ABSTRACT

Graphical models are widely used to model complex interactions amongst components of biological and social networks. A graphical model defines a probability distribution over a graph, or network, with p random variables. An important task is then to learn, on the basis of, e.g. n observations, the edges of the network based on the postulated probability distribution. In recent years, significant progress has been made in developing penalized estimation methods for addressing this problem in high dimensional networks. Existing literature primarily focuses on the homogenous setting, where the observations are sampled from a single population, and random variables represented on the nodes of the network are similarly distributed, e.g. they are all conditionally Gaussian. However, in many settings, including high throughput biological experiments, observations are drawn from heterogenous populations, and high dimensional networks consist of heterogenous random variables. For instance, in biology, we are interested in estimating the joint network of interactions among genes, proteins and metabolites, based on samples collected from patients with different subtypes of cancers. In such settings, estimated networks from existing approaches may not correctly capture the relationships among variables. In this talk, I will discuss extensions of penalized estimation methods for learning high dimensional graphical models with heterogeneity in observations and variables.