D I M A C S

Center for Discrete Mathematics and Theoretical Computer Science

DIMACS SPECIAL YEAR ON LARGE SCALE DISCRETE OPTIMIZATION: 1998 ANNUAL REPORT

DIMACS Center • Rutgers University • 96 Frelinghuysen Road • Piscataway, NJ 08854-8018 • USA
TEL: 732-445-5928 • FAX: 732-445-5932 • EMAIL: center@dimacs.rutgers.edu
WWW: http://dimacs.rutgers.edu/ • TELNET: telnet info.rutgers.edu 90

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs - Research, Bell Labs, NEC Research, and Telcordia Technologies (formerly Bellcore)

Overview:

The Special Year on Large Scale Discrete Optimization began in September 1998. Some pre-special year activities were held during 1997-98. The year builds on the 1992-93 Special Year on Combinatorial Optimization.

Organizers:

- Endre Boros, Rutgers University (local chair)
- William Pulleyblank, IBM
- Michael Trick, Carnegie Mellon University (general chair)
- Bob Vanderbei, Princeton University
- Vijay Vazirani, Georgia Tech

Sponsors and Collaborating Institutions:

The National Science Foundation
The New Jersey Commission on Science and Technology
AT&T Labs - Research
Bell Labs
Bellcore
Carnegie Mellon University
Georgia Institute of Technology
IBM Research
Institute for Advanced Study, Princeton
NEC Research Institute
Rutgers University
Princeton University

Theme of Program:

This special year is motivated by developments in the past ten years. There have been three simultaneous advances: at the algorithmic level, with new and interesting algorithms for solving large scale discrete problems both exactly and approximately; at the implementation level, providing codes and systems for solving such problems; and at the practical level, with interest in the fast solution of real world problems. Examples of such applications include airline crew scheduling, chemical process design, and telecommunication network design. Our goal in this year is to unite groups working at these three levels. The confluence of exciting recent work on approximation and other algorithms and greatly improved software/hardware for optimization makes large scale discrete optimization a practical, useful approach to solving problems of interest in applications.

The special year organizers have provided forums to increase research contacts between groups that do not frequently come in contact - for example practitioners and academics, operations researchers and computer science researchers, theoretically oriented researchers (in areas such as interior point IP methods, combinatorial geometry approaches, approximation methods, and group formulations) and computationally oriented researchers, including those interested in AI techniques such as constraint programming.

Workshops and Special Events, Feb. 1998 - Nov. 1998

- **Joint DIMACS-CMU-Georgia Tech. Workshop on Large Scale Discrete Optimization**
  - Dates: May 27-29, 1998
  - Location: Rutgers University
  - Attendance: 113
  - Organizers: R. Ravi, Carnegie Mellon University, and Vijay Vazirani, Georgia Tech

Very large problems in a variety of areas ranging from telecommunications to airline scheduling map to questions in discrete optimization. With a growing number of applications seeking to solve very large discrete optimization problems, impressive advances in the design of algorithms for solving such problems exactly and approximately, and the development of large computer systems that can support such activity, there is tremendous interest in this topic in industry and academia. The aim of this workshop then was to bring practitioners and theoreticians with insight into specific areas where large scale discrete optimization methods are used, to talk about central issues as well as outstanding open questions in their respective areas. Examples of such areas include telecommunication network design and routing, airline crew and flight scheduling, scheduling and production. The focus was to identify problems for which fundamental insight into their structure and solution will translate into practical gains in their application areas. Keeping such high impact problems in mind, a mix of leaders from different areas of application and development were invited to lecture in the workshop.

The talks were roughly organized under two broad themes, "Applications" and "Tools." While the Applications talks introduced several traditional areas where discrete optimization is applied in practice (e.g., airline industry, production planning, network survivability) as well as advances in several methodological areas of development in optimization (e.g., flows, Markov chain methods), the Tools talks focused on design and functionalities in popularly available tools (e.g., CPLEX, LEDA) to solve generic optimization problems.

