Argumentation interfaces to facilitate human-machine collaboration in scientific discovery: A preliminary exploration

Yuanxi Fu[1,*], Jodi Schneider[1]

[1]*School of Information Sciences, University of Illinois at Urbana-Champaign, 501 E. Daniel Street, Champaign, IL 61820 USA*

Abstract

Computational methods are pillars of today's data- and AI-driven scientific enterprise, but their misuse trouble many scientists. We propose an application called an argumentation interface to improve the quality of a scientist's application of computational methods to their task. Using a toy example, we demonstrated how an abstract argumentation framework can be used to model the reasoning behind choosing one method over another for ranking players in a chess tournament. Many hypotheses need to be tested to prove that the argumentation interface is feasible, and we start with two fundamental ones: One concerns whether we can use the scientific literature as a source of information, and the other concerns the applicability of argumentation frameworks. We plan to use a case involving two community detection algorithms—the Louvain algorithm and the Leiden algorithm—to test them. Preliminary results show that the scientific literature is permeated with uncritical usage of computational methods, which justified our motivation for developing the argumentation framework but cast doubt on whether an unfiltered scientific literature can be used as a source of information. Further analysis of the case is underway.

Keywords

argumentation in science, argumentation interface, argument models, methods, Argumentation Framework, Louvain algorithm, Leiden algorithm, community detection algorithms

1. Introduction

Computational methods are pillars of today's data- and AI-driven scientific enterprise [1], but their misuse trouble many scientists [2, 3, 4, 5]. Particularly, easy access to software packages and the society-wide enthusiasm towards AI fuel the chance of misuse since many scientists are eager to adopt new methods without educating themselves to a sufficient level of competence[5]. Moreover, flaws within a method can surface years after the method's wide adoption [6, 7], which creates two problems: first, how to assess results produced without knowledge about the flaw [8], and second, how to share the information about uncovered flaws with new generations of users in a timely manner.

We propose an application called an argumentation interface that mediates between a scientist and one or several computational methods to improve the quality of the scientist's application of computational methods to their task. First, we use a toy example to illustrate the potential of argumentation research and prepare readers to understand the design of our argumentation interface. Second, we outline the argumentation interface we envision and list two fundamental hypotheses we want to test first. Lastly, we describe the preliminary work we have done using a case of community detection algorithms to test our hypotheses.

CMNA'24: The 24th International Workshop on Computational Models of Natural Argument, September 17, 2024, Hagen, Germany
*Corresponding author.
fu5@illinois.edu (Y. Fu); jodi@illinois.edu (J. Schneider)
https://yuanxifu.site/ (Y. Fu); https://jodischneider.com/jodi.html (J. Schneider)
0000-0003-2726-0999 (Y. Fu); 0000-0002-5098-5667 (J. Schneider)

2. A toy example to illustrate the potential of argumentation research

2.1. Landau's PageRank

In 1895, German mathematician Edmund Landau proposed a new method to rank players in a round robin chess tournament. In this type of tournament, each player must play with everyone else. Traditionally, a player's score at the end was calculated by (1) giving 1 point to each winner, 0 to each loser, and 1/2 to each player in a draw and (2) summing up all points. Landau sensed a fairness issue with such a naive scoring system: players who won against stronger players should get more points than those who won against weaker players. He proposed "relative Wertbemessung" (relative score) as an alternative, which in mathematical essence is the same as Page and Brin's famous PageRank algorithm for ranking webpages. Thus, we will call it "Landau's PageRank" for easy remembrance. However, Landau also pointed out that his method, although fairer in evaluation, introduced the incentive for players to strategically play NOT to win. He used a three-player scenario to illustrate his concern, where a player can be better off at the end by not winning certain games.

2.2. Modeling arguments in Landau's PageRank case with the abstract argumentation framework

The story behind Landau's PageRank was described by Sinn and Zeigler [9]. We the arguments regarding choosing between the traditional scoring system and Landau's PageRank, which can be found in pages 1 to 3 of [9], into an abstract argument framework (AF) [10] shown in Figure 1. We also added one argument of our own. Landau's concern about manipulation was illustrated by a three-player scenario. However, when the number of players becomes sufficiently large (although we do not know the bound), it will be difficult to straticially play not to win. The reasons are twofold. The first is about randomness and chance: A player cannot always control the outcome of the games even if they do not try to win. Second is complexity: A player must also consider the outcomes of other games, which is likely out of reach for humans. We added this argument into Figure 1.

