10261 Abstracts Collection
Algorithm Engineering
— Dagstuhl Seminar —

Giuseppe F. Italiano[1], David S. Johnson[2], Petra Mutzel[3] and Peter Sanders[4]

[1] Università di Roma "Tor Vergata", IT
italiano@disp.uniroma.it
[2] AT&T Research - Florham Park, US
dsj@research.att.com
[3] Technische Universität Dortmund, DE
petra.mutzel@cs.tu-dortmund.de
[4] Universität Karlsruhe, DE
sanders@ira.uka.de

Abstract. From June 27 to July 2, the Dagstuhl Seminar 10261 "Algorithm Engineering " was held in Schloss Dagstuhl – Leibniz Center for Informatics. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Links to extended abstracts or full papers are provided, if available.

Keywords. Experimental algorithmics, Game theory, Parallel and distributed algorithms, Multi-core

10261 Executive Summary – Algorithm Engineering

Algorithm engineering (AE) consists of the design, theoretical analysis, implementation, and experimental evaluation of algorithms, with the aim of bridging the gap between theory and practice in the area of algorithms. In the last decade, this approach to algorithmic research has gained increasing attention.

The aim of this seminar was to bring together researchers with different backgrounds, e.g., from combinatorial optimization, algorithmic theory, and algorithm engineering, in order to strengthen and foster collaborations in the area of algorithm engineering and to identify key research directions for the future.

Keywords: Experimental algorithmics, Game theory, Parallel and distributed algorithms, Multi-core

Joint work of: Giuseppe Italiano; David Johnson; Petra Mutzel; Peter Sanders

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2010/2796

Multicore and Manycore for Algorithm Engineers

David A. Bader (Georgia Institute of Technology, US)

This talk presents the opportunities and challenges for an algorithm engineer in the era of multicore and manycore computing.

Full Paper:
http://www.cc.gatech.edu/~bader

From A to B with Bus and TGV

Hannah Bast (Universität Freiburg, DE)

Joint work with Erik Carlsson, Robert Geisberger, Chris Harrelson, Veselin Raychev and Fabien Viger.

We present the new routing algorithm behind Google Transit and some of the context in which it was developed at Google in Zurich. This new algorithm is the first to achieve fast query times (on the order of milliseconds) on very large and realistically modelled public transportation networks. It is based on two key observations: (1) many shortest paths share the same transfer pattern, i.e., the sequence of stations where a change of vehicle occurs; (2) direct connections without change of vehicle can be looked up quickly. We precompute the respective data; in practice, this can be done in time linear in the network size, at the expense of a small fraction of non-optimal results.

We also highlight that more than this algorithm was needed to turn this into a product used by millions of people worldwide.

An extended abstract of our algorithmic results has been accepted to ESA 2010.

Joint work of: Bast, Hannah; Eigenwillig, Arno

Pricing Lotteries

Patrick Briest (Universität Paderborn, DE)

Randomized mechanisms, which map a set of bids to a probability distribution over outcomes rather than a single outcome, are an important but ill-understood area of computational mechanism design.

We investigate the role of randomized outcomes ("lotteries") in the context of a fundamental and archetypical multi-parameter mechanism design problem: selling heterogeneous items to unit-demand bidders. To what extent can a seller improve her revenue by pricing lotteries rather than items, and does this modification of the problem affect its computational tractability? We show that the answers to these questions hinge on the exact model of consumer behavior we deploy and present several tight bounds on the increase in revenue obtainable via randomization and the computational complexity of revenue maximization in these different models. This is joint work with Shuchi Chawla, Bobby Kleinberg, and Matt Weinberg.

Multi-Core: Trends, Processors and Tools

Roman Dementiev (Intel GmbH - Feldkirchen, DE)

We discuss trends in processor performance including: trade-offs between power consumption and core frequency, development of single-threaded and multi-threaded performance, Intel research compute platforms. In the main part of the talk we dive into the current architecture codenamed Nehalem/Westmere: instruction level parallelism, data level parallelism, SMT (Intel(r) HyperThreading), cache hierarchy, non-uniform memory access (NUMA, Intel(r) QuickPath Interconnect), Intel(r) Turbo Boost Technology. We overview the set of tools that Intel offers to simplify software development on modern CPU architectures (performance profilers and analysers, compilers, libraries, correctness checking tools).

Keywords: Processor architecture, multithreading, caches, memory hierarchy, threading, software tools

Highway and VC Dimensions: From Practice to Theory and Back

Andrew V. Goldberg (Microsoft Research - Mountain View, US)

Highway dimension and shortest path covers (SPCs) [Abraham et al., SODA 2010] were introduced to give a theoretical explanation to the excellent practical performance of recent shortest path heuristics. We discuss our on-going work on better theoretical understanding of these new concepts, including the relationship to VC-dimension. This suggests an interesting relationship to computational geometry and learning theory, and leads to better theoretical bounds for building near-optimal SPCs.

