61.85
Italy	63	24.79
Japan	35	163.44
Iran	35	10.79
Singapore	35	11.8
Correlation coefficient	**0.93**	

From the country visualization, we can find the development path of cloud computing algorithms and identify the current hot areas. Meanwhile, the research connections among countries can also be detected. For example, Canada and France have close academic cooperation on cloud computing algorithms. The formulation of these academic connections is mainly due to the exchange and cooperation among related institutions.

Behind the countries, research and application institutions have a specific impact on the development of cloud computing algorithms. In order to reveal which institutions make influential contributions to cloud computing algorithms, we also conducted the institution visualization. By choosing the node type as institution, we got the institution visualization of 1833 studies after 2321 iterations, as Figure 5 and Table 5 show (The threshold of citation frequency is 15).

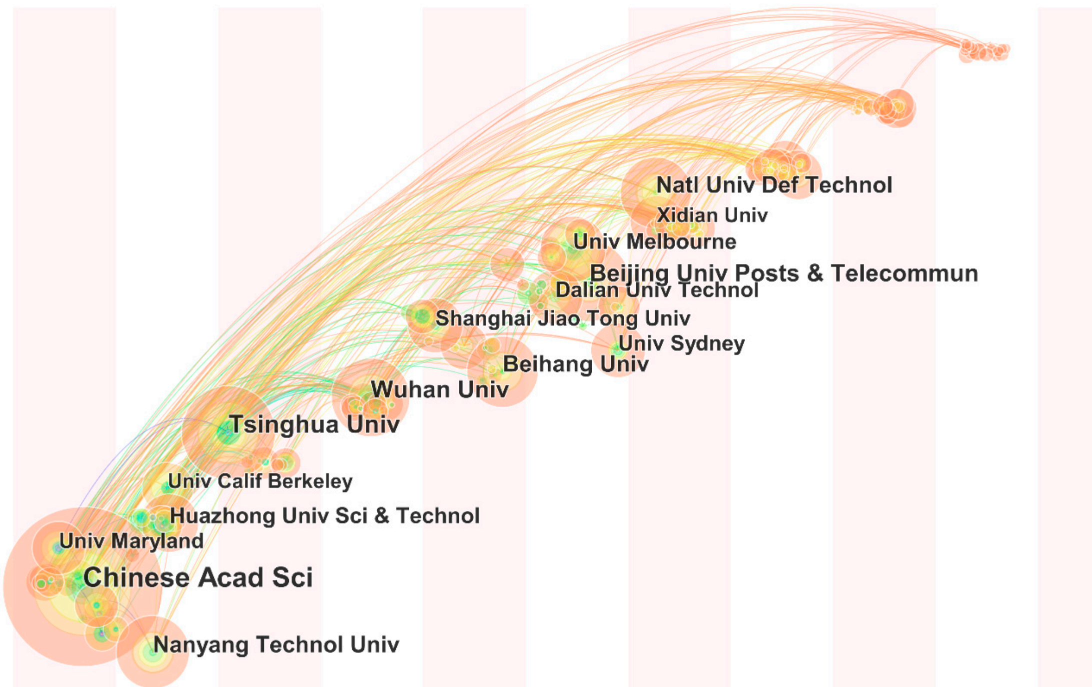

Figure 5. Institution visualization.

Table 5. The identified institutions.

