

MASTER INTERNSHIP PROPOSAL

Laboratory: Inria Sophia Antipolis - Méditerranée and ENS Lyon

Supervisors:

Frederic.Havet@cnrs.fr (Inria, COATI, <https://team.inria.fr/coati>)

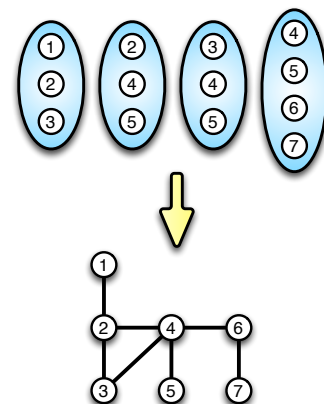
Dorian.Mazaauric@inria.fr (Inria, ABS, <http://team.inria.fr/abs>)

Remi.Watrigant@univ-lyon1.fr, (ENS Lyon, MC2, <http://www.ens-lyon.fr/LIP/MC2/>)

GRAPH ALGORITHMS

FOR LOW RESOLUTION MODEL OF LARGE PROTEIN ASSEMBLIES

Graph Problems. Let H be a hypergraph with $V(H)$ its set of vertices and $E(H)$ its set of hyperedges. The Minimum Connectivity Inference (MCI) problem consists in finding a smallest set of edges E satisfying the following constraints: the set of nodes of every hyperedge of H must induce a connected subgraph in $G = (V(H), E)$. For example, the top figure represents the four hyperedges of an instance and the bottom figure describes an optimal solution composed of seven edges: every hyperedge induces a connected subgraph. This problem is motivated by a problem in bioinformatics, see Context below.



Research programme. The aim of this internship is to develop algorithms for some generalized versions of MCI handling combinatorial constraints reflecting biophysical properties (bounded maximum degree, constraints on the diameter, ...). For each variant of the problem, the first aim is to determine the complexity of the problem (polynomial-time solvable, NP-hard).

In (the very likely) case the problem is NP-hard, the goal is then to develop efficient approximation algorithms or to prove that this problem is hard to approximate (APX-hard). We also plan to develop parameterized algorithms and/or moderately exponential algorithms. The same study for different instance classes is also planned in order to obtain faster and/or more accurate algorithms for specific problems (e.g. by integrating biophysical assumptions).

Context. A macromolecular assembly is composed of subunits (e.g. proteins). We assume that the composition, in terms of individual subunits, of selected complexes of the assembly is known. Indeed, a given assembly can be chemically split into complexes by manipulating chemical conditions. A node represents a subunit and there is an edge between two nodes if the two corresponding subunits are in contact in the assembly. The hypergraph represents the different selected complexes. *The MCI problem consists in finding a smallest set of contacts satisfying some the connectivity constraints on complexes.*

Background. Theoretical computer science and/or bioinformatics and/or applied maths.

Misc. Ideally, the MSc will be followed-up by a PhD thesis. Based on the results for inference problems, we will then develop novel methods to design atomic resolution models of molecular assemblies.

References

- [1] F. Alber, F. Forster, D. Korkin, M. Topf, and A. Sali. Integrating diverse data for structure determination of macromolecular assemblies. *Ann. Rev. Biochem.*, 77:11.1–11.35, 2008.
- [2] T. Taverner, H. Hernández, M. Sharon, B.T. Ruotolo, D. Matak-Vinkovic, D. Devos, R.B. Russell, and C.V. Robinson. Subunit architecture of intact protein complexes from mass spectrometry and homology modeling. *Accounts of chemical research*, 41(5):617–627, 2008.
- [3] D. Agarwal, J.-C. S. Araujo, C. Caillouet, F. Cazals, D. Coudert, and S. Pérennes. Connectivity inference in mass spectrometry based structure determination. In *ESA 2013*, volume 8125 of *LNCS*, pages 289–300. 2013.
- [4] D. Agarwal, C. Caillouet, D. Coudert, and F. Cazals. Unveiling contacts within macro-molecular assemblies by solving minimum weight connectivity inference problems. *Molecular and Cellular Proteomics*, 14:2274–2282, 2015.
- [5] I. D. Mantas. The subset interconnection design problem: algorithms and special cases. Master thesis, 2015.