The Phase II data for the network of statisticians

Jiashun Jin

Carnegie Mellon University
E-mail: jiashun@stat.cmu.edu

Abstract: We have collected a data set for the networks of statisticians, consisting of titles, authors, abstracts, MSC numbers, keywords, and citation counts of more than 80K papers published in representative journals in 36 journals in statistics and related fields, spanning about 41 years. The data set provides a fertile ground for research in social networks, text mining, and knowledge discovery, and motivates an array of interesting problem in statistics and machine learning. In this talk, we discuss several problems including overall productivity of statisticians, journal ranking, journal clustering, citation patterns, citation prediction, co-authorship network communities, co-authorship network mixed-memberships, dynamic networks, topic estimation, and dynamic topic estimation. Our analysis uses an array of new methods in network analysis, topic estimation, and neural networks.