Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 19 (2013) 586 – 593

Procedia
Computer Science

The 4th International Conference on Ambient Systems, Networks and Technologies (ANT 2013)

EXPLORING THE E-SCIENCE KNOWLEDGE BASE THROUGH CO-CITATION ANALYSIS

Navonil Mustafee[a,*], Nik Bessis[b], Simon J. E. Taylor[c], Stelios Sotiriadis[b]

[a]Centre for Innovation and Service Research, Exeter Business School, University of Exeter, Exeter EX4 4PU
[b]School of Computing and Mathematics, University of Derby, Derby DE22 1GB
[c]School of Information Systems, Computing and Mathematics, Brunel University, Uxbridge UB8 3PH

Abstract

E-Science is the "science of this age"; it is realized through collaborative scientific enquiry which requires utilization of non-trivial amounts of computing resources and massive data sets. In this paper we explore the e-Science knowledge base through co-citation analysis of extant literature. Our objective is to use the knowledge domain visualization software CiteSpace to identifying the turning point articles and authors. In other words, our analysis is not solely based on tabulating the frequency of co-cited articles and authors, but the identification of landmark articles and authors irrespective of their co-citation count. The dataset for this analysis is downloaded from the ISI Web of Science and includes approx. 1000 articles. It is expected that this paper will be an important source of reference for academics and researchers working in the area of e-Science and its three technology enablers - grid computing, desktop grids and cloud computing.

Selection and peer-review under responsibility of Elhadi M. Shakshuki

Keywords: e-Science, co-citation analysis, grid computing, desktop grid computing, cloud computing

1. Introduction

E-Science can be defined as science that necessitates the utilization of non-trivial amounts of computing resources and massive data sets to perform scientific enquiry; science that requires access to remote scientific instruments and distributed software repositories; science that generates data that may demand analysis from experts belonging to multiple organizations and specialists in different knowledge domains – such science is usually carried out in highly distributed environments by exploiting advance hardware and software technologies [1]. Grid computing has the potential to provide users on-demand access to large amounts of computing power [2] and this focus on large-scale computing makes grid computing an enabling technology for e-Science [3]. There are numerous examples of public-funded e-Science projects in various disciplines that rely on grid computing, for example, the earth system grid (ESG) project that focuses on climatology [4], the Large Hadron Collider (LHC) project devoted to particle physics [5], the biomedical informatics research network project in medicine [6] and the network

Selection and peer-review under responsibility of Elhadi M. Shakshuki
doi:10.1016/j.procs.2013.06.078

for earthquake engineering simulation project (NEESGrid) in earthquake engineering [7]. Refer to [8] for other examples. Although it can be argued that the majority of e-Science applications in use today are predominantly executed over dedicated and high-performance cluster (referred to as cluster-based grid computing), it is also true that two other forms of large scale computing, namely, desktop grid computing and cloud computing, are increasingly being considered as viable alternatives for executing e-Science applications. Furthermore, there is research which specifically focuses on interoperability amongst the aforementioned technologies thereby making it possible for e-Science applications to be transparently scheduled over grid, desktop grid and cloud resources. Notable examples include utilizing EGEE (Enabling Grids for E-sciencE – now superseded by the European Grid Infrastructure; the largest grid infrastructure in the world) for desktop grids [9] and gluing grids and clouds together [10, 11].

In this paper we explore the e-Science knowledge base through co-citation analysis of literature. The purpose of this analysis is to identify the turning point authors and articles, irrespective of their of co-citation frequency. The identification of the significant authors and articles (also referred to as "turning point" authors and articles) is made possible through use of knowledge domain visualization software called CiteSpace [12]. As has been stated previously, grid computing, desktop grids and cloud computing are the technology enables of e-Science and as such our study will implicitly incorporate them in our analysis since the majority of the scholarly publications on e-Science make reference to these underlying technologies.

The remainder of the paper is organized as follows. Section 2 presents an overview of grid computing, desktop grid computing and cloud computing. Section 3 provides a brief introduction to co-citation analysis and reviews literature that have used such analysis in other disciplines. This is followed by a section on research methodology (Section 4) where we describe the dataset and the software used. We have generated two separate co-citation networks (for articles and authors) and the findings from these have been presented in Sections 5. This is followed by discussions and conclusions in Section 6.

