DEPARTMENT OF STATISTICS
COLLOQUIA SERIES

Monday October 24th, Talk: 4:15 PM — Science Center 300H
Reception: 3:50 PM

Guest Speaker: Tracy Ke
Assistant Professor
Department of Statistics
University of Chicago

Social Networks for Statisticians

Abstract:

We have collected a data set for the social networks of statisticians. The data set consists of the meta information (e.g., authors, abstracts, citation counts) of about 70,000 papers in 36 representative journals in statistics and related fields, from 1984-2015. Our data collection project (which we may call it the Phase II) is a continuation of the recent data collection project by Ji and Jin (which we may call the Phase I). The Phase I data set consists of similar meta information, but only for 4 journals in a 10-year period.

The Phase I data set has been ready for analysis since 2014, and Phase II data set is not yet but about to be ready for full analysis.

We investigate the Phase II data and report some Exploratory Data Analysis (EDA) (a term introduced by Tukey, 1977) results. In particular, we discuss the overall productivity, journal-journal citation exchanges, and citation patterns of individual papers.

The data sets also pose many problems, to solve which we need more sophisticated methods, and one such problem is mixed membership estimation. We propose Mixed-SCORE as a new spectral method for mixed membership estimation. At the heart of Mixed-SCORE is a (tall by very skinny) matrix of entry-wise ratios, formed by dividing the first few eigenvectors of the network adjacency matrix over the leading eigenvector of the same matrix in an entry-wise fashion. The main surprise is that, the rows of the entry-wise ratio matrix form a cloud of points in a low-dimensional space with the silhouette of a simplex, and the simplex carries all information we need for estimating the memberships.

We apply Mixed SCORE to a Coauthorship network and a (symmetrized) Citation network constructed with the Phase I data, and obtain meaningful results. We propose a Degree-Corrected Mixed Membership model, and use it to solidify our discoveries with delicate spectral analysis and Random Matrix Theory.