In the Applications area, George Nemhauser and Ellis Johnson surveyed their long-standing work on applying optimization in airline integrated planning and operations and Jitendra Malik introduced graph partitioning and matching problems arising in computer vision, and current methods used to tackle them. Methodological advances in optimization Algorithms were featured when Prabhakar Raghavan described an exciting microeconomic view of customer segmentation and collaborative filtering and the problems arising therein; Rajeev Motwani described recent work on high-dimensional similarity search in large databases; and Eva Tardos surveyed methods for regulating flows in communication networks. Laurence Wolsey described large production planning problems focusing on modeling and software issues; Gary Miller discussed the use of graph embedding techniques in scientific computation; and Dana Randall surveyed Markov chain methods in physics. Finally, Craig Tovey, Georgia Tech., described his intriguing work on self-organizing social structures in fish groups and his attempts at modeling and predicting their behavior; Michael Trick described his adventures in sports scheduling; and Milena Mihail surveyed optimization issues in network survivability and network evolution.

In talks devoted to Tools, Kurt Mehlhorn, a co-developer of the popular LEDA software package, talked about Secondary Memory Algorithms for Networks and Geometry, and capabilities to handle such methods in LEDA; Andrew Goldberg reported on insights on implementing maximum flow methods and describing new techniques and algorithms that arose from such implementations; John Hooker talked about the connections between logical and linear programming and proposed a hybrid stronger than either of the two. Also, Bob Bixby described recent developments in the new version 6 of the popular CPLEX package; Martin Savelsbergh described the central design issues in PARINO, a PARallel INteger Optimizer developed in Georgia Tech; and Sebastian Ceria described the idea of lift-and-project cuts, an efficient solution method for mixed integer programs.

- **DIMACS Workshop on Constraint Programming and Large Scale Discrete Optimization**
  - Dates: September 14-17, 1998
  - Location: Rutgers University
  - Attendance: 75
  - Organizers: Richard J. Wallace and Eugene C. Freuder, University of New Hampshire

  This workshop was concerned with all aspects of constraint programming related to large scale discrete optimization, including contributions from artificial intelligence, discrete mathematics, operations research and programming languages. The intention of the organizers was to bring together people interested in constraint programming approaches to large scale discrete optimization (LSO) problems, coming from the areas just mentioned and from both academia and industry, so that new results could be presented, problems posed, and the current state of the field clarified. To further these ends, the workshop included tutorial talks from well-known figures in key areas related to the constraint programming and LSO, who reviewed work within their fields of expertise.

  There were tutorial talks on "Constraint programming languages for combinatorial optimization" by Pascal Van Hentenryck and "Integration of artificial intelligence and operations research techniques" by Ken McAloon, Carol Tretkoff and Carla Gomes.

  There was a session on local search methods for LSO and invited talks on optimization methods in NASA's autonomous Deep Space probe by Brian Williams and on search methods for quantum computing by Tad Hogg. Some sessions were devoted to an important European project, CHIC2, that has pioneered the development of "hybrid algorithms" that combine procedures derived from constraint satisfaction algorithms and from operations research (integer and linear programming). Papers were given by Tibor Kokeny and Vassilis Liatsos on applications for two different large-scale industrial problems, involving scheduling and, in one case, inventory management, and by Carmen Gervet on guidelines, or "methodologies", that were spawned by the CHIC2 project, for developing new hybrid algorithms for industrial problems. There was an invited talk by David Johnson on "Average case behavior of bin packing algorithms," which covered an important case of approximation methods for NP-hard problems. A session presented AI, or constraint satisfaction, techniques for complex applications ("course of action analysis" in a military context), as well as new techniques for solving constraint satisfaction problems. A session organized by Thomas Schiex was a review of the "valued constraint satisfaction" framework that can represent various types of optimization problem as overconstrained constraint satisfaction problems. There was an "industrial session" with presentations by James Crawford of I2 Technologies, Mark Boddy of Honeywell, and Helmut Simonis of Cosytec on LSO problems that these companies have been involved with and some of the applications that have been developed for these problems. Finally, a half-day was devoted to presentations on applications based on constraint programming for scheduling problems, including sports scheduling, crew scheduling for train crews, bus drivers, and nurses, satellite scheduling and scheduling of broadcast commercials.

