2017-00185 - Graph matching by SAT solving (research internship)

**Level of qualifications required**: Bachelor's degree or equivalent

**Fonction**: Internship Research

**About Inria**
Inria, the French National Institute for computer science and applied mathematics, promotes “scientific excellence for technology transfer and society” Graduates from the world's top universities, Inria's 2,700 employees rise to the challenges of digital sciences. With its open, agile model, Inria is able to explore original approaches with its partners in industry and academia and provide an efficient response to the multidisciplinary and application challenges of the digital transformation. Inria is the source of many innovations that add value and create jobs.

**About the research centre or Inria department**
Located at the heart of the main national research and higher education cluster, member of the Université Paris Saclay, a major actor in the French Investments for the Future Programme (Idex, LabEx, IRT, Equipex) and partner of the main establishments present on the plateau, the centre is particularly active in three major areas: data and knowledge; safety, security and reliability; modelling, simulation and optimisation (with priority given to energy).

The 450 researchers and engineers from Inria and its partners who work in the research centre's 31 teams, the 100 research support staff members, the high-level equipment at their disposal (image walls, high-performance computing clusters, sensor networks), and the privileged relationships with prestigious industrial partners, all make Inria Saclay Île-de-France a key research centre in the local landscape and one that is oriented towards Europe and the world.

**Context**
This research internship is offered at Inria Saclay IdF (https://www.inria.fr/en/centre/saclay) in the LIFEWARE project-team (http://lifeware.inria.fr). This team works in computational systems biology and develops the Biochemical Abstract Machine (BIOCHAM http://lifeware.inria.fr/biocham4) software for modeling, analyzing and now synthesizing biochemical reaction networks (CRNs) using methods from fundamental computer science and mathematics. The software developments are expected to be integrated in BIOCHAM.

The internship will be supervised by François Fages who has supervised 29 Ph.D. theses in his career, the last defended in May 2016, and who currently is 100% available with no Ph.D. student.

**Assignment**
Graph matching problems concern the detection of similarities between two graphs G and H. The exact graph matching problem is the problem of existence of a graph isomorphism between G and H. The subgraph isomorphism problem (SISO), i.e. the...
existence of a graph isomorphism between a subgraph of G and H, is an NP-complete problem. SISO provide a standard notion of graph motif and are widely used in pattern recognition in computer vision for instance. In the context of systems biology, the question of detecting model reduction relationships between biochemical reaction networks (CRN bipartite directed graphs) in large model repositories, led us to the notion of subgraph epimorphism (SEPI), i.e. subgraph morphism surjective on vertices and edges. The existence of a SEPI from G to H is indeed equivalent to the existence of a sequence of deletions or mergings of vertices in G that lead to a graph isomorphic to H, which, in the context of CRNs, are basic operations for simplifying a model by neglecting or merging some species or reactions. We have shown that the SEPI problem is NP-complete but that its encoding as a constraint satisfaction problem, and the use of constraint programming or SAT solvers, provide a sufficiently efficient solution in practice with few exceptions.

The goal of this research internship is to develop this approach in mainly two directions:

- investigate the theory, and the implementation with a SAT solver, of a refinement of SEPIs in which only adjacent vertices can be merged (at distance two in the bipartite graph of a CRN), and differences with graph minors,
- investigate the theory, and implementation with SAT, of minimal SEPI upper bounds, and maximal SEPI lower bounds, with the restrictions above.

Main activities

Our previous work on this subject is described in


The theoretical work will consist in investigating similar properties for the notion of restricted SEPIs. The implementation work using SAT solvers will be expected to be integrated in BIOCHAM and evaluated on the repository of models BioModels.

Skills

This subject requires common and basic knowledge in graph theory, algorithmics and programming.

There is no specific prerequisite for this internship. However, some specific knowledge about SAT solvers, constraint programming, Prolog, or systems biology will be a plus.

Benefits package

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

Remuneration

500 euros/month + living compensation for foreigners