Network Approaches to Data-Driven Problems: Fundamental Limits, Scalable Algorithms, and Applications

Abstract:

Network models provide a unifying framework for understanding dependencies among variables in medical, biological, and other sciences. Networks can be used to reveal underlying data structures, infer functional modules, and facilitate experiment designs. In practice however, size, uncertainty and complexity of underlying associations render these problems challenging. In this talk, I will illustrate the use of spectral, combinatorial, and statistical inference techniques in learning the network topology and subsequent network analysis.

First, we introduce Network Maximal Correlation (NMC), a multivariate measure of nonlinear association suitable for large datasets. NMC infers transformations of variables to reveal underlying nonlinear dependencies among them. We illustrate applications of NMC in learning graphical models when variables have unknown nonlinear dependencies. Moreover we highlight NMC's utility in inference of nonlinear association gene networks and modules in cancer datasets and validate them using survival times of patients. Next, we discuss the problem of network alignment that aims to find a bijective mapping across two graphs so that, if two nodes are connected in one graph, their images are also connected in the other graph. This problem has a broad range of applications such as comparative analysis of gene or protein-protein interaction networks across different species. To solve this combinatorial problem, we present a new scalable spectral algorithm, and establish its efficiency, theoretically and experimentally, over several synthetic and real networks.