- **DIMACS/IAS Workshop on Randomized and Derandomized Algorithms for Discrete Structures**
  - Dates: November 2-4, 1998
  - Location: Institute for Advanced Study
  - Attendance: 41
  - Organizers: Bruce Reed, University Pierre et Marie Curie; Noga Alon, Tel Aviv University and IAS, Alan Frieze, CMU

  The workshop was devoted to two techniques which have wide application in theoretical discrete optimization: the semi-random method and Szemeredi's Regularity Lemma. The semi-random method is a technique for generating combinatorial objects via repeated applications of the probabilistic method; a typical example of applying the method is the so-called Rődl Nibble. The Regularity Lemma states roughly that the vertices of any large enough graph can be split up into pieces so that the bipartite graph spanned by the edges between each pair of pieces behaves essentially like a random bipartite graph. Both of these techniques apply mainly to large structures and have numerous recent algorithmic applications.

  The workshop began with a historical survey of the uses of the Regularity Lemma by Miki Simonovits. Gabor Sarkozy showed how together with the "blow-up" lemma it was being used to solve several graph packing problems. Raphy Yuster discussed decomposing Hypergraphs with Hypertree. Andrzej Rucinski discussed some packing problems and Yoshiharu Kohayakawa discussed some Ramsey problems in random graphs. Jeong Han Kim showed how the semi-random method could be used to find small complete arcs in projective planes and Vijay Vazirani talked about the approximation of #P-complete problems. There was a talk by Andrew Thomason on generalized graph colouring. Penny Haxell discussed the use of the algorithmic version of the Regularity Lemma in finding integer and fractional packings in dense graphs and Lubos Thoma discussed dependent arcs in acyclic orientations. Joel Spencer analyzed an on-line algorithm for packing random rectangles. Sarmad Abbasi discussed some questions related to spanning subgraphs of dense graphs. Ravi Kannan described randomized algorithms for finding Low Rank Approximations to matrices and applications thereof. Van Vu showed how the semi-random method could be used to bound the list-chromatic number of locally sparse graphs. Michael Krivelevich described an efficient algorithm for testing bipartiteness in dense graphs. Also, Janos Komlos discussed tiling Turan theorems, Hanno Lefmann talked about algorithms for finding independent sets in graphs and hypergraphs and Andrzej Czygrinow discussed the algorithmic aspects of the hypergraph regularity lemma.

Workshops and Miniworkshops Planned for the Rest of 1998-99

- **DIMACS Workshop on Semidefinite Programming and Large Scale Discrete Optimization**
  - Dates: January 7-9, 1999
  - Location: Princeton University
  - Organizers: Robert Vanderbei, Princeton University and Yinyu Ye, University of Iowa
- **DIMACS Workshop on Large Scale Discrete Optimization in Logistics**
  - Dates: February 8-10, 1999
  - Location: Rutgers University

- Organizers: George Nemhauser and Martin Savelsbergh, Georgia Tech

- **DIMACS Workshop on Large Scale Discrete Optimization in Robotics and Vision**
  - Dates: March 21-23, 1999
  - Location: Rutgers University
  - Organizers: Jitendra Malik, University of California - Berkeley; Jean Ponce, University of Illinois - Urbana Champaign

- **DIMACS Workshop on Large Scale Discrete Optimization, the State of the Art**
  - Dates: Tentative July 1999
  - Location: Rutgers University (tentative)
  - Organizers: Peter Hammer, Rutgers University

- **DIMACS - IBM Conference in Honor of Ralph Gomory's 70th Birthday**
  - Dates: August 2-4, 1999
  - Location: IBM Watson Labs
  - Organizer: Bill Pulleyblank

- **Computational Challenge Preworkshop**
  - Dates: Tentative TBA
  - Location: TBA
  - Organizers: David Johnson, AT&T Labs - Research

Special Year Tutorial Series
Tutorials Feb. 1998 - Nov. 1998

Models and Algorithms for Railway Crew Management

Paolo Toth, DEIS, University of Bologna, Italy
October 30, 1998

Crew Management is concerned with building the work schedules (rosters) of crews needed to cover a given set of timetabled trips. Toth gave a general definition of the problem and showed possible operational constraints and objective functions arising in real-world applications. In practice, the overall problem is decomposed into two subproblems, called Crew Scheduling (CSP) and Crew Rostering (CRP). In CSP the short-term schedule of the crews is considered, and a convenient set of duties covering all the trips is constructed. In CRP the selected duties are sequenced to obtain the final rosters. Several mathematical programming and graph-theory formulations, and different exact and heuristic algorithms proposed in the literature for the solution of the two subproblems, were presented. Advantages and disadvantages of the proposed formulations and algorithms were discussed. Two main solution approaches were illustrated in more detail. CSP is solved by generating a very large number of potential duties, and by solving an associated Set Covering Problem to select a minimum-cost set of duties covering all the trips. CRP is solved by using a Lagrangian relaxation of a graph-theoretical formulation of the problem. Extensive computational results on real-world instances arising in the Italian railway company, Ferrovie dello Stato SpA, were presented.

