Internship Proposal:
Algorithms for finding quasi-cliques in large graphs

Mauro Sozio
sozio@telecom-paristech.fr

Quasi-cliques are subgraphs that are “almost” fully connected, more precisely they contain at least a constant fraction of the number of edges in a clique with the same number of nodes. Quasi-cliques are relatively rare, and therefore they might represent some interesting regions in the input graph. As a result, algorithms for finding quasi-cliques find application in bioinformatics, social network analysis and virtually anything that can be represented as a graph. The state-of-the-art algorithm for finding cliques is the algorithm developed in [1], which can enumerate all cliques of up to 10 nodes in graphs containing billions of edges. On the other hand, the problem of finding quasi-cliques is much less understood (see for example [2]).

Main Goal. The main goal of the internship is to develop an algorithm for efficiently finding quasi-cliques in large graphs. The algorithm will be evaluated against other approaches on real-world graphs, such as graphs from social networks (Facebook, Twitter, etc.) and biological data (e.g. protein-protein interaction graph). The project requires excellent coding skills in C, Python or Java. The project is mainly practical, however, related theoretical problems could also be studied during the internship.

Location and supervision. The internship will last about 4-6 months and it will be supervised by Mauro Sozio (Associate Professor at Telecom Paris University). Telecom Paris is located in Palaiseau (south of Paris). The internship can lead to a PhD thesis (fundings on this or related topics are available).

References
