French Polytech network form for PhD Research Grants from the China Scholarship Council

This document describes one of the PhD subjects proposed by the French Polytech network. The network is composed of 15 engineering schools/universities. The document also provides information about the supervisor. Please contact the PhD supervisor by email for further information regarding your application.

Supervisor information	
Family name	Néron
First name	Emmanuel
Email	emmanuel.neron@univ-tours.fr
Web reference	https://www.researchgate.net/profile/Emmanuel-Neron
Lab name	EA 6300 LIFAT, ERL CNRS 7002 ROOT
Lab web site	https://lifat.univ-tours.fr/
Polytech name	Polytech Tours
University name	Université de Tours
Country	France

PhD information	
Title	Graph-based solving methods for metabolic networks analysis
Main topics regards to CSC list (3 topics at maximum)	Network calculation Understanding Models and Intelligent Systems

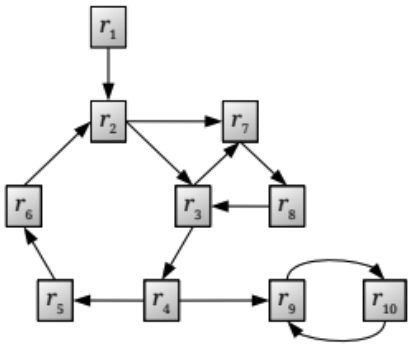
Required skills in science and engineering	Operations research, AI, graph theory, mathematical programming and constraint programming. Be interested in biological problems. Be able to discuss with non computer scientists.

Subject description (two pages maximum including biblio)

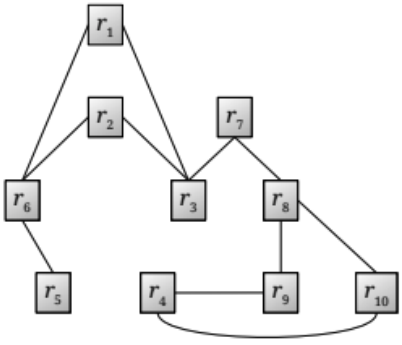
The successful candidate will join the Laboratory of Fundamental and Applied Computer Science of Tours (LIFAT) (EA 6300, ERL-CNRS 7002). LIFAT's scientific concerns are to design and develop models, methods and algorithms, and to provide resources and software for extracting information, deriving knowledge from data, integrating human-machine interaction and solving combinatorial optimization problems with the desire to achieve good results in good computation time. The unit is currently organized into 3 research groups: Databases and natural language processing (BDTLN), Operations research, scheduling and transport (ROOT, ERL CNRS 7002), and Pattern recognition and image analysis (RFAI). Due to regional specificities, the preferred fields of application are Digital human sciences and Health and disability.

As part of a previous thesis, in partnership with the University of Nouakchott, Mauritania, we have worked in recent years on the implementation of solving methods for the comparison of heterogeneous graphs, with applications in bioinformatics. We believe that continuing this work is of scientific interest: on the one hand to develop new efficient resolution methods and test them on real data; on the other hand, to improve our understanding of the underlying biological issues.

The problem dealt with here, known to be strongly NP-hard, is defined as follows. Let D be a directed graph (representing a metabolic network) and G a non-oriented graph (representing the proximity of the genes producing the proteins involved in these reactions), both defined on the same set of vertices.



Metabolic graph (D)



Proximity graph (G)

The problem is to find a longest path or trail in D, such that the subgraph of G generated by the vertices of this path or trail is connected. We are therefore looking for a longest metabolic pathway involving proteins produced by neighboring genes. These paths and trails constitute markers of the species and can be used for their comparison.

The thesis will focus on:

(1) Improving methods developed so far (such as mathematical and constraint-based formulations), proposing new methods (e.g. using decomposition techniques), and applying these new methods to real datasets.

(2) Better modeling criteria used for comparing species in the longest path and longest trail problems, in order to find a family of paths or trails more suitable for this comparison. It is envisaged that two criteria will be considered simultaneously (length and disparity of the paths or trails for example).

(3) Open this work to other problems concerning metabolic networks modeled by graphs. Initial contacts have been established with colleagues from the Biomolecules Biotechnologies Vegetables laboratory) working on these problems. Contacts have also been established with research teams from the Pasteur Institute (Paris). The problem of enumerating minimal precursor sets in genome-wide metabolic networks will be considered.

References

[1] Mohamed Lemine Ahmed Sidi, Ronan Bocquillon, Hafedh Mohamed Babou, Cheikh Dhib, Emmanuel Néron, Ameur Soukhal, Mohamedade Farouk Nanne, Improved approaches to solve the One-To-One SkewGraM problem, Computers and Operations Research, October 2021.

https://doi.org/10.1016/j.cor.2021.105584

[2] Hafedh Mohamed Babou, Comparaison de réseaux biologiques, thèse de doctorat, Université de Nantes, 2012. https://tel.archives-ouvertes.fr/tel-00767578

[3] Kévin Billet, Benjamin Houillé, Thomas Dugé de Bernonville, Sébastien Besseau, Audrey Oudin, Vincent Courdavault, Guillaume Delanoue, Laurence Guérin, Marc Clastre, Nathalie Giglioli- Guivarc'h and Arnaud Lanoue, Field-based metabolomics of Vitis vinifera L. stems provides new insights for genotype discrimination and polyphenol metabolism structuring, Frontiers in Plant Science, 2018.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6021511/

[4] Guillaume Fertin, Christian Komusiewicz, Hafedh Mohamed Babou and Irena Rusu, Finding supported paths in heterogeneous networks, Algorithms, 2015.

https://www.mdpi.com/1999-4893/8/4/810

[5] Alexandra Zaharia, Bernard Labedan, Christine Froidevaux and Alain Denise, CoMetGeNe: mining conservedneighborhood patterns in metabolic and genomic contexts, BMC Bioinformatics, 2019.

https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2542-2