Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250331

A Semi-Automated Approach Based on Network Analysis to Suggest New Collaborations and Foster Multisite Clinical Trials

Giovanna NICORA[a,1], Enea PARIMBELLI[a], Eva BRIONNE-BORDA[a], Irene Giovanna APRILE[b], and Silvana QUAGLINI[a]
[a] *Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy*
[b] *IRCCS Fondazione Don Carlo Gnocchi ONLUS, Florence, Italy*

Abstract. Several research institutions nowadays collaboratively conduct many scientific projects. Within a national Italian initiative on robotic rehabilitation, this study aims to develop new collaborations that can support the project's missions. Bibliographic information from previous publications realized by the centers that are part of the initiative allowed us to model existing partnerships among research institutions as a network of collaboration. We then designed and implemented an algorithm able to suggest new collaborations among the institutions based on the bibliograph-ic-based network. The results show that using previous publications to suggest new collabora-tions between research centers can be a promising way to develop new partnerships within a group of research centers. However, the number of suggestions can be significant, so criteria should be defined to prioritize the relevant suggestions according to the needs.

Keywords. Collaborative Network, Recommendation, Rehabilitation, Robotics

1. Introduction

As of today, many research projects are conducted by different institutions that collaborate together on topics of interest. Scientific collaboration is an interaction between several scientists and fosters the generation of new practices, rules, and technologies [1]. Various approaches have been proposed to analyze collaborative efforts, including social network and co-authorship analysis tools [2], [3]. Network analysis can be performed at different levels: at a macroscopic one, researchers have studied the evolution of scientific collaboration across the world [4], or at a microscopic level, for example [5], where authors describe the collaboration of researchers of the Italian Institute of Technology. Identifying potential collaborators with similar or complementary research interests can be a major problem, especially when researchers come from many different centers. While different works analyze existing collaborative networks [4, 5], few studies focus on the development of methods to suggest new collaboration. The aim of this paper is to describe an approach that leverages information

[1] Corresponding Author: Giovanna Nicora; E-mail: giovanna.nicora@unipv.it.

extracted from peer-reviewed publications to recommend new scientific collaborations in an existing Consortium. In particular, new collaborations will be suggested to research institutions taking part in the Italian initiative Fit4MedRob. The initiative aims to revolutionize current Italian rehabilitative protocols using robots and allied technologies by implementing pragmatic multi-center clinical trials that will evaluate the effectiveness of rehabilitation enhanced by commercially available robotics devices in comparison with standard care. Different patient populations will be recruited across the various trials, from stroke patients to children with disabilities. Currently, the Fit4MedRob consortium consists of 11 research centers, 11 healthcare institutions for a total of 24 clinical centers (some institutions have centers spread throughout the country), and 3 private companies (https://www.fit4medrob.it). The full potential of expertise within the Consortium must be exploited through intense collaborations. For example, we aim to promote collaboration within clinical centers working on the same patient population (for instance, stroke), or with bioengineers working on innovative solutions and healthcare professionals that can test those solutions. In addition, as many centers have international collaborations, new institutions outside Italy could be involved. Although Fit4MedRob is an Italian Initiative, its vision is supported by a "twin" flagship Initiative, called Swiss Neurorehab [https://www.swissneurorehab.ch/]. Therefore, the method proposed in this paper can be useful to identify potential collaborators for future projects.

Here, we have developed a simple method that can recommend new collaborations (a) among centers part of the Fit4MedRob consortium and (b) between Fit4MedRob centers and Italian institutions that are not part of the Consortium, but whose expertise is aligned with the Initiative, and that could be involved in further phases of the Initiative. To do so, we have created a network gathering information from past publications from researchers in the Consortium. We extracted keywords and authors affiliations, and we created a network linking together affiliations that worked together. Also, affiliations (i.e., centers) are linked to keywords, that represent the various fields of study. New links among institutions are suggested based on (a) common fields of study or (b) a common collab-orator. A programming solution has been developed to reach this goal and an analysis of the resulting suggestions has been realized.

