

A simple algorithm for finding clusters in a random environment

Van Vu, Yale University

In this talk, we present a natural, simple and fast algorithm for finding clusters in a random environment.

One model problem is as follows. Let X be a large vertex set which is partitioned into k parts, X_1, \dots, X_k . Draw random edges between the vertices in X_i and X_j , with density depending varies with i and j . Given one instance of such a graph, we want to identify the clusters X_1, \dots, X_k . (Notice that vertices in the same cluster behave in the same way, with respect to others).

This problem contains several well studied problems, such as finding a hidden clique, finding a hidden coloring, finding a hidden bipartition in a random graphs. We will discuss several earlier results on the subject, in particular a general result by Mc Sherry from 2001.

The key ingredient in the analysis of our algorithm is the fact the length of the orthogonal projection of a random vector onto a given subspace is very strongly concentrated. This is easy to verify if the random vector is gaussian, but not entirely so in the general case. Time allows, we will present some applications of this tool, one concerns random matrices, the other concerns random quadratic forms, extending a classical result of Hanson and Wright