Trends in computational science: natural language processing and network analysis of 23 years of ICCS* publications

Lijing Luo[1], Sergey Kovalchuk, Valeria Krzhizhanovskaya[1], Maciej Paszynski[2], Clélia de Mulatier[1], Jack Dongarra[3], and Peter M.A. Sloot[1]

[1] University of Amsterdam, The Netherlands
[2] AGH University of Krakow, Poland
[3] University of Tennessee, USA

lijing.luo@student.uva.nl, sergey.v.kovalchuk@gmail.com, v.krzhizhanovskaya@uva.nl, paszynsk@agh.edu.pl, c.m.c.demulatier@uva.nl, dongarra@icl.utk.edu, p.m.a.sloot@uva.nl

Abstract. We analyze 7826 publications from the International Conference on Computational Science (ICCS) between 2001 and 2023 using natural language processing and network analysis. We categorize computer science into 13 main disciplines and 102 sub-disciplines sourced from Wikipedia. After lemmatizing full texts of these papers, we calculate the similarity scores between the papers and each sub-discipline using vectors built with TF-IDF evaluation. Among the 13 main disciplines, machine learning & AI have become the most popular topics since 2019, surpassing parallel & distributed computing, which peaked in the early 2010s. Modeling & simulation, and algorithms & data structure have always been popular disciplines in ICCS over the past 23 years. The most frequently researched sub-disciplines, on average, are algorithms, numerical analysis, and machine learning. Deep learning shows the most rapid growth, while parallel computing has declined over the past 23 years in ICCS publications. The network of sub-disciplines exhibits a scale-free distribution, indicating certain disciplines are more connected than others. We also present correlation analysis of sub-disciplines, both within the same main disciplines and between different main disciplines.

Keywords: natural language processing, topic modelling, computational science, graph theory, network analysis, scientometrics, ICCS

1 Introduction

The continuous growth and digitization of scientific publications offer extensive research opportunities in scientometrics. As an integral part in science of science, scientometrics plays a crucial role in guiding policies related to scientific development [1]. Additionally, it enables exploration of the progress within

* International Conference on Computational Science: http://www.iccs-meeting.org/

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

current scientific research fields [2]. The necessity to utilize quantitative methods for modeling and analyzing the progress of science has emerged as a key area of research [3]. The International Conference on Computational Science (ICCS) is an annual conference in the field that provides a prestigious platform for researchers, scientists, and engineers to explore computational disciplines encompassing mathematics and computer science [4,5,6]. Computational science is inherently interdisciplinary, offering advanced computing methodologies for addressing problems, identifying new issues, and shaping future directions in physics, chemistry, social sciences, and other fields. Since its inception in 2001, ICCS has consistently attracted an average of 340 highly cited papers per year [4,5,6]. This remarkable achievement establishes it as one of the most influential events within the field of computational science.

As a noteworthy asset in the field of computational science, the rapidly expanding proceedings series serve as a valuable corpus for quantifying scientific advancements. In this study, we apply a topic modeling technique to model and analyze the content of research papers. Building upon the concept that documents consist of various topics corresponding to specific disciplines, we apply text classification techniques and natural language processing methods to discover and analyze these topics. Utilizing a standardized corpus categorized by discipline, we conduct an annual analysis to explore changes in the distribution of research fields over 23 years. Simultaneously, we investigate emerging and declining research fields based on their popularity. Additionally, static and dynamic network analyses are conducted to examine how correlations between disciplines evolve and how network structures change over time. We answer the following questions: Which disciplines are gaining prominence or diminishing in computational science research? How have popular disciplines emerged or disappeared over the past 23 years? What is the structure of disciplinary networks and how does it evolve? How do correlations between disciplines change? The general methodology and tools we present can be applied to other fields of science.

This paper will be divided into seven sections presenting our studies: Section 2 summarizes related work; Section 3 describes data collection, pre-processing, and relevant methodologies; Section 4 demonstrates results about first-level disciplines; Section 5 provides an analysis from the perspective of second-level disciplines; Section 6 analyzes disciplinary networks; finally, Section 7 presents conclusions and future work.

2 Related work

As the two most commonly used topic modelling methods, Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) are widely used [7,8,10]. Blei et al. first introduced the Latent Dirichlet Allocation (LDA) as a generative probabilistic model to collect discrete tokens to provide an explicit representation of a document [7]. Similar to the LDA method, Greene et al. proposed that the NMF method can be used to model topics in documents [8]. However, some studies pointed out the disadvantages of NMF. In particular,

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

Wang et al. demonstrated that NMF based topic modelling may suffer from optimization and high computational complexity issues [9]. Pan also indicate that the Non-Uniqueness of NMF would cause multiple different factorization for a given input [10]. This may lead to the interpretation and comparison of results being more challenging. The feature selection and the result of tokens in topics would also cause differences in results, which causes ambiguity and mislabeling.

In the analysis of ICCS publication activity in 2017, Abuhay et al. used the NMF topic modelling method and classified the corpus into 13 high-level topics [12,13]. The authors found that modelling, HPC and e-science were the most popular topics between 2001 and 2017 [13]. However, as the disadvantage stated from previous research: NMF topic modelling is not unique and requires manually labelling the extracted keywords in the topics. It may cause ambiguity, non-exclusive, and cannot be extended to other fields or subjects. On the other hand, the research focus on computer science has changed since 2017. The Council of Europe and the European Union reported that machine learning and artificial intelligence experienced a rapid increase after 2016 and had a profound impact on society [14,15]. The application and research on machine learning increased rapidly after 2018 [15]. It is necessary to re-evaluate the most popular topics in computational science after 2017 and see if new topics have emerged.