These results, in turn, suggest potentially practical approaches to computing good SPCs.

Joint work with Ittai Abraham, Daniel Delling, Amos Fiat, and Renato Werneck.

Keywords: Shortest paths, algorithm analysis, algorithm engineering, highway dimension

Joint work of: Goldberg, Andrew V.; Abraham, Ittai; Delling, Daniel; Fiat, Amos; Werneck, Renato

MIP Domination

Zonghao Gu (Gurobi Optimization - Alpharetta, US)

This talk considers the identification and use of domination information in MIP models. This information can be used to simplify a model in presolve, and to avoid unnecessary search in the branch-and-bound algorithm. We discuss different MIP domination techniques, such as variable and node domination, disjoint subtrees, and symmetry breaking, and present computational results to show that these techniques can significantly improve the performance of a MIP solver.

Keywords: MIP

Joint work of: Gu, Zonghao; Bixby, Robert; Rothberg, Ed

Resilient Algorithms and Data Structures

Giuseppe F. Italiano (Università di Roma "Tor Vergata", IT)

Large and inexpensive memory devices may suffer from faults, where some bits may arbitrarily flip and corrupt the values of the affected memory cells. The appearance of such faults may seriously compromise the correctness and performance of computations. In recent years, many algorithms for computing in the presence of memory faults have been introduced in the literature: in particular, an algorithm or a data structure is called resilient if it is able to work correctly on the set of uncorrupted values. We contribute carefully engineered implementations of recent resilient algorithms and data structures.

Experimentalists Anonymous: A Cautionary Tale

David S. Johnson (AT&T Research - Florham Park, US)

I discuss research-in-progress on an optimization problem arising in a scheme of [Gu et al., 2008] for measuring packet loss on paths between routers in a network. Given a directed graph $G = (V, A)$ with weights on the arcs, together with subsets C (customers) and F (potential facility locations) of V, find a subset F' of F such that, for every c in C, either c is in F' or there exist two vertices f, f' in F' such that no shortest path from c to f shares any vertex other than c with any shortest path from c to f' (Set-Disjoint Problem) or there exist shortest paths from c to f and c to f' which share no vertices other than c (Path-Disjoint Problem). The Set-Disjoint version applies when routing is done by the OSPF protocol with path-splitting; the Path-Disjoint version applies when we are able to specify the precise routes. In addition to the monitoring application, there

is also a potential application of these problems to distribution of time-critical streaming content.

Assuming a widely-believed conjecture, no polynomial-time approximation algorithms with good worst-case guarantees exist for these problems.

However, we have devised heuristics that found optimal solutions on over half of our synthetic and real-world instances, and averaged within a few percent of optimal overall. They also confirmed that the proposed monitoring scheme would provide a major savings in hardware over the naive scheme.

In this talk, I discuss the history of the work, with its ever-increasing list of new experiments that needed to be done. These included (1) pushing the limits of MIP-based optimization, (2) characterizing the differences between our synthetic instances, built using a standard tool for generating synthetic Wide Area Networks (WAN's) plus algorithms for OSPF weight optimization, and our real-world instances, based on actual ISP topologies and OSPF weights, and (3) determining the advantage of using Path-Disjoint solutions instead of Set-Disjoint solutions. We also recognized, late in the game, that the real bottleneck for obtaining solutions may well be in the instance preprocessing phase, where the network problem is converted into an instance of the "Cover-by-Pairs" problem. In our initial experiments we underestimated the importance of this step, and so performed the conversions inefficiently and did not keep track of the time involved.

We have now come up with algorithms that do the conversions in time $O(n^3)$ for both the Set-Disjoint or Path-Disjoint cases, but these remain to be implemented and tested.

(Joint work with Lee Breslau, Ilias Diakonikolas, Nick Duffield, Yu Gu, Mohammadtaghi Hajiaghayi, Howard Karloff, Mauricio Resende, and Subhabrata Sen)

Keywords: Set cover, approximation algorithms, content distribution

Solving Two-Stage Stochastic Steiner Tree Problems by Two-Stage Branch-and-Cut

Ivana Ljubic (Universität Wien, AT)

We consider the Steiner tree problem under a 2-stage stochastic model with recourse and finitely many scenarios (SSTP). Thereby, edges are purchased in the first stage when only probabilistic information on the set of terminals and the future edge costs is known. In the second stage, one of the given scenarios is realized and additional edges are purchased to interconnect the set of (now known) terminals. The goal is to choose an edge set to be purchased in the first stage while minimizing the overall expected cost of the solution.