2. Technology Enablers of E-Science: Computational Grid, Desktop Grid and Cloud Computing

2.1. Grid Computing

Computational grid was defined by Ian Foster and Carl Kesselman in their edited book *The Grid: The Blueprint for a New Computing Infrastructure*, as a hardware and software infrastructure that provides access to high-end computational resources [13]. It was further stated that this access should be dependable, consistent, pervasive and inexpensive. It was expected that computational grids would not only serve the scientific communities but also the government. Thus, national-level scenarios, e.g. response to environmental and man-made disasters, climate change simulations, could gain through use of the nation's fastest computers, data archives and shared intellect. Solving such problems usually necessitates investment in high-end computing resources. However, in concert with such investments, it is judicious that underutilized computation facilities within research organizations and universities be leveraged to derive maximum utilization of existing resources. Desktop grid computing makes this possible.

2.2. Desktop Grid Computing

While much of grid computing is focused on meeting the needs of large Virtual Organizations like academic institutions and R&D centers engaged in e-Science, desktop grid computing or desktop grids addresses the potential of harvesting the idle computing resources of desktop PCs [14]. These resources can be part of the same local area network (e.g. network of PCs in a university) or can be geographically dispersed and connected via a wide area network such as the internet (e.g. computers in different university campuses connected through research and educational networks, like JANET in the UK <www.ja.net/>). Studies have shown that desktop PCs can be underutilized by as much as 75 per cent of the time [15]. Furthermore, spare capacities are available on an hour-to-hour, day-to-day, and month-to-month basis, not only during the evening hours and on weekends, but during the busiest times of normal working hours [16]. This coupled with the widespread availability of desktop computers and the fact that the power of network, storage and computing resources is projected to double every nine, 12 and 18 months, respectively [17], represents an enormous computing resource that can be used for e-Science.

Yet another form of desktop grid computing is Volunteer Computing (VC) wherein the underlying grid infrastructure is composed of computational resources that are donated by the users, e.g., SETI@Home project. Although VC was originally conceived for large-scale scientific computation requiring millions of volunteer PCs, more recently however, VC grids such as SZTAKI Desktop Grid [18] have been used for meeting the computation needs of an organization [19].

2.3. Cloud Computing

Cloud computing is defined as a type of "parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers" [20]. The cloud infrastructure usually comprises of commodity hardware and specialized programs for virtualization and resource management. The software components enable dynamic provisioning of hardware resources based on the requirements of the user. Furthermore, through use of virtualization technology, several instances of virtual machines are started over the provisioned physical resource (every virtual machine has its own copy of the operating system and is sandboxed). Finally, user applications are executed within the confines of the virtual machines. The use of the cloud resources by the user application is usually monitored using accounting applications (e.g. CPU cycles used, data storage utilized and network bandwidth consumed). The clouds implement a pay-per-use model and therefore accounting is a very important part of this distributed computing technology. Providers of commercial cloud computing platform include Amazon Elastic Compute Cloud EC2 <aws.amazon.com/ec2/>, Amazon Simple Storage Service S3 <aws.amazon.com/s3/>) and Eucalyptus Eucalyptus <www.eucalyptus.com>. Since cloud computing is a comparatively new distributed computing technology, the uptake of cloud computing as an infrastructure platform for executing e-Science applications has been justifiably low. One notable e-Science project that has used cloud technology is CARMEN – a system that allows neuroscientists to share, integrate and analyze data [21].

