Reverse Engineering Evolving Networks Underlying Developing Systems

Eric Xing

Carnegie Mellon University

Estimating rewiring gene regulatory networks over developing biological systems, such as proliferating cells, growing embryos, and differentiating cell lineages, is central to a deeper understanding of how cells evolve during development. However, one challenge in estimating such evolving networks is that their host cells are not only contiguously evolving, but also branching over time. For example, stem cells evolve into two more specialized daughter cells at each division, forming a tree of networks. Another example is in a laboratory setting: a biologist may apply several different drugs to a malignant cancer cell to analyze the changes each drug has produced in the treated cells. Each treated cell is not directly related to another treated cell, but rather to the malignant cancer cell that it was derived from. This posts an interesting question on multiplicity control.

We propose two interesting statistical frameworks, one builds on a generalization of kernel density estimator for regularized nonparametric estimation of inverse-coveriance matrix from non-iid high-dimensional samples, and the other one builds on the L1 plus total variation penalized graphical logistic regression, to effectively estimate multiple evolving gene networks corresponding to cell types related by either a linear-sequence or a tree-genealogy, based on only a few samples from each cell type. Our methods take advantage of the similarity between related networks along the biological lineage, while at the same time exposing differences between the networks. We demonstrate that our methods perform significantly better than existing methods via simulation, and enjoy strong statistical guarantees unlike other heuristic based approaches. We explore an application to a breast cancer analysis. Based on only a few microarray measurements, our algorithm is able to produce biologically valid results that provide insight into the progression and reversion of breast cancer. Finally, I will discuss a few additional complex scenarios for network estimation, where graphs are directional, have missing value, or are multi-attributes, and ideas for consistent structure estimation.