We provide a new semi-directed cut-set based integer programming formulation that is stronger than the previously known undirected model. To solve the formulation to provable optimality, we suggest a two-stage branch-and-cut framework, facilitating (integer) L-shaped cuts. The framework itself is also applicable to a range of other stochastic problems.

As SSTP has yet been investigated only from the theoretical point of view, we also present the first computational study for SSTP, showcasing the applicability of our approach and its benefits over solving the extensive form of the deterministic equivalent directly.

Keywords: Stochastic Steiner tree, stochastic integer programming, branch-and-cut, Benders decomposition

Joint work of: Bomze, Immanuel; Chimani, Markus; Jünger, Michael; Ljubic, Ivana; Mutzel, Petra; Zey, Bernd

Experimental Procedures for Tuning Algorithms

Catherine McGeoch (Amherst College, US)

This talk covers two topics: (1) building a catalog of algorithm tuning techniques, organized by algorithm paradigm, by data structure access pattern, and by architecture and high level language considerations; and (2) applying techniques of statistical design of experiments to problems in algorithm tuning. A case study using an algorithm for all pairs shortest paths is used to illustrate these ideas.

Keywords: Algorithm engineering, experimental analysis of algorithms, experimental algorithmics

On Dynamic Graph Partitioning and Graph Clustering using Diffusion

Henning Meyerhenke (Universität Paderborn, DE)

Load balancing is an important requirement for the efficient execution of parallel numerical simulations. In particular when the simulation domain changes over time, the mapping of computational tasks to processors needs to be modified accordingly. State-of-the-art libraries for this problem are based on graph repartitioning. They have a number of drawbacks, including the optimized metric and the difficulty of parallelizing the popular repartitioning heuristic Kernighan-Lin (KL).

Here we further explore the very promising diffusion-based graph partitioning algorithm DIBAP (Meyerhenke et al., JPDC 69(9):750–761, 2009) by adapting DIBAP to the related problem of load balancing. The presented experiments with graph sequences that imitate adaptive numerical simulations demonstrate the applicability and high quality of DIBAP for load balancing by repartitioning. Compared to the faster state-of-the-art repartitioners PARMETIS and parallel JOSTLE, DIBAP's solutions have partitions with significantly fewer external edges and boundary nodes and the resulting average migration volume in the important maximum norm is also the best in most cases.

We also prove that one of DIBAP's key components optimizes a relaxed version of the minimum edge cut problem. Moreover, we hint at a distributed algorithm based on ideas used in DIBAP for clustering a virtual P2P supercomputer.

Keywords: Dynamic graph partitioning/clustering, disturbed diffusion, load balancing, relaxed cut optimization

Joint work of: Meyerhenke, Henning; Gehweiler, Joachim

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2010/2798

Unstructured overlay networks with guarantees

Marina Papatriantafilou (Chalmers UT - Göteborg, SE)

We propose a framework to facilitate reasoning about unstructured overlays; the framework combines widely used randomized techniques such as random walks with elementary electric circuit theory and overlay network techniques. We discuss examples of the framework's usage and trace its possible use in other settings such as the construction phase of these networks.

We present a simple distributed approximation algorithm that enables peers with preferences to collaborate in an overlay network, while at the same time achieving a guaranteed level of quality for their requested connections (guaranteed termination and satisfaction level for matching with preferences).

This method forms also a step into enabling users to collaborate despite having different goals and interests and opens the road to further study, e.g. from a game-theoretic point of view (if e.g. strategic decisions and coalitions are included in the behavior of the users).

It is possible to combine the two approaches, framework and algorithm, to gain insights on the mechanisms of topology creation when network nodes (or their users) have autonomy to decide on their own connections based on individual preferences.

Keywords: Overlay networks, matching with preferences, approximation algorithms

Joint work of: Georgiadis, G; Papatriantafilou, M.

See also: Georgiadis, G., and Papatriantafilou M., "A Least Resistance Path in Reasoning about Unstructured Overlay Networks", In Proceedings of the 15th International Euro-Par conference on Parallel Processing, Lecture Notes in Computer Science, Volume 5704/2009, 797-808, Springer Verlag. Georgiadis, G., and Papatriantafilou M., "Overlays with preferences: Approximation algorithms for matching with preference lists", In Proceedings of the 24th IEEE International Parallel and Distributed

An Experimental Comparison of Speed Scaling Algorithms with Deadline Feasibility Constraints

Kirk Pruhs (University of Pittsburgh, US)

We consider the first, and most well studied, speed scaling problem in the algorithmic literature: where the scheduling quality of service measure is a deadline feasibility constraint, and where the power objective is to minimize the total energy used.

Four online algorithms for this problem have been proposed in the algorithmic literature.

Based on the best upper bound that can be proved on the competitive ratio, the ranking of the online algorithms from best to worst is: qOA, OA, AVR, BKP.