3. Co-Citation Analysis and Related Work

The aim of this study is to explore the e-Science knowledge base through an analysis of co-citations. *Co-citation analysis identifies clusters of "co-cited" references by creating a link between two or more references when they co-occur in the reference lists of citing articles* [22]. The resultant co-citation networks provide important insights into knowledge domains by identifying frequently co-cited papers, authors and journals related to the domains in question, and which could have been overlooked if only conventional citation analysis techniques were used. In a citation-based analysis the significance of an article is often measured on the basis of the number of citations it has had. However it can be argued that there may exist certain articles that can be considered high-impact even though the number of citations it has received are comparatively less (for example, papers that have been cited a few times but across domains; papers that have been cited consistently through the years; papers that have been published recently). The opposite of this may also be true (for example, self-citations will usually increase the number of citations for a paper). Furthermore, it usually takes at least 5-6 years for a paper to build up its citation count. Thus, complimenting traditional citation-based metrics with co-citation analysis is arguably a superior approach for identifying articles that hold promise and which represent the grounded knowledgebase of a discipline. Through the investigation of extant e-Science literature, the objective of this paper is to demonstrate this added value of using co-citation analysis (as compared with citation-only analysis) when it comes to undertaking bibliometric research. Furthermore, we have identified prominent authors and articles in this field irrespective of their co-citation frequencies. This has been made possible through use of CiteSpace (see Section 4.2).

3.1. Existing Work

Co-citation analysis and profiling studies are two different forms of bibliometric and meta-data analysis. Profiling studies are usually conducted in relation to a particular journal [23], studies that compare between journals [24], or indeed those that aim to methodologically study the contribution of specific research fields, e.g., Modelling and Simulation (M&S), with regard to particular application domains, e.g., application of M&S in healthcare [25]. Such profiling studies help to identify currently under-explored research issues, and select theories and methods appropriate to their investigation, all of which are recognized in Information Systems as important issues for conducting fruitful, original and rigorous research [23, 26]. It can be argued that the same holds true for research in e-Science, and indeed, most other research areas. Studies that have used co-citation analysis include the study of the Information Science discipline [27], the studies on the intellectual structure of Management Information Systems [28, 29], Operations Management [30] and general management [31, 32]. However there are presently no studies that have investigated the e-Science knowledge base through co-citation analysis, which is the subject of this paper.

4. Research Method

4.1 Dataset

For the purposes of this study we conducted a search on ISI Web of Science. We used the keyword *"e-Science"* to conduct a search on article title, abstract, author keywords and ISI keywords plus®. The search was conducted on all available ISI Web of Science citation databases, e.g., Science Citation Index Expanded (SCI-EXPANDED) and Conference Proceedings Citation Index- Science (CPCI-S), and the timespan selected was "All Years". The specific search criterion that was used was as follows: *Topic= (e-Science); Databases= (SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH); Timespan = (All Years).* The search retrieved a total of 1009 articles, of which we only considered conference proceedings, journal articles and review papers (we ignored editorial material and news items). The total number of articles included for the review was thus 969. The next step involved downloading the ISI-format data associated with the 969 records in plain-text format, which was then input to the tool we used for the co-citation analysis – CiteSpace [12].

4.2 CiteSpace

Through co-citation analysis CiteSpace identifies turning points associated with articles and authors irrespective of their co-citation count. This is achieved through use of the full feature set of CiteSpace, including visual identifications of significant articles and authors through innovative visualization techniques [12]. The use of CiteSpace requires careful selection of a multitude of options, and an acceptable options' combination frequently requires learning through "trial and error" as well as knowledge of the underlying research domain. For the purpose of ensuing repeatability of this exercise, we present the specific option values that were selected in CiteSpace. (a) *Time interval of analysis:* 1995-2013inclusive; (b) *The unit of analysis:* 1 year per time slice; (c) *C, CC, CCV for the earliest time slice:* 2,3,20; (d) *C, CC, CCV for the middle time slice:* 3,3,20; (e) *C, CC, CCV for the last time slice:* 4,3,20; (f) *Pruning and merging:* Pathfinder network scaling [33] is used to prune the merged networks; (g) *Visualization:* A merged network cluster view has been selected.