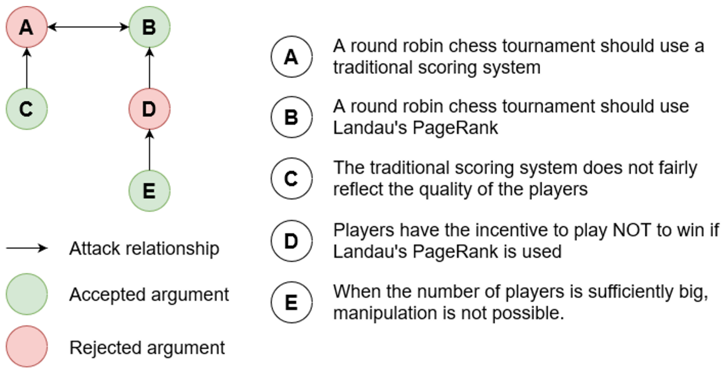

Figure 1: Using an abstract argumentation framework (AF) to represent arguments regarding choosing between the traditional scoring system and Landau's PageRank for a round robin chess tournament

Figure 1 shows one solution of the argument framework matching our preferences. Since argument **E** is proposed by us, we view it as accepted, and we also agree with Landau's concern about fairness (i.e., accepting Argument **C**). In this solution, we accept the method update (Argument **B**) from the common scoring system to Landau's PageRank. This is only a toy example since we only considered a small number of arguments found in one paper [9] and one argument we came up with (Argument **E**). However, this toy example demonstrates that AF has the potential to structure and visualize the underlying reasoning of method choices for a specific task.

2.3. An argumentation interface to improve the quality of a scientist's application of computational methods to their task

We propose an argumentation interface to improve the quality of a scientist's application of computational methods to their task (Figure 2). The argumentation interface mediates between a pool of computational methods and a scientist. The interface takes relevant information about a method (1) from the documents describing the computational method and (2) from the scientific literature where the utilization and critiques of the method may be described. It also converse with the scientist to learn about the characteristics of their task. At the end, the argumentation interface assembles all information into an AF, visualize AF, and provide a summary of suggestions. When the AF can arrive at a stable solution [10] that accomendate all of user's belief (e.g., our Argument E) as well as hard-coded rules (e.g., users should not commit statistical fallcies). It will provide the solution to the user. We do not yet know what to do if a stable solution cannot be arrived but the scientist must proceed with one method. A possibility is to make compromises in the AF and state the compromises as limitations in the manuscript.

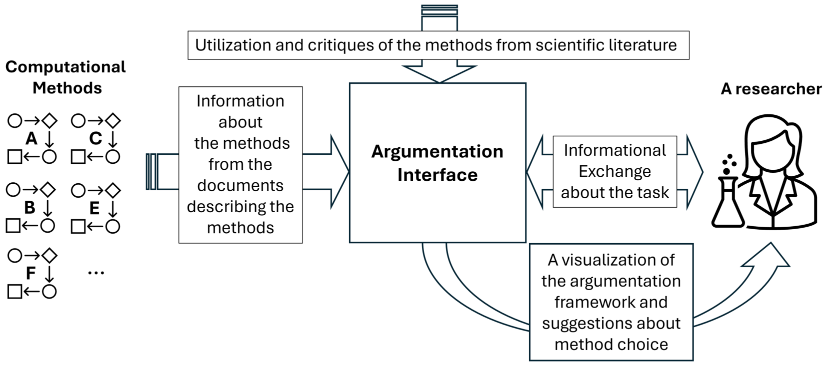

Figure 2: A sketch of the argumentation interface

Many hypotheses need to be tested to prove that the argumentation interface is feasible. We start with two fundamental ones, one concerns the source of information (i.e., the scientific literature) and the other concerns the applicability of argumentation frameworks. We will use a case involving community detection algorithms to test them.

H1 The scientific literature can provide essential information to support rational choice of computational methods, such as tasks where a method was applied to, the conditions for utilization of a method, and the critiques of a method.

H2 The reasoning in choosing a computational method for a research task can be modeled by AFs.

3. Testing our hypotheses with a case

3.1. The case: from Louvain to Leiden

Community detection algorithms are a class of network science methods widely used by multiple scientific disciplines [11]. A community detection algorithm detects communities in a network (also called a graph). Currently, more than 50 community algorithms exist [12]. We limit the subject of study to two prominent community detection algorithms, the Louvain algorithm [13] and the Leiden algorithm [7].

The Louvain algorithm was proposed in 2008 by Blondel et al. [13] and has received more than 11000 citations as of August 2024. The Leiden algorithm, proposed in 2019 by Traag et al. [7], presented a necessary update due to a significant flaw uncovered in the Louvain algorithm: It can produce arbitrary and disconnected “communities”. As of August 2024, the Traag et al. paper has received more than 1700 citations.

