CSCI2950-C: Topics in Computational Biology: Fall 2010

Meeting time: Tu/Th 10:30-11:50 CIT 241
Instructor: Ben Raphael, CIT 505, braphael@brown.edu
Website: http://cs.brown.edu/courses/csci2950-c/

Description

We will examine new and classic problems in computational biology including: genome assembly, genome rearrangements, phylogeny, and cellular interaction networks. We will introduce algorithmic, modeling, and machine learning challenges that emerge in these topics, and explore applications on these techniques to human and cancer genome data.

The course will be organized in seminar style. Following introductory lectures on (topics below), students will read and present recent research papers, and will undertake a research project. The project will include a written proposal, midterm report, and final presentation.

Prerequisites

An undergraduate-level background in algorithms and/or probability/statistics.

Grading

Papers: 40% total
- Reviews 20%
- Presentations 20%

Project: 50% total
- 10% Written proposal.
- 10% Written midterm report.
- 20% Final written report.
- 10% Final presentation.

Participation: 10%

Papers

Reviews: You will submit a written review for each discussion paper, with the exception of those that you present. Reviews are due via email (to braphael@cs.brown.edu) BEFORE the start of class on the day that the paper is presented. Reviews will be graded on a 3-point scale: 2 = mastered the key issues introduced in the paper 1 = read the paper and understood the basics; 0 otherwise.

Presentations: Each student will present at least two papers from the reading list. The exact number will depend on class size. Whiteboard talks presentation, the student will outline the presentation. The talk slides or a PDF summary (if a chalkboard talk is given) will be submitted on the day of the presentation to be posted on the web page. A long presentation is expected to be about 40 minutes, and a short presentation 20 minutes with the remaining class time for questions and discussion. Students are *strongly* encouraged to discuss presentation with instructor beforehand.

Project

The project is a semester-long effort to further study one of the class topics. The project could range from theoretical (e.g. designing a new algorithm and proving its correctness), to the practical

(a software implementation) depending on the interest of the student. Projects can be undertaken in groups of 1-3 students. Students are encouraged to propose a project matching their research interests and discuss with instructor. A list of suggested projects will also be distributed.

Participation

All students are expected to contribute to paper discussions by asking questions, making observations, identifying strengths and weaknesses of the approaches under discussion. Critical discussion and analysis are key prerequisites for research.

Course Credits

PhD: Area B (Algorithms)
ScM: "Theory" or "Practice"* course.
Significant Programming*
*With appropriate class project.

Course Outline (Subject to change)

I. Introduction -- Biology review
Genomes. Central dogma: DNA, RNA, and proteins.

II. DNA sequencing, genome Assembly and Resequencing (~ 3 lectures, 3 papers)
Shortest common superstring.
Hamiltonian and Eulerian path formulations.
Next-generation sequencing technologies and assembly.
Variant discovery.

III. Genome Rearrangements (~ 2 lectures, 2 papers)
Sorting by reversals
Breakpoints and breakpoint graph
Hannenhalli-Pevzner theory for Inversions and Translocations

IV. Phylogeny and Ancestral Genome Reconstruction (~ 3 lectures, 2 papers)
Phylogenetic Trees
Parsimony and Likelihood Models for Ancestral Reconstruction
Orthology and paralogy; Gene trees and species trees.

V. Cancer Genomes (~ 2 lectures, 3 papers)
Measuring Mutations in Cancer
Models of Cancer Progression and Evolutionary Dynamics

VI. Cellular Interaction Networks (~ 4 lectures, 4 papers)
Regulatory and Protein Interaction Networks
Network Alignment and Motifs
Models for signaling networks
Network Perturbations in Cancer