As a test case on the effectiveness of competitive analysis to predict the best online algorithm, we report on an experimental "horse race" between these algorithms using instances based on web server traces.

Our main conclusion is that the ranking of our algorithms based on their performance in our experiments is identical to the order predicted by competitive analysis. This ranking holds over a large range of possible power functions, and even if the power objective is temperature.

Keywords: Scheduling, Speed Scaling, Experimental Algorithms, Power Management

Joint work of: Abousamra, Ahmed; Bunde, David; Pruhs, Kirk

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2010/2797

Algorithm Engineering

Peter Sanders (KIT - Karlsruhe Institute of Technology, DE)

I give an attempt at a definition of algorithm engineering with example from the DFG priority program 1307 on the same topic.

Keywords: Algorithms, experiments, methodology

Full Paper:
http://www.springerlink.com/content/5405480m781251r3/

See also: P. Sanders, Algorithm Engineering – An Attempt at a Definition, Efficient Algorithms, Springer, Lecture Notes in Computer Science 5760, 2009, pages 321–340.

A "Survey" on Stochastic Programming

Ruediger Schultz (Universität Duisburg-Essen, DE)

We review models and algorithms for risk management in stochastic programming. Risk aversion in the objective can be accomplished by mean-risk models involving different risk measures. On finite probability spaces, the equivalent optimization problems are large-scale mixed-integer programs. Depending on the risk measure, these models obey linearity properties and show block structures which are amenable to decomposition algorithms. Risk management in the constraint scan be modeled via stochastic orders, i.e., partial orders for comparing relevant random variables with preselected stochastic benchmarks. Again, with finite probability spaces, we arrive at structured mixed-integer linear programs that give rise to decomposition. Case studies from power production and trading serve to illustrate the superiority of the presented decomposition methods over general-purpose mixed-integer linear programming solvers.

Keywords: Stochastic programming, risk measures, decomposition methods

Lock-Free Data Structures

Philippas Tsigas (Chalmers UT - Göteborg, SE)

Lock-free data structures offer several advantages over their blocking counterparts, such as being immune to deadlocks and convoying and, more importantly, being highly concurrent.

In the first part an overview of the area is presented. Implementation and performance evaluation issues of lock-free data structures are considered.

In the second part a lock-free methodology for composing highly concurrent linearizable data structures together by unifying their linearization points is presented. This methodology makes it possible to relatively easily introduce atomic lock-free move operations to a wide range of concurrent data structures. Experimental evaluation has shown that the operations originally supported by the data structures keep their performance behavior under our methodology.

A Robust PTAS for Machine Covering and Packing

Jose Verschae (TU Berlin, DE)

In general, combinatorial optimization problems are unstable: slight changes on the instance of a problem can render huge changes in the optimal solution. Thus, a natural question arises: Can we achieve stability if we only maintain approximate solutions? In this talk I will first explain some models that try to formalize these ideas, and then show some results on the parallel machine covering problem. Through examples, I will explain how to get a robust PTAS, i.e., I will show how to construct a solution that is not only (1-epsilon)-approximate, but is also stable. That is, if the instance is changed by adding or removing a job, then we can construct a new near-optimal solution by only slightly modifying the previous one.

This is joint work with Martin Skutella.

Keywords: Approximation algorithms, migration factor, online

Clustering Changing Graphs

Dorothea Wagner (KIT - Karlsruhe Institute of Technology, DE)

The talk gives an overview of dynamic algorithms we developed to compute clusterings of changing graphs. On one hand, we give a dynamic version of clustering based on the minimum cut tree. On the other hand we present experiments for dynamic versions of different modularity clustering algorithms. Besides the running time for the dynamic update and the quality we focus on the smoothness of the clusterings we derive.

Keywords: Graphs, clustering, modularity,minimum cut tree, social networks, dynamic graphs

Joint work of: Wagner, Dorothea; Hartmann, Tanja; Görke, Robert

Full Paper:
http://i11www.ira.uka.de/extra/publications/gmsw-mdcdg-10a.pdf

Full Paper:
http://i11www.ira.uka.de/extra/publications/ghw-dmctc-09.pdf

See also: 1. Robert Görke, Pascal Maillard, Christian Staudt and Dorothea Wagner. Modularity-Driven Clustering of Dynamic Graphs. In: Proceedings of the 9th International Symposium on Experimental Algorithms (SEA'10), volume 6049 of Lecture Notes in Computer Science. Springer, May 2010. 2. Robert Görke, Tanja Hartmann and Dorothea Wagner.Dynamic Graph Clustering Using Minimum-Cut Trees. In: Algorithms and Data Structures, 11th International Workshop, volume 5664 of Lecture Notes in Computer Science. Springer, August 2009.