Progress Report: ONR Award N00149710589 "Heuristic Approaches to Optimization with Applications"

Principal Investigator: Steven S. Skiena
Department of Computer Science
State University of New York at Stony Brook

August 30, 2001

1 Results

This progress report covers the twelve months of funding on this grant since my previous progress report (August 2000). This grant covers two distinct lines of work: (1) the development of algorithms and implementations of heuristic search with applications, and (2) theoretical research on combinatorial algorithms and optimization. As detailed below, substantial progress has been made in both areas since my last progress report.

This grant is currently supporting two PhD students starting September 1, meeting the scheduled expenditure for this fiscal year. Our work continues to prove interesting to a large number of Navy personnel, as documented by WWW hits. Since March 2001, our pages have received over 6,700 ".mil" hits, with navy.mil being the largest identifiable component.

2 Environments for Combinatorial Computing

We have made progress this year on several projects related to combinatorial computing:

- *Enhancements to Combinatorica* – We [3] have completed our effort to build an improved version of our combinatorial computing environment *Combinatorica*. Our primary goals are improved algorithmic performance and a better user interface. We have produced a beta-release of the software, as well as documentation of the new features; all available at http://www.cs.sunysb.edu/~skiena/combinatorica. This site also includes representative graphics produced by the system.

  The new *Combinatorica* is now capable of solving interesting problems on thousands or even hundreds of thousands of vertices, as opposed to the previous limit of about 100 vertices. The new system supports arbitrary edge/vertex colors, labels, and shapes; enhancing its ability to visualize graphical content and structure. Finally, we have provided powerful new algorithms for set partitions, Polya's theorem, and graphical enumeration.

- *A Model for Analyzing Heuristic Search* – In [1], we introduce a new framework for analyzing the performance of heuristics for black-box optimization. Black-box heuristics such as gradient descent, simulated annealing, and tabu search use local search techniques to provide quick-and-dirty solutions for computationally hard problems. We analyze the performance of several local search and backtracking heuristics on different classes of local neighborhoods. Our analysis, based on random graphs and combinatorial dominance, illustrates the tradeoff between search time and solution quality, and provides insight in identifying the proper heuristic to use for a given problem.

- *Discropt: A System for Time-Sensitive Combinatorial Optimization* – We are developing the first general system for combinatorial optimization based on heuristic search which adjusts the search procedure in response to the time alloted to produce a solution. Our system, *Discropt*, contains implementations for such diverse combinatorial optimization problems as max-cut, vertex cover, traveling salesman, maximum satisfiability, and shortest common superstring. By automatically identifying the right flavors of greedy heuristics, simulated annealing, and gradient descent search for the given problem, instance, and allotted time, we get good performance without tuning parameters. See http:/www.cs.sunysb.edu/~discropt for more details.

3 Graph and Combinatorial Algorithms

We have also worked on a variety of other theoretical and applied problems in graph, string, and geometric algorithms:

- *Sheet Metal Assembly Problems* – We [5] have studied the algorithmic complexity of certain geometric problems which arise in assembling structures from sheet metal and wires. In [5], we solved a long-standing open problem on the hardness of constructing an assembly sequence for a sheet structure from prescribed set of folds.

- *Backtranslation Problems in Biology* – Genes are DNA sequences which code for proteins. Due to redundancy in the triplet code, any protein of given length can be coded for in an exponential number of distinct ways. In [2], we have established that the genomes of bacteria to exploit this redundancy by designing genes whose intermediate RNA either maximizes or minimizes amount of structure. This idea of designing biological structures has some relevance to nanotechnology, and our techniques of using dynamic programming and backtracking to construct an optimal structure are of independent algorithmic interest.

  In [8], we exploit the triplet code in a different way, demonstrating that it is possible to significantly protect bacteriophages against the restriction-enzyme protection system employed by bacteria. This potentially points a way to help turn bacteriophages into a effective therapeutic, thus combating the growing and important problem of antibiotic resistance.

As to the impact on education human and resources, in this period of ONR funding I have advised four Ph.D students (Barry Cohen, Vinhthuy Phan, Pavel Sumazin, and Vladimir Filklov), who are in various stages of their doctoral work (third through fifth years). Three of these (Barry Cohen, Vinhthuy Phan, and Pavel Sumazin) are American citizens or permanent residents.

Among other accomplishments, I continued to serve as Associate Editor for the *ACM Journal of Experimental Algorithmics.* I also received the 2001 IEEE Computer Science and Engineering Undergraduate Teaching Award.

A list of recent publications by the PI acknowledging ONR grant support follows. These papers are available on-line at http://www.cs.sunysb.edu/~skiena/papers.html.

References

[1] V. Phan, P. Sumazin, and S. Skiena. A Model for Analyzing Black-Box Optimization, Submitted to *Thirteenth ACM-SIAM Symp. on Discrete Algorithms* (SODA 2002), January 6-8, 2002.

[2] B. Cohen and S. Skiena, Evidence of Selection on Stability among mRNA Sequences, Submitted to *Nucleic Acids Research.*

[3] S. Skiena and S. Pemmaraju, *Computational Discrete Mathematics: Combinatorics and Graph Theory with Mathematica*, Cambridge University Press, 2002

[4] S. Skiena, *The Algorithm Design Manual*, Telos/Springer-Verlag, New York, 1998.

[5] E. Arkin, M. Bender, E. Demaine, M. Demaine, J. Mitchell, S. Sethia, and S. Skiena, When Can You Fold a Map? *Proc. 7th Workshop on Algorithms and Data Structures (WADS)*, Providence, RI, USA August 8-10, 2001. Springer-Verlag LNCS 2125, 401-413. Submitted to *Computational Geometry: Theory and Applications.*

[6] M. Bender, G. Pemmasani, S. Skiena and P. Sumazin, Least Common Ancestors on Directed Acyclic Graphs, *ACM-SIAM Symp. on Discrete Algorithms* (SODA 2001), January 7-9, 2001, pp. 845-854.

[7] V. Filkov, S. Skiena, and J. Zhi, Analysis Techniques for Microarray Time-Series Data *Fifth International Conf. on Computational Molecular Biology (RECOMB 01)* Montreal, Quebec, April 21-24, 2001.

[8] S. Skiena, Designing Better Phages, To appear in *Bioinformatics.* Also *Ninth International Conf. on Intelligent Systems for Molecular Biology (ISMB 2001)*, Copenhagen, Denmark, July 21-25, 2001.

[9] V. Phan and S. Skiena. Dealing with Errors in Interactive Sequencing by Hybridization To appear in *Bioinformatics* special issue of *Third Georgia Tech-Emory International Conference on Bioinformatics*, Atlanta, Georgia, November 15-18, 2001.

[10] S. Skiena, *Calculated Bets: Computers, Gambling and Mathematical Modeling to Win*, Cambridge University Press, New York, 2001.

[11] S. Skiena, Combinatorial Dominance Guarantees for Heuristic Algorithms, manuscript in preparation.