

Final-Project

October 22, 2018

1 Final Project

CBIO (CSCI) 4835/6835: Introduction to Computational Biology

This is an opportunity for you to combine a problem in biology you find interesting with some of the concepts you've learned in CBIO 4835/6835.

Put another way, it's an opportunity for you to inflict pain and suffering on your instructor!

Ideally you will come up with something related to your own research or interests that includes real data, but if you need project ideas I can provide them. You may use any Python packages as long as they can be installed with a package manager (pip or conda).

You are required to provide:

- A writeup with sufficient detail for a student to understand and implement the assignment in a common word processing format (Word, HTML, LaTeX, Jupyter notebook).
- The python code to the solution. The solution should require 20-100 lines of code.
- At least three example inputs including the necessary commandline to get the output. Make sure it is okay to publicly release the data and include any appropriate acknowledgements/citations.

1.0.1 Some General Guidelines

You are allowed to work either individually or in pairs. Be aware that if you work with a partner, you will both receive the same grade and your team will be expected to accomplish more than if you were working alone.

The amount of work should be around 1 to 1.5 homework assignments; for 2 people, this would obviously double. That said, don't overextend yourselves!

You may not 'recycle' papers or reports that you wrote for some other class or as part of a research project.

A project is **required for students in 6835**, but **optional for students in 4835**. Anyone who does the project is exempt from Assignment 6, and students in 4835 can do the project for extra credit!

There are **three components** to the final project; they are as follows.

1.0.2 1: Proposal

Due Friday, October 26 at 11:59pm

Proposals should be no more than 1 page (not including references) and should contain the following:

- **Background.** Introduction that describes the biological context of your question and any previous work.
- **Goal.** A succinct state of the question that you will address and how you will use computational tools to address it.
- **Approach.** Description of what work you will perform and what tools you will use. Be as specific as possible here. Think in terms of the figures that you might obtain as results.
- **Assessment.** Describe how you can verify / validate your results. How would you convince someone that they are correct? What does “victory” look like for your project? This may entail envisioning possible pitfalls or failure modes of your project.
- **References.** Provide references to all papers that are providing, background, models, or data that you are using for your project.

1.0.3 2: Presentation

Tuesday, November 27

The presentation should be about 10-15 minutes (followed by 3-5 minutes of questions) per project.

(this may vary depending on how many presentations we have, but aim for the 15-minute mark)

They should accurately summarize the problem and its background as put forth in your proposal, as well as demonstrate the progress you have made.

Your project doesn't have to be 100% complete. **But DO NOT misconstrue this—it should be AT LEAST 70-80% finished.** Maybe you're waiting on one more experiment, but the bulk of the work should be finished at this point.

1.0.4 3: Write-up

Due Thursday, December 6 at 11:59pm

This is the write-up alluded to at the beginning. It can take any readable format you'd like (e.g. Word, Jupyter notebook) but should be **readable** and **understandable** to fellow students.

You don't have to follow this structure exactly, but a fail-safe organizational strategy would involve:

- **Introduction:** Here is where you will establish what questions you are asking and why they are important. You will state concretely the objectives of your work. You will also provide sufficient background that a general scientific reader can understand what you are doing and why.
- **Related Work:** You will describe briefly and give relevant citations to other work that addresses the specific systems and mechanisms you are investigating. If you are building on an existing model, you should cite that model and give a brief description here, or possibly in the Introduction.
- **Methods:** You will describe the methods, code, and software that you used to develop the model and simulations results you are presenting.
- **Results:** Here you will describe the results of your simulations and other analyses. If you developed a new model you can also describe the content of that model in the first section of Results (you would describe the code and software though in Methods). For example, if you developed an ODE model using SciPy, you would briefly mention and cite SciPy and ODE's in the Methods section, but present the actual model—species and reactions—in the Results section.

- **Discussion:** Discuss and interpret your findings, paying particular attention to any results that you think may be incorrect with an explanation of what you think might have gone wrong and how you would fix it going forward.
- **Conclusions and Future Work:** Summarize and state whether you did or did not achieve the initial goals of the project.
- **References:** Be sure to include complete list of authors and the title of the paper.

You should also include any code or data; if you want to use GitHub to store your code (where [all the materials for this course](#) are hosted), please let me know and I will create a repository for you.

1.0.5 Some [potential] Project Titles

- Agent-based Modeling of the Zombie Apocalypse
- Multiple sequence alignment for predicting structure from sequence
- A Network Approach to Global Scale Epidemic Spreading - Ebola Case Study
- Analysis of the kinetics of Influenza A infections in humans via mathematical modeling
- Automated cell counting and tracking
- Simulating evolution of yeast pheromone signaling pathway
- Modeling of rapid activation of an apoptosis pathway in response to DNA damage
- Pharmacological analysis of a dynamic apoptosis-autophagy model

1.0.6 Some Starting Points

- Nature, Science, PLOS Biology
- Cell
- Molecular Systems Biology, Science Signaling, Nature Systems Biology
- PLOS Computational Biology
- Biophysical Journal
- Interface
- Molecular Biosystems
- BMC Systems Biology
- PLOS One
- arxiv.org/q-bio (A lot of interesting work appears here in preprint form)

1.0.7 A final note on “success”

tl;dr: **It is 100% acceptable if your project does not pan out.**

What matters to me is your *process*: you demonstrated an understanding of the problem, a good grasp on how a computational approach might help, and you executed it as best you could.

Maybe the effect you were hoping for just wasn't there. **That is completely OK!** I just want to see that you have an idea of how you could incorporate some of what you've learned in this class into a real-world research problem.

1.0.8 Questions?

Post in the Slack #questions channel!