

Topics for Algorithmic Systems Biology

Vingron part. Version 1, 1. Juli 2010

Networks

- Experimental methods for protein-protein interaction networks
- Characteristics of protein-protein interaction networks
- Scale-free networks vs. random networks (definition)
- Connections between hub proteins (degree-degree interaction)
- Spoke-matrix model in analysis of purification data
- Models for PPI (directed, undirected, bipartite)
- Evaluating interaction under the hypergeometric distribution
- Principles of genetic interactions

Bayesian statistics

- Basics of bayesian statistics
- Prior, posterior, joint probability

Network models

- Euclidian distance based networks
- Principles of bayesian networks
- Conditional independence and equivalence classes

Petri nets

- Definitions
- Basic modeling

Statistics

- Entropy and conditional entropy
- Mutual information
- Multiple testing corrections
- False discovery rate
- Beta distribution (pdf, cdf, parameters, scope)
- Gamma distribution (pdf, cdf, parameters, scope, definitions)
- Hypergeometric distribution (parameters, scope, definitions)
- Fisher exact test
- Geneset enrichment analysis

Markov Chain Monte Carlo

- Metropolis algorithm (goals, principles)
- Proposal and target distribution (see exercise)

Clustering

- Principles of clustering, classification and semi-supervised learning
- Advantages and problems with k-means and variants of hierarchical clustering
- Cluster evaluation measures (Rand index etc.)

Additional exercises

In random order for self study.

- Sketch the Metropolis algorithm.
- Calculate the mutual information of two distributions using different discretization.
- A TAP purification of a bait protein retrieves 5 other protein. In another experiment 50 proteins are identified. How many interactions are discovered in each experiment? Comment on the plausibility of an individual interaction.
5 and 50 interactions under the spoke model. 10 and 1225 under the matrix model. The matrix model interactions have a high false positive rate, many contacts are indirect. The spoke model has a high false negative rate - many of the interactions are not identified.
- Enumerate all directed networks with 3 nodes and 2 graphs and group them into equivalence classes.
- Review ROC curves (false-positive rate, TPR etc.)
- Review Fisher's exact test and the approximation with the χ^2 test. Why is the approximation useful? Calculate the probabilities for a small example.
- Rather tough exercises for Petri nets are provided at the end of the book chapter in the material provided on the website. Check what you can solve with the material presented in the lecture only.