Level of qualifications required: Graduate degree or equivalent

Fonction: PhD Position

About the research centre or Inria department

The Inria Sophia Antipolis - Méditerranée center counts 37 research teams and 9 support departments. The center’s staff (about 600 people including 400 Inria employees) is composed of scientists of different nationalities (250 foreigners of 50 nationalities), engineers, technicians and administrators. 1/3 of the staff are civil servants, the others are contractual. The majority of the research teams at the center are located in Sophia Antipolis and Nice in the Alpes-Maritimes. Six teams are based in Montpellier and a team is hosted by the computer science department of the University of Bologna in Italy. The Center is a member of the University and Institution Community (ComUE) “Université Côte d’Azur (UCA)”.

Context

A macromolecular assembly is composed of subunits (e.g. proteins or nucleic acids). 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, and the composition of these complexes can then be inferred using native mass spectrometry. A node represents a subunit, while the hyperedges represent the different complexes. The MCI problem consists in inferring the contact graph of these subunits, where an edge between two nodes means that the two corresponding subunits are in contact in the assembly.

Hence, the MCI problem consists in finding a smallest set of contacts satisfying the connectivity constraints on complexes. In a second time, this contact graph will define the input graph for the domino problem that consists in determining the high resolution structure of a given assembly. A configuration of a node is so a conformation of the corresponding protein (that is a position of each of its atoms in R3) and is obtained by X-ray crystallography. This process involves two main steps, namely computing conformations of subunits compatible with low resolution data [CDM+15, RDRC16], and assembling these conformations to build the assembly [ACW15].

Both steps can be phrased as optimization and enumeration problems involving graphs. See also [AFK+08, THS+08]. To summarize, the goals of the two graph problems considered in this thesis are the determination of low resolution structure of given assemblies and the reconstruction of atomic resolution models of the large protein assemblies.

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

Assignment

Research programme. The aim of this PhD thesis is to develop algorithms for some generalized versions of both problems handling combinatorial constraints reflecting biophysical properties (bounded maximum degree, constraints on the diameter of the graph sought, ...). For each variant of the problem, the rst aim is to determine the complexity of the problem (polynomial-time solvable, NP-hard). In (the very likely) case the problems are 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).

This thesis is under the supervision of Dorian Mazauric (EPI ABS) and Frédéric HAVET (EPI COATI)

Main activities

Skills
Theoretical computer science and/or bioinformatics and/or applied mathematics.

Benefits package
- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

Remuneration
Duration: 36 months
Location: Sophia Antipolis, France
Gross Salary per month: 1982€ brut per month (year 1 & 2) and 2085€ brut/month (year 3)