The various CiteSpace options are now briefly discussed. An extensive discussion of these variables is outside the scope of this paper and the reader is referred to [33]. Nodes and links are the building blocks of a co-citation network and in our analysis we have created a total of 19 individual co-citation networks with the unit of analysis being 1 year (b). The resultant number of nodes and links are 176 and 653 respectively for document co-citation network and 295 nodes and 1592 links for author co-citation network. We have selected a merged network cluster view (g) and are therefore presented with a single cross-cluster co-citation network visualization, wherein each of the 19 individual co-citation networks (i.e., one for each year) are merged and the resultant merged network is pruned using the pathfinder network scaling algorithm (f). In this study the co-citation network is related to either articles or authors. The C, CC and CCV values in (c), (d) and (e) refer to the citation threshold - for example, C=2 for (c) implies that only those articles/authors that have been cited at least two times will be considered in the co-citation network for the earliest time slice of 2002-2003; the co-citation threshold, for example, CC=2 for (d) implies that for any two papers (or authors) to be included in the resultant co-citation network for the middle time slice of 2006-2007, they should have been co-cited at least two times; finally, the co-citation co-efficient threshold CCV is a normalized co-citation association strength, for example, for the last time slice of 2010-2011 the CCV=20 (e) and the remaining time slices CiteSpace will assign interpolated threshold values automatically. A comparison of the threshold values for (c), (d) and (e) shows an increasing trend. This is in line with the generally accepted fact that the number of citations for an article usually increases with time.

4.3 Variables for Analysis

In this study we have used the CiteSpace options discussed in the previous section in order to generate co-citation networks pertaining to cited articles – this is also referred to as Document Co-citation Network (DCN) and cited authors - Author Co-citation Network (ACN). The resultant networks (see figure 1) have been used for answering the following two questions:

- *Question 1:* Which are the turning point articles (i.e., articles with high significance, irrespective of the article co-citation count)? We will use DCN to answer this.
- *Question 2:* Who are the turning point authors (i.e., authors with high impact, irrespective of author co-citation count) - We will use ACN to answer this.

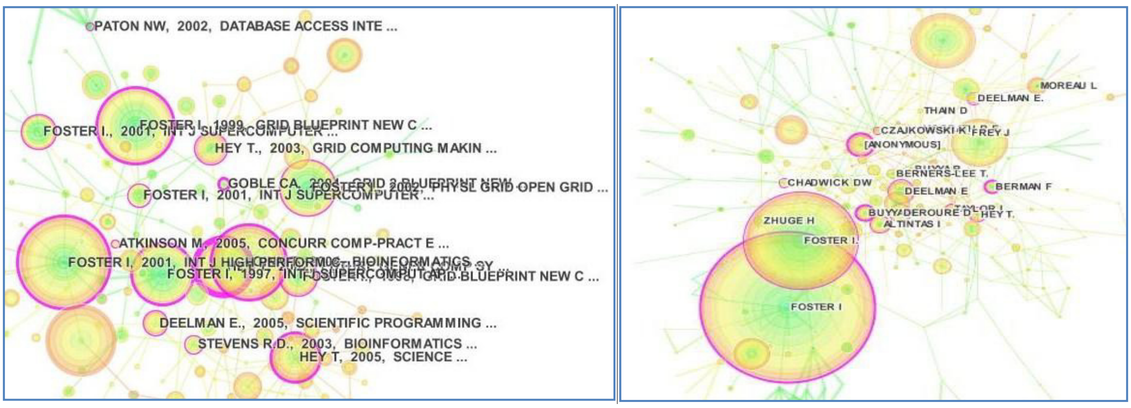

Fig. 1. (left) Document Co-Citation Network - DCN; (right) Author Co-citation Network -ACN

5. Turning Point Articles and Authors

In turning point analysis an article/author with comparatively low number of co-citations can be recognized as a landmark/turning point article/author since CiteSpace identifies potentially important articles/authors in a co-citation network through *landmark nodes* (a node with extraordinary attributes), *hub nodes* (widely co-cited article) and *pivot nodes* (common nodes that are shared between two co-citation network or gateway nodes that are interconnected by inter-network links), and by enhancing the visual features of such nodes it makes it easier to detect them through visual inspection [12]. In this study we analyse the top ten turning point articles and 15 such authors (refer to figure 1, tables 1 and 2), identify the co-citation count associated with each and reflect on why these articles and authors are a significant part of the e-Science knowledge base.