Tutorials Planned for the Rest of 1998-99

Progress in Linear Programming Based Branch-and-Bound: Modeling and Methodology

George L. Nemhauser, Georgia Institute of Technology
December 9 and 10, 1998

In the last decade, the use of mixed-integer programming (MIP) models has increased dramatically. Fifteen years ago, main frame computers were required to solve problems with a hundred integer variables. Now it is possible to solve problems with thousands of integer variables on a pc and to obtain provably good approximate solutions to problems such as set partitioning with millions of binary variables. These advances have been made possible by developments in modeling, algorithms, software and hardware. In the first of these two expository talks, Professor Nemhauser will focus on applications, effective modeling and preprocessing. The second talk will deal with the methodologies of branch-and-cut and branch-and-price, which are the techniques that make it possible to treat problems with either a very large number of constraints or a very large number of variables.

Metaheuristics

Fred Glover, Jim Kelly, Manuel Laguna,
University of Colorado, Boulder
April 29 and 30, 1999

An Introduction to Metaheuristics will place a primary focus on Tabu Search, from the standpoint of adaptive memory methods, and on Scatter Search and Path Relinking, from the standpoint of evolutionary methods. Comparisons and contrasts with Genetic Algorithms and other approaches will be included. An Introduction to OptQuest will highlight demonstrations of real world applications, especially those applications that involve significant elements of uncertainty and nonlinearity, including discreteness.

Optimal Connected Partitions of Graphs

Bruno Simeone, La Sapienza University, Italy
Spring 1999

This will be a series of five lectures on the topics: Path equipartition: trimming the shortest path network; greedy algorithms, semimodular lattices, and supermodular functions; shifting algorithms for tree partitioning; how to make a polynomial movie out of pseudo-polynomially many frames: continuous shifting algorithms; and connected k-partitions in k-connected graphs: a topological approach.

Discrete Optimization in Telecommunications and VLSI Design

Michel Minoux, Universitè Paris 6, France
May 1999

Eigenvalue and Semidefinite Techniques in Discrete Optimization

Henry Wolkowicz, University of Waterloo, Canada

May 1999

Local Search Algorithms in Discrete Optimization
Toshihide Ibaraki, Kyoto University, Japan
May 1999

Branch and Bound Algorithms on Distributed Processors
Ewald Speckenmeyer, University of Kőln, Germany
June 1999

Special Year Graduate Course

Professor Vijay Vazirani, visiting DIMACS from Georgia Tech, has been giving a special year graduate course on Approximation Algorithms. The course has been attended by about 50 people from Rutgers, Princeton, AT&T Labs, Bell Labs, Bellcore, and NEC Research. Besides presenting key results in the area of approximation algorithms, the course has been giving a feeling for the important open problems. Topics include combinatorial algorithms, primal-dual schema based algorithms, multicommodity flows, semidefinite programming, estimating network reliability, PTAS for scheduling and Euclidean TSP, FPAS for knapsack, and hardness of approximation. The course is based on a preliminary draft of a book on Approximation Algorithms written by the instructor.

Selected Research Results

A Self Organizing Bin Packing Heuristic Suppose there is a random process that generates items of various integer sizes which we must pack into fixed size bins, and we wish to minimize the total wasted space in the non-empty bins of the packing. If N is the number of items generated, it is known that the total waste in the optimal packing must grow either as $\Omega(N)$, $\Omega(\sqrt{N})$, or $O(1)$, depending on the probability distribution. Standard bin packing algorithms such as Best Fit do not typically perform optimally. This is true even for offline algorithms such as Best Fit Decreasing. Although in general it is NP-hard to determine the optimal growth rate given a distribution, Janos Csirik, David Johnson (PM, ATTL), Claire Kenyon (V), Peter Shor (PM, ATTL), and Richard Weber showed that this problem can be solved in pseudo-polynomial time (polynomial in the bin size) using linear programming. They provided a remarkably simple online algorithm that is conjectured to perform "optimally" for ALL distributions, based on extensive empirical tests, which demonstrates its apparently "self-organizing" method of operation.