2. Materials and Methods

At the beginning of project, each center was asked to provide a list of papers, relevant for the initiative, published in the last five years by its researchers. The survey was implemented using Kobotoolbox (https://www.kobotoolbox.org/) and collected 426 entries. We used the Application Programming Interface (API) Pybliometrics [6] to query Scopus and to extract the affiliations of the authors and the keywords from each publi-cation's DOI (see https://github.com/GiovannaNicora/recommend_colab_network). Using these data, the program defines all the existing links between two centers (which means that the centers have collaborated to write a paper) or between a center and a keyword (which means that a center worked on that particular subject). If the same link appears several times, the weight of the link will be increased. The links are undirected because they represent collaboration. Therefore, the more two institutions have published papers together, the higher the weight, since the weight represents the number of papers in which the two institutions appear as co-authors. If we consider the link between an institution and a keyword, highest weights indicate that the center has work heavily on that specific subject. Using the software Gephi [7], an analysis of the collaborative network obtained was realized, where each node represented either a center or a keyword. We then suggested new collaborations between two node centers if they have a common link to a third node, which can be either a center or a keyword. To reduce the computational time due to the high number of links and nodes of the dataset, we exploited the simmetry of the table of existing links, and we considered only the triangular matrix. Manual pre-processing of the retrieved affiliations and keywords was needed to remove inconsistencies such as wrong affiliation names. Filters have been created to remove all the links that are not relevant to the project: only suggestions of links between two centres have been kept.

3. Results

3.1. General analysis of the Fit4MedRob collaborative network

The resulting network has 1750 nodes, each representing a center or a keyword, and 17499 edges. An edge between two centers indicates that researchers from the two centers have been co-authors in at least one publication. An edge between a center and a keyword means that the researcher(s) in that center had worked on the specific topic represented by the keyword. The number of connections that a node has to other nodes is the degree of the node. We then performed the community detection using Gephi. A network has a community structure if its nodes can be easily grouped into (potentially overlapping) sets of nodes that are densely connected internally. The color of the nodes in Figure 1 reflects the community detection realized, based on graph modularity. Modularity measures how easily a network can be separated into communities. A network with a high modularity indicates strong relationships within a community and weaker ones across different communities. In our case, 13 communities have been detected and the modularity equals 0.457. Modularity is a measure of the network structure designed to measure the strength of division of a network into submodules, and can assume values between -1 and 1. This positive value indicates that the number of edges within communities exceeds the number expected on the basis of chance.
Through network analysis, we determined the nodes that have many collaborations. There are 731 "center" nodes and 8896 edges between any two centers. In our case, 462 nodes (63.2%) have more than 10 relations. Half of the nodes have less than 15 relations. The average degree is 20 and the center that has the highest number of relations is Fondazione Don Gnocchi – Milan, with a degree equal to 326.

3.2. Suggestion of new collaborations

After applying the approach described in the Methods section, we were able to suggest 677 new collaborations. To focus on the collaboration needs of Fit4MedRob, the results obtained have been considered regarding two perspectives: the suggestions including the centers that are part of the Initiative and the suggestions generated from a node that represents a disease of interest of the project. In Table 1, a summary of the results are presented. For three diseases of interest, we have selected nodes that were referring to this disease. For example, for the keyword "stroke" and the clinical center IRCCS Maugeri (located in Pavia, Italy), the algorithm suggests 111 potential collaborations, 3 with centers that are already part of the Initiative (Inail, Fondazione Campus Biomedico

Roma, and Fondazione Don Gnocchi Firenze). This result highlights that the aforementioned Fit4MedRob centers have experience with stroke, and they might benefit from a collaboration, for example in multicenter clinical trials where stroke patient recruitment is required.

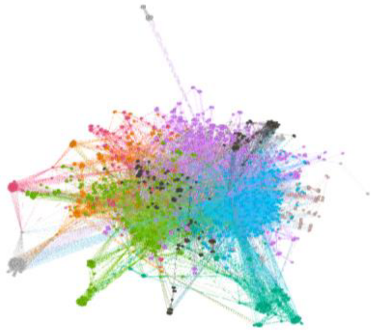

Figure 1. Representation of the Fit4MedRob network, obtained with Gephi.

