RUTGERS
School of Arts and Sciences

Department of
Statistics and Biostatistics

DEPARTMENT OF STATISTICS AND BIOSTATISTICS

Jiashun Jin
Department of Statistics
Carnegie Mellon University

Coauthorship and Citation Networks of Statisticians

October 4, 2017
3:20 – 4:20pm
Light refreshments will be served
110 Frelinghuysen Road
Hill Center, Room 552

Abstract: We have collected a data set for the networks of statisticians, consisting of titles, authors, abstracts, MSC numbers, keywords, and citation counts of papers published in representative journals in statistics and related fields. In Phase I of our study, the data set covers all published papers from 2003 to 2012 in Annals of Statistics, Biometrika, JASA, and JRSS-B. In Phase II of our study, the data set covers all published papers in 36 journals in statistics and related fields, spanning 40 years. We report some Exploratory Data Analysis (EDA) results including productivity, journal-journal citations, and citation patterns. This part of result is based on Phase II of our data set (ready for use not very long ago).

We also discuss two closely related problems: network community detection, and network membership estimation. We attack these problems with the recent approach of Spectral Clustering On Ratioed Eigenvectors (SCORE), reveal a surprising simplex structure underlying the networks, and explain why SCORE is the right approach.

We apply SCORE to the Coauthorship and Citation networks of statisticians (based on Phase I of our data set), and present several communities including “Large- Scale Multiple Testing”, “Variable Selection”, “Carroll-Hall”, and “North Carolina”.

Bio: Jiashun Jin received his Ph.D in Statistics from Stanford University in 2003. He was trained in statistical inference for Big Data, specializing in dealing with the most challenging regime where the signals are both Rare and Weak. His earlier work was on large-scale multiple testing, focusing on (Tukey's) Higher Criticism and practical False Discovery Rate (FDR) controlling methods. His more recent interest is on complex graphs, social networks, and sparse PCA and Random Matrix Theory. He has developed a number of new methods, among which are the Graphlet Screening (GS) for high dimensional variable selection, IF-PCA for dimension reduction and high dimensional clustering, and SCORE for network community detection.