

Course Proposal: CMSC 702 - Computational Systems Biology

September 2010

Bioinformatics Field Committee

Motivation and background

This course will focus on the computational analysis of biological systems beyond genome sequence, and cover topics in functional genomics, population genetics, interaction networks and epigenetics. Computational methods studied for this type of analysis include: network and graph algorithms, supervised and unsupervised learning, large data/network visualization, statistical modeling and inference, probabilistic graphical models, sparse methods in data analysis, numerical optimization (linear, integer and semidefinite programming). These methods are complementary to those used for analyzing biological sequences (on which CMSC 701 is focused).

Course Description

An introduction to the fundamental concepts in the computational analysis of biological systems with applications to: functional genomics, population genetics, interaction networks, epigenetics. Computational concepts covered: network and graph algorithms, machine learning, large data/network visualization, statistical modeling and inference, probabilistic graphical models, sparse methods in data analysis, numerical optimization. No knowledge of biology required.

Prerequisites

CMSC 423, or Computer Science or Applied Mathematics graduate student, or permission of instructor.

Textbooks:

There is currently no single textbook that covers all the topics presented in the course. The following textbooks contain a number of topics discussed and will be listed as recommended reading. The relevant material from these books will be made available to the students as handouts.

Bjorn Junker and Falk Schreiber. *Analysis of Biological Networks*. Wiley-Interscience (2008). ISBN: 978-0470041444

Aidong Zhang. *Advanced analysis of gene expression microarray data*. World Scientific (2006). ISBN 981-256-645-7

Frank Emmert-Streib and Matthias Dehmer (eds.). *Analysis of Microarray Data: A Network-Based Approach*. Wiley-VCH (2008). ISBN: 978-3527318223

Tentative Syllabus

- Introduction (1 week)
 - A molecular biology for computer scientists refresher.
- Functional Genomics (3 weeks)
 - Differential analysis of gene expression by probabilistic modeling and statistical inference
 - Supervised and unsupervised learning models for expression data classification and clustering
 - Deriving medical diagnosis and prognosis models from expression data using sparse methods in machine learning
 - Probabilistic graphical models of gene co-regulation
- Networks (6 weeks)
 - Experimental and computational methods for detecting molecular interactions within the cell
 - Gene function prediction via network clustering, clique-finding, information flow heuristics
 - Identification of functional interaction patterns using network alignment and subgraph isomorphism
 - Database search of conserved pathways
 - Network evolution and random models of biological networks
- Genetics and epigenetics (3 weeks)
 - Genotype-phenotype association analysis by probabilistic modeling and statistical inference
 - Genome-Environment-Clinical data integration models
- Emerging topics (1 week, time permitting), e.g. sequencing technologies, nanotechnology.