Algorithm Design and Engineering: 2009 Report

Research activity regarding design and engineering of computer algorithms and computational complexity analysis has been developed at DIS since when the Department has been created in the early Eighties. In the first years the emphasis has been on theoretical aspects such as those related to the notion of approximation preserving reductions among optimization problems and the classification of optimization problems based on their approximability properties. Subsequently, research activities have evolved in various directions according to the evolution of information technology and of application domains. New emerging topics have been addressed such as dynamic graph algorithms, on line algorithms, external memory, and streaming algorithms for massive data sets. Also the emphasis of the approach has changed moving from traditional worst case analysis to experimental performance analysis.

The most relevant recent results include contributions in the following areas:

- Principles of Design and Analysis of Algorithms
- Experimental Algorithmics
- External Memory and Streaming Algorithms for Massive Data Processing
- Incremental Algorithms and Dynamic Data Structures
- Approximation and On-line Algorithms
- Algorithmic Game Theory

The ADE group is currently cooperating with several prestigious research institutions: Max Planck für Informatik (Saarbrücken, Germany), University of Paderborn (Germany), CTI-Patras and University of Athens (Greece), ETH (Zurich, Switzerland), Université de Paris (Dauphine, France), Tel-Aviv University (Israel), AT&T - Research Labs (Florham Park, NJ, USA), Polytechnic University of Catalunya, ICSI-Berkeley (USA), Brown University (Providence, RI, USA), Carnegie-Mellon University (Pittsburgh, PA, USA), Microsoft Research (Mountain View, CA, USA), Yahoo! Research (Barcelona, Spain).

In 2009 the AE group was involved in the following research projects:

- AEOLUS: “Algorithmic principles for building overlay computers”, EU VI framework program, 2005-2010, 170K Euro, 40% in research area.
- SIMBIOSI: “INRIA associated team”, INRIA, France, 2009-2011, 30K Euro, 100% in research area.
• “Decentralized algorithms for mining and maintaining large information networks”, Progetto Ateneo Federato Scienza e Tecnologia, Sapienza University of Rome, 2008-2009, 5K Euro, 100% in research area.

• “New Methodologies for Pervasive and Data-intensive Computing”, Progetto Ateneo Federato Scienza e Tecnologia, Sapienza University of Rome, 2009-2010, 10K Euro, 100% in research area.

Group members

Faculty members: Giorgio Ausiello, Luca Becchetti, Fabrizio D’Amore, Camil Demetrescu, Stefano Leonardi, Alberto Marchetti-Spaccamela, Umberto Nanni, Andrea Vitaletti.

Post-docs: Aris Anagnostopoulos, Vincenzo Bonifaci, Luigi Laura, Andrea Ribichini, Fabiano Sarracco, Piotr Sankowski.

PhD students: Donatella Firmani, Ilaria Bordino.

Principles of Design and Analysis of Algorithms. In [1], we consider several complexity issues arising in metabolic networks; in fact by making very simple assumptions such as that the system is at steady-state and some reactions are irreversible, and without requiring kinetic parameters, general properties of the system can be derived. A central concept in this methodology is the notion of an elementary mode (EM for short) which represents a minimal functional subsystem. The computation of EMs still forms a limiting step in metabolic studies and several algorithms have been proposed to address this problem leading to increasingly faster methods. However, although a theoretical upper bound on the number of elementary modes that a network may possess has been established, surprisingly, the complexity of this problem has never been systematically studied. In our paper, we give a systematic overview of the complexity of optimization problems related to modes.

In [15], we consider the problem of placement of regenerators in optical networks that has attracted the attention of recent research works in optical networks. In this problem we are given a network, with an underlying topology of a graph G, and with a set of requests that correspond to paths in G. There is a need to put a regenerator every certain distance, because of a decrease in the power of the signal. We investigate the problem of minimizing the number of locations to place the regenerators. We present analytical results regarding the complexity of this problem, in four cases, depending on whether or not there is a bound on the number of regenerators at each node, and depending on whether or not the routing is given or only the requests are given (and part of the solution is also to determine the actual routing). These results include polynomial time algorithms, NP-complete results, approximation algorithms, and inapproximability results.
Applications of advanced algorithmic techniques in the optimization of transporta-
tion systems and in support of cooperative work in building design have also been in-
vestigated. The results have been presented in various national and international confer-
ences [11, 13, 14, 16].

