Advanced Algorithmics (6EAP)  
Project proposals

Jaak Vilo  
2012 Spring
Key info

- Project = 1-2-3 person teams
- Essay – May 24
- Poster & submission: May 30th
- Poster Session
  - Delay: 1 day = 20% of max points

- Prerequisite for exam
Expectations:

• Study the problem
• Implement, Evaluate, Compare, Measure, ...
• Your task is to make the project interesting to others: right questions; cool applications; novel ideas; desire to read; materials to complement next year courses.
• Find a clear objective and focus, state it, study it!
• 20-40h
• Report – Poster
Tasks

• Here is a list of some proposals

• You can propose your own.

• Or select some on your own
  – from international competitions
    • e.g. IOI (ACM) olympics finals series
    • implementation challenges from DIMACS, etc.
    • etc.
• Compare some alternative algorithms, and try to improve on them
• Take a problem and try to apply algorithmic problems to solve it
• Take an algorithm from the literature, implement and test
• Find cool ways to study/visualise algorithmic ideas presented in the course.
Combinatorial search

• Optimal solution from a (or any) state

• “Discover” the short assembly step algorithms

• Provide solutions
Bloom filter(s)

- Bloom filter storage of text using de Bruijn graphs (see video)
Graph layout

• Graph layout
  – “Physical Spring model” with some extra added constraints or specialised nodes for stars, cliques, connection strength, etc.

• Create a nr of criteria and try to minimize nr of crossings, area of graph, etc.
Constrained Spring Embedding Layout

• Define certain areas (or lines, etc) that “attract” nodes. Allow graph to “layout” itself dynamically.
Visit all cities... - physically!

http://cswww.essex.ac.uk/staff/sml/gecco/PTSPComp.html
http://algoval.essex.ac.uk/ptsp/ptsp.html
942, 941
• 652, 652
• 648, 636
Seriation


- Serialise matrices

- (2-way)
Biclustering

• Ordering rows and columns to reveal modules/areas of high “coherence”

• Example: A. Tanay, R. Sharan, R. Shamir: Discovering statistically significant biclusters in gene expression data. Bioinformatics 18, Suppl.1, 136-44, 2002
Query of OCT4 (POU5F1) (210265_X_AT)

StdDev < 0.29
Query of OCT4 (POU5F1)
(210265_X_AT: 50 top StdDev datasets in query)

http://eid.ee/b4
Some algorithmic competition

• Test your skills on some algorithmic competition
Finally, 15,000 pages later:

\begin{verbatim}
  -7  260  0
  7  -260  0
  1072  1070  0
  -15  -14  -13  -12  -11  -10  0
  -15  -14  -13  -12  -11  10  0
  -15  -14  -13  -12  11  -10  0
  -15  -14  -13  -12  11  10  0
  -7  -6  -5  -4  -3  -2  0
  -7  -6  -5  -4  -3  2  0
  -7  -6  -5  -4  3  -2  0
  -7  -6  -5  -4  3  2  0
  185  0
\end{verbatim}

Combinatorial search space of truth assignments: HOW?

\[ 2^{50000} \approx 3.160699437 \cdot 10^{15051} \]

Current SAT solvers solve this instance in approx. 1 minute!
Robust Regression
(Differential Evolution)

Fitting a regression line using minimum median error as a measure.

\[ aX + bY + c = 0 \]

\[ Y = aX + c \]

Find \( a \) and \( c \)
Robust Regression

least quantile of squares

(Gilli, Maringer and Schumann, 2011)

$$\min_{\beta} e^2_{(\alpha N)} \quad \text{where} \quad e = X\beta - y$$

$$e^2_{(j-1)} \leq e^2_{(j)} \quad j = 2..N$$
Differential Evolution

Fit any polynomial, use mean or median, add MDL based identification of the degree of polynomial

\[ A_n X^n + A_{n-1} X^{n-1} + \ldots + A_1 X + A_0 \]
Traveling Salesman

- Experiment with different meta-heuristics
Your own projects

• Ask a question
• Study literature
• Propose solution
• Implement
• Experiment and report results of experiments