Fast Network Community Detection by SCORE

Jiashun Jin

Department of Statistics
Carnegie Mellon University

Johnson Center 326-Meeting Room B
4400 University Drive, Fairfax, VA 22030

Time: 11:00 A.M. - 12:00 P.M.

Date: Friday, Apr 18, 2014

Abstract

Consider a network where the nodes split into $K$ different communities. The community labels for the nodes are unknown and it is of major interest to estimate them (i.e., community detection). Degree Corrected Block Model (DCBM) is a popular network model. How to detect communities with the DCBM is an interesting problem, where the main challenge lies in the degree heterogeneity.

We propose Spectral Clustering On Ratios-of-Eigenvectors (SCORE) as a new approach to community detection. Compared to existing spectral methods, the main innovation is to use the entry-wise ratios between the first a few leading eigenvectors for community detection. The central surprise is, the effect of degree heterogeneity is largely ancillary, and can be effectively removed by taking such entry-wise ratios.

We have applied SCORE to the well-known web blogs data and the statistics co-author network data which we have collected very recently. We find that SCORE is competitive both in computation and in performance. On top of that, SCORE is conceptually simple and has the potential for extensions in various directions. Additionally, we have identified several interesting communities in statisticians, including what we call the “Object Bayesian community”, “Theoretic Machine Learning Community”, and the “Dimension Reduction Community”.

We develop a theoretic framework where we show that under mild regularity conditions, SCORE stably yields consistent community detection. In the core of the analysis is the recent development on Random Matrix Theory (RMT), where the matrix-form Bernstein inequality is especially helpful.