Full Institution Names	Abbreviated Institution Names	Citation Frequency	Countries
Chinese Academy of Sciences	Chinese Acad Sci	49	China
Tsinghua University	Tsinghua Univ	29	China
Wuhan University	Wuhan Univ	25	China
Beijing University of Posts and Telecommunications	Beijing Univ Posts & Telecommun	25	China
Nanyang Technological University	Nanyang Technol Univ	23	Singapore
Beihang University	Beihang Univ	23	China
National University of Defense Technology	Natl Univ Def Technol	22	China
University of Melbourne	Univ Melbourne	20	Australia
Huazhong University of Science & Technology	Huazhong Univ Sci & Technol	19	China
University of Maryland	Univ Maryland	17	USA
The University of Sydney	Univ Sydney	17	Australia
Dalian University of Technology	Dalian Univ Technol	17	China
Shanghai Jiao Tong University	Shanghai Jiao Tong Univ	17	China
University of California, Berkeley	Univ Calif Berkeley	16	USA
Xidian University	Xidian Univ	16	China
University of Electronic Science and Technology of China	Univ Elect Sci & Technol China	15	China
Zhejiang University	Zhejiang Univ	15	China
Nanjing University	Nanjing Univ	15	China
South China University of Technology	S China Univ Technol	15	China

As Figure 5 and Table 5 show, the most influential institutions focusing on cloud computing algorithms are mainly from China, USA, Australia, and Singapore. The top five Chinese institutions are Chinese Academy of Sciences, Tsinghua University, Wuhan University, Beijing University of Posts and Telecommunications, and Beihang University. The top American institutions are University of Maryland and University of California at Berkeley. The top Australian institutions are University of Melbourne and The University of Sydney.

As we can see, the institution visualization result is consistent with the cited reference visualization and country visualization. Two institutions which produce five of the most influential studies, that is, University of California at Berkeley and University of Melbourne, are detected.

4.4. Author Visualization

By selecting the node type as author, we got the author visualization of the 1833 studies, as Figure 6 shows (The citation frequency threshold is set as 50). Based on the author visualization, we can find the most influential researchers on cloud computing algorithms.

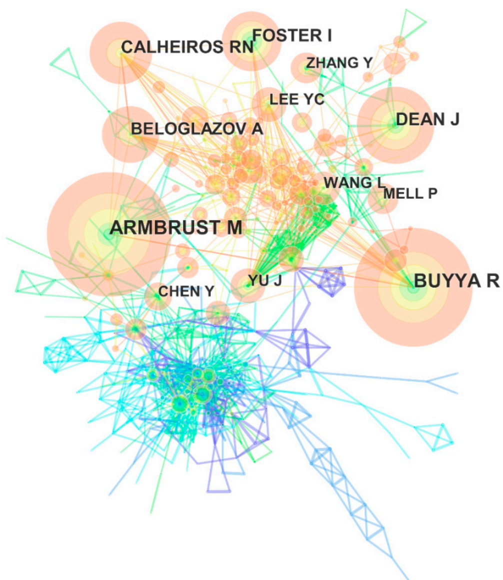

Figure 6. Author visualization.

Michael Armbrust ranks on the top of the list. He is from University of California, Berkeley. As we observed above, Prof. Armbrust's two overview papers [6,8] have been highly cited, so here he is identified. However, he does not present specific contributions on cloud computing algorithms. Prof. Rajkumar Buyya, from University of Melbourne, ranks the second. Actually, several of Prof. Buyya's cooperators are also identified as the most influential scholars on cloud computing algorithms, such as Rodrigo N. Calheiros, Anton Beloglazov, and Young Choon Lee, respectively working at University of Melbourne, IBM Research-Australia, and Macquarie University. Jeff Dean, from Google Inc., ranks the third. As stated in Section 4.1, all the above scholars have been identified as contributors of the most influential studies. This consistency reveals the impact of key studies on the position of one scholar.

Ian T. Foster, working at University of Chicago, is another influential scholar focusing on cloud computing algorithms. Prof. Foster had focused on grid computing before he got involved in cloud computing. The top record on cloud computing from Prof. Foster is "Cloud Computing and Grid Computing 360-Degree Compared" [39]. This work presented a side-by-side comparison between cloud computing and grid computing from various angles such as business model, architecture, resource management, programming model, application model, and security model.