In terms of text similarity comparison, past research indicates the method of comparing the cosine similarity of TF-IDF vector to measure similarity between papers [16,17,18,19]. Gunawan et al. demonstrated that a measure of cosine Similarity could be implemented to classify papers into subject types from text keywords [16]. The study gives a method to classify the research fields of the document from a series of keywords. When comparing highly specialized terminologies or disciplines, a corpus that is standardized into a unique document-term matrix shows its advantages in labeling documents. Wang et al. pointed out the shortcomings of the traditional "Bag of Words" (BoW) representation and introduced a method of using Wikipedia to apply content-based measure to compare the similarity between two texts [19]. Although Wang et al. indicated that Wikipedia's category structure does not form a tree taxonomy, but a directed acyclic graph in which multiple classification schemes coexist simultaneously [19]. This suggests that the method can be improved by building tree structures of domains with parents and subclasses. By applying the TF-IDF vector and the cosine similarity, terminologies within the disciplines of computer science can be compared. Nastase et al. [20] also pointed out that the knowledge base of Wikipedia could be transformed into a large-scale multilingual concept network. On the other side, the Association for Computing Machinery (ACM) classified the entire underlying disciplines of computer and computational science into 17 bodies of knowledge and dozens of sub-disciplines [21]. Curlie also provides related classification libraries in the field of computer science [22].

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

3 Data collection & preprocessing

In this study, the data is divided into two corpora: ICCS corpus and classification corpus. We collected all the papers published in the ICCS proceedings by Springer Lecture Notes in Computer Science (LNCS) (2001 - 2009 & 2018 - 2023), as well as Elsevier Procedia Computer Science from 2010 to 2017 [4,12,13]. The links to published volumes can be found in the conference webpage. The ICCS corpus encompasses 7826 papers over the twenty-three years, which is 340 papers on average each year with an average length of ten pages each. For the text classification corpus, we referred to the computer science curricula for 2023 provided by the Association for Computing Machinery (ACM) [21] and Curlie's outline of computer science [22] to classify computer science into thirteen first-level disciplines and one hundred and two second-level disciplines (see Table 1). We utilized Wikipedia's public API known as English Wikipedia API to extract the textual content as the second-level discipline standard classification library [19,20].

We perform the following pre-processes to the text content:
1) We removed the HTML tags and the unrelated content from the classification corpus to create a standardized classification corpus for each second-level discipline.
2) In the ICCS corpus, we first standardized all the documents into text. We only keep the main content of the papers (from abstract to conclusion & discussion).
3) We removed all the English stop words, punctuation marks, and numbers, which contain no topical information. We extracted and excluded information such as place name, names, organization, etc. which may interfere with the classification process.
4) We chose lemmatisation instead of Porter stemmization to standardise different forms of words used by authors for grammatical reasons without changing the information the word contains [13].

4 The first-level disciplines of ICCS papers

After applying the TF-IDF vectorizer to generate the document-term matrices, we computed the cosine similarity between ICCS papers and the corpus of each second-level discipline. Initially, we examined the topical structure of first-level disciplines. Similar to most topic modelling problems encountered, it is crucial to determine the number of disciplines (K) in each paper. The exclusive classification between the disciplines prevents the disadvantage of numerous highly similar topics. On the other hand, a large number of chosen disciplines will lead to irrelevant topics being included. Consequently, we conducted five preliminary experiments with K = 10, 20, 30, 40, 50 respectively. Based on our experiment results, we decided to adopt K = 20 as it represents a significant number of strongly relevant disciplines (see Table 2). Comparing this result with that from 2017 revealed a high degree of similarity between both methods which further validated our approach [13].

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

Table 1. Discipline classification structure

First-level disciplines	Number of second-level disciplines
Mathematical foundations	8
Modelling and simulation	6
Algorithms and data structures	7
Machine learning & AI	10
Network and security	7
Computer architecture	5
Computer graphics	5
Parallel, and distributed systems	8
Database	5
Programming languages and compilers	10
Scientific computing & Interdisciplinary	14
Software engineering	11
Theory of computation	6

Table 2. Comparison of new results with the previous analysis [12,13]: Pearson correlation coefficient R for different number of topics K. The p-values are indicated by stars: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Old & new disciplines	K = 10	K = 20	K = 30	K = 40	K = 50
Machine Learning	0.252	0.578***	0.161	0.086	-0.310
Network & Security	0.919***	0.919***	0.817***	0.886***	0.809***
HPC	0.803***	0.820***	0.829***	0.832***	0.822***
Programming	0.610***	0.638***	0.636***	0.607***	0.760***

We then calculated the popularity score for each year by aggregating the similarity scores across all first-level disciplines. To ensure comparability among different years' popularity scores, we standardized them into average popularity scores per every 100 papers. After smoothing the data to observe long-term trend by calculating the rolling mean using a window size of 2, thirteen first-level disciplines displays different trends during the past 23 years (see Fig 1). Simultaneously, we calculated and visualized the proportion distribution among these thirteen first-level disciplines (see Fig 2).

