Spectral clustering and the high-dimensional Stochastic Block Model

Bin Yu

Departments of Statistics and EECS University of California at Berkeley binyu@stat.berkeley.edu

Abstract

In recent years network analysis have become the focus of much research in many fields including biology, communication studies, economics, information science, organizational studies, and social psychology. Communities or clusters of highly connected actors form an essential feature in the structure of several empirical networks. Spectral clustering is a popular and computationally feasible method to discover these communities.

The Stochastic Block Model is a social network model with well defined communities. This talk will give conditions for spectral clustering to correctly estimate the community membership of nearly all nodes. These asymptotic results are the first clustering results that allow the number of clusters in the model to grow with the number of nodes, hence the name high-dimensional. (Based on joint work with Karl Rohe and Sourav Chatterjee at UC Berkeley.)