Peter Mell, who is from the National Institute of Standards and Technology and gave a standard definition of cloud computing together with T. Grance [26], is also identified on the list. As mentioned above, the visualization results by CiteSpace are based on the citation relationships among the used records. Peter Mell is detected out because enough studies on cloud computing algorithms have cited his work, although the NIST definition by P. Mell and T. Grance does not focus on algorithms.

Four Chinese scholars formulate another group of influential researchers on cloud computing algorithms: Jian Yu, Lizhe Wang, Yinqian Zhang, and Yu Chen, who are now working at Auckland University of Technology, The Chinese Academy of Sciences, Ohio State University, and Binghamton University, respectively.

4.5. Keyword and Category Visualization

The keyword term can be used to cluster intellectual bases and find research directions. After 2521 iterations, 29 keywords were identified (The citation frequency threshold is set as 30), as Figure 7 shows.

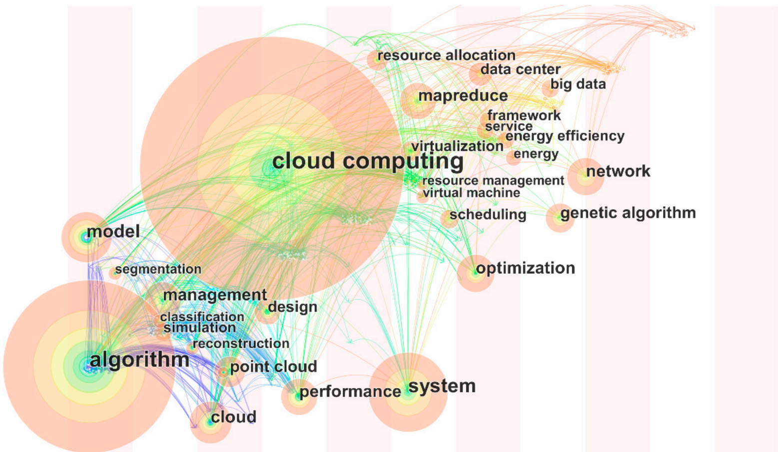

Figure 7. Keyword visualization.

These 29 keywords can be divided into four categories:

(1) Cloud computing and its characteristics

The identified keywords in this category include cloud computing, cloud, virtual machine, point cloud, virtualization, and big data. Obviously, this category takes the biggest proportion in the identified keywords, since almost every study in the records will place the "cloud" or "cloud computing" into the keywords list. It is worth noting that only the virtualization and big data among the characteristics of cloud computing are identified. This, to some extent, shows the most remarkable characteristics of cloud computing.

(2) Cloud algorithms

The identified keywords in this category include algorithm, MapReduce, and Genetic Algorithm. The algorithm keyword ranks the second on the list, which reveals that designing suitable algorithms is the key to implementing cloud computing. Among the exiting algorithms, two specific algorithms—that is, MapReduce and Genetic Algorithm—are detected. This observation can provide a guidance for researchers focusing on the development and application of cloud computing algorithms.

(3) Cloud issues

The cloud issues refer to the considered problems in the literature related to cloud computing. General keywords in this category include model, system, network, framework, management, design, optimization, classification, simulation, computation, and service. Specific keywords include resource allocation, resource management, scheduling, segmentation, reconstruction, and data center. These keywords can provide some research directions for interested scholars on cloud computing.

(4) Cloud computing performance

The keywords in this category include performance, energy, and energy efficiency. As stated by Khanghahi and Ravanmehr [40], nowadays the term "performance" includes more extensive concepts such as reliability, energy efficiency, and scalability. Khanghahi and Ravanmehr [40] also listed some specific cloud computing performance indicators such as average response time per unit time, network capacity per second, average waiting time per unit time, and percentage of CPU utilization. Nevertheless, only the energy efficiency is detected in our work, which reveals that more and more attention is paid to energy saving and environmental friendliness in the research of cloud computing algorithms, as stated by Tseng et al. [20].