We look into some key time points to study the change of proportions of the first-level discipline (See Fig 1 & 2). Combining the two sets of figures, we can draw some results:

1. Machine learning & artificial intelligence did not garner significant attention prior to 2016. It was only in 2017 that topics related to machine learning & artificial intelligence began exhibiting a growth trajectory. In 2019, it surpassed modelling and simulation to emerge as the most prominent topics in ICCS. Fig 2 illustrates an escalating proportion of machine learning & artificial intelligence across the entire corpus, encompassing a substantial share of total topics at approximately 26.2% by 2023.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

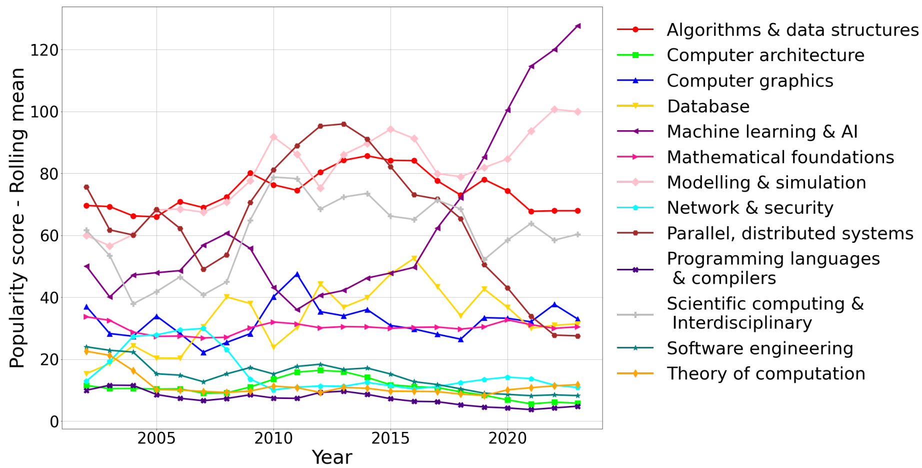

Fig. 1. The trend of popularity scores for the first-level disciplines

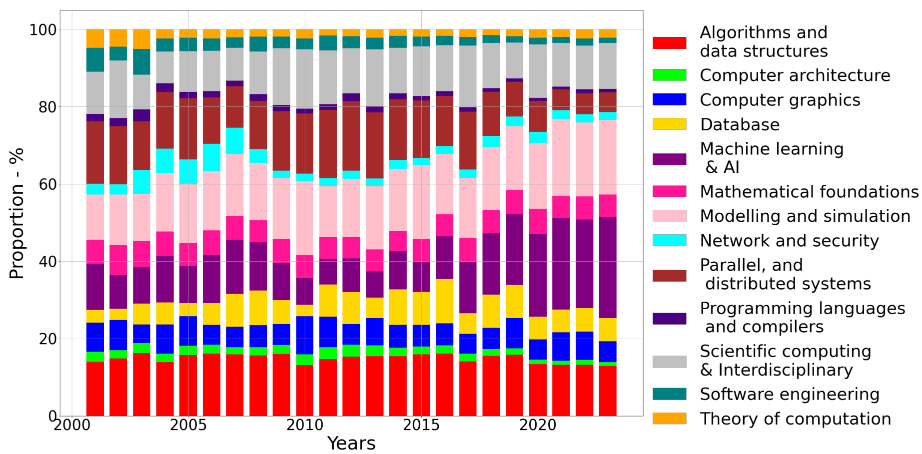

Fig. 2. The percentage of each discipline

2. Parallel and distributed computing has undergone significant transformations over the past 23 years. It exhibited a fluctuating and declining trend from 2001 to 2007, followed by rapid growth after 2008 that culminated in its peak in 2012 at 18%. During the period between 2011 and 2014, it remained the most sought-after topic. However, its proportion gradually decreased after 2015, accounting for only 5.4% of the total research popularity in 2023. This observation aligns closely with the intense competition of supercomputers between 2009 and 2016 when there was a two-order-of-magnitude improvement in the performance of the fastest supercomputer (Jaguar:1.759 PFLOPS - Sunway TaihuLight:93.01

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

PFLOPS) [23]. In other periods, high-performance computing has never experienced such remarkable progress.

3. Under the 23-year time frame, modelling & simulation and algorithms & data structures have always been popular among all the ICCS papers. The proportion illustrates that the two disciplines cover more than 30% of the topic (see Fig 2). It implies the essential contribution of these two disciplines toward computer science and computational science. Among other theoretical disciplines, such as the theory of computation and mathematical foundation, they show a stable trend with no significant increase or decrease.

4. Notably, we notice the Network & Security experienced a rise from 2003 to 2008 and reached its peak in 2007 at 6.9%. This finding aligns with the results of the 2017 study which shows the same sudden increase [12,13]. The explanation is the growing interest in early IPv6 deployment within universities [24]. These academic institutions provided a testing platform for evaluating and pre-commercializing IPv6 products and networks.

