3 Research Activity

3.1 Computer Science

Algorithm Engineering Group

Site report 2003

The research activity of the group of Algorithm Engineering (AE) is concerned with the design, the engineering, the theoretical and experimental performance analysis of combinatorial algorithms for problems arising in modern Computer Systems and Networks, and in applications related to complex resource management problems. Our main research interests deal with the solution of optimization problems and the design of efficient data structures, with special emphasis on those applications involving large data sets. In particular we concentrate on:

1. algorithms that perform efficiently in a dynamically changing environment;
2. models and methodologies for the analysis and design of algorithms for information retrieval;
3. the efficient management of communication and information delivery and recovery in Wireless Networks and on the Internet;
4. the design and analysis of approximation algorithms for NP-hard optimization problems;
5. the design of on-line algorithms that work with incomplete information on the input instance;
6. the efficient solution of problems arising in geometric applications with emphasis on numeric robustness;
7. the design and implementation of tools and platforms for the experimental analysis and visualization of the behavior of algorithms and data structures.

The achievements of the AE group are widely recognized. Giorgio Ausiello is Editor in Chief of Theoretical Computer Science, Series A, Algorithms and Complexity. Members of the AE group are continuously involved in the Program Committees of prestigious International Conferences. Alberto Marchetti-Spaccamela has been co-chair of the 3rd Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS'03) and has been in the Program Committee of the 29th Workshop on Graph Theoretic Concepts in Computer Science (WG 03). Stefano Leonardi has been in the Program Committees of the 15th ACM Symposium on Parallel Algorithms and Architectures (SPAA'03), the 2nd Workshop on Algorithms and Models for the Web-Graph (WAW'03), the 6th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'03), the 30th International Colloquium on Automata, Languages and Programming (ICALP'03), and the 14th ACM-SIAM Symposium on Discrete Algorithms (SODA'03).

The AE group has recently organized several international scientific events. In particular it has organized the "Conference on Growing Networks and Graphs in Statistical Physics, Finance, Biology and Social Systems" in Rome in September 2003, School of Engineering, Università "La Sapienza".

A regular Seminar Program, the Interdepartmental Seminar on Algorithms (SIA), is also organized in cooperation with the Department of Computer Science of this University (see `http://www.dis.uniroma1.it/sia/`).

The AE group is currently cooperating with several prestigious research institutions: Max Planck für Informatik (Saarbrücken, Germany), CTI-Patras (Greece), ETH (Zurich, Switzerland), Université de Paris (Dauphine, France), Tel-Aviv University (Israel), AT&T - Research Labs (Florham Park, USA), ICSI-Berkeley (USA), Brown University (Providence, USA).

The AE group is presently involved in the following research projects: EU-IST ALCOM-FT "Algorithms and Complexity in Future Technologies"; EU IST "Approximation and On-line Algorithms (APPOL2)."; EU-RTN AMORE "Algorithmic Methods for Optimizing the Railways in Europe"; MIUR "Resource Allocation in Computer Networks"; MIUR National Project "Rete multimediale nell'evoluzione verso UTMS - Linea di ricerca Applicazione ai beni culturali"; EU-IST "Coevolution and self-organization in dynamical networks (COSIN)"; EU Contract 001907 "Dynamically Evolving Large Scale Information Sytems" (DELIS); MIUR National Project ALINWEB "Algorithmics for Internet and the Web"; FIRB National Projects - WEBMINDS and VICOM. Very recently a national committee of the MIUR has approved a financial support of 1M-Euro in three years for the institution at the University of Rome "La Sapienza" of an *Excellence Centre* that aims at creating new scientific and technical synergies in the area of transportation and logistics. A major role in this project is played by the Algorithm Engineering group at DIS.

Group members. Giorgio Ausiello, Luca Becchetti, Fabrizio d'Amore, Camil Demetrescu, Paolo Giulio Franciosa, Luigi Laura, Stefano Leonardi, Alberto Marchetti-Spaccamela, Umberto Nanni, Andrea Vitaletti.

Post-docs. Jochen Koenemann, Alex Kesselman, Panayiotis Tsaparas, Marco Gartner.

Graphs and Combinatorial Algorithms Part of our effort was devoted to studying efficient algorithms for dynamic graph problems. In particular, in [5] we study novel combinatorial properties of graphs that allow us to devise a completely new approach to dynamic all pairs shortest paths problems. Our approach yields a fully dynamic algorithm for general directed graphs with non-negative real-valued edge weights that supports any sequence of operations in $O(n^2 \log^3 n)$ amortized time per update and unit worst-case time per distance query, where n is the number of vertices. We can also report shortest paths in optimal worst-case time. These bounds improve substantially over previous results and solve a long-standing open problem. Our algorithm is deterministic and uses simple data structures.