Table 1. Turning point articles identified using Document Co-Citation Network

Num	Article	Co-citation count
1	Hey T, Trefethen AE. The UK e-Science Core Programme and the Grid. Future Generation Computer Systems 2002;18(8):1017–31.	33
2	Foster I, Kesselman C, Tuecke S. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. International Journal of High Performance Computing Applications 2001;15(3):200-22.	56
3	Foster I, Kesselman C. The Grid: Blueprint for a New Computing Infrastructure, 1st ed., San Francisco, CA: Morgan Kaufmann; 1999.	46
4	Oinn T, Addis M, Ferris J, et al. Taverna: a tool for the composition and enactment of bioinformatics workflows. Bioinformatics 2004;20(17):3045-54.	44
5	Hey T, Trefethen AE. Cyberinfrastructure for e-Science. Science 2005;308(5723):817-21.	30
6	Goble CA, Roure DD, Shadbolt NR, et al. Enhancing Services and Applications with Knowledge and Semantics. In: Foster I, Kesselman C, editors. The Grid 2: Blueprint for a New Computing Infrastructure, 2nd ed., San Francisco, CA: Morgan Kaufmann; 2004, Pages 431–58.	5
7	Foster I, Kesselman C. Globus: a Metacomputing Infrastructure Toolkit. International Journal of High Performance Computing Applications 1997;11(2):115-28.	38
8	Hey T, Trefethen A. The Data Deluge: An e-Science Perspective. In: Berman F, Fox G, Hey T, editors. Grid Computing: Making the Global Infrastructure a Reality, Chichester, UK: John Wiley & Sons. Chapter 36.	20
9	Foster I, Kesselman C, Nick J, Tuecke S. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. Open Grid Service Infrastructure WG, Global Grid Forum, June 22, 2002.	35
10	Paton NW, Atkinson MP, Dialani V, et al. Database Acccess and Integration Services on the Grid, UK DBTF working paper, 2002.	4

As can be seen from Table 1, most of the paper titles make reference to grid computing either explicitly (e.g., Open Grid Services Architecture; paper 9) or implicitly (e.g., Globus; paper 7). This is not surprising since grid computing has traditionally been the distributed computing infrastructure of choice, especially when considering large-scale and collaborative e-Science research. Indeed, the term Grid predates e-Science, with the latter term being credited to John Taylor, the former Director General of UK Research Councils in the UK Office of Science and Technology [34]. The result of our co-citation analysis shows that four of the research artefacts have been co-authored by Ian Foster and Carl Kesselman; these publications occupy four of the top five places in terms of the number of co-citations. This is hardly surprising. Ian Foster is frequently referred to as the "Father of Grid Computing" [35]; frequently this honour is shared with Carl Kesselman and Steve Tuecke, all of whom are considered as pioneers in this field and were also the co-founders of the Globus project. Globus (paper 7) was a research project that was mainly based in the Argonne National Laboratory, USA, and developed the basic mechanisms and infrastructure for grid. Table 1 further shows that there are three papers that have e-Science in their title, and all of which have been co-authored by Tony Hey and Anne E. Trefethen. Although it is not surprising that the aforementioned papers appear in the top-10 article list in our co-citation analysis (the search term used in ISI was "e-Science"), further look at the papers reveal additional reasons for their appeal. Paper 1 describes the £120 million UK e-Science core program and it was written by the authors in their capacity as the Director (T. Hey) and Deputy Director (A. E. Trefethen) of the core e-Science programme respectively. Paper 8 is a foresight paper that looks into the deluge of e-Science data and its implications for the Grid. Paper 5 is a viewpoint point paper that puts the case for service-oriented infrastructure for e-Science and, in the context of e-Infrastructure, enumerates UK's vision for "plug-and-play composable services" and the role of semantics in it.

Table 2. Turning point authors identified using Author Co-Citation Network

Num	Authors	Co-citation count	Num	Authors	Co-citation count	Num	Authors	Co-citation count
1	Foster I	374	6	Czajkowski K	15	11	Mcgrath RE	3
2	Buyya R	27	7	Deelman E	23	12	Deelman E	35
3	Berman F	21	8	Hey T	27	13	Moreau L	32
4	Deroure D	35	9	Altintas I	27	14	Berners-Lee T	23
5	Taylor I	20	10	Chadwick DW	16	15	Thain D	14

Our ACN analysis is limited to the first author of the cited reference. The top-15 list of turning point authors (see table 2) includes those whose papers were identified in table 1, e.g., Ian Foster and Tony Hey. Rajkumar Buyya is identified as the second most prominent author. Again this is not surprising since he is considered a pioneer of economic paradigm for utility-oriented distributed computing and has been instrumental in the development of grid, desktop grid and cloud computing middleware – all of which serve as the technological backbone for e-Science. Sir Tim Berners-Lee, the father of the World Wide Web, is also recognized as a turning point author. Most of his scholarly contributions relate to the WWW and the semantic Web. His inclusion in top-15 list perhaps shows the importance of WWW and allied technologies (like Web2.0; portlets) for e-Science.