In addition, we also got the category visualization to show the contribution of various subjects, as Figure 8 shows (The citation frequency threshold is 60). The top five subjects are Computer Science, Engineering, Computer Science-Theory & Methods, Engineering-Electrical & Electronic, and Computer Science-Software Engineering.

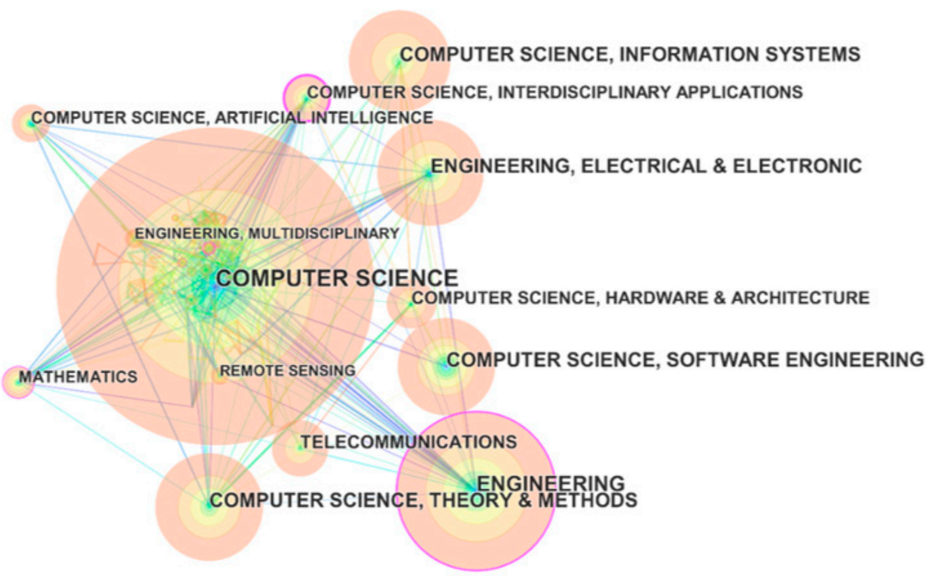

Figure 8. Category visualization.

5. Conclusions

In this work, we collected 1833 records on cloud computing algorithms from Web of Science, and conducted a visualization review on these studies using one mainstream co-citation analysis tool—CiteSpace. From the visualization results, we got some enlightening observations for supporting interested researchers to make further studies on cloud computing:

(1) The most influential studies as well as two specific algorithms (namely, MapReduce and CloudSim) on cloud computing are identified by the work, which are contributed by Armbrust's team at University of California, Berkeley; Buyya's team at the University of Melbourne; and Google Inc. Besides some specific contributions, overview papers can also have an underlying profound influence on the motivation and significance of the studies related to cloud computing.

(2) The top five journals in terms of the contribution to cloud computing algorithms are Lecture Notes in Computer Science, Future Generation Computer Systems, Communications of the ACM, IEEE Transactions on Parallel and Distributed Systems, and IEEE Transactions on Pattern Analysis and Machine Intelligence. Two of the above five journals, that is, Future Generation Computer Systems and Communications of the ACM, reported the most influential studies. Thus, the key studies with high citations take an important role in improving the position of academic journals.

(3) China, USA, Taiwan, Australia, Canada, South Korea, England, France, and then other countries or regions make decreasing contributions to the intellectual bases of cloud computing algorithms. However, these countries have different hot periods. The contribution of one country to the research in one specific field has a positive relationship with its amount of supporting funds, and publications with high citations in one institution can help to improve the position of the institution.

(4) Author visualization finds the most influential authors on cloud computing, including Prof. Buyya from the University of Melbourne as well as his team members; Prof. Armbrust from University of California, Berkeley; Jeff Dean from the Google Inc.; Prof. Foster from University of Chicago; Peter Mell from The National Institute of Standards and Technology; and four Chinese scholars: Jian Yu, Lizhe Wang, Yinqian Zhang, and Yu Chen.

