

CSDS/SYBB 459: Bioinformatics for Systems Biology

Course Objectives

In the last two decades, there has been an explosion in generation of biological data that provide genome-scale information on cellular processes. These data include genomic sequences, chromatin structure, DNA methylation, gene expression, protein expression, protein-protein interactions, post-translational modifications, protein-DNA interactions, metabolite concentrations, and metabolic fluxes, among others. Such data are often analyzed within the framework of biological network models, which provide a large-scale view of the functional relationships among multiple molecules in the cell.

This course provides a detailed review of computational approaches that are developed to analyze these novel sources of biological data. Topics include prediction of protein-protein interactions, genetic regulatory network inference, analysis of network topology, network motifs, network alignment, network-based functional annotation, gene set enrichment analysis, and network-based analysis of genomic, functional genomic, and proteomic data with applications to the systems biology of complex diseases. It is expected that, upon completion of this course, the students will accomplish the following learning objectives:

- Become familiar with existing tools and resources for computational analysis of biological networks.
- Develop awareness of the computational problems that arise in the modeling and analysis of cellular systems.
- Understand fundamental abstractions and computational approaches used to formulate and address these problems.
- Be able to use, manipulate, and extend existing computational infrastructure for analyzing systems biology data.

Class Meeting

- MW 12:45-2:00 PM, Olin 408.
- Zoom Meeting ID: 972 9913 8510, Passcode on course Canvas site.

Course Personnel

- **Instructor:** Mehmet Koyutürk
Office: Olin 512

Phone: 368-2963
e-mail: mxk331@case.edu
Office hours: By appointment

- **Teaching Assistant:** Sean Maxwell
e-mail: stm@case.edu
Office hours: By appointment

Course Material

No textbook is required. Class notes and reading materials, mostly composed of recent review and research articles, will be provided by the instructor.

Course Website

Canvas: <https://canvas.case.edu/courses/31421>

Course Work & Grading

Topics: The course is organized into 5 chapters/modules. For each chapter/module, a set of slides and a set of required/recommended reading will be provided on Canvas. During the discussion of each chapter, there will be exercises that will be done collectively in class. There will also be discussions centered around the required reading (which are mostly survey/review papers).

The chapters/modules that will be covered will be the following:

- Chapter 1: Systems Biology, Omics, Bioinformatics - Conceptual introduction to systems biology and computational approaches to studying systems biology.
- Chapter 2: Network Biology and Graph Theory - What biological networks represent, where they come from, and what are computational/mathematical approaches to their basic topological analyses.
- Chapter 3: Omics, Enrichment, Network Inference - Molecular data sources that are available, how these data are generated, basic analysis tasks involving these datasets, the use of molecular data to make systems-level inferences, including enrichment analysis and inference of associations and regulatory interactions among biological entities.
- Chapter 4: Machine Learning in Biology - High-level discussion of considerations in applying machine learning to biomedical data analysis and translational research, basic framework for developing and validating machine learning workflows.
- Chapter 5: Graph Machine Learning for Biology - More advanced approaches to network biology with applications to predictive tasks, including network propagation, network embeddings, and graph neural networks.

Engagement: (10%) The students are expected to review the material to be covered before each class meeting, attend the class meetings, and actively participate in class discussions. It is imperative that students ask questions whenever something is not clear to them in the class. The term project requires students to be engaged with each other's projects, ask questions, and offer ideas to each other.

Awareness of diversity and inclusive behavior is expected. All students (including racial minorities, ethnic minorities, religious minorities, women, LGBTQ+, students with disabilities, non-native speakers of English, and other under-represented groups) should know that they are all welcome and they are part of this classroom. It is also important to note that this is a highly interdisciplinary course in which students can come from very different academic backgrounds and something that is trivial for many can be highly complicated and difficult to understand for others. Students are expected to take this into account and try to communicate effectively across disciplinary boundaries. Students are encouraged to reach out to the instructor whenever they feel left out or feel like the classroom does not provide an inclusive learning environment.

Assignments (60%): There will be 4 assignments, each asking students to explore a dataset and answer specific questions. The objective of the assignments is to align the concepts and ideas covered in the class with basic programming/scripting tasks, and provide exposure to biological data analysis with emphasis on comparison of different approaches, testing of hypotheses, validation of methods, and visualization.

The four assignments will be on the following topics:

- Assignment #1: Topological Analysis of Protein-Protein Interaction Networks.
- Assignment #2: Analysis of Single-Cell Gene Expression Data.
- Assignment #3: Network Inference.
- Assignment #4: Graph Machine Learning for Network Biology.

Students are allowed to complete the assignments individually or in teams of up to 3 members.

Term Project (30%): All students will complete a term project by analyzing data to test a specific hypothesis. The hypothesis must be clearly formulated, the data sources must be identified, and the data analysis plan must be clearly formulated before students move forward with the implementation.

In the first half of February, all students will present their ideas for the hypothesis they will test, also showing their preliminary research into identifying the data sources and computational tools/approaches they will use to test their hypothesis. These presentations will be informal, with the understanding that the entire class will provide input and offer ideas on the project. At this stage, the students may have already formed teams or they can form teams based on the ideas presented (e.g., a computational person approaches to a life science person who presents an idea they like or vice versa). Once the students receive feedback from the instructor and the class, they will finalize their plans and execute the project. The results will be presented at the end of the semester.

Teams (which can be formed ahead of time or during/after proposal presentations) are required to be interdisciplinary (all students in a team cannot be enrolled in the same program). Students are allowed to execute their projects individually if they choose to do so.

Accommodations for Disabilities

In accordance with federal law, if you have a documented disability, you may be eligible to request accommodations from Disability Resources. In order to be considered for accommodations you must first register with the Disability Resources office. Please contact their office to register at 216.368.5230 or get more information on how to begin the process. Please keep in mind that accommodations are not retroactive.

Plagiarism Policy

Zero-tolerance policy on plagiarism is enforced. In case of doubt, the students are responsible for checking with the TA or the instructor on what is allowed and what is not. If plagiarism is detected, action will be taken in accordance with the university's Academic Integrity policy: <https://bulletin.case.edu/undergraduatestudies/academicintegrity/>.