6. Discussion and Conclusions

In this paper we have presented a preliminary exploration of the e-Science knowledge base using co-citation analysis. We have identified top-10 turning point papers and top-15 turning point authors from the extant literature. Not surprisingly, most of the papers and authors' work relate to grid computing and grid-based technologies as this has traditionally been the distributed computing technology of choice for e-Science research. However, with the advancement of Grid-DesktopGrid-Cloud interoperability technologies (e.g., [9-11]), this overt reliance on Grid may gradually pave the way for Cloud and Desktop Grid-based execution of e-Science applications. The restriction on page size for the conference has meant that we have not been able to provide a more comprehensive analysis. This is for a future paper. A direction for future research could be the development of algorithms that perform cluster analysis on cited data (clusters may pertain to authors, journals, keywords, etc.) from scholarly literature being accessed by researchers in real-time; such an application could be delivered from a browser-based application [36].

References

[1] Mustafee N. Exploiting grid computing, desktop grids and cloud computing for e-Science: future directions. *Transforming Government: People, Process and Policy 2010*;**4**(4):288-98.

[2] Baker M, Buyya R, Laforenza D. Grids and grid technologies for wide-area distributed computing. *Software – Practice & Experience 2001*;**32**(15):1437-66.

[3] Hey T, Trefethen AE. The UK e-science core programme and the grid. *Future Generation Computer Systems 2002;* **18**(8):1017-31.

[4] Bernholdt D, Bharathi S, Brown D, Chanchio K, Chen M, Chervenak A, et al. The earth system grid: supporting the next generation of climate modelling research. *Proceedings of the IEEE 2005*;**93**(3):485-95.

[5] Lamanna M. The LHC computing grid project at CERN", *Nuclear Instruments and Methods in Physics Research (Section A: Accelerators, Spectrometers, Detectors and Associated Equipment) 2004*;**534**(1/2):1-6.

[6] Ellisman M, Peltier S. Medical data federation: the biomedical informatics research network. In: Foster I, Kesselman C, editors. *The Grid 2: Blueprint for a New Computing Infrastructure, 2nd ed*., San Francisco, CA: Morgan Kaufmann; 2004, 109-120.

[7] Spencer B, Finholt TA, Foster I, Kesselman C, Beldica C, Futrelle J, et al. NEESgrid: a distributed collaboratory for advanced earthquake engineering experiment and simulation. In Proceedings of the 13th World Conference on Earthquake Engineering 2004;Paper 1674.

[8] Bessis N. *Grid technology for maximizing collaborative decision management and support: advancing effective virtual organizations*. Hershey, New York: Information Science Reference (IGI Global); 2009.

[9] Urbah E, Kacsuk P, Farkas Z, Fedak G, Kecskeméti G, Lodygensky O, et al. EDGeS: bridging EGEE to BOINC and XtremWeb. *Journal of Grid Computing 2009*;**7**(3):335–54.

[10] Anjum A, Hill R, McClatchey R, Bessis N, Branson A (2012). Glueing grids and clouds together: a service-oriented approach. *International Journal of Web and Grid Services 2012*;**8**(3):248-265.

[11] Huang Y, Bessis N, Norrington P, Kuonen P, Hirsbrunner B. Exploring decentralized dynamic scheduling for grids and clouds using the community-aware scheduling algorithm. *Future Generation Comp. Syst. 2013*:**29**(1):402-415.

[12] Chen C. Searching for intellectual turning points: progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences, USA, 2004*;**101**(Suppl.),pp.5303–310.

[13] Foster I, Kesselman C. Computational grids. In: Foster I, Kesselman C, editors. *The Grid: Blueprint for a New Computing Infrastructure,* San Francisco, CA: Morgan Kaufmann; 1998, Chapter 2.

