

Title: Workshop on Challenging problems in Statistical Learning

Authors: Nathalie Villa-Vialaneix (1,2) and Fabrice Rossi (3)

(1) Institut de Mathématiques de Toulouse, Université de Toulouse, Toulouse - France

(2) Université de Perpignan Via Domitia, IUT, Département STID, Carcassonne - France

(3) SAMM, Université Paris 1 - France

A growing number of applicative fields generate data that are pairwise relations between the objects under study instead of attributes associated to every object: social networks (relations between persons), biology (interactions between genes, proteins), www (relations between websites or blogs), marketing (relations between customers and services)... To help understanding and interpreting such data, specific data analysis tools have been extended from the classical multivariate data analysis: visualization, clustering, classification ... This talk deals with an exploratory methodology: a common way to help understanding a graph is to cluster its vertices into relevant groups and then to represent the (simplified) graph of clusters. As will be explained, these two objectives (clustering and representation) can be somehow contradictory. Two approaches related to self-organizing maps will be presented and compared on real-world data to solve this issue.