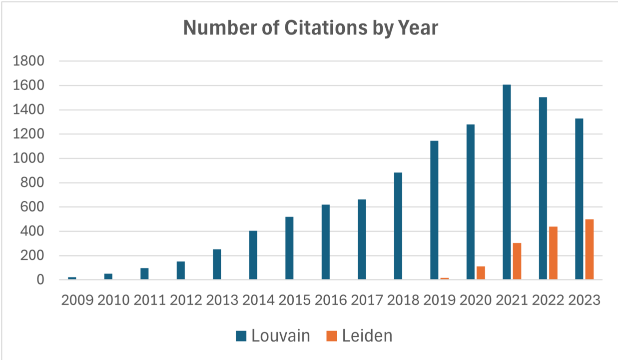

Figure 3: Number of citations per year of the documents describing the Louvain[13] and Leiden algorithms [7] from 2009 to 2023.

We retrieved all citations of Blondel et al., 2008 [13] and Traag et al., 2019 [7] from Web of Science on August 21, 2024. We will refer to these two paper as “the Louvain paper” and “the Leiden paper” from now on. Figure 3 shows that citations of the Louvain paper peaked in 2021 while citations of the Leiden paper are still growing. Yet, the Louvain paper’s total number of annual citations still surpasses the Leiden paper’s. They are also cited by quite different communities. The top five most common Web of Science categories citing the Louvain paper are: Computer Science Information Systems (17.0%), Computer Science Artificial Intelligence (14.0%), Computer Science Theory Methods (13.4%), Multidisciplinary Sciences (10.4%), and Engineering Electrical Electronic (9.8%). In contrast, citations of the Leiden paper concentrates in the biology and interdiciplinary domains. The top five most common Web of Science cateogries are: Multidisciplinary Sciences (19.3%), Cell Biology (10.8%), Biochemical Research Methods (8.2%), Mathematical Computational Biology (8.0%), and Computer Science Information Systems (8.0%).

3.2. Data and method

We collected from Web of Science 483 publications citing both the Louvain and the Leiden paper since 2019 on August 21, 2024. Using this corpus, we will test H1 and H2. However, we only have preliminary work on testing H1.

We identify the top 3 most common Web of Science categories for the 484 publications: Multidisciplinary Sciences, Computer Science, Artificial Intelligence, and Biochemical Research Methods. We randomly sample one research paper (document type as “article”) and one review paper (document type as “review”) from each Web of Science category. We made this differentiation because we suspect these two types provide different information. The “article” type will be more likely to provide concrete information about method utilization whereas “review” type publications critiques of a method. Because no “review” type publication was retrieved under “Multidisciplinary Sciences” category, we sampled and analyzed 5 publications instead of 6. We describe our preliminary findings in section 3.3.

3.3. Preliminary results

Three publications applied the Louvain and Leiden algorithms to discern spatial gene expression patterns [14, 15, 16]. One applied them to cluster cells [17], and one used them for scientometrics analysis [18]. Only one publication [18] mentioned the difference between the Louvain and the Leiden algorithms using information from the Leiden paper [7]. Four publications mentioned [14, 17, 15, 16] Louvain and Leiden as if they are equals, although one of them [14] applied Louvain and Leiden to different datasets without justifying why.

These findings reveal a concerning lack of criticality in computational methods application. Authors gave no justification for why Louvain and Leiden were chosen amongst many different types of community detection algorithms. The only critique about the two algorithms we found was inherited from the Leiden paper. While our findings justified our motivation to develop the argumentation interface, it cast doubts on whether unfiltered scientific literature can be used as a source of information as it is possibly permeated with uncritical usage of computational methods. Further analysis of the corpus is underway.

4. Conclusions

We proposed argumentation interface to address the issue of computational methods misuse and improve the quality of a scientist's application of computatational methods to their research task. We sketched out the design of the argumentation interface whose feasiblity will depend on two hypotheses: (H1) the scientific literature can provide essential information to support rational choice of computational methods and (H2) the reasoning in choosing a computational method for a research task can be modeled by AFs. We show very prelimnary work in testing H1. Our findings reveal a concerning lack of criticality in computational methods application in the scientific literature, which cast doubt on whether the unfiltered scientific literature can be used as a source of information to support rational methods choices. Yet it also justified our motivation to develop the argumentation interface.