**Experimental Algorithmics.** This line of research is concerned with the implementation
and engineering of advanced algorithms and data structures for graph problems. In [3],
we report the results of an extensive experimental analysis of efficient algorithms for
computing graph spanners in the data streaming model, where an \((\alpha, \beta)\)-spanner of a
graph \(G\) is a subgraph \(S\) of \(G\) such that for each pair of vertices the distance in \(S\) is at most
alpha times the distance in \(G\) plus beta. To the best of our knowledge, this is the first
computational study of graph spanner algorithms in a streaming setting. We compare
experimentally the randomized algorithms proposed by Baswana (B) and by Elkin (E)
for general stretch factors with the deterministic algorithm presented by Ausiello et al.
(AFI), designed for building small stretch spanners. All the algorithms we implemented
work in a data streaming model where the input graph is given as a stream of edges
in arbitrary order, and all of them need a single pass over the data. Differently from
algorithm AFI, algorithms B and E need to know in advance the number of vertices in
the graph. The results of our experimental investigation on several input families confirm
that all these algorithms are very efficient in practice, finding spanners with stretch and
size much smaller than the theoretical bounds and comparable to those obtainable by
off-line algorithms. Moreover, our experimental findings confirm that small values of the
stretch factor are the case of interest in practice, and that the algorithm by Ausiello et al.
tends to produce spanners of better quality than the algorithms by Baswana and Elkin,
while still using a comparable amount of time and space resources.

**External Memory and Streaming Algorithms for Massive Data Processing.** This re-
search line focuses on external-memory and streaming algorithms for very large graph
problems. In particular, Giorgio Ausiello, Donatella Firmani, and Luigi Laura studied the
problem of the computation of all the biconnectivity properties of a graph (i.e., connected
and biconnected components, articulation points, and bridges) in the semi-streaming
model. The problem has been studied from both a theoretical and an experimental point
of view, and the algorithm has been tested against massive graph samples. The results
have been preliminary presented by Donatella Firmani, in november 2009, at the Italian
Conference on Theoretical Computer Science (ICTCS) and have been submitted to the
journal “Networks”.

Further results were obtained in the field of streaming algorithms. In particular, we
investigated streaming problems on both undirected and directed graphs. A \(t\)-spanner
\(S\) of a given graph \(G\) is a sparse subgraph of \(G\) in which the shortest path between two
nodes is guaranteed to be never greater than \(t\) times the shortest path between the same
nodes in \(G\). During the year 2009 the study of algorithms for determining a \(t\)-spanners
in the streaming model have been continued. In this model we assume that a very large
graph \(G\) is presented edge by edge and we do not even know beforehand the number
of vertices and of edges of $G$. The algorithms that have been devised have been both analysed in theoretical terms in [5] and also experimentally tested in [3] on a variety of real world graphs of very large size, and compared against the main on line and off line algorithms known from the literature.

Surprisingly, no algorithm with both sublinear space and passes is known for natural graph problems in classical read-only streaming. In [7], we show that the use of intermediate temporary streams is powerful enough to provide effective space-passes tradeoffs for natural graph problems. In particular, for any space restriction of $s$ bits, we show that single-source shortest paths in directed graphs with small positive integer edge weights can be solved in $O((n \log^3 n) / \sqrt{s})$ passes. The result can be generalized to deal with multiple sources within the same bounds. This is the first known streaming algorithm for shortest paths in directed graphs. For undirected connectivity, we devise an $O((n \log n) / s)$ passes algorithm. Both problems require $\Omega(n/s)$ passes under the restrictions we consider. We also show that the model where intermediate temporary streams are allowed can be strictly more powerful than classical streaming for some problems, while maintaining all of its hardness for others.

In [10], we consider the problem of maintaining a fixed number of items observed over a data stream, so as to optimize the maximum value over a fixed number of recent observations. Unlike previous approaches, we use the competitive analysis framework and compare the performance of the online streaming algorithm against an optimal adversary that knows the entire sequence in advance. We consider the problem of maximizing the aggregate max, i.e., the sum of the values of the largest items in the algorithm’s memory over the entire sequence. For this problem, we prove an asymptotically tight competitive ratio, achieved by a simple heuristic, called partition-greedy, that performs stream updates efficiently and has almost optimal performance. In contrast, we prove that the problem of maximizing, for every time $t$, the value maintained by the online algorithm in memory, is considerably harder: in particular, we show a tight competitive ratio that depends on the maximum value of the stream. We further prove negative results for the closely related problem of maintaining the aggregate minimum and for the generalized version of the aggregate max problem in which every item comes with an individual window.

