CMSE 410/890 - Bioinformatics & Computational Biology

Arjun Krishnan

arjun@msu.edu | @compbiologist | thekrishnanlab.org

Day 01: Introduction and Overview

Course overview

- Introductions
- Course website
- Communication
- Course activities
- Tentative Schedule
- Course topics
- Wrap-up

Introductions

- arjun@msu.edu | @compbiologist | thekrishnanlab.org
- Assistant Professor
  - Dept. Computational Mathematics, Science, and Engineering
  - Dept. Biochemistry and Molecular Biology
- Research Interests: Computational genomics, Biomedical data science, Biological networks, Natural language analysis, Data integration, Machine learning

Introduce yourself to one other person in this class you've not met before:

Say your name, department/program

Introductions

Introduce yourself on the #welcome channel on Slack with:

- Name:
- Preferred pronoun:
- Three words/phrases to describe you/your-interests:
- Research/interests in emojis:

If you have not joined Slack, now is a good time to do it.

Come & talk to me if you've not received an invitation.

What you should get out of this course

How to become a practicing computational biologist in 60+h?

- Introduction to the inner-workings of methods in bioinformatics and computational biology:
  - Analytical techniques, algorithms, and statistical/machine-learning approaches developed to address key questions in biology and medicine.

- At the end you should be able to:
  - Critically read bioinformatics / computational-biology literature.
  - Apply the methods you have learned to problems both within & outside biology.

Modules & Topics

1. Genome assembly & annotation	• de Bruijin graphs; Hidden Markov models
2. Sequence alignment & pattern finding	• Dynamic programming; Substitution matrices; Fast Local Alignment
3. Comparative genomics; Phylogenomics	• Suffix trees; Tree construction
4. Genetic variation & quantitative genetics	• Regularized linear regression; Statistical inference, Multiple testing
5. Regulatory genomics	• Gibbs sampling; Expectation-Maximization
6. Functional genomics	• Two-sample tests; Hypergeometric test; Unsupervised/supervised learning
7. Single-cell genomics	• Missing value imputation; Dimensionality reduction; Trajectory inference
8. Molecular dynamics; Structure prediction	• Molecular simulation; Maximum entropy modeling
9. Modeling cellular pathways	• Dynamical simulation, State Space, Bifurcation; Discrete/Boolean modeling
10. Whole-cell models; Digital evolution	• Wiring diagrams; Constraint-based modeling; Artificial life; Whole-cell models
11. Biological networks	• Measuring associations; Network inference; Graph theory, Label propagation

Modules & Topics

1. Genome assembly & annotation
2. Sequence alignment & pattern finding
3. Comparative genomics; Phylogenomics
4. Genetic variation & quantitative genetics
5. Regulatory genomics
6. Functional genomics
7. Single-cell genomics
8. Molecular dynamics; Structure prediction
9. Modeling cellular pathways
10. Whole-cell models; Digital evolution
11. Biological networks

- de Bruijin graphs; Hidden Markov models
- Dynamic programming; S
- Suffix trees; Tree constr
- Regularized linear regres
- Gibbs sampling; Expecta
- Two-sample tests; Hype
- Missing value imputatio
- Molecular simulation; Ma
- Dynamical simulation, S
- Wiring diagrams; Constr
- Measuring associations;

MW: Arjun presents

Each week:

- Lecture
- Paper discussion

F: You present/lead as team of two.

Prerequisites & Expectations

- CMSE 201 and two semesters of introductory biology (LB 144 and 145 OR BS 161 and 162 OR BS 181H and 182H, or equivalent).
  - Statistics at the level of STT 231 is strongly recommended.

- Basically, it would be assumed that you:
  - know how to code in one of the mainstream languages like Python or R,
  - have an understanding of basic statistics and probability, and
  - have studied basic genetics, molecular biology, and cellular biology.

Course website

bit.ly/compbio2020

- Contact information
- Course outline and materials
- Schedule, location, calendar, and office hours
- Website and communication
- Course activities
- Grading information
- Attendance, conduct, honesty, and accommodations

- Lecture slides
- Learning materials
- Assignments

Communication

compbio2020.slack.com

- The primary mode of communication in this course (including major announcements) will be the course Slack account.
- All of you should have invitations to join this account in your MSU email.

