Inferring the T-cell signaling network from flow cytometry data

Biological signaling networks
- Consist of interacting proteins, genes, and small molecules
- Underlie the major functions of living cells
- Can be computationally represented as Bayesian networks
Goal: learn personalized Bayesian signaling models from biological data

Learning Bayesian networks

Expression data
- Networks vs. Nodes

Approaches
- Find the network $G$ that is most likely given the data, i.e., find $G$ that maximizes $P(G | data)$
- Impossible to solve maximization problem directly (see table)
- Two heuristic procedures
  - Markov Chain Monte Carlo (MCMC) samplings of Pr(G | data)
  - Direct search of graphs guided by Pr(G | data)

Problems
- Flat posterior landscape: many graphs $G$ with similar values of $P(G | data)$
- Graphs often over-fitted to data
- Small perturbations to data yield large perturbations in graphs
- No computational validation procedure for assessing quality of learned graphs

T-cell signaling network

Flow cytometry data
- Data gathered using 10-color flow cytometry measuring phosphoprotein and phospholipid
- 11 biomolecules, 5400 observations, 5 general perturbations, 6 specific perturbations
- First analyzed by Sachs, et al. using MCMC and simulated annealing

Complete processing pipeline

Our learning algorithm
- Model averaging by bootstrap aggregation, 2500 resamples
- Greedy hill-climbing search with randomized restarts and Friedman’s sparse candidate algorithm
- Dichotomized Graph Prior with λ = 1

Conclusions
- Our computational results point to incompleteness of the current theory of T-cell signaling
- Transitive edges (Raf to Akt, Mek to Akt) are new regulatory edges confirmed by recent literature
- These new edges are unlikely to be artifacts of our computations
- We are planning biological validation experiments to confirm the high probability edges
- There appears to be significant crosstalk between major cellular signaling pathways: the Jnk to P38 edge is an especially interesting one to experimentally validate
- The real value of our methods will be realized when we reconstruct major signaling networks of cancer patients from flow cytometry data

Acknowledgements
- Funded by Doer A. Foundation Grant (Gulf Coast Center for Computational Cancer Research)
- Eaton and Murphy “Exact Bayesian structure learning from uncertain interventions.” AI & Statistics. 2007
- Koch, Broom, and Subramanian. “A Search-Based Approach for Bayesian Inference of the T-cell Signaling Network.” BIOIT 2007 conference