Table 1. Number of suggestions available for each disease of interest of the initiative

Disease of interest corresponding to a node	# of suggestions	# of suggestions that includes at least 1 center of the Initiative
Stroke	14636	1407
Multiple sclerosis	8089	689
Parkinson disease	2561	319

4. Discussion and Conclusions

This study presents a simple approach to help identify partners and reference centres with experience in the same pathology. The proposed technique is obviously useful when dealing with well-studied diseases, but it is especially crucial when studying rare or understudied pathologies. The Fit4MedRob Initiative is probably an excellent instance of how to carry out this type of study given its enormous size. Indeed collaborative networks are a powerful tool for studying the interactions between research centers and researchers within an organization, particularly in large-scale, interdisciplinary projects like Fit4MedRob. Our approach not only maps existing collaborations but also suggests new, potentially fruitful partnerships within a group of research centers. For example, our algorithm identified 10 new collaboration opportunities within the context of amputation research, with 4 of these involving a center already part of the Fit4MedRob initiative. These suggestions are grounded in the fact that while the institutions have previously published work on amputation, they have not yet collaborated directly on this topic, making these potential partnerships particularly relevant in multicenter clinical trials, where their collaboration can allow for the recruitment of a high number of patients. To enhance the practical utility of these suggestions, we propose visualizing the new collaborations through a geographical network, facilitating easier identification and exploration of these opportunities. However, the main challenge encountered in this project is the inconsistency in the naming conventions for research centers, as extracted from the authors' affiliations in the publications. Available APIs could extract incomplete affiliations on some papers. This issue necessitated a manual preprocessing step to standardize the data, leading to a semi-automatic approach. Moving forward, the development of standardized naming conventions for affiliations is crucial to reduce processing time and minimize errors. While our algorithm has successfully suggested a significant number of potential collaborations, it is essential to establish criteria for selecting the most relevant ones for the project. The next step will involve refining our approach to prioritize these suggestions, ensuring that they align closely with the goals of the Fit4MedRob initiative. For instance, keywords could be grouped into categories to reduce redundant keywords. However, this step would add an additional manual step from experts [8], proving once again the importance of standardization and the benefits of encoding keywords in pre-defined relationships and taxonomies. Collecting feedback from stakeholders to implement and assess the suggested partnerships could provide evidence to support the effectiveness of the recommendations. By doing so, we aim to foster stronger and more effective collaborations that can significantly enhance research productivity and contribute to the success of the project. Being based on bibliographic data, this framework could also be applied to enhance collaboration around a topic of interest.

Acknowledgments: This work was supported by the Italian Ministry of Research, under the complementary actions to the NRRP "Fit4MedRob - Fit for Medical Robotics" Grant (# PNC0000007)

References

[1] Lawrence T, Hardy C, Phillips N. Institutional effects of interorganizational collaboration: the emergence of proto-institutions [Internet]. Acad Manag J. 2002 [cited 2024 Aug 21]. Available from: https://www.semanticscholar.org/paper/Institutional-effects-of-interorganizational-The-of-Lawrence-Hardy/a2029c534a93d1ea1e04354fb14054e536a0c325

[2] Chinchilla-Rodríguez Z, Moya-Anegón F, Vargas-Quesada B, Corera-Álvarez E, Hassan-Montero Y. Inter-institutional scientific collaboration: an approach from social network [Internet]. 2008 [cited 2024 Aug 21]. Available from: https://www.semanticscholar.org/paper/Inter-institutional-scientific-collaboration%3A-an-Chinchilla-Rodr%C3%ADguez-Moya-Aneg%C3%B3n/c61c1f8a11ff025bb2dff2ac5b79c9f4ceeff85a

[3] Toral S, Bessis N, Martínez-Torres MR. External collaboration patterns of research institutions using shared publications in the Web of Science [Internet]. Program. 2013 [cited 2024 Aug 21]. Available from: https://www.semanticscholar.org/paper/External-collaboration-patterns-of-research-using-Toral-Bessis/e54a419433d5c7f287da13330edbea93a2ef4077

[4] Gui Q, Liu C, Du D. Globalization of science and international scientific collaboration: a network perspective. Geoforum. 2019 Oct;105:1–12. doi:10.1016/j.geoforum.2019.06.017

[5] di Bella E, Gandullia L, Preti S. Analysis of scientific collaboration network of Italian Institute of Technology. Scientometrics. 2021 Oct;126(10):8517–39. doi:10.1007/s11192-021-04120-9

[6] Rose ME, Kitchin JR. pybliometrics: scriptable bibliometrics using a Python interface to Scopus. SoftwareX. 2019 Jul;10:100263. doi:10.1016/j.softx.2019.100263

[7] Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. Proc Int AAAI Conf Web Soc Media. 2009 Mar;3(1):Art. no. 1. doi:10.1609/icwsm.v3i1.13937

[8] Pesta B, Fuerst J, Kirkegaard EOW. Bibliometric keyword analysis across seventeen years (2000–2016) of Intelligence articles. J Intell. 2018 Oct;6(4):46. doi:10.3390/jintelligence6040046....