5 The second-level disciplines

After analysing the topical trends of the first-level disciplines, we then study the evolution of the second-level disciplines. We first look into the average rank of the second-level disciplines (see Table 3). Throughout the entire span of 23 years, algorithms, numerical analysis, machine learning, mathematical models, and computer simulation emerge as consistently popular research topics that align closely with computational science. Subsequently, we explore the most prevalent secondary disciplines in 2023 (see Table 3). Notably, machine learning and artificial intelligence-related disciplines dominate six positions. This finding corroborates our earlier observations from first-level discipline trends: Machine learning and artificial intelligence-related topics are progressively gaining popularity.

We then analyze the evolution of the second-level disciplines. The disciplines are sorted based on changes in popularity scores, and we rank the ten most increased and decreased second-level disciplines (see Fig 3). According to Fig 3, deep learning shares the highest contribution in research popularity growth within machine learning & artificial intelligence. Deep learning is the most important driving force behind the rapid growth of artificial intelligence. It is followed by cross-validation, machine learning, reinforcement learning, and artificial intelligence. In terms of data science categories, data mining undergoes rapid development from 2005 to 2012. Agent-based modeling (ABM) stabilizes after a period of rapid growth between 2003 and 2010. Concerning scientific computing & interdisciplinary applications, computational social science contributes significantly to the growth of its first-level disciplines. Additionally, mathematical modeling exhibits a strong increase after 2017.

We then look into Fig 3 bottom to explore the top ten decreasing disciplines. The second-level disciplines associated with parallel computing, distributed computing, and supercomputers have witnessed a significant decrease in popularity.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

Table 3. Top 20 second-level disciplines for 23 years and in 2023

Top 20 in 23 years	Average rank	Top 20 in 2023	Rank
Algorithm	2.83	Machine learning	1.00
Numerical analysis	4.13	Deep learning	2.00
Machine learning	4.61	Mathematical model	3.00
Mathematical model	5.83	Artificial intelligence	4.00
Computer simulation	7.96	Algorithm	5.00
Data & information visualization	8.35	Numerical analysis	6.00
Parallel computing	10	Computational sociology	7.00
Computational sociology	10.78	ABM	8.00
Mathematical optimization	10.82	Data & information visualization	9.00
Genetic algorithm	11.26	Data mining	10.00
GPGPU	11.52	Stochastic simulation	11.00
Stochastic simulation	11.95	Mathematical optimization	12.00
ABM	13.48	Cross-validation	13.00
Data mining	13.96	Computer simulation	14.00
Distributed computing	14.57	Scientific modelling	15.00
Computational science	15.70	Genetic algorithm	16.00
Data structure	17.96	Reinforcement learning	17.00
Artificial intelligence	18	Computational chemistry	18.00
Deep learning	20.04	GPGPU	19.00
Database	20.52	Natural language processing	20.00

It is noteworthy that these three disciplines were once the most popular and crucial topics in computer and computational science between 2010 and 2015. However, starting from 2016, this group of research areas has experienced a substantial downturn. In conjunction with Table 3, it becomes evident that the current research focus within parallel and distributed computing lies in general-purpose graphics processing units (GPGPU). Furthermore, we observe a diminishing presence of programming (Computer programming) and operating systems (software engineering), which have declined by 81.06% and 77.88%, respectively. Additionally, topics such as computational complexity and computational geometry enjoyed popularity prior until 2004 but gradually faded away after 2005.

6 The dynamic analysis of networks

We then proceed to investigate the interrelationships between topics. To construct the network of disciplines, we utilize a methodology akin to previous studies involving co-occurrence: if two disciplines are mentioned in the same paper, it signifies a correlation between them [12,13]. Subsequently, we assign weights to edges based on the similarity scores obtained for each discipline and generate twenty-three undirected networks corresponding to twenty-three years. In these networks, nodes represent disciplines, edges signify correlations among disciplines, and edge weights indicate the strength of co-occurrence [25]. Simultaneously, we conduct three experiments by pruning edges below 0.01, 0.05, and

