

PhD Research proposal

Dynamic community detection in node-attributed networks

Skills and profile:

- Master degree in Computer Science, especially in Optimization and/or Data Mining
- Very good programming skills are mandatory
- Knowledge in bioinformatics could be a great help
- Oral and written english is required. French is not mandatory

Laboratory: I3S, <https://www.i3s.unice.fr>

PhD supervision:

- Claude Pasquier, CR CNRS HDR, <http://claud.pasquier.net>
- Denis Pallez, Associate professor, <http://denispallez.i3s.unice.fr>

Scientific description:

Many complex systems in nature and society can be described in terms of networks where the entities under study are represented by nodes, the links between these entities by edges, and the characteristics of the entities by attributes on each node, so called *node-attributed graphs*. The existence of a community structure is an important property of real-world networks because it often helps to explain the functionality of the system. Despite the ambiguity in the definition of community, many techniques have been developed for effective and efficient community detection, often considering the problem as clustering nodes in the graph. However, the algorithms generally rely on the use of only one source of information: the network structure. In recent years, community detection algorithms that also consider the attributes of the nodes have been proposed, with varying degrees of success (Chunaev, 2020). As community detection is NP-hard and an exhaustive search over all possible solutions is usually intractable, nature-inspired techniques have been employed and compared (Osaba et al., 2020) but on small or medium datasets (around 1000 nodes). Moreover, works mainly focused on the different criteria to combine in a multi-objective context and less on the representation of solution themselves. However, some progress was made in this area (Garza-Fabre et al., 2018; Luo et al., 2021).

On the opposite, some advances have been made in finding communities where structure of the graph is changing over time (Yin et al., 2021). It is often the case in social networks or protein-protein interactions in biology. However, techniques proposed in this context do not consider attributed graphs and often apply community detection on several static networks corresponding to temporal snapshots of the dynamic network. This approach generates communities that change too drastically over time, which does not reflect the reality.

The goal of this doctoral thesis is to tackle the problem of dynamic community detection in node-attributed networks by combining a bio-inspired algorithm with a local search. To do so, we envisage to use new insights (Tian et al., 2021) to scale up a first attempt we have made for active module detection in biology (Correa et al.,

2019) and combine it with a greedy approach (Pasquier et al., 2021) in order to have better convergence capacity. This work will have concrete applications in the industrial project NewgenTOXiv that we have just started which aims to develop new *in vitro* toxicological tests.

References:

- Chunaev, P., 2020. Community detection in node-attributed social networks: A survey. *Computer Science Review* 37, 100286.
<https://doi.org/10.1016/j.cosrev.2020.100286>
- Correa, L., Pallez, D., Tichit, L., Soriani, O., Pasquier, C., 2019. Population-based meta-heuristic for active modules identification, in: *Proceedings of the Tenth International Conference on Computational Systems-Biology and Bioinformatics, CSBio '19*. Association for Computing Machinery, New York, NY, USA, pp. 1-8. <https://doi.org/10.1145/3365953.3365957>
- Garza-Fabre, M., Handl, J., Knowles, J., 2018. An Improved and More Scalable Evolutionary Approach to Multiobjective Clustering. *IEEE Transactions on Evolutionary Computation* 22, 515-535.
<https://doi.org/10.1109/TEVC.2017.2726341>
- Luo, W., Lin, X., Zhang, J., Preuss, M., 2021. A Survey of Nearest-Better Clustering in Swarm and Evolutionary Computation, in: *2021 IEEE Congress on Evolutionary Computation (CEC)*. Presented at the 2021 IEEE Congress on Evolutionary Computation (CEC), pp. 1961-1967.
<https://doi.org/10.1109/CEC45853.2021.9505008>
- Osaba, E., Del Ser, J., Camacho, D., Bilbao, M.N., Yang, X.-S., 2020. Community detection in networks using bio-inspired optimization: Latest developments, new results and perspectives with a selection of recent meta-heuristics. *Applied Soft Computing* 87, 106010.
<https://doi.org/10.1016/j.asoc.2019.106010>
- Pasquier, C., Guerlais, V., Pallez, D., Rapetti-Mauss, R., Soriani, O., 2021. Identification of active modules in interaction networks using node2vec network embedding. <https://doi.org/10.1101/2021.09.22.461345>
- Tian, Y., Si, L., Zhang, X., Cheng, R., He, C., Tan, K.C., Jin, Y., 2021. Evolutionary Large-Scale Multi-Objective Optimization: A Survey. *ACM Comput. Surv.* 54, 174:1-174:34. <https://doi.org/10.1145/3470971>
- Yin, Y., Zhao, Y., Li, H., Dong, X., 2021. Multi-objective evolutionary clustering for large-scale dynamic community detection. *Information Sciences* 549, 269-287.
<https://doi.org/10.1016/j.ins.2020.11.025>