

# COMPUTATIONAL SYSTEMS BIOLOGY

## CS/GBCB 5854

Home Department: Computer Science

### I Catalog Description

Phenomenological and data-driven models of molecular interaction networks. Applications of graph theory, discrete algorithms, data mining, and machine learning to the modeling and analysis of molecular interaction networks. Biological applications. Interaction between biological and computational disciplines in systems biology. Pre: 4104 or 5046 or graduate standing in CSA; GBCB 5314 or equivalent (3H,3C).

Course Number: 5854 (GBCB 5854)

ADP Title: Computational Systems Biology

### II Learning Objectives

Having successfully completed this course, the student will be able to:

- describe the utility of phenomenological models of molecular interaction networks;
- apply basic concepts of graph theory, machine learning, clustering, and data mining to compute biclusters in gene expression data; integrate molecular profiles with interaction networks; predict physical and genetic interactions;
- use algorithms to probe large-scale topological properties of interaction networks; predict gene, protein, and microRNA functions; and detect groups of co-expressed, co-regulated, or interacting groups of genes in multiple species;
- identify methods for integrating massive quantities of different types of functional genomic data.

### III Justification

Dramatic advances in genome sequencing and high-throughput biological experiments are yielding massive quantities of diverse types of data about cellular processes. Sophisticated computational methods are needed to process these data sources in an effort to analyse and unravel the complex biological phenomena that take place in a cell. The field of computational systems biology has emerged in recent years to tackle this fundamental challenge. Rather than analyse molecular interactions a gene or protein at a time, systems biology aims to understand them in terms of their modular and hierarchical structures, modelling relationships between the modules, and how these relationships change with time, under different external conditions, and across organisms. No course offered at Virginia Tech covers the algorithms and computational systems being developed to address the challenges raised by data-driven modelling and analysis of cellular interaction networks.

The course is at the 5000 level because it requires a level of maturity and experience in *both* computer science and biology. In particular, students in this course must be prepared for work in sophisticated computational algorithms as well as understanding specific applications of these algorithms to questions related to molecular interaction networks. The computational background for such work is supplied either by an undergraduate degree in computer science or by the specified pre-requisites. The life science background for this course is supplied by GBCB 5314 or equivalent.

## IV Prerequisites and Corequisites

Graduate standing in CSA or CS 4104 or 5046 is required to insure that students have the proper computer science background necessary for this course. GBCB 5314 is required to insure the students have the proper life-science background for the motivating problems in the course.

## V Texts and Special Teaching Aids

The following textbooks are recommended:

1. Isaac S. Kohane, Alvin Kho, Atul J. Butte, *Microarrays for an Integrative Genomics*, The MIT Press, 326 pages, ISBN: 026211271X.
2. Feng, Jianfeng; Jost, Jürgen; Qian, Minping (Eds.), *Networks: From Biology to Theory*, Springer Verlag, 2006, 340 pages, ISBN-10: 1-84628-485-6.
3. David Hand, Heikki Mannila, and Padhraic Smyth, *Principles of Data Mining*, MIT Press, 2001, 550 pages, ISBN: 026208290X.

These textbooks will be supplemented by reading material from the literature. These papers will be selected from journals such as *Journal of Computational Biology*, *Bioinformatics*, *PLoS Computational Biology*, *IEEE Transactions on Knowledge and Data Engineering*, *Algorithms for Molecular Biology* and *BMC Bioinformatics* and annual conferences such as *Research in Computational Molecular Biology*, *Intelligent Systems for Molecular Biology*, and *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

## VI Syllabus

	Topic	Percent
1.	Introduction to models in systems biology	5
2.	Algorithms for biclustering gene expression data	10
3.	Applications of biclustering to disease classification	5
4.	Applications of biclustering to cross-species gene expression analysis	5
5.	Techniques for integrating multiple types of functional genomic data	20
6.	Prediction of gene functions	20
7.	Prediction of molecular interactions	15
8.	The large scale topological structure of molecular interaction networks	20
		100

**VII Old (Current) Syllabus**

N/A

**VIII Core Curriculum Guidelines**

N/A