Research Activity in Computational Physics utilizing High Performance Computing: Co-authorship Network Analysis

Sul-Ah Ahn and Youngim Jung
Korea Institute of Science and Technology Information, 245 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea
acorn@kisti.re.kr

Abstract. The research activities of the computational physicists utilizing high performance computing are analyzed by bibliometric approaches. This study aims at providing the computational physicists utilizing high-performance computing and policy planners with useful bibliometric results for an assessment of research activities. In order to achieve this purpose, we carried out a co-authorship network analysis of journal articles to assess the research activities of researchers for high-performance computational physics as a case study. For this study, we used journal articles of the Scopus database from Elsevier covering the time period of 2004-2013. We extracted the author rank in the physics field utilizing high-performance computing by the number of papers published during ten years from 2004. Finally, we drew the co-authorship network for 45 top-authors and their coauthors, and described some features of the co-authorship network in relation to the author rank. Suggestions for further studies are discussed.

1. Introduction

The purpose of bibliometrics is to map the literary production in a given field and to determine the structure of a field of study. Bibliometric analyses in various research fields have been conducted during decades. Lately, we also conducted a bibliometric analysis in the high-performance computing field [1]. For the computational physics field utilizing high-performance computing (ComPhy-HPC), it is of interest that bibliometric analyses are known to be very rare in ComPhy-HPC and related fields, because the high-performance computing is an emerging field of computational research at the intersection of science and computation [2]. By use of bibliometric approaches, we analyzed the research activities of the computational physicists utilizing high-performance computing. Our work includes the results of a bibliometric study of the computational physics utilizing high-performance computing, featuring in co-authorship network analysis of journal articles. In our knowledge, this is the first time that the co-authorship analysis is performed on journal publication data in the computational physics filed utilizing high-performance computing.

Earlier, a concept of scientific collaboration network is firstly introduced by Newman [3]. He took note on networks of scientists in which two scientists are considered connected if they have jointly coauthored a paper. After this, the affiliation networks of scientist, or co-authorship network, in which a link between two scientists is established by their co-authorship of one or more scientific papers are constructed [4]. It is assumed in the further study that co-authorship indicates a level of scientific
collaboration [5]. Recently, Morel et. al. utilized the co-authorship network analysis as a powerful tool in order to plan strategically research, development and capacity building programs on neglected diseases [6]. More recently, Ding studied scientific collaboration and endorsement by using network analysis of co-authorship and citation networks [7].

2. Our Work
We carried out a co-authorship network analysis of journal articles to assess the research activities of computational physicists utilizing high-performance computing as a case study. The data sources used for this study are journal articles of the Scopus database from Elsevier covering the time period of 2004-2013 for journal articles. We extracted author rank in the high performance computing field by the number of papers published during ten years from 2004. We calculated the weight of co-authorship and draw the co-authorship network for top-45 authors and their coworker.

2.1. Data
Computational physics utilizing High Performance Computing (ComPhy-HPC) was selected as the test area. Papers were collected from the Scopus database covering the time period of 2004-2013. Based on a set of search terms related to ComPhy-HPC, the following queries were formed: SUPERCOMPUT* PHYSICS, HIGH PERFORMANCE COMPUT* PHYSICS, HIGH THROUGHPUT COMPUT* PHYSICS, CLUSTER COMPUT* PHYSICS, GRID COMPUT* PHYSICS

2.2. Methodology for Analyzing Co-author Network
Analysis of co-authorship network reveals collaboration using published papers. Refinement of the source data was necessary because the first names of the authors are abbreviated and the authors who have multiple affiliations are duplicated in the Scopus database.
We generated a pathfinder network using symmetric square matrix whose elements are correlation value between authors. Pathfinder network algorithm is helpful to describe the structure of the network, because it simplifies the complex network in matrix representation [8]. For the visualization of the generated coauthor network, NodeXL has been utilized.

3. Results and Discussions

3.1. Author Ranks
In total, 4,403 papers were gathered. The corresponding authors of all articles have not been identified. Only 4,197 articles among 4,403 articles, the corresponding author of each article are identified. The number of authors involving in authoring those articles is 11,460. More than 78.7% authors, tallying 9,022 authors in the author list have authored only once. Only 46 scholars in this field have authored at least ten articles during this period. ComPhy-HPC is emerging and novel technology in physics field and the limited number of scholars have studied related with this topic. Figure 1 illustrates the number of articles published and indexed in the Scopus database for the last ten years.
Table 1 presents the author rank in the target field by the number of papers published from 2004 to 2013.

| Rank | Author Name       | Number of papers | Percentage over publication by top 45 authors |
|------|-------------------|------------------|---------------------------------------------|
| 1    | Gauss Jürgen      | 26               | 8.7                                         |
| 2    | Schaefer III Henry F. | 22           | 7.4                                         |
| 3    | Maeno T           | 18               | 6.0                                         |
| 3    | Sfiligoi Igor     | 17               | 5.7                                         |
| 5    | Graciani R        | 14               | 4.7                                         |
| 5    | Saiz P            | 14               | 4.7                                         |
| 7    | Wenaus T          | 13               | 4.3                                         |
| 7    | Bunčič P          | 13               | 4.3                                         |
| 7    | Bartlett Rodney J | 13               | 4.3                                         |
| 10   | Tsaregorodtsev A  | 12               | 4.0                                         |
| 10   | Szalewicz Krzysztof | 12           | 4.0                                         |
| 10   | Garzoglio G       | 12               | 4.0                                         |
| 10   | Grigoras C        | 12               | 4.0                                         |

Ranked as the top 1 author, Gauss, Jürgen published 26 papers in ComPhy-HPC field during decade. His affiliation is the Johannes Gutenberg University of Mainz, well known as the top tier of high performance computing. His work is focused on the application of the quantum chemical calculation and the development of new quantum chemical methods.
Maeno T is ranked as the 3rd best publishing author, with 18 papers published. His affiliation is the Brookhaven National Laboratory, well-known for nuclear and high-energy physics. The main topics in his publication includes the design, development and commissioning of innovative new distributed computing software critical to the ATLAS experiment.