(5) Twenty-nine keywords are identified to reveal the intellectual bases and focuses of cloud computing algorithms in the literature, and the mainstream subjects focusing on cloud computing algorithms are also detected. The emerging intellectual foci from the literature related to cloud computing consist of four aspects: cloud computing and its characteristics, cloud algorithms, cloud issues, and cloud computing performance.

Although the view has observed visual insights related to cloud computing algorithms in last decade, some deficiencies—which are difficult to avoid—still exist. For example, we cannot make sure all the valid records have been collected, and some intrinsic causes still need further analysis.

Acknowledgments: The authors gratefully acknowledge the editor and reviewers for their constructive comments which greatly improved the work. This work was supported by National Natural Science Foundation of China (No. 71471158, 71503199), the Hong Kong Scholars Program Mainland-Hong Kong Joint Postdoctoral Fellows Program (No. G-YZ87), the Natural Science Basic Research Project in Shaanxi Province (No. 2016JQ7005), Doctoral Scientific Research Foundation of Northwest A&F University (No. 2452015325), and China Ministry of Education Social Sciences and Humanities Research Youth Fund Project titled "Delivery Optimization of Maturity-Based Fruit B2C Ecommerce-Taking Shaanxi Kiwifruit as the Example" (No. 16YJC630102).

Author Contributions: Junhu Ruan and Felix T. S. Chan designed the work; Junhu Ruan and Fangwei Zhu collected and analyzed the data; Xuping Wang and Jing Yang performed the experiments; Felix T. S. Chan and Fangwei Zhu contributed introduction/related studies; Junhu Ruan and Xuping Wang contributed methodology/visualization results; Junhu Ruan wrote the paper; Felix T. S. Chan and Jing Yang made the read proof.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sookhak, M.; Gani, A.; Talebian, H.; Akhunzada, A.; Khan, S.U.; Buyya, R.; Zomaya, A.Y. Remote Data Auditing in Cloud Computing Environments: A Survey, Taxonomy, and Open Issues. *ACM Comput. Surv.* **2015**, *47*, 1–34. [CrossRef]
2. Tan, X.; Di, L.; Deng, M.; Fu, J.; Shao, G.; Gao, M.; Sun, Z.; Ye, X.; Sha, Z.; Jin, B. Building an Elastic Parallel OGC Web Processing Service on a Cloud-Based Cluster: A Case Study of Remote Sensing Data Processing Service. *Sustainability* **2015**, *7*, 14245–14258. [CrossRef]
3. Botta, A.; de Donato, W.; Persico, V.; Pescapé, A. Integration of Cloud Computing and Internet of Things: A Survey. *Future Gener. Comput. Syst.* **2016**, *56*, 684–700. [CrossRef]
4. Zhou, L.; Chen, N.; Chen, Z. A Cloud Computing-Enabled Spatio-Temporal Cyber-Physical Information Infrastructure for Efficient Soil Moisture Monitoring. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 81. [CrossRef]
5. Vouk, M.A. Cloud Computing—Issues, Research and Implementations. *J. Comput. Inf. Technol.* **2008**, *16*, 235–246. [CrossRef]
6. Armbrust, M.; Fox, A.; Griffith, R.; Joseph, A.D.; Katz, R.H.; Konwinski, A.; Lee, G.; Patterson, D.A.; Rabkin, A.; Stoica, I.; et al. Above the Clouds: A Berkeley View of Cloud Computing. 2009. Available online: https://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf (accessed on 11 May 2016).
