

Topics for Algorithmic Systems Biology

Vingron part. Version 2, 6. July 2011

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)
- 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

Statistics

- Entropy and conditional entropy

- Mutual information
- Multiple testing corrections
- False discovery rate
- Hypergeometric distribution (parameters, scope, definitions)
- Fisher exact test
- Geneset enrichment analysis

Unsupervised learning

- Principles of clustering, classification and semi-supervised learning
- Advantages and problems with k-means and variants of hierarchical clustering
- Principle of spectral clustering

Machine learning

- Supervised learning
- Least squares
- Regression and classification
- Linear discriminant analysis
- Support vector machines
- Kernel functions

Graphical models

- Bayes' theorem
- Bayesian networks

Additional exercises

In random order for self study in addition to the exercises given.

- 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 Fisher's exact test and the approximation with the χ^2 test. Why is the approximation useful? Calculate the probabilities for a small example.