In [1], we propose a fully-dynamic distributed algorithm for the all-pairs shortest paths problem on general networks with positive real edge weights. If Δ_σ is the number of pairs of nodes changing the distance after a single edge modification σ(insert, delete, weight decrease, or weight increase) then the message complexity of the proposed algorithm is $O(n\Delta_\sigma)$ in the worst case, where n is the number of nodes of the network. If $\Delta_\sigma = o(n^2)$, this is better than recomputing everything from scratch after each edge modification. Up to now only a result of Ramarao and Venkatesan was known, stating that the problem of updating shortest paths in a dynamic distributed environment is as hard as that of computing shortest paths.

In [5], we propose a new solution for the fully dynamic single source shortest paths problem in a directed graph $G = (N, A)$ with arbitrary arc weights, that works for any digraph and has optimal space requirements and query time. If a negative-length cycle is introduced in the subgraph of G reachable from the source during an update operation, then it is detected by the algorithm. Zero-length cycles are explicitly handled. We evaluate the cost of the update operations as a function of a structural property of G and of the number of the output updates. We show that, if G has a k-bounded accounting function (as in the case of digraphs with genus, arboricity, degree, treewidth or page number bounded by k), then the update procedures require $O(minm, k \cdot nA \cdot \log n)$ worst case time for weight-decrease operations, and $O(\min\{m \cdot \log n, k \cdot (nA + nB) \cdot \log n + n\})$ worst case time for weight-increase operations. Here, $n = |N|$, $m = |A|$, nA is the number of nodes affected by the input update, that is the nodes changing either the distance or the parent in the shortest paths tree, and nB is the number of nonaffected nodes considered by the algorithm that also belong to zero-length cycles. If zero-length cycles are not allowed, then nB is zero and the bound for weight-increase operations is $O(\min\{m \cdot \log n, k \cdot nA - \log n + n\})$. Similar amortized bounds hold if we perform also insertions and deletions of arcs.

In [1], we study the problem of maintaining a t-spanner of a general weighted undirected graph under a sequence of edge insertions and deletions.

Finally, in [1] we survey some of the best known techniques for dynamic graph problems, discussing efficient dynamic algorithms for classical problems such as connectivity, minimum spanning tree, transitive closure, and shortest paths.

Information Retrieval In [6] we have considered the problem of collaborative filtering, introducing a random planted model of bi-categorical data. We have adapted the ideas of an algorithm due to Condon and Karp to develop a simple linear time algorithm to discover the underlying hidden structure of a graph generated according to the planted model with high probability. Our experimental analysis has shown that the algorithm might work quite well in practice.

The Latent Semantic Indexing (LSI) proved to be an effective technique in the field of Information Retrieval. Its drawbacks are the time needed to compute the SVD decomposition and to answer queries, since the query must be compared against each document in the collection. On the other hand, fast and powerful IR techniques for Web documents based on the analysis of the hyperlinked structure of the Web have been proposed over the last few years. In [3] we propose an approach that integrates Web IR technique with latent semantic analysis for collections of documents missing a hyperlinked structure. Our experiments show that this approach achieves both fast answer to on-line queries and retrieval effectiveness.

In [2] a different approach is presented. This technique, that we call *approximated LSI*, uses the information computed by the traditional LSI to provide a fast online answer to the users. Informally, if compared to the traditional LSI, we can say that instead of returning the documents that are related to the concepts in the query, we return the documents that contain terms that are related to the concepts in your query.

Distributed and network algorithms In [6], we study the multicast routing and admission control problem on unit-capacity trees and meshes. The objective is to maximize throughput. The offline version of the problem is MAX-SNP hard. We give the first constant-factor approximation algorithm for trees.

Sensor networks demand for the design of energy conserving and scalable routing protocols. We are currently studying the impact of clustering and data aggregation on routing energy consumption. We assess the advantages of these techniques by means of a comparative NS2-based performance evaluation. In particular, we have compared the performances of the routing algorithm Directed Diffusion with a variant of the Distributed Clustering Algorithm in which Directed Diffusion Greedy is used for routing over the backbone. Finally we have introduced Directed Diffusion Light that results in significant savings in terms of exchanged control messages without compromising the protocol robustness. Results in this topics are expected to be published in 2004.

Approximate and on-line algorithms The completeness in differential approximability classes has been studied in [1]. In differential approximation, the quality of an approximation algorithm is the measure of both how far is the solution computed from a worst one and how close is it to an optimal one. We have defined natural approximation preserving reductions and proved completeness results for the class of the NPO optimization problems (class NPO), as well as for DAPX, the differential counterpart of the class APX.