7. Kim, W. Cloud Computing: Today and Tomorrow. *J. Object Technol.* **2009**, *8*, 65–72. [CrossRef]
8. Armbrust, M.; Fox, A.; Griffith, R.; Joseph, A.D.; Katz, R.; Konwinski, A.; Lee, G.; Patterson, D.; Rabkin, A.; Stoica, I.; et al. A View of Cloud Computing. *Commun. ACM* **2010**, *53*, 50–58. [CrossRef]
9. Zhang, Q.; Cheng, L.; Boutaba, R. Cloud Computing: State-of-the-Art and Research Challenges. *J. Internet Serv. Appl.* **2010**, *1*, 7–18. [CrossRef]
10. Hu, F.; Qiu, M.; Li, J.; Grant, T.; Tylor, D.; McCaleb, S.; Butler, L.; Hamner, R. A Review on Cloud Computing: Design Challenges in Architecture and Security. *J. Comput. Inf. Technol.* **2011**, *19*, 25–55. [CrossRef]
11. Subashini, S.; Kavitha, V. A Survey on Security Issues in Service Delivery Models of Cloud Computing. *J. Netw. Comput. Appl.* **2011**, *34*, 1–11. [CrossRef]
12. Khorshed, M.T.; Shawkat Ali, A.B.M.; Wasimi, S.A. A Survey on Gaps, Threat Remediation Challenges and Some Thoughts for Proactive Attack Detection in Cloud Computing. *Future Gener. Comput. Syst.* **2012**, *28*, 833–851. [CrossRef]
13. Dinh, H.T.; Lee, C.; Niyato, D.; Wang, P. A Survey of Mobile Cloud Computing: Architecture, Applications, and Approaches. *Wirel. Commun. Mob. Comput.* **2013**, *13*, 1587–1611. [CrossRef]
14. Jula, A.; Sundararajan, E.; Othman, Z. Cloud Computing Service Composition: A Systematic Literature Review. *Expert Syst. Appl.* **2014**, *41*, 3809–3824. [CrossRef]
15. Hashem, A.T.; Yaqoob, I.; Anuar, N.B.; Mokhtar, S.; Gani, A.; Khan, S.U. The Rise of Big Data on Cloud Computing: Review and Open Research Issues. *Inf. Syst.* **2015**, *47*, 98–115. [CrossRef]
16. Liu, N.; Chen, Q.; Lu, X.; Liu, J.; Zhang, J. A Charging Strategy for PV-Based Battery, Switch Stations Considering Service Availability, and Self-Consumption of PV Energy. *IEEE Trans. Ind. Electron.* **2015**, *62*, 4878–4889. [CrossRef]
17. Dean, J.; Ghemawat, S. MapReduce: Simplified Data Processing on Large Clusters. *Commun. ACM* **2008**, *51*, 107–113. [CrossRef]
18. Calheiros, R.N.; Ranjan, R.; Beloglazov, A.; De Rose, C.A.F.; Buyya, R. CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms. *Softw. Pract. Exp.* **2011**, *41*, 23–50. [CrossRef]
19. Ali, M.; Khan, S.U.; Vasilakos, A.V. Security in Cloud Computing: Opportunities and Challenges. *Inf. Sci.* **2015**, *305*, 357–383. [CrossRef]
20. Tseng, Y.-C.; Lee, D.-S.; Lin, C.-F.; Chang, C.-Y. The Energy Savings and Environmental Benefits for Small and Medium Enterprises by Cloud Energy Management System. *Sustainability* **2016**, *8*, 531. [CrossRef]
21. Wang, N.; Liang, H.; Jia, Y.; Ge, S.; Xue, Y.; Wang, Z. Cloud Computing Research in the IS Discipline: A Citation/Co-citation Analysis. *Decis. Support Syst.* **2016**, *86*, 35–47. [CrossRef]
22. Regalado, A. Who Coined Cloud Computing. 2011. Available online: http://www.technologyreview.com.br/printer_friendly_article.aspx?id=38987 (accessed on 12 May 2016).
23. Geelan. Twenty One Experts Define Cloud Computing. 2009. Available online: http://www.virtualizationconference.com/node/612375 (accessed on 12 May 2016).