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

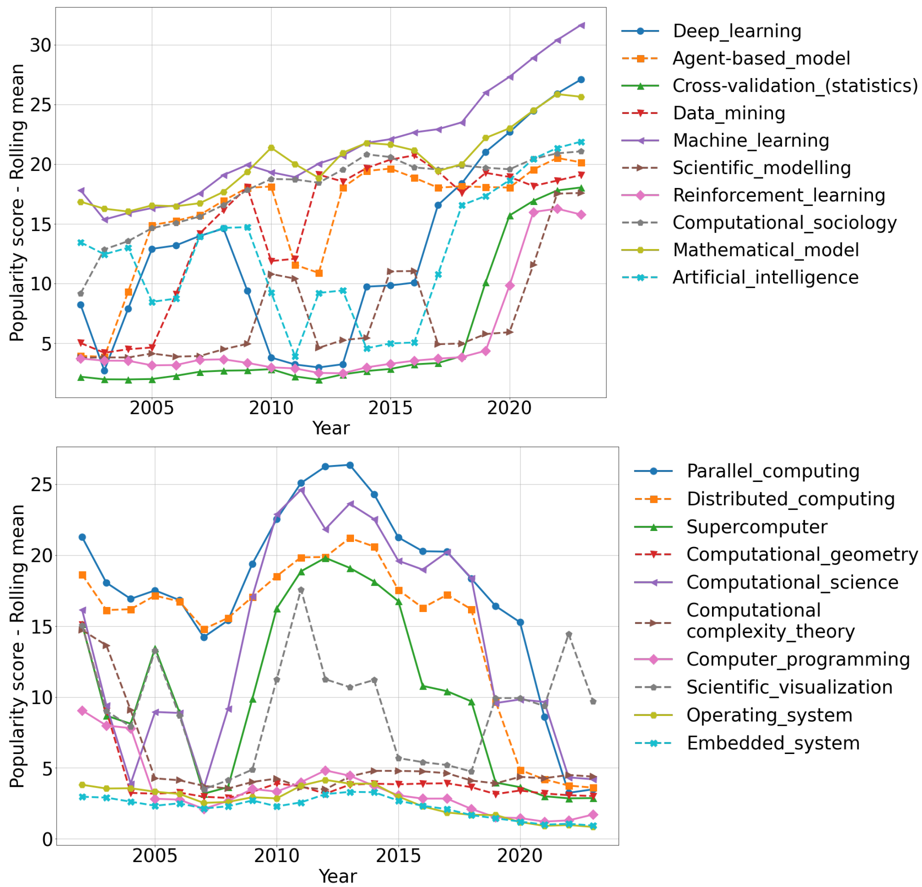

Fig. 3. The top 10 second-level disciplines. Top: rising disciplines, bottom: falling disciplines.

0.09 respectively. We set the threshold at 0.05 for constructing networks based on the distribution of edge weights (see Fig 4). Using Gephi software, we visualize the network matrix of 2022 as an example (see Fig 5). The larger the network node, the higher the research popularity in this disciplines. And a wider edge means a stronger correlation between the two topics.

After analyzing the network structure, we observed that the fraction of nodes with degree k follows a power-law distribution ($f(x) = cx^{-\alpha-1}$) where the exponent parameter $\alpha > 1$, and $p > 0.1$ (see Table 4 & Fig 4). This indicates that the networks are scale-free [26,27]. Additionally, we notice a decreasing trend in both the average clustering coefficient and the total number of edges (see Table 4). Coupled with an increase in α (The exponent of the power), this suggests a decline in network density and decentralization of the network.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

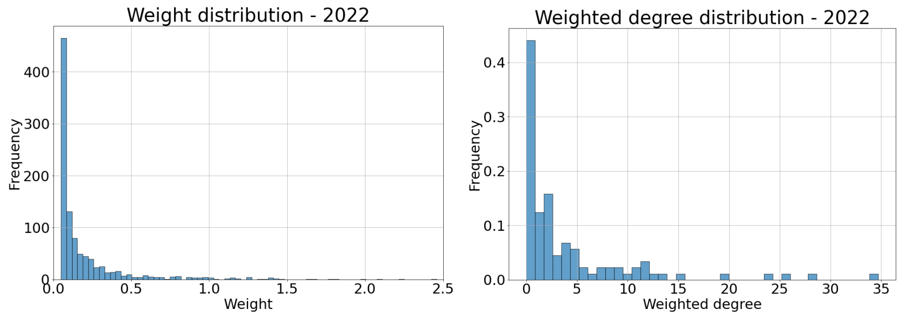

Fig. 4. Network analysis. Left: node degree distribution, right: edge weight distribution for year 2022.

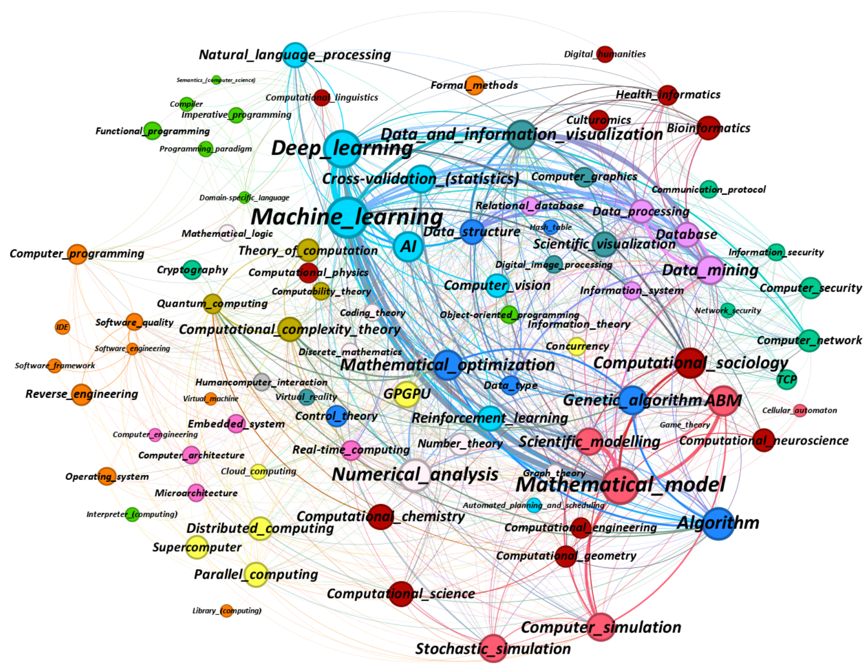

Fig. 5. The visualisation of the network - 2022.

