Improving maximum biclique detection

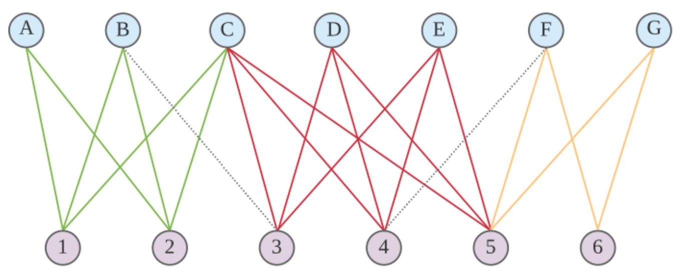
supervision: Maximilien Danisch, Lionel Tabourier (LIP6 – CNRS / SU)
location: Sorbonne Université – 4 place Jussieu, Paris

stages@complexnetworks.fr, http://complexnetworks.fr

Context:

Bipartite graphs are largely used when modeling real-world networks [GL06]. They are suited to represent systems such as online platforms, as they allow to represent users selecting items (watching videos, clicking on links, buying a product, etc). They are also used to represent individuals collaborating in projects, or groups of people taking part to a same event. They may also represent interaction data in other fields, such as in biological interactions between proteins and DNA. Actually, bipartite networks may represent many systems where the interacting entities are of two different kinds.

Identifying dense subgraphs in these networks has attracted much interest as it reveals for example users with common interests, or proteins related to the same region of the DNA molecule. A biclique is the densest subgraph in bipartite networks: it is a subgraph in which all entities of type A are connected to all entities of type B. As such, bicliques are relevant units in order to decompose the structure of a network, as schematized in the Figure below. That's why detecting bicliques – especially in large graphs – has been the focus of much work, e.g. [Dam14]. This internship tackles the problem of detecting efficiently these subgraphs. It is part of the ANR JCJC project LiMass http://bit.ly/LiMass.



Biclique decomposition of a bipartite graph (each set of links in a given color is a maximal biclique of the graph).

Goal:

In a previous work [DBS18], it has been shown that it was practically possible to list all k-cliques (complete subgraph of k nodes) in very large real-world networks. The purpose of the internship is to adapt the method to the context of bicliques in bipartite graphs in order to count and list them, even for very large graphs (say billions of nodes and links).

A direct application of fast biclique detection is bipartite graph compression. Indeed, bicliques may be encoded in a tripartite graph, the third level representing the existence of a biclique connecting the nodes [TT19]. Supposing that a biclique connects x nodes from family A to y nodes of family B, encoding it in a third level allows to represent the biclique with $x + y$ links instead of $x \cdot y$. So this representation allows to reduce significantly the size of a graph in memory, especially for real networks which exhibit a lot of such structures.

Requested profile:

This internship is directed at students with a strong background in algorithmics and programming (especially graph algorithmics), who are also interested in multidisciplinary applications. Code efficiency is essential for this internship, so even if the programming language is not imposed, it should be chosen with this purpose in mind.

The intern will be part of the Complex Networks team of the LIP6 (SU-CNRS), located in Paris on Jussieu Campus. The internship will be supervised by Maximilien Danisch and Lionel Tabourier.

References

[Dam14] Peter Damaschke. Enumerating maximal bicliques in bipartite graphs with favorable degree sequences. *Information Processing Letters*, 114(6):317–321, 2014.

[DBS18] Maximilien Danisch, Oana Balalau, and Mauro Sozio. Listing k-cliques in sparse real-world graphs. In *Proceedings of the 2018 World Wide Web Conference*, pages 589–598, 2018.

[GL06] Jean-Loup Guillaume and Matthieu Latapy. Bipartite graphs as models of complex networks. *Physica A: Statistical Mechanics and its Applications*, 371(2):795–813, 2006.

[TT19] Fabien Tarissan and Lionel Tabourier. A random model that relies on maximal bicliques to preserve the overlaps in bipartite networks. In *8th International Conference on Complex Networks and their Applications*, 2019.