24. Vaquero, M.; Rodero-Merino, L.; Caceres, J.; Lindner, M. A Break in the Clouds: Towards a Cloud Definition. *ACM SIGCOMM Comput. Commun. Rev.* **2009**, *39*, 50–55. [CrossRef]
25. Wang, L.; von Laszewski, G.; Younge, A.; He, X.; Kunze, M.; Tao, J.; Fu, C. Cloud Computing: A Perspective Study. *New Gener. Comput.* **2010**, *28*, 137–146. [CrossRef]
26. Mell, P.; Grance, T. The NIST Definition of Cloud Computing. 2011; Available online: http://csrc.nist.gov/publications/nistpubs/800-145SP800-145.pdf (accessed on 13 May 2016).
27. Park, S.; Hwang, M.; Lee, S.; Park, Y.B. A Generic Software Development Process Refined from Best Practices for Cloud Computing. *Sustainability* **2015**, *7*, 5321–5344. [CrossRef]
28. Thomson Reuters. Web of Science Advanced Search. Available online: http://apps.webofknowledge.com/WOS_AdvancedSearch_input.do?SID=Z1JGkfrSAEhxCHs8hIf&product=WOS&search_mode=AdvancedSearch (accessed on 12 September 2016).
29. White, H.D.; Griffith, B.C. Author Cocitation: A Literature Measure of Intellectual Structure. *J. Am. Soc. Inf. Sci.* **1981**, *32*, 163–171. [CrossRef]
30. Jeong, Y.K.; Song, M.; Ding, Y. Content-Based Author Co-citation Analysis. *J. Inf.* **2014**, *8*, 197–211. [CrossRef]
31. Chen, C.M. Searching for Intellectual Turning Points: Progressive Knowledge Domain Visualization. *Proc. Natl. Acad. Sci. USA* **2004**, *101* (Suppl. 1), 5303–5310. [CrossRef] [PubMed]
32. Chen, C.M. CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. *J. Am. Soc. Inf. Sci. Technol.* **2006**, *57*, 359–377. [CrossRef]
33. Chen, C.M. CiteSpace: Visualizing Patterns and Trends in Scientific Literature. Available online: http://cluster.ischool.drexel.edu/~cchen/citespace/download/ (accessed on 2 October 2016).
34. Buyya, R.; Murshed, M. GridSim: A Toolkit for the Modeling and Simulation of Distributed Resource Management and Scheduling for Grid Computing. *Concurr. Comput. Pract. Exp.* **2002**, *14*, 1175–1220. [CrossRef]
35. Buyya, R.; Yeo, C.S.; Venugopal, S.; Broberg, J.; Brandic, I. Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility. *Future Gener. Comput. Syst.* **2009**, *25*, 599–616. [CrossRef]
36. Beloglazov, A.; Abawajy, J.; Buyya, R. Energy-Aware Resource Allocation Heuristics for Efficient Management of Data Centers for Cloud Computing. *Future Gener. Comput. Syst.* **2012**, *28*, 755–768. [CrossRef]
37. Barroso, A.; Hölzle, U. The Case for Energy-Proportional Computing. *Computer* **2007**, *40*, 33–37. [CrossRef]
38. The Industrial Research Institute. 2016 R&D Trends Forecast. Available online: https://www.iriweb.org/sites/default/files/2016GlobalR%26DFundingForecast_2.pdf (accessed on 15 May 2016).
39. Foster, I.; Zhao, Y.; Raicu, I.; Lu, S. Cloud Computing and Grid Computing 360-Degree Compared. In Proceedings of the 2008 Grid Computing Environments Workshop, Austin, TX, USA, 16 November 2008.
40. Khanghahi, N.; Ravanmehr, R. Cloud Computing Performance Evaluation: Issues and Challenges. *Int. J. Cloud Comput. Serv. Arch.* **2013**, *3*, 29–41. [CrossRef]