The structure of the network is undergoing changes, and throughout the timeline, there have been shifts in the central points and essential edges within the network. To explore the correlation between various disciplines in the networks, we first look into the trend between first-level disciplines. By regarding the division of first-level disciplines as network communities, Fig 6 shows stan-

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

Table 4. The network structure & node degree distribution(2013 - 2023)

Parameters	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Total edges (N)	1296	1279	1328	1058	1004	823	1025	1095	794	1006	809
Cluster coefficient	0.65	0.67	0.66	0.60	0.62	0.58	0.63	0.62	0.56	0.60	0.56
The exponent α	1.83	1.96	1.77	2.60	1.68	2.24	1.85	3.68	1.80	3.00	3.20
KS statistics	0.124	0.117	0.132	0.117	0.137	0.095	0.103	0.131	0.131	0.103	0.097
p-value	0.417	0.219	0.631	0.762	0.297	0.655	0.854	0.665	0.744	0.982	0.221

dardized average intra-community and inter-community edges of each first-level disciplines [29]. The same result as metrics in Table 4 indicates, the number of both intra- and inter-community edges are decreasing. Each first-level disciplines presents different characteristics. Disciplines such as algorithms & data structure, modelling & simulation, and database show strong external correlations. Fig 6 shows that these three first-level disciplines are the highly connected hubs of multiple disciplines and provide collaborative bridges for multidisciplinary interactions. On the other side, disciplines such as programming languages & compilers, software engineering, and scientific computing & interdisciplinary application present a strong internal correlation. It suggests that these communities are more independent than other communities. Papers in these communities have begun to form a tight internal structure which connect outwards through few key nodes. For example, the only junction between programming languages community and software engineering community is computer programming (see Fig 5).

We then examine the evolution of weighted edges from the first-level disciplines' perspectives. We select 2023 as an example to illustrate the heat map of the network for both first and second-level disciplines (see Fig 7). We found that machine learning and artificial intelligence is highly correlated with other disciplines. Machine learning & AI is actively collaborated with computer graphics and scientific computing & interdisciplinary applications. This indicates that machine learning has significant impact on computer graphics and interdisciplinary application research in 2023. Additionally, the heat maps reveal high activity levels for database, algorithms & simulation, modeling & simulation and theory of computation. From the perspective of computer and computational science, data, algorithms, modelling and computational theory are the foundations of research and the bridge between disciplines. They play a very important role among the research of computer and computational science.

7 Conclusion & Future work

In this paper, we have discussed the changes in popular disciplines within ICCS research over the past 23 years. The most prominent disciplines are machine learning & AI, modelling & simulation, and algorithms. From 2005 to 2008, there was a minor peak in computer network due to the rising interest in

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

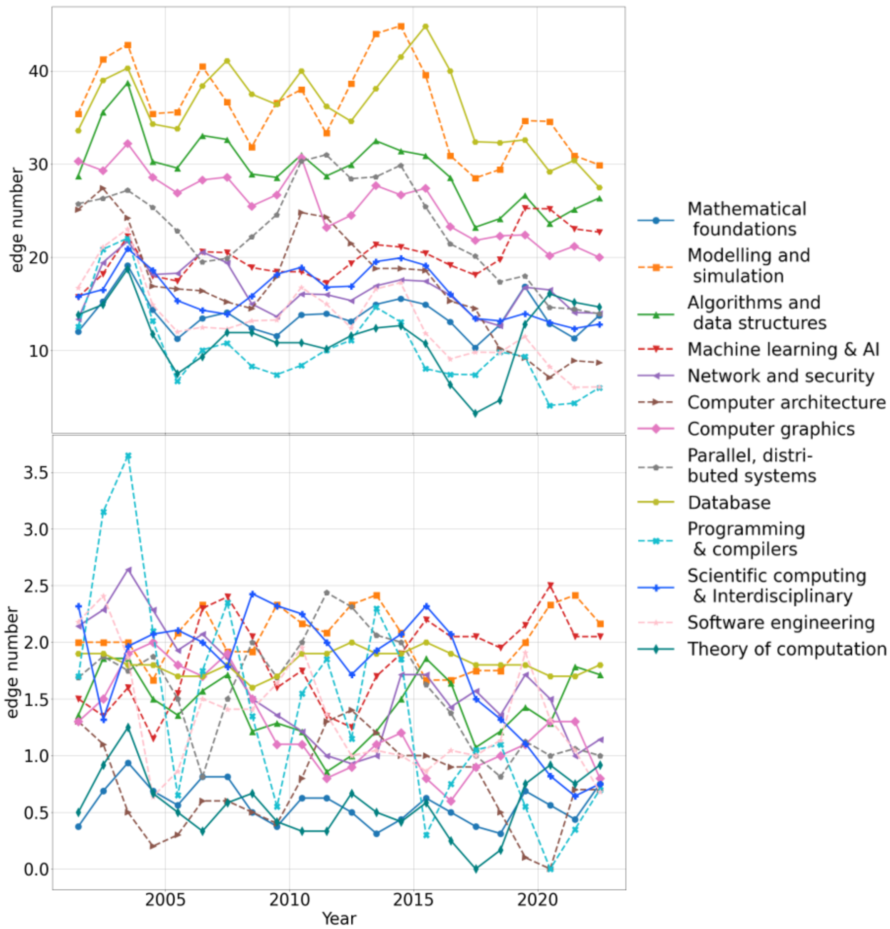

Fig. 6. The standardised N of community edges by first-level disciplines. Top:inter-community Bottom: intra-community

application of IPv6. Between 2009 and 2016, research focused on parallel and distributed computing reached its zenith, which correlates with the rapid growth of supercomputer computing power. Simultaneously, the popularity of parallel and distributed computing on ICCS has declined after 2017. As academia and institutions recognizes 2016 as the rise of artificial intelligence [14,15], we provide evidence that machine learning and artificial intelligence-related research experienced rapid growth thereafter. By 2019, it surpassed modelling and simulation as the most popular discipline.

