Announcements 5/2

- Read Lathrop and Smith (1996)
  - Protein threading
- HW4 updates
  - Part 1: use generalized suffix tree
  - Part 3A: test example_graph.txt with k=7
  - Part 3A: multiple equivalent solutions with cost of 26
- Project report guidelines posted
  - Emphasis on interpretation of results
  - Pay attention to the rubric

Final exam

- Final exam Sunday May 7 at 2:45 PM
- Similar style as midterm exam
- No printed materials allowed
- Will not execute algorithms by hand
  - Do need to understand optimization, objectives, etc.
- Almost all emphasis is on material not covered on midterm
  - 7 themes: Noncoding variants to protein threading
  - May need prior material for comparisons, context
- *Slightly* more emphasis on material not covered on homework

Final exam

- Content includes (but is not limited to) material from lecture and required reading
  - Syllabus lists topics and require reading

Large-Scale and Whole-Genome Sequence Alignment

- topics: large-scale alignment, whole-genome alignment, suffix trees, k-mer tries, longest increasing subsequence problem, MUMmer
- required reading
  - A. Delcher, S. Kasif, R. Fleischmann, J. Peterson, O. White and S. Salzberg. Alignment of Whole Genomes. *Nucleic Acids Research* 27(11):2369-2376, 1999.

  - Focus on high-level concepts for generalized HMMs, pair HMMs, multiple large-scale sequence alignment, RNA structure energy-minimization
  - Optional reading is also helpful

Final exam

- Focus on concepts, strengths/weaknesses, algorithmic strategies
  - Why or in what cases would we use one method or experimental technology instead of another?
  - How does method X compare to method Y and what are the unique advantages of each?
  - What is important concept Z?
  - What does a method optimize?
  - What assumptions does a method make and what input does it require?

Final exam

- Examples
  - What is the difference in the objective function for the Nussinov and energy-minimization dynamic programming algorithms? Why is one preferable?
  - What is the difference between two Markov model or hidden Markov model variants?
  - Given these types of data, how do we use algorithms and ideas from lecture to model them?

Final exam

- Class evaluation incentive
- Will post specific sub-problem topics at particular completion thresholds
  - 2 problems for 13 / 16 completion
  - 2 problems for 16 / 16 completion
- Example from midterm:
  - Problem 2A: MEME and Gibbs sampling
  - Problem 2E: multiple testing
  - Problem 3A: Dirichlet prior