In [2] we have introduced the notion of smoothed competitive analysis of online algorithms. Smoothed analysis has been proposed by Spielman and Teng to explain the behaviour of algorithms that work well in practice while performing very poorly from a worst case analysis point of view. We apply this notion to analyze the Multi-Level Feedback (MLF) algorithm to minimize the total flow time on a sequence of jobs released over time when the processing time of a job is only known at time of completion. The initial processing times are integers in the range $[1, 2^k]$ and they are smoothened by changing the k least significant bits under a quite general class of probability distributions. We show that MLF admits a smoothed competitive ratio that exponentially decreases as the variance of the distribution increases. A direct consequence of our result is also the first average case analysis of MLF. We show a constant expected ratio of the total flow time of MLF to the optimum under several distributions including the uniform distribution.

In [3] we explore the the quality of service (QoS) that is achievable by *semi-clairvoyant* online scheduling algorithms, which are algorithms that only require approximate knowledge of the initial processing time of each job, on a single machine. We give a semi-clairvoyant algorithm that is $O(1)$-competitive with respect to average flow time on one single machine. We also show a semi-clairvoyant algorithm on parallel machines that achieves up to contant factors the best known competitive ratio. It is known that the clairvoyant algorithm SRPT is optimal with respect to average flow time and is 2-competitive with respect to average stretch. Thus it is possible for a clairvoyant algorithm to be simultaneously competitive in both average flow time and average stretch. In contrast we show that no semi-clairvoyant algorithm can be simultaneously $O(1)$-competitive with respect to average stretch and $O(1)$-competitive with respect to average flow time. Thus in this sense one might conclude that the QoS achievable by semi-clairvoyant algorithms is not competitive with clairvoyant algorithms.

Given a weighted directed graph $G = (V, A)$, the minimum feedback arc set problem consists of finding a minimum weight set of arcs $A' \subseteq A$ such that the directed graph $(V, A \setminus A')$ is acyclic. Similarly, the minimum feedback vertex set problem consists of finding a minimum weight set of vertices containing at least one vertex for each directed cycle. Both problems are NP-complete. In [2], we present simple combinatorial algorithms for these problems that achieve an approximation ratio bounded by the length, in terms of number of arcs, of a longest simple cycle of the digraph.

Experimentation and visualization Algorithm Engineering is concerned with the design, analysis, implementation, tuning, debugging and experimental evaluation of computer programs for solving algorithmic problems. It provides methodologies and tools for developing and engineering efficient algorithmic codes and aims at integrating and reinforcing traditional theoretical approaches for the design and analysis of algorithms and data structures. In [3] we survey some relevant contributions to the field of Algorithm Engineering and we discuss significant examples where the experimental approach helped in developing new ideas, in assessing heuristics and techniques, and in gaining a deeper insight about existing algorithms.

In [7], we present an experimental study of the properties of web graphs. We study a large crawl from 2001 of 200M pages and about 1.4 billion edges made available by the WebBase project at Stanford, and synthetic graphs obtained by the large scale simulation of stochastic graph models for the Webgraph. This work has required the development and the use of external and semi-external algorithms for computing properties of massive graphs, and for the large scale simulation of stochastic graph models. We report our experimental findings on the topological properties of such graphs, describe the algorithmic tools developed within this project and report the experiments on their time performance.

In [4], we consider the one-sided crossing minimization problem (CP): given a bipartite graph G and a permutation x_0 of the vertices on a layer, find a permutation x_1 of the vertices on the other layer which minimizes the number of edge crossings in any straightline drawing of G where vertices are placed on two parallel lines and sorted according to x_0 and x_1. Solving CP represents a fundamental step in the construction of aesthetically pleasing layouts of hierarchies and directed graphs, but unfortunately this problem has been proved to be NP-complete. We first address the strong relation between CP and the problem of computing minimum feedback arc sets in directed graphs and we devise a new approximation algorithm for CP, called PM, that exploits this dependency. Then, we experimentally and visually compare the performance of PM with the performance of well-known algorithms and of recent attractive strategies. Experiments are carried out on different families of randomly generated graphs, on pathological instances for CP, and on real test sets. Performance indicators include both number of edge crossings and running time, as well as structural measures of the problem instances. We found CP to be a very interesting and rich problem from a combinatorial point of view. Our results clearly separate the behavior of the algorithms, proving the effectiveness of PM on most test sets and showing tradeoffs between quality of the solutions and running time. However, if the visual complexity of the drawings is considered, we found no clear winner. This confirms the importance of optimizing also other aesthetic criteria such as symmetry, edge length, and angular resolution.

