Syllabus for *New Directions Computational Systems Biology* CS 6824, Spring 2011

Meeting times	MW 2:30PM–3:45PM, 208 Randolph Hall
Instructor	T. M. Murali
Office	2160B Torgerson Hall
Office hours	9:30am-11:30am Wednesdays and by appointment
Email	`murali AT cs DOT vt DOT edu`
Mailing list	`CS5854_12061@listserv.vt.edu`
WWW	`http://courses.cs.vt.edu/~cs6824`

From the Graduate Catalogue

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: Graduate standing in CSA (3H,3C).

Description

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* (CSB) 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.

This graduate course will introduce students to CSB, especially approaches developed to study fundamental biological questions such as:

- What are the properties of the molecular interaction networks in a cell?

- How do these networks interact with each other?

- How does a biological system behave over time and under various conditions?

- How does disease result from disruption of normal cellular activities?

This course will emphasise a data-driven approach to systems biology. This approach is in contrast to physics-driven methods that construct models of (small) biological networks and simulate them. The data-driven methodology, or at the least what we focus on in this course, tries to uncover and understand the large-scale structure of molecular interaction networks by analysing and integrating massive amounts of different types of data. In the case of the vast majority of data-driven methods, manually-intensive examination of the results is necessary in order to propose hypotheses that can be validated in the wet lab. *In this course, we will focus on some of the latest approaches that have been designed from the outset to more easily suggest experimental hypotheses and validation studies.* This course will discuss such methods, mainly We will use a selection of papers from the literature to discuss these topics. Lectures by me, student presentations, and one or two invited lectures will drive the course.

Intended Topics

We will discuss research papers in the following topics:

- the analysis of DNA microarray data: basic clustering algorithms; computing condition-specific patterns of co-expression; identifying sets of co-regulated genes.

- high-throughput biological data: experimental techniques for detecting protein-protein interactions, protein-DNA binding, protein phosphorylation, metabolite concentrations; assessing data quality.

- regulatory networks: determining their basic building blocks and structural properties.

- interaction networks underlying infectious diseases: prediction and analysis.

- obtaining biological insights by combining diverse sources of data: functional annotation of proteins; improving the reliability of the data; refining our current knowledge about biochemical pathways.

The course web-site will contain a detailed list of papers on each of these topics that we will select from for discussing in the course. Students can suggest other topics and papers to me. If you have any such ideas, talk to me as soon as possible so that we can include them in the schedule before it is settled.

Pre-requisites

You must have graduate student standing in the Department of Computer Science, a Life Science department, the Department of Mathematics, the Department of Statistics, or the Genetics, Bioinformatics, and Computational Biology Program. The course is also open to undergraduate students; please speak to me first.

Computer Science students

- Required: Data and Algorithm Analysis (CS 4104).
- Required: Biological Paradigms for Bioinformatics (GBCB/PPWS 5314).
- Preferred: Algorithms in Bioinformatics (CS 5124).
- Useful: Applied Combinatorics (MATH 3134).

Life Science students

- Required: Undergraduate or graduate training in biology, genetics, biochemistry, or a similar subject.
- Required: Computation for the Life Sciences (CS 5045/5046).

Other students should meet me to discuss your qualifications. I will accept coursework at the undergraduate or graduate level that is equivalent to these pre-requisites. Computational students should also have prior programming experience and should be able to create a software package from a given project specification.

Course Format and Grading Policy

Lectures, presentations of papers from the literature by individual students or groups of students, and some invited lectures will drive the course. **Class participation, i.e., active participation in class discussions is extremely important.** *Merely attending class does not constitute class participation.* Your group presentation will make up 30%, your class participation 30%, and your final project or term paper 40% of your grade. This ratio is commensurate with the amount of effort I expect you to put in each of these items.

I encourage students to form groups of 2–3 people before the second class of the semester, go over the list of papers on the course website, and carefully consider what topics and the papers you want to present. Choose at least two or three topics in case your choice conflicts with somebody else's. We will assign papers and topics to groups in the second or third class (January 24 or January 26). Student presentations will begin about halfway through the semester.

Presentation The group of students assigned to present the papers should prepare a formal presentation and lead the discussion. Prepare your presentation like how you would for a conference or symposium. Use a set of prepared slides to drive your presentation but use the blackboard if necessary. I would like a copy of each group's slides. Please provide them to me in PDF format.

Participation Each class will have one–two (and perhaps three) assigned papers. I also expect every student to carefully read each paper assigned to each class and participate in the discussion. Given the interdisciplinary nature of CSB, it is critical that both life and computational scientists contribute to our understanding of the research we will all learn about.

Final project Students have a choice of a research project or term paper. **Although I call this a "final" project, successful work will require sustained effort throughout the semester.**

Research Project I will describe the projects in class on January 24 or 26. Each group and I will decide on the specifications of the project. A significant component of your project must involve performing biological analysis of your results; if your group contains a life-student student, this analysis will primarily be his or her responsibility. A high-quality software project can result in a publication in a premier conference like ISMB, RECOMB, PSB, and/or an appropriate journal. **Only groups of highly motivated students with some research experience should attempt research projects.**

Term paper Each student choosing this option should pick a topic not covered in class and a set of papers on that topic. The student should write a term paper on these papers: describe the problem area, the methods used, the results found; compare these papers; and discuss future research directions. Write the paper as if you were submitting a review paper to a forum like *ACM Computing Surveys* or one of the *Nature Reviews* journals. If your paper is well-written, thoughtful, and broad in scope, you may indeed be able to submit it to such a journal.

Announcement

If any student needs special accommodations because of a disability, please contact me during the first week of classes.