Our network analysis revealed that the degree distribution of second-level discipline networks are scale-free networks. This also suggests a preferential attachment phenomenon where new research disciplines tend to align with existing

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

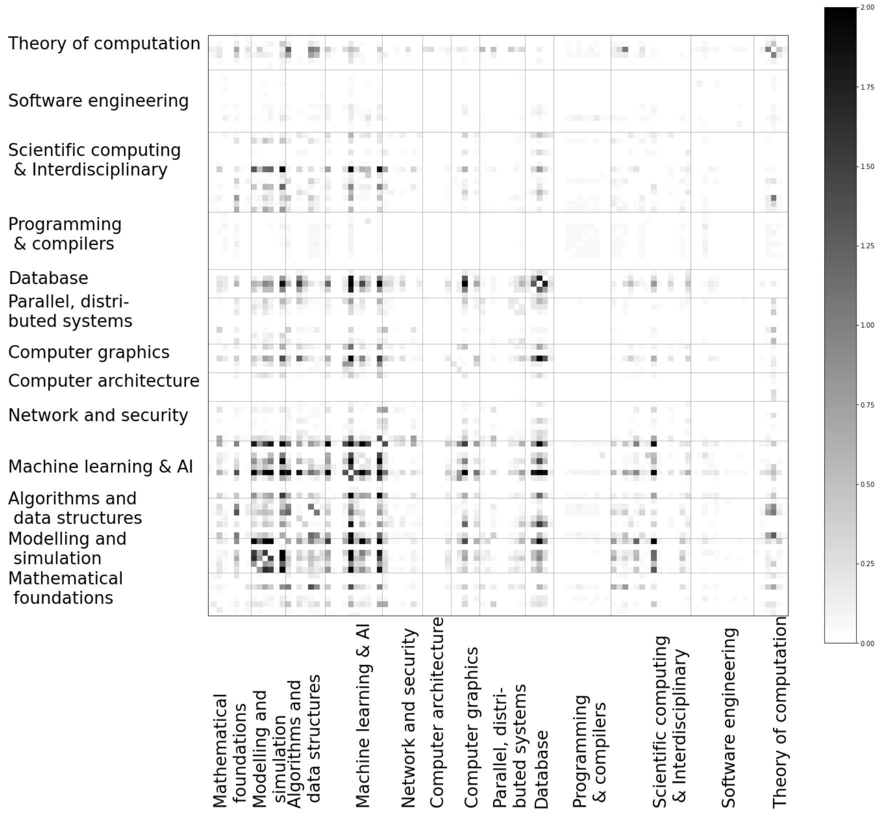

Fig. 7. The heatmap of edge correlations - 2023

popular disciplines. Regarding the network structure itself, the decrease of clustering coefficient and the increase of α imply that the ICCS research network has become more decentralized over the past decade and a few nodes will have more connections. This indicates that a few popular subjects are widely utilised in more research, while the overall cluster structure of the research network is gradually weaken.

At the same time, the network analysis also suggests that machine learning & AI are interacting with multiple disciplines: we see strong correlations with modelling & simulation, algorithms & data structures, computer graphics, and interdisciplinary applications. The communities formed by each first-level discipline also demonstrate different intra- and inter-community characteristics: Algorithms, modelling & simulation and database show strong external and internal correlations, while programming language & compilers and software engineering present strong internal but weak external correlation.

However, we only discussed the correlation rather than the causal relationship between disciplines due to the methodology. Future research can apply more ad-

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

vanced natural language processing techniques to explore the causal relationship for more accurate directed correlations.

Finally, it is important to note that although ICCS holds prominence within the computational science conferences, the results may be biased due to its limited coverage. At the same time, our classification of disciplines can be further refined. The accuracy of the results can be improved by more detailed division of disciplines. A larger corpus with wider coverage could depict the full picture of the entire computational science research. Towards this end, we will analyze the corpus of publications of the Journal of Computational Science (JoCS).

