

## CSC 2417 Algorithms for Genome Analysis Project Handout

### Don't Panic

In lieu of a final exam, the class will have a term project. For the project you are strongly encouraged (but not required) to team up with one other person. Groups of three people will be allowed in rare circumstance if there is a very good reason to have them. The topic for the project is defined as “anything in Computational Biology”, with Computational Biology defined as broadly as possible. You are welcome to work on anything including applying known algorithms to different data sets (very applied), developing novel algorithms (mix of application & theory) and proving lower or upper bounds on certain formulations of biological problems (very theoretical). One requirement is that your project, while not necessarily ground-breaking or publishable, should address areas that are of current interest to people in the field (see “Proposal” section for how to demonstrate this). You are strongly encouraged to speak with the instructor about your ideas about topics, but you should come up first with some ideas of your own.

There will be three deliverables for your project: an initial project proposal, and once your work is done a presentation before the class and a final write-up of your results. For the written parts only one writeup is expected from each group. Each of the three deliverables is equally important to having a successful project, but there will not be a firm division of the grade – rather the grade will be based on a “total impression” from all three, with a greater weight for the final write-up. It will look bad if one section is done poorly, but even worse if all of them are mediocre.

#### **Project proposal (due November 2)**

In the proposal you should outline what research you intend to undertake. You should summarize existing work on both the problem you are planning to solve and the method you plan to use – for example you may say “Problem A is important in biology and has been attacked using computational techniques X and Y. Method Z is a complicated algorithmic (or machine learning, or database) technique that has been successfully used for problems B and C. However no one has applied Method Z to problem A, and we think it will do better than X and Y because it models the parameters of problem A more accurately.” Analogously, if you are working on a more theoretical paper you should outline how you plan to tackle it (“Reduction from problem A to problem B”). You should explain some of the challenges you expect to encounter and the possible solutions. Finally you should be explicit about how you plan to test your method, whether it is using real or simulated data, and mention any gold-standard data-sets that currently exist for your problem. You are also welcome to speculate why no one has applied method Z to problem A. To demonstrate that the problem and/or methodology you are working on is of current interest you are expected to cite and discuss at least two, and if possible more papers related to your work that have been published in 2004 or later. While there is no hard length requirement for the proposal, you should make it as short as possible while still giving a through background to both the problem and methodology and an

explanation of how you plan to apply one to the other. The expected length of the proposal is in the range of 3-5 pages.

### **Presentations (Dec 5 & 7)**

You will be expected to present the results of your work in a short (10-15 minute) talk before the class. The current plan is for groups of two to present for 15 minutes, while individual presenters will get 10 minutes. The current plan is for the presentations to be on December 5 & 7, in place of our usual class. Groups will be expected to “present” jointly, either by dividing up the talk into two parts or by sharing responsibilities in some way. Talks where only one person presents for the whole group are very strongly discouraged.

### **Final report (due mid December; exact date TBA)**

While the work you do may not be publishable (you are not expected to improve on any current results – though if you do this would be great), the final report should in its manner of presentation resemble a paper ready to be submitted to a conference or a journal. You should introduce both the problem and techniques, demonstrate your results, compare your results to previously known ones, and possibly suggest some future directions for research. Regardless if your algorithm did or did not do as well as you expected you should speculate why it achieved the results it did. The exact length is again open-ended, but you are encouraged to keep it short – probably no submission should be more than 10 pages.

### **Advice**

Once you have some ideas about projects I strongly suggest you talk with me about them by scheduling an appointment or during office hours. Do not leave this till the last second, as you may find me impossible to find! I also strongly encourage you to talk with other students both about potential topics and any problems you encounter while solving them. Think of the class as a large research group – people working on different, but related problems who can give each other useful advice. Finally we have (or should have soon) a class newsgroup. You are more than welcome to post any concerns there and also to respond to other people’s messages.