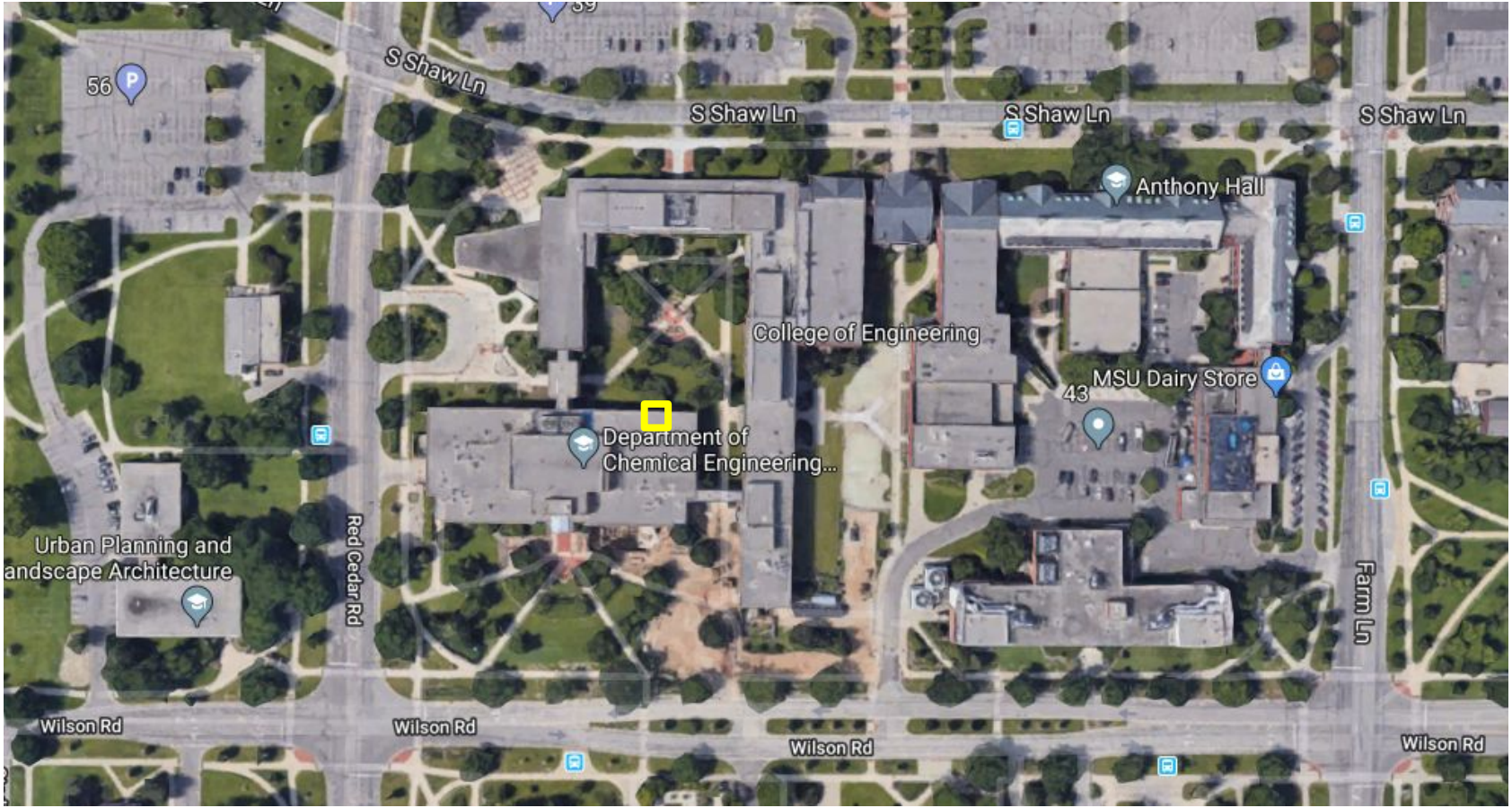
#syllabus-schedule	#assignments-reading
#announcements	#project
#slides-materials	#random

Tell us about you

- Name
- Major or graduate program
- Research interests and/or topics you're especially interested in learning about
- Background in programming, algorithms, statistics, biology
- Office hours

Survey: bit.ly/compbio2020_incoming

My office: 2507H Engineering Building (2nd floor)



Course activities

- Assignments: ~35%
- Class participation: ~15%
- Project: ~50%

Assignments

- For each topic, you will be given an assignment after the topic's first "Lecture" class on **Monday**:
  - A paper to read, summarize, and critique.
  - A software tool to explore.
  - An exercise to work on.
- Submit a report *before* the topic's "Paper discussion" class on **Friday**.
- See class website for details on what this report needs to contain.

Class participation

- Do the assignments and additional readings.
- Show up to class and set your phones asides.
- Work in groups during in-class discussion sessions.
- No one will have the perfect background.
    - Ask questions about computational or biological concepts.
- Correct me when I am wrong.

- Paper presentation (see the class website).

Project

- A major goal of this course is to prepare your ability to perform original research in computational biology, and to present your ideas and research.
- Can be one of:
  - Design and implement a new computational method for a task in biology.
  - Improve an existing method.
  - Perform an evaluation of several existing methods.
  - Develop a fully-reproducible codebase for an existing analysis in a paper.

Project

- Project profile due Wed, Jan 15.
- Project topic due Fri, Jan 31.
- Project pre-proposal due Fri, Feb 07.
- Project proposal due Wed, Feb 19.
- Proposal reviews due Fri, Feb 28.
- Mid-course project report due Fri, Mar 27.
- Final project report due Fri, Apr 26.
- Final project poster presentation on Thu, Apr 30.

Groups @ MSU doing (a sigf. amount of) Bioinfo & Compbio

- Jianrong Wang
- Yuying Xie
- Alex Dickson
- Michael Feig
- George Mias
- Ana Vasquez
- Gustavo de los Campos
- Wen Huang
- Eran Andrechek
- Robin Buell
- Erik Goodman
- Kevin Liu
- Mark Reimers
- Sudin Bhattacharya
- Jiayu Zhou
- Shinhan Shiu
- MANY OTHERS

The **MSU_Bioinfo-Compbio-Investigators_2020.md** file in the class website contains a full(er) list.

You are welcome to work with me as well!

What you need to do before the next class

- Read the course website: bit.ly/compbio2020
- Fill out the incoming survey: bit.ly/compbio2020_incoming
- Sign-up for scribing and paper presentation: check Slack for link
- Introduce yourself on Slack