Using Bibliometric Analysis to Measure and Understand the Gender Gap in Published Computing Books:
Gender Gap in Computer Science

Arbana Kadriu, SEE University, Macedonia

iD https://orcid.org/0000-0003-4922-4753

Kosovare Sahatqija, SEE University, Macedonia

Lejla Abazi-Bexheti, SEE University, Macedonia

ABSTRACT

The purpose of the research presented in this paper is the investigation of the gender gap in published computing books. The book titles from the DBLP computer science bibliography were the basis for this investigation. The conducted research involves co-authorship network exploration using social network analysis methods, as well as content learning by keyword extraction and ranking from book titles. The findings show that female authors tend to publish fewer books in computing than their male colleagues, and there is a huge gap of women regarding the collaboration. There are just two women names within the 50 author names with the highest social network top metrics, indicating collaboration. Regarding the extracted keywords, though there are differences, results do not show some huge divergences when it comes to the used language for computing titles.

KEYWORDS

Co-Authorship, DBLP, Gender Gap, Information Retrieval, Keyword Extraction, Scholarly Data, SNA Metrics, Text Mining

INTRODUCTION

Machine learning algorithms are ubiquitously used to provide inside information on human behavior in different contexts. Particularly, parts of these algorithms are employed aiming to understand the relationship between different researchers in a particular field, as well as the features that characterize their work. As the gender gap is a matter of social concern and investigation for decades, different machine learning algorithms have been utilized seeking to comprehend the situation in various fields such as research, politics, education, health, etc.

Machine learning has been used in order to increase understanding of various gender gap issues such as correlations between gender biases and statistical gender gaps (Friedman et al., 2019);

gender earnings gap (Saavedra, 2019); risk of automation gender biases (Brussevich et al., 2019); news commenters variations by gender (Lee & Ryu, 2019); electric vehicles purchase preferences by gender (de Rubens, 2019); dissimilarities in investment decision making arising from the gender (Silva et al., 2019); risk of wage gender inequality (Karimian et al., 2019); gender bias in relation extraction (Gaut et al., 2019); gender gap in parties (Mancosu & Bobba, 2019); gender diversity in research (Stathoulopoulos & Mateos-Garcia, 2019); effects of gender performance in online product reviews (Sikdar et al., 2020); and others. In the research presented in this paper, we want to understand the gender gap in publishing computing books from a historical perspective, exploring the DBLP dataset. Our research is trying to explore two aspects of this gender share in computing. In the first analysis, we make use of social network analysis to discover some kind of implicit information that is new and can be beneficial, through metrics calculations and their visualizations. These kinds of relationships can lead us to other conclusions, which will allow us to see which author is usually collaborating with whom, and then with further investigation, we can understand their interests and specific field of their collaboration (Sahatqija & Kadriu, 2019). We explain the results gained from main social network analysis metrics, with a focus on gender share. In our second analysis, we extract and investigate the keywords that identify the field of interest for both genders, using book titles and information retrieval algorithms.

The paper is organized as follows: the first chapter introduces the field, followed by the second chapter that gives an overview of the related work from other authors on different interrelated topics. Then a deeper explanation for the methodology is given, including the gained data, network analysis metrics and keyword extraction, which are going to be included in our analysis. The next chapter is about the obtained results and their interpretation. Finally, the conclusion of this paper concludes the research and gives an outline for further work.

RELATED WORK

The field of social network analysis has received a lot of interest from different authors, with a crucial role in analyzing and at the same time forecasting future relations. Several authors investigated this topic by analyzing books' collaboration networks.

Sun et al. (2011) proposed an approach in order to predict future collaboration relationships regarding heterogeneous networks. Their *PathPredict* model is based on the topological features in such networks, using logistic regression-based supervised learning methods to learn the coefficients associated with each feature. A related approach used by Yan et al. (2010) is the weighted PageRank algorithm, which is considered to provide reliable results in measuring author impact in collaborations. Another research similar to this one is made by Zhou et al. (2007). They propose a framework for co-ranking entities of different kinds in a heterogeneous network connecting researchers and publications they produce. Since the co-authorship domain got very huge interest, also Liu et al. (2005) have investigated the co-authorship network of the Digital Libraries of the research community in the conferences. They pointed out that the densest shapes (collaborations) include authors from the same institution or working on the same project. Lu & Feng (2009) have made a measure of authors' centrality in co-authorship networks. The experiment that they implemented showed that the extensity centrality and betweenness have a relatively high correlation and the correlation between extensity centrality and degree is low. An advantage that we can get because of examining these kinds of relations is finding the appropriate team members and their similar collaborative abilities. Cheatham et al. (2006) applied social network analysis to analyze this issue. They collected the data of people in their company and their collaborative relations. Using those relations, they found out the way of creating teams that will have a helpful collaboration. On the other hand, Huang et al. (2006) visualized the co-authorship network via an internal tool that they developed which showed the weight of the co-authorship.

Related Content

Assessment of the Purpose of the Use of GAS: A Perspective of Internal Audit Functions in Australia
Louis Smidt, Leandi Steenkamp, Aidi Ahmi, D. P. van der Nestand David Schalk Lubbe (2021). *International Journal of Information Systems in the Service Sector (pp. 65-82).*
www.irma-international.org/article/assessment-of-the-purpose-of-the-use-of-gas/273645

Flow-based Adaptive Information Integration
Dickson K.W. Chiu, Thomas Trojer, Hua Hu, Haiyang Hu, Yi Zhuangand Patrick Hung (2010). *Electronic Services: Concepts, Methodologies, Tools and Applications (pp. 525-545).*
www.irma-international.org/chapter/flow-based-adaptive-information-integration/43968

Telecommunication Industry: Storage and Mobility
Fredrik Solsvikand Michel Dao (2013). *Data Intensive Storage Services for Cloud Environments (pp. 145-157).*
www.irma-international.org/chapter/telecommunication-industry-storage-mobility/77436

Building a Knowledge Experience Base for Facilitating Innovation
Pasquale Ardimento, Vito Nicola Convertiniand Giuseppe Visaggio (2013). *International Journal of E-Services and Mobile Applications (pp. 40-49).*
www.irma-international.org/article/building-a-knowledge-experience-base-for-facilitating-innovation/100395

MoBiFlow: Principles and Design of a Workflow System for Molecular Biology
Markus Held, Wolfgang Küchlinand Wolfgang Blochinger (2013). *Best Practices and New Perspectives in Service Science and Management (pp. 300-313).*
www.irma-international.org/chapter/mobiflow-principles-design-workflow-system/75001