[14] Choi S, Baik M, Hwang C, Gil J, Yu H. Volunteer availability based fault tolerant scheduling mechanism in desktop grid computing environment, *Proceedings of the 3rd IEEE International Symposium on Network Computing and Applications, Cambridge, MA, 2004*;pp.366-71.

[15] Mutka MW. Estimating capacity for sharing in a privately owned workstation environment. *IEEE Transactions on Software Engineering 1992*;**18**(4):319-28.

[16] Mutka MW, Livny M. The available capacity of a privately owned workstation environment. *Performance Evaluation 1991*;**12**(4):269-84.

[17] Casanova H. Distributed computing research issues in grid computing. *ACM SIGACT News 2002*; **33**(3):50-70.

[18] Kacsuk P, Kovacs J, Farkas Z, Marosi AC, Gombas G, Balaton Z. SZTAKI Desktop Grid (SZDG): a flexible and scalable desktop grid system. *Journal of Grid Computing 2009*;**7**(4):439-461.

[19] Taylor SJE, Ghorbani M, Kiss T, Farkas D, Mustafee N, Kite S, et al. Distributed computing and modeling and simulation: speeding-up simulations and creating large models. Advance Tutorial. *Proceedings of the Winter Simulation Conference, Phoenix, Arizona, 2011*;pp.161-75.

[20] Buyya R, Yeo CS, Venugopal S. Market-oriented cloud computing: vision, hype, and reality for delivering IT services as computing utilities. *Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications, Dalian, 2008*;pp.5-13.

[21] Watson P, Lord P, Gibson F, Periorellis P, Pitsilis G. Cloud computing for e-science with CARMEN. *Proceedings of the 2nd Iberian Grid Infrastructure Conference Proceedings, Porto, 2008*;pp.3-14.

[22] Raghuram S, Tuertscher P, Garud R. Mapping the field of virtual work: A co-citation analysis. *Information Systems Research 2009*;**21**(4):983-99.

[23] Palvia P, Pinjani P, Sibley EH. A profile of information systems research published in Information and Management. *Information & Management 2007*;**44**(1):1–11.

[24] Claver E, Gonzalez R, Llopis J. An analysis of research in information systems (1981–1997). *Information and Management 2000*;**37**(4):181–95.

[25] Mustafee N, Katsaliaki K, Taylor SJE. Profiling Literature in Healthcare Simulation. *SIMULATION: Transactions of the Society of Modelling and Simulation International 2010*;**86**(8-9):543-58.

[26] Galliers RD, Whitley EA, Paul RJ. Guest editorial: the European information systems academy. *European Journal of Information Systems 2007*;**16**(1):3–4.

[27] White HD, McCain KW. Visualizing a discipline: an author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science 1998*;**49**(4):327-55.

[28] Culnan MJ. The intellectual development of management information systems, 1972–1982: a co-citation analysis. *Management Science 1986*;**32**(2):156-172.

[29] Mustafee N. Evolution of IS research based on literature published in two leading IS journals - EJIS and MISQ. *Proceedings of the 19th European Conference on Information Systems, Helsinki, Finland 2011*;Paper 228.

[30] Pilkingtona A, Meredith J. The evolution of the intellectual structure of operations management—1980–2006: a citation/co-citation analysis. *Journal of Operations Management 2009*;**27**(3):185–202.

[31] Nerur SP, Rasheed AA, Natarajan V. The intellectual structure of the strategic management field: an author co-citation analysis. *Strategic Management Journal 2008*;**29**(3):319–36.

[32] Acedo FJ, Casillas JC. Current paradigms in the international management field: An author co-citation analysis. *International Business Review 2005*;**14**(5):619–39.

[33] Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology 2006;***57**(3):359-77.

[34] Hey T, Trefethen AE. The UK e-Science Core Programme and the Grid. *Future Generation Computer Systems 2002*;**18**(8):1017–31.

[35] Braverman AM. Father of the grid. *University of Chicago Magazine 2004*;**96**(4).

[36] Marin A, Pop F. Intelligent Web-History Based on a Hybrid Clustering Algorithm for Future-Internet Systems. *Proceedings of the 13th International Symposium Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania 2011*;pp.145-152.