Bayesian Nonparametric Graph Clustering

Veera Baladandayuthapani, UT MD Anderson Cancer Center

We present clustering methods for multivariate data exploiting the underlying geometry of the graphical structure between variables. As opposed to standard approaches that assume known graph structures, we first estimate the edge structure of the unknown graph using Bayesian neighborhood selection approaches, wherein we account for the uncertainty of graphical structure learning through model-averaged estimates of the suitable parameters. Subsequently, we develop a nonparametric graph clustering model on the lower dimensional projections of the graph based on Laplacian embeddings using Dirichlet process mixture models. In contrast to standard algorithmic approaches, this fully probabilistic approach allows incorporation of uncertainty in estimation and inference for both graph structure learning and clustering. More importantly, we formalize the arguments for Laplacian embeddings as suitable projections for graph clustering by providing theoretical support for the consistency of the eigenspace of the estimated graph Laplacians. We develop fast computational algorithms that allow our methods to scale to large number of nodes. Through extensive simulations we compare our clustering performance with standard clustering methods. We apply our methods to a novel pan-cancer proteomic data set, and evaluate protein networks and clusters across multiple different cancer types.