EUROPT 2015

Modularity maximization clustering with cohesion conditions

Sonia Cafieri

École Nationale de l'Aviation Civile, University of Toulouse, France

joint work with

A. Costa, Singapore University of Technology and Design, and P.Hansen, GERAD, HEC Montréal

The study of systems composed of interacting components through their representation as graphs is attracting a growing attention in a wide variety of domains. A modular structure characterizes many of these systems, that means that they contain subgroups of entities sharing some common properties. Therefore, a topic of particular interest is the identification of modules, called clusters or communities. The problem can be formulated using mathematical programming, where the idea is to optimize a function expressing the quality of the clustering partition. The most studied of these functions is the so-called *modularity*. We consider modularity maximization, formulated as a mixed-integer quadratic optimization problem, that has a convex continuous relaxation.

We address the problem of combining different clustering criteria and the corresponding impact on the optimization problem. Defining a good clustering criterion is in fact difficult. On the one hand, quality functions to be optimized have been proposed, and on the other hand, properties to be satisfied by each cluster of a partition have been suggested. It has recently been observed that one of the best known such properties, i.e., the *weak condition* [Proc. Natl. Acad. Sci. USA, 2004] was not satisfied by one or more clusters in a partition which maximizes some of the best known criteria. We consider five cluster-defining conditions, that we call *cohesion conditions* (including the weak one). We then modify the modularity optimization problem to add these conditions, one at a time, as constraints. These can be expressed as linear constraints (applying, for one of the considered conditions, a Fortet's linearization). Then, the obtained optimization models are still mixed-integer quadratic optimization problems, that we solve by exact methods. We thus discuss the solution of the new models, that enable to understand the impact of cohesion conditions on modularity maximization.