In [4], we describe Leonardo Web, a collection of tools for building animated presentations that can be useful for teaching, disseminating, and e-learning. Presentations can be created via the combined use of a visual editor and a Java library. The library allows it to generate animations in a batch fashion directly from Java code according to an imperative specification style. Batch-generated animations can then be refined and customized using the editor. Presentations can be finally viewed with a simple Java player, which ships both as a stand-alone application for off-line deployment and as a Java applet embedded in a Web page. The player supports step-by-step and continuous execution, reversibility, speed selection, and smooth animation.

Journals

[1] S. Cicerone, G. Di Stefano , D. Frigioni, and U. Nanni. A fully dynamic algorithm for distributed shortest paths. Theoretical Computer Science (Elsevier), 297:83–102, 2003.

[2] C. Demetrescu and I. Finocchi: Combinatorial Algorithms for Feedback Problems in Directed Graphs, Information Processing Letters 86 (IPL), pp. 129-136, 2003.

[3] C. Demetrescu, I. Finocchi, and G.F. Italiano: Algorithm Engineering, in The Algorithmics Column (J.Diaz), Bulletin of the EATCS 79, pp. 48-63, February 2003.

[4] C. Demetrescu and I. Finocchi: Removing Cycles for Minimizing Crossings, ACM Journal on Experimental Algorithmics (JEA), 6(1). Special Issue devoted to selected papers from the 2nd International Workshop on Algorithms Engineering and Experiments.

[5] D. Frigioni, A. Marchetti-Spaccamela, and U. Nanni. Fully dynamic shortest paths in digraphs with arbitrary arc weights. Journal of Algorithms (Academic Press), 49(1):86–113, 2003.

[6] M. Henzinger and S. Leonardi. Scheduling multicasts on unit-capacity trees and meshes.*Journal of Computer and System Sciences*, 66(3): 567-611, 2003.

Conference Proceedings

[1] G. Ausiello, C. Bazgan, M. Demange, V. Th. Paschos, Completeness in differential approximation classes, Proceedings of MFCS 2003, LNCS, Springer 2003

[2] L. Becchetti and S. Leonardi and A. Marchetti-Spaccamela and G. Schäfer and T. Vredeveld, Average Case and Smoothed Competitive Analysis of the Multi-Level Feedback Algorithm, Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS), pp. 462-471, 2003.

[3] L. Becchetti and S. Leonardi and A. Marchetti-Spaccamela and K. Pruhs, Semi-clairvoyant Scheduling, Proceedings of the 11th Annual European Symposium on Algorithms (ESA'03), pp.67-77, 2003.

[4] B.A. Colombo, C. Demetrescu, I. Finocchi, and L. Laura: A System for Building Animated Presentations over the Web, Proceedings of the AICCSA'03 Workshop on Practice and Experience with Java Programming in Education, Tunis, July 2003.

[5] C. Demetrescu and G.F. Italiano: A New Approach to Dynamic All Pairs Shortest Paths, Proceedings of the 35th Annual ACM Symposium on Theory of Computing (STOC'03), San Diego, California, June 2003.

[6] D.Dubashi, L. Laura and A. Panconesi. Collaborative Filtering in Planted Partition Models. Proceedings of the 23rd conference on Foundations Of Software Technology And Theoretical Computer Science (FSTTCS '03). Also ALCOM-FT Technical Report 03-186.

[7] L. Laura, S. Leonardi, S. Millozzi, U. Meyer and J. Sibeyn. Algorithms and Experiments for the Webgraph. Proceedings of the *11th Annual European Symposium on Algorithms (ESA03)*, Lecture Notes in Computer Science 2832 (Springer-Verlag, 2003).

Book

[1] C. Demetrescu, I. Finocchi, and G.F. Italiano: Dynamic Graph Algorithms, in Handbook of Graph Theory, chapter 10.2, J. Yellen e J.L. Gross eds., CRC Press, Discrete Mathematics and Its Applications, ISBN 1-58488-090-2, December 2003.

Submitted papers, technical reports and others

[1] G. Ausiello, P. G. Franciosa and G. F. Italiano, Fully dynamic maintenance of t-spanners on general graphs, ALCOMFT-TR-03-23, April 2003.

[2] L. Laura, U. Nanni, and F. Sarracco. Query Transformation through approximated LSI computation ALCOM-FT Technical Report 03-185.

[3] L. Laura, S. Leonardi, S. Millozzi and S. Portelli. Information Retrieval via LSI-induced link Analysis. Technical Report Dis 06-03.

3.2 System Science

3.3 Management Science