CSCI 2840 - Advanced Algorithms in Computational Biology and Medical Bioinformatics

Prof. Sorin Istrail
Department of Computer Science
Brown University

Course meeting: Tues./Thurs. 2:30-3:50pm, CIT 241 (SWIG)

Course website: `https://cs.brown.edu/courses/csci2840/spring-2025`

Staff:

- Professor Istrail. `sorin_istrail@brown.edu`
- Colin baker (HTA). `colin_baker@brown.edu`
- Pranav Mahableshwarkar (UTA). `pranav_mahableshwarkar@brown.edu`

1 Course Description

The aim of this course is to examine the impact of mathematics and computer science on the field of medical bioinformatics. We will cover a number of areas in computer science with profound medical applications, with a focus on how they impact research on mental disease.

Course Topics

The course is organized into five chapters:

1. SNPs and Haplotypes: Linkage Disequilibrium and the Haplotype Phasing Problem
2. Protein Folding, Misfolding, and Disease
3. Markov Chain Monte Carlo and Spectral Graph Theory, with Applications to Population Stratification
4. Missing Heritability, Genetic Heterogeneity, and Rare and Common Genetic Variants
5. Polygenic Risk Scores and GWAS

Each chapter is devoted to a fundamental field of research and mathematics in medical bioinformatics. We will explore not only the foundational algorithms that built up to famous notions like GWAS and AlphaFold but also how they can be applied to specific diseases, including Alzheimer's, autism, and more. As in each of Professor Istrail's undergraduate courses, we will come to understand together the most elegant of "beautiful" algorithms in the medical arena.

While we will still emphasize rigorous mathematical and statistical concepts in computer science like runtime and convergence as they apply to our algorithms, the course will be focused on their applications in medicine. In each chapter, we will devote our time in roughly equal proportions to selected algorithms and their medical relevance. Although there are many applications to discuss in each chapter, we will primarily consider mental disease through unique datasets and examples for each assignment.

Prerequisites

The course is designed for graduate students and upper-level undergraduates. It is open to computer science and math students, as well as biological and medical students. While there are no formal prerequisites for the course, you should have a strong background in at least one of these two areas. Please contact the professor if you are unclear as to whether you have the necessary prerequisites for the course.

2 Course Format

Meeting times and place:

Tuesday and Thursday, 2:30-3:50pm in the SWIG, CIT 241. **We will additionally meet at some other time during the week for paper-reading presentations, TBD**.

You are expected to attend all classes. Class lecture notes will be made available as they are compiled. Since a unified set of notes will be new this year, please be patient as we compile them on a regular basis.

Assignments

Homeworks and paper-reading presentations will alternate throughout the semester. There will be three of each.

Each homework (HW) will focus on the conceptual frameworks discussed in one or two recent chapters. They will consist of some shorter conceptual questions, in addition to a few longer programming objectives. Homeworks will generally be available over a 1-3 week period and will be evaluated on both correctness and justification.

Each paper-reading presentation (PP) will involve presenting a modern paper in an area related to the most recent homework. A week in advance of each presentation date, we will collect (via google form) the paper each student will present. If a selected paper is not in the correct research area, we will ask the relevant student to choose a new paper prior to the presentation date (although we will be quite flexible, within reason). Presentations will be done alone or in pairs; individuals will be expected to present for 10 minutes, while pairs will be expected to present for 15 minutes. These limits are subject to change as course enrollment reaches equilibrium. Presentations will be evaluated on understanding of content (40%), presentation skills (30%), Q&A (15%), and engagement with other presentations (15%). The engagement portion is a binary value depending on whether or not a student asks a question of another group.

During the last 6 weeks of the course, each student will embark on a final research project, which we will allow to overlap their existing research, as long as there is some presentable piece of code that is also related to the course content. We will host final presentations during the reading period, to be scheduled in the coming months.

Grading

- Homework 30%
- Literature review presentations 30%
- Final project 40%

Grades will be determined by your overall performance according to these metrics. At the end of the class, a *Pastiche Pie* award will be given to the student(s) with the overall most impressive performance in the class as judged by the TAs and the professor. All final grades will be determined by the professor.

Literature

There is no textbook for this course, although "Principles of Population Genetics (Fourth Edition, 2007) Daniel L. Hartl and Andrew G. Clark" may be an interesting supplement. Suggested readings will be provided in the resources page of the course website to complement the lecture content of the class. Any required readings will be made clear in one of the course homeworks.

3 Course Policies

Collaboration Policy

In addition to Brown's Academic Code, CS 2840 follows the collaboration policy below:

- You may discuss **conceptual problems** with other students in the class; however, all solutions must be written up independently and reflect your own understanding of the material.
- You may discuss **programming problems** and compare output with other students in the class; however, all code must be written up independently. You may not examine code written by other students.

- For both **conceptual and programming problems**, you **may not use generative AI**. If we suspect the use of generative AI (*which is much more apparent than you think!*), we reserve the right to report any infraction to the university.
- You will be required to accept this collaboration policy electronically at the beginning of the semester as a prerequisite for receiving grades for all subsequent assignments.

The course staff takes violations of the collaboration policy seriously and will prosecute with the standing committee on the academic code as necessary.

Late Handin Policy

You will receive 3 late days for use throughout the course. As all handins will be electronic, you may use these late days at your discretion, with two caveats:

- You may use a maximum of 2 late days per individual assignment
- You may only use late days on homework assignments

Extra late days will be additively penalized 15% each. Additional extensions on homeworks and projects will only be granted by the professor under extenuating circumstances and at his discretion. TAs cannot grant extensions.

Coursework Hours

Students will spend 3 hours in class for 14 weeks for a total of 42 hours. Homework assignments are expected to take 15 hours each (5 conceptual, 10 programming) for a total of 45 hours. Literature reading and presentation assignments are expected to take 15 hours (3 selection, 10 reading and preparing, 2 presenting) each for a total of 45 hours. The final project, including research and presentation preparation, is expected to take 48 hours (25 research, 20 preparing a presentation, 3 presenting). The total amount of time for the entire course will be 180 hours.

Diversity, Inclusion, Accessibility & Accommodations

Brown is committed to the full inclusion of all students, and CS 2840 strives to be a welcoming and inclusive place for the diverse student body. Please reach out to the professor if you have any concerns regarding inclusivity, accessibility, or SEAS accommodations.