A Data Structure for Maintaining a Partition in VLSI Design Gruia Calinescu (PD) worked on data structures for maintaining a partition. In VLSI design, it would be useful to solve efficiently the following problem: update a partition of the ground set $\{1, 2, ..., n\}$ after adding or deleting a bipartition (a bipartition is a partition of the ground set into two parts). Using the credit technique, Calinescu proved that a straightforward data structure obtains the best possible amortized time bound.

The Effect of Slack on Competitiveness for Admission Control Michael Goldwasser (PD) worked on the effect of slack on competitiveness for admission control. Imagine that a company rents use of a single resource to customers, charging a fixed amount per minute of use. Job requests arrive in an online manner, each request specifying the length of time for which the customer will use the resource as well as a deadline by which the job should be completed. The company is not required to complete each job. However, it will only be paid for jobs which are completed on or before their deadline and it may not interrupt a job once started. Therefore, a scheduling algorithm provides admission control as it decides which jobs to run, as well as when to run them, and the goal is to maximize the successful use of the resource and thus the profit for the company. Goldwasser analyzed the performance of an algorithm using competitive analysis, comparing the schedule produced by an algorithm to the optimal schedule produced by an algorithm with full knowledge of future jobs. This model arises in many applications of real-time systems, with the most typical motivations involving communication-based applications. In this setting, the resource represents a channel of communication and a job (or call) represents a request for delivery of text, sound, or video through the channel. The *slack* of a job is equal to the amount of time between the job's arrival and the last possible time at which it could be started while still meeting its deadline. Goldwasser focused on the effect of slack on the competitiveness of scheduling algorithms in this model. His main result was that if each job is willing to wait an amount of time at least proportional to the job's length, the competitiveness for online scheduling algorithms can be significantly improved. When arbitrary slacks are allowed previous work has shown that no bounded competitiveness is possible if arbitrary job lengths are allowed. However, Goldwasser showed for this model of slack, that a simple deterministic algorithm is $(2 + \frac{1}{\omega})$-competitive, thereby providing a constant competitive ratio which does not depend on the range of job lengths. He also provided a matching lower bound showing that this is the best possible deterministic result, even when all jobs have one of three distinct lengths. In the special case where all jobs are required to have the same length (e.g., packets in an ATM network), the natural greedy algorithm is $(1 + \frac{1}{\omega+1})$-competitive, and again that is the best possible deterministic result. This result generalizes a previous bound for the case of arbitrary slack. In the case where jobs have one of two distinct lengths, Goldwasser gave tight upper and lower bounds which improve slightly on the $(2+\frac{1}{\omega})$ bound for arbitrary lengths.

A Polyhedral Analysis of the p-cycle Polytope Ozgur Ozluk (GS, U. North Carolina) worked on a polyhedral analysis of the p–cycle polytope which is the convex hull of the incidence vectors of all the p–cycles (simple cycles consisting of p arcs) of the complete directed graph. His main motivation was to use the results of this study as a first step in a polyhedral approach to the Traveling Circus Problem (TCP) which combines the location aspects of the p–median problem with the routing aspects of the Traveling Salesman Problem (TSP). The TCP has many potential applications. One such application is the design of the distributed computer networks in which case cities on the tour would represent concentrators connected by a tour (a ring network). A very interesting feature of the TCP is that it can be used to solve many extensions and generalizations of the TSP such as the Time-Dependent TSP and Generalized TSP. The p–cycle polytope should not be expected to have a simple characterization, since it contains the Asymmetric TSP polytope on p nodes as a face. It is also clear that the problem of finding a minimum length p–cycle is *NP*-hard. Ozluk developed an integer programming formulation of the p–cycle polytope, and derived some useful properties of its set of equations. He then determined the dimension of the p–cycle polytope, and proved two lifting results.