3.2. Co-authorship and Co-authorship Network in ComPhy-HPC

In the co-authorship network, nodes represent authors and an edge shows that linked authors (nodes) have worked together at least for publishing one paper. The width of the edge represents the weight of the co-authorship between the two authors. The weight of the co-authorship is calculated by the cosine coefficient [9] of the two authors who authored articles published from 2004 to 2013 together. Only 45 top authors and their coauthors who involved in authoring at least ten publication among 11,460 authors have been selected for calculating the cosine coefficient. Table 2 shows the cosine coefficient, or weight of co-authorship among 45 top authors and their coauthors in ComPhy-HPC.

| Co-authoring Rank | Author Name       | Author Name       | Weight of Co-Authorship (Cosine Coeff.) |
|-------------------|-------------------|-------------------|----------------------------------------|
| 1                 | Negri A           | Vercesi V         | 0.9987                                 |
| 2                 | Luminari L        | Vercesi V         | 0.99422                                |
| 2                 | Luminari L        | Werner P          | 0.99422                                |
| 4                 | George S          | Vercesi V         | 0.99065                                |
| 4                 | George S          | Werner P          | 0.99065                                |
| 6                 | Buncic P          | Harutyunyan A     | 0.98705                                |
| 7                 | Grigoras C        | Betev L           | 0.98644                                |
| 8                 | Sciaba A          | Flix J            | 0.89658                                |
| 9                 | Saiz P            | Grigoras C        | 0.80007                                |
| 10                | Maeno T           | Vercesi V         | 0.79691                                |
| 10                | Maeno T           | Werner P          | 0.79691                                |
| 12                | Schaefer III Henry F | Allen Wesley D   | 0.78716                                |
| 13                | Graciani R        | Tsaregorodtsev A  | 0.74148                                |
| 14                | Wenaus T          | Maeno T           | 0.71652                                |
| 15                | Furano F          | Carminati F       | 0.67777                                |
| 16                | Stewart G A       | Wenaus T          | 0.6049                                 |
| 17                | Andreeva J        | Saiz P            | 0.55235                                |
| 18                | Sfiligoi Igor     | Bonacorsi D       | 0.20799                                |
| 19                | Krylov Anna I     | Head Gordon Martin| 0.16717                                |
| 20                | Bartlett Rodney J | Hirata So         | 0.14447                                |
| 21                | Gauss Jürgen      | Schaefer III Henry F | 0.14207                        |
| 22                | Garzoglio G       | Sfiligoi Igor     | 0.11282                                |
Negri A and Vercesi V are ranked as the top-1 co-author couple, who have the strongest correlation for the co-authorship among 45 top authors and their coauthors in HPC.

Figure 2, the co-authorship network among 45 top authors and their coauthors in ComPhy-HPC is represented using NodeXL.

**Figure 2.** Co-authorship network of 45 top authors and their coauthors in ComPhy-HPC (2004-2013)

Gauss Jürgen and Schaefer III Henry F, the top-1 and top-2 author according to number of publishing, seem to have relatively weak correlation for the co-authorship among 45 top authors and their coauthors in HPC. George S and Werner P are not listed in the top 10 list in the author rank, but they show relatively strong correlation for the co-authorship (top 4 in table 2). As shown in figure 2, Werner P is also connected to Maeno T (top 3 in table 1), which reveals collaboration between authors may easily propagate through dominant authors.

4. Summary
We performed a co-authorship network analysis of journal articles to evaluate the research activities in the computational physics field utilizing high-performance computing (ComPhy-HPC). The data sources used for this study are journal articles of the Scopus database from Elsevier covering the time period of 2004-2013 for journal articles. We extracted author rank in the ComPhy-HPC field by the number of publication during ten years. We calculated the weight of co-authorship and draw the co-authorship network for 45 top-authors and their coauthors. High ranked authors by publication are not also strong in co-authorship among 45 top-authors and their coauthors. Some authors, not much high ranked authors by publication, show relatively more strong correlation for the co-authorship. In order to provide significant bibliometric results for an evaluation of research activities to computational physicists utilizing high-performance computing and policy planners, further studies using co-authorship network analysis or citation network [10] are suggested to be done for more than 45 top-authors in the ComPhy-HPC field.

References
[1] Ahn S-A and Jung Y 2015 ICIC Express Letters, Part B: Applications 6 1437
[2] 2013 High-Performance Computing Using FPGAs, ed W Vanderbauwhede and K Benkrid (New York: Springer-Verlag)
[3] Newman M E J 2001 Proc. Natl. Acad. Sci. USA 98 404
[4] Newman M E J 2001 Physical Review E 64 016131.
[5] Newman M E J 2004 Proc. Natl. Acad. Sci. USA 101(suppl. 1) 5200.
[6] Morel C M, Serruya S J, Penna G O and Guimariaes R 2009 PLOS/Neglected Tropical Diseases 3 e501
[7] Ding Y 2011 J. Informetr. 5 187
[8] Lee J 2006 Journal of the Korean Society for Library and Information Science 40 333
[9] Egghe L and Leydesdorff L 2009 J. Am. Soc. Inf. Sci. Technol.60 1027
[10] Erman N and Todorovski L 2009 Mapping the e-Government Research with Social Network Analysis EGOV 2009 (LNCS vol 5693) ed M A Wimmer et al (Heidelberg: Springer) pp 13-25