References

1. Santo Fortunato, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, Alessandro Vespignani, Ludo Waltman, Dashun Wang, Albert László Barabási: Science of science. Science **359**(6349), (2018). https://doi.org/10.1126/science.aao0185
2. John Mingers, Loet Leydesdorff: A review of theory and practice in scientometrics. European Journal of Operational Research **246**(1), 1–19 (2015). https://doi.org/10.1016/j.ejor.2015.04.002
3. Dashun Wang, Albert-László Barabási: The Science of Science. The Science of Science (2021). https://doi.org/10.1017/9781108610834
4. Derek Groen, Clélia De Mulatier, Maciej Paszynski, Valeria V Krzhizhanovskaya, Jack J Dongarra, Peter M A Sloot: The Computational Planet. Preface to ICCS 2022. https://www.iccs-meeting.org/iccs2022/
5. Maciej Paszynski, Dieter Kranzlmüller, Valeria V Krzhizhanovskaya, Jack J Dongarra, Peter M A Sloot: Computational Science for a Better Future. Preface to ICCS 2021. https://www.iccs-meeting.org/iccs2021/
6. Valeria V Krzhizhanovskaya, Gábor Závodszky, Michael Lees, Jack Dongarra, Peter M A Sloot, Sérgio Brissos, João Teixeira: Twenty Years of Computational Science. Preface for ICCS 2020. https://www.iccs-meeting.org/iccs2020/
7. David M Blei, Andrew Y Ng, Jordan: Latent Dirichlet Allocation Michael I. Jordan. Journal of Machine Learning Research **3**, 993–1022 (2003)
8. Derek Greene, James P. Cross: Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach. (2016) http://arxiv.org/abs/1607.03055
9. Jianyu Wang, Xiao Lei Zhang: Deep NMF topic modeling. Neurocomputing **515**, 157–173 (2023). https://doi.org/10.1016/j.neucom.2022.10.002
10. W. Pan, F. Doshi-Velez: A Characterization of the Non-Uniqueness of Nonnegative Matrix Factorizations. http://arxiv.org/abs/1604.00653
11. Sang Woon Kim, Joon Min Gil: Research paper classification systems based on TF-IDF and LDA schemes. Human-centric Computing and Information Sciences **9**(1), (2019) . https://doi.org/10.1186/s13673-019-0192-7
12. Tesfamariam M. Abuhay, Sergey V. Kovalchuk, Klavdiya O. Bochenina, George Kampis, Valeria V. Krzhizhanovskaya, Michael H. Lees: Analysis of Computational Science Papers from ICCS 2001-2016 using Topic Modeling and Graph Theory.Procedia Computer Science **108**, 7–17 (2017). https://doi.org/10.1016/j.procs.2017.05.183

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2

13. 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 modelling and graph theory. Journal of Computational Science **26**,193–204 (2018). https://doi.org/10.1016/j.jocs.2018.04.004
14. European Union: Artificial intelligence, blockchain and the future of Europe: How disruptive technologies create opportunities for a green and digital economy (2021) https://www.eib.org/attachments/thematic/artificial_intelligence_blockchain_and_the_future_of_europe_report_en.pdf
15. The council of Europe: History of Artificial Intelligence (2024) https://www.coe.int/en/web/artificial-intelligence/history-of-ai
16. D. Gunawan, C. A. Sembiring, M. A. Budiman: The Implementation of Cosine Similarity to Calculate Text Relevance between Two Documents. Journal of Physics: Conference Series **978**(1), (2018) . https://doi.org/10.1088/1742-6596/978/1/012120
17. Baoli Li and Liping Han: LNCS 8206 - Distance Weighted Cosine Similarity Measure for Text Classification. LNCS, **8206**(1), 611-618 (2013).
18. Ritika Singh, Satwinder Singh: Text Similarity Measures in News Articles by Vector Space Model Using NLP. Journal of The Institution of Engineers (India): Series B **102**(2), 329–338 (2021). https://doi.org/10.1007/s40031-020-00501-5
19. Pu Wang, Jian Hu, Hua Jun Zeng, Zheng Chen: Using Wikipedia knowledge to improve text classification. Knowledge and Information Systems **19**(3), 265–281.Springer London (2009). https://doi.org/10.1007/s10115-008-0152-4
20. Vivi Nastase, Michael Strube: Transforming Wikipedia into a large scale multilingual concept network. Artificial Intelligence **194**, 62–85 (2013). https://doi.org/10.1016/j.artint.2012.06.008
21. Association for Computing Machinery (ACM), IEEE-Computer Society (IEEE-CS), Association for Advancement of Artificial Intelligence (AAAI): Computer Science Curricula 2023. (2023) https://csed.acm.org/wp-content/uploads/2023/09/Version-Gamma.pdf
22. Curlie: Computers. https://curlie.org/en/Computers/
23. Top500: NOVEMBER 2017. (2018) https://www.top500.org/lists/top500/2017/11/
24. Lorenzo Colitti, Steinar H Gunderson, Erik Kline and Tiziana Refice: Evaluating IPv6 Adoption in the Internet. PAM 2010 (2010) http://www.pam2010.ethz.ch/papers/full-length/15.pdf
25. Jiming Hu, Yin Zhang: Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization. Scientometrics **112**(1), 91–109 (2017). https://doi.org/10.1007/s11192-017-2383-1
26. Anna D. Broido, Aaron Clauset: Scale-free networks are rare. Nature Communications **10**(1), (2019) . https://doi.org/10.1038/s41467-019-08746-5
27. Jinho Choi, Sangyoon Yi, Kun Chang Lee: Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. Information and Management **48**(8), 371–381 (2011). https://doi.org/10.1016/j.im.2011.09.004
28. Intel: Building Optimized High Performance Computing (HPC) Architectures and Applications. https://www.intel.com/content/www/us/en/high-performance-computing/hpc-architecture.html
29. Yoonsuk Kang, Jun-Seok Lee, Won-Yong Shin, Sang-Wook Kim: Community reinforcement: An effective and efficient preprocessing method for accurate community detection. Knowledge-Based Systems **237**(8), 107741 (2022). https://doi.org/10.1016/j.knosys.2021.107741

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_2