Acknowledgments

Ideation of this work came from research supported by Alfred P. Sloan Foundation (G-2020-12623; G-2022-19409). YF thanks Prof. Scott Jacobs and Prof. Sally Jackson for their introductory courses on argumentation theories that lay the foundation for this work. Thanks to Prof. Michael Twidale and Dr. Heng Zheng for comments. Thanks to Prof. Bertram Ludäscher for providing the essential Landau's PageRank case.

References

[1] C. Goble, S. Cohen-Boulakia, S. Soiland-Reyes, D. Garijo, Y. Gil, M. R. Crusoe, K. Peters, D. Schober, FAIR computational workflows 2 (2020) 108–121. doi:10.1162/dint_a_00033.
[2] L. Peel, T. P. Peixoto, M. De Domenico, Statistical inference links data and theory in network science, Nature Communications (2022). doi:10.1038/s41467-022-34267-9.
[3] J. Gardenier, D. Resnik, The misuse of statistics: concepts, tools, and a research agenda 9 (2002). doi:10.1080/08989620212968.
[4] C. L. Soskolne, S. Kramer, J. P. Ramos-Bonilla, D. Mandrioli, J. Sass, M. Gochfeld, C. F. Cranor, S. Advani, L. A. Bero, Toolkit for detecting misused epidemiological methods 20 (2021) 90. doi:10.1186/s12940-021-00771-6.
[5] J. Calder, R. Coil, J. A. Melton, P. J. Olver, G. Tostevin, K. Yezzi-Woodley, Use and misuse of machine learning in anthropology 2 (2022) 102–115. doi:10.1109/MBITS.2022.3205143.
[6] J. B. Neupane, R. P. Neupane, Y. Luo, W. Y. Yoshida, R. Sun, P. G. Williams, Characterization of Leptazolines A–D, polar oxazolines from the cyanobacterium Leptolyngbya sp., reveals a glitch with the "Willoughby–Hoye" scripts for calculating NMR chemical shifts, Organic Letters 21 (2019) 8449–8453. doi:10.1021/acs.orglett.9b03216.
[7] V. A. Traag, L. Waltman, N. J. van Eck, From louvain to leiden: guaranteeing well-connected communities 9 (2019) 5233. doi:10.1038/s41598-019-41695-z.
[8] Y. Fu, J. Schneider, Towards knowledge maintenance in scientific digital libraries with the keystone framework, in: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, JCDL '20, Association for Computing Machinery, 2020, pp. 217–226. doi:10.1145/3383583.3398514.

[9] R. Sinn, G. M. Ziegler, Landau on chess tournaments and google's PageRank (2022). URL: http://arxiv.org/abs/2210.17300.

[10] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games 77 (1995) 321–357. doi:10.1016/0004-3702(94)00041-X.

[11] S. Fortunato, Community detection in graphs 486 (2010) 75–174. URL: https://www.sciencedirect.com/science/article/pii/S0370157309002841. doi:10.1016/j.physrep.2009.11.002.

[12] D. Jin, Z. Yu, P. Jiao, S. Pan, D. He, J. Wu, P. S. Yu, W. Zhang, A survey of community detection approaches: from statistical modeling to deep learning 35 (2023) 1149–1170. doi:10.1109/TKDE.2021.3104155.

[13] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks 2008 (2008) P10008. doi:10.1088/1742-5468/2008/10/P10008.

[14] C. Zhang, X. Li, W. Huang, L. Wang, Q. Shi, Spatially aware self-representation learning for tissue structure characterization and spatial functional genes identification 24 (2023) bbad197. doi:10.1093/bib/bbad197.

[15] S. Prabhakaran, C. Gatenbee, M. Robertson-Tessi, J. West, A. A. Beg, J. Gray, S. Antonia, R. A. Gatenby, A. R. A. Anderson, Mistic: An open-source multiplexed image t-SNE viewer 3 (2022) 100523. doi:10.1016/j.patter.2022.100523.

[16] L. Wang, M. Maletic-Savatic, Z. Liu, Region-specific denoising identifies spatial co-expression patterns and intra-tissue heterogeneity in spatially resolved transcriptomics data 13 (2022) 6912. doi:10.1038/s41467-022-34567-0.

[17] S. K. Ata, M. Wu, Y. Fang, L. Ou-Yang, C. K. Kwoh, X.-L. Li, Recent advances in network-based methods for disease gene prediction 22 (2021) bbaa303. doi:10.1093/bib/bbaa303.

[18] J.-V. Rodríguez, I. Rodríguez-Rodríguez, W. L. Woo, On the application of machine learning in astronomy and astrophysics: A text-mining-based scientometric analysis 12 (2022) e1476. doi:10.1002/widm.1476.