**Incremental Algorithms and Dynamic Data Structures.** In [9], we formulate and study a new computational model for dynamic data. In this model, the data changes gradually and the goal of an algorithm is to compute the solution to some problem on the data at each time step, under the constraint that it only has a limited access to the data each time. As the data is constantly changing and the algorithm might be unaware of these changes, it cannot be expected to always output the exact right solution; we are interested in algorithms that guarantee to output an approximate solution. In particular, we focus on the fundamental problems of sorting and selection, where the true ordering of the elements changes slowly. We provide algorithms with performance close to the optimal in expectation and with high probability.

In [17] we studied automatic incremental change propagation algorithms in the con-
text of the one-way dataflow constraint programming paradigm. In this model, the solution of a computational problem is specified as a set of constraints of the form $y = f(x_1, x_2, \ldots, x_k)$, where $y$ is automatically re-evaluated whenever any of variables $x_1, \ldots, x_k$ change. We discussed a general approach for integrating one-way dataflow constraints into imperative languages, showing a concrete embodiment of this approach into the C programming language. We also studied convergence properties of systems of dataflow constraints based on fixpoint reasoning. As a case study, we considered an implementation of the increase-only single-source shortest paths problem based on one-way dataflow constraints, matching the best known worst-case bounds for the problem. According to a preliminary experimental investigation on dynamic US road networks, our implementation can update distances orders of magnitude faster than recomputing from scratch using a state-of-the-art static implementation.

**Approximation and On-line Algorithms.** During 2009 further studies have been conducted on approximation and on line algorithms for the classical travelling salesman problem. In particular, in [6] the problem of routing $k$ vehicles in order to serve on line requests has been addressed. Possible objective functions include minimizing the makespan ($k$-Traveling Salesman Problem) and minimizing the sum of completion times ($k$-Traveling Repairman Problem). Competitive algorithms, resource augmentation results and lower bounds for $k$-server routing problems in a wide class of metric spaces have been provided.

The travelling salesman problem has also been studied from the so-called re-optimization point of view. In such case we assume to know the optimum solution for a given instance and we try to achieve the best possible approximate solution for instances obtained by slight modification of the given one. Interesting results that show that under these circumstances the classical Chrostofides’ approximation ratio can be outperformed are presented in [4].

The set covering problem has also been addressed in the on line model. In this case we show in [2] that, under suitable hypotheses, an optimum $\sqrt{n}$ competitive ratio can be achieved by a deterministic algorithm.

In [8], we analyse the size of a balanced cut in a random geometric graph obtained when nodes are uniformly distributed in the unit square. We obtain tight bounds on the number of edges to be removed in order to disconnect the graph in two subgraphs each one with a constant fraction of the total number of nodes.

In recent years considerable attention has been devoted to scheduling algorithms that are energy concerned. In [12] we study online nonclairvoyant speed scaling to minimize total flow time plus energy. We first consider the traditional model where the power function is $P(s) = s^\alpha$. We give a nonclairvoyant algorithm that is shown to be $O(\alpha^3)$-competitive. We then show an $\Omega(\alpha^{1/3-\varepsilon})$ lower bound on the competitive ratio of any nonclairvoyant algorithm. We also show that there are power functions for which no nonclairvoyant algorithm can be $O(1)$-competitive.
Algorithmic Game Theory. In [18] we focus on mechanism design for multi-objective optimization problems. In this setting we are given a main objective function, and a set of secondary objectives which are modeled via budget constraints. Multi-objective optimization is a natural setting for mechanism design as many economical choices ask for a compromise between different, partially conflicting, goals. The main contribution is showing that two of the main tools for the design of approximation algorithms for multi-objective optimization problems, namely approximate Pareto curves and Lagrangian relaxation, can lead to truthful approximation schemes. By exploiting the method of approximate Pareto curves, we devise (probabilistically) truthful FPTASs for multi-objective optimization problems whose exact version admits a pseudo-polynomial-time (Monte-Carlo) algorithm, as for instance the multi-budgeted versions of minimum spanning tree, shortest path, maximum (perfect) matching, and matroid intersection. Our construction also applies to multi-dimensional knapsack and multi-unit combinatorial auctions for unknown multi-minded bidders. Our FPTASs compute a $(1 + \epsilon)$-approximate solution violating each budget constraint by a factor $(1 + \epsilon)$. For the case of (non-perfect) matchings we also present a PTAS (not violating any constraint), which combines the approach above with a novel monotone way to guess the heaviest edges in the optimum solution.

Journals


Conference proceedings


**Articles in books**


**Submitted papers, technical reports and others**
