Title: Analysis of Publication Activity of Computational Science Society in 2001-2017 Using Topic Modeling and Graph Theory

Authors: Tesfamariam M. Abuhay, Sergey V. Kovalchuk, Klavdiya Bochenina, Gali-Ketema Mbogo, Alexander A. Visheratin, George Kampis, Valeria V. Krzhizhanovskaya, Michael H. Lees

PII: S1877-7503(18)30246-1
DOI: https://doi.org/10.1016/j.jocs.2018.04.004
Reference: JOCS 852

To appear in:

Received date: 10-3-2018
Accepted date: 2-4-2018

Please cite this article as: Tesfamariam M.Abuhay, Sergey V.Kovalchuk, Klavdiya Bochenina, Gali-Ketema Mbogo, Alexander A.Visheratin, George Kampis, Valeria V.Krzhizhanovskaya, Michael H.Lees, Analysis of Publication Activity of Computational Science Society in 2001-2017 Using Topic Modeling and Graph Theory, Journal of Computational Science https://doi.org/10.1016/j.jocs.2018.04.004

Analysis of Publication Activity of Computational Science Society in 2001-2017 Using Topic Modeling and Graph Theory

Tesfamariam M. Abuhay[1], Sergey V. Kovalchuk[1], Klavdiya Bochenina[1], Gali-Ketema Mbogo[1], Alexander A.Visheratin[1], George Kampis[1,3], Valeria V. Krzhizhanovskaya[1,2], Michael H. Lees[1,2]

[1]ITMO University, Russia
[2]University of Amsterdam, The Netherlands
[3]Eötvös University, Hungary
tesfaabeba@gmail.com, kovalchuk@mail.ifmo.ru, k.bochenina@gmail.com, ketema.galy@gmail.com, alexvish91@gmail.com, kampis.george@gmail.com, v.krzhizhanovskaya@uva.nl, m.h.lees@uva.nl

Abstract
This paper presents results of topic modeling and network models of topics using the ICCS corpus, which contains domain-specific (computational science) papers over seventeen years (a total of 5982 papers). We discuss topical structures of ICCS, how these topics evolve over time in response to the topicality of various problems, technologies and methods, and how all these topics relate to one another. This analysis illustrates multidisciplinary research and collaborations among scientific communities, by constructing static and dynamic networks from the topic modeling results and the authors' keywords. The results of this study give insights about the past and future trends of core discussion topics in computational science and how "computational thinking" propagates into different fields. We used the Non-negative Matrix Factorization (NMF) topic modeling algorithm to discover topics and labeled and grouped results hierarchically. We used Gephi to study static networks of topics, and an R library called DyA to analyze the dynamic networks of topics. We also analyzed the conference as a platform for potential collaboration development through the perspective of collaboration networks. It occurred that authors of ICCS papers continue to actively collaborate after the conference - on the average authors collaborate with three other authors, - which suggests that ICCS is a valuable platform for collaboration development.

Keywords: topic modeling, natural language processing, ICCS, computational science, graph theory, collaboration networks

1 Introduction

Our aggregate knowledge continues to be digitized and filed away as news, scientific articles, and books [1]. The major sources of topical scientific knowledge are conferences and scientific journals. The International Conference