Algorithm Engineering Group

Site report 2002

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 Chairman of the Technical Committee on Foundations of Computer Science of the International Federation of Information Processing (IFIP – TC 1) since 1997 and Editor in Chief of Theoretical Computer Science, Series A, Algorithms and Complexity. Members of the AE group are continuously involved in the Program Committes of prestigeous International Conferences. Giorgio Ausiello was in the Program Committee of the 5th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'02). Camil Demetrescu was in the Program Committee of the 10th European Symposium on Algorithms - Applied Track (ESA'02). Alberto Marchetti-Spaccamela was in the Program Committee of 10th European Symposium on Algorithms - Theory Track (ESA'02). Stefano Leonardi has been recently involved in several Program Committees, in-

cluding the 15th ACM Symposium on Parallel Algorithms and Architectures (SPAA'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 ALGO 2002, where the European Symposium on Algorithms (ESA 2002), the Workshop on Approximation Algorithms for Combinatorial Optimization (APPROX 2002), and the Workshop on Algorithms in Bioinformatics (WABI 2002) have been co-located, in Rome in September 2002. Besides it has organized the Workshop on Algorithms and Models for the Web in Udine (June 2002). A regular Seminar, the Interdepartmental Seminar on Algorithms (SIA), is also organized in cooperation with the Department of Computer Science of this University.

The AE group is currently cooperating with several prestigeous 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 APPOL "Approximation and On-line algorithms"; EU-RTN AMORE "Algorithmic Methods for Optimizing the Railways in Europe"; MURST "Resource Allocation in Computer Networks"; MURST National Project "Rete multimediale nell'evoluzione verso UTMS - Linea di ricerca Applicazione ai beni culturali"; EU-IST COSIN "Coevolution and self-organization in dynamical networks"; MURST National Project ALINWEB "Algorithmics for Internet and the Web".

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.

Graphs and Combinatorial Algorithms Part of our effort was devoted to studying dynamic path problems on directed graphs. In particular, we have considered the problem of maintaining all pairs shortest paths in a weighted directed graph. We have shown how to obtain effective query/update trade-offs for dynamic all pairs shortest paths, by introducing two new families of algorithms that improve the best known update bounds

at the price of a non-constant query time. More recently, we have studied novel combinatorial properties of graphs and we have devised a completely new approach to the problem. Our approach yields a fully dynamic algorithm for general directed graphs with real edge weights that supports any sequence of operations in $O(n^2 \log^2 n)$ amortized time per update and unit worst-case time per distance query, where n is the number of vertices. This solves a long-standing open problem, improving substantially over previous results, and being the first algorithm that solves the dynamic all pairs shortest paths problem in its full generality. We have also studied the problem of constructing compact data structures for representing information about all pairs shortest paths in a weighted directed graph. In particular, for a directed graph G we have considered queries of the form: "What is the shortest path distance from vertex x to vertex y in G avoiding a failed link (u, v)?" We have shown that an oracle for such queries can be stored in $O(n^2 \log n)$ space with a constant query time. No non-trivial solution was known for this problem. These results appear in [7, 8, 9]. During 2002 an investigation on the fully dynamic maintenance of sparse spanners in a graph has also been started.

Information Retrieval Qiu and Frei showed that the use of a similarity thesaurus can be very effective if compared to other query expansion techniques. We have proposed a novel approch to calculate the similarity thesaurus, based on the Latent Semantic Indexing (LSI). We have also shown experimental evidence of the retrieval effectiveness of our approach. In a different work, 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. We have also given applications to the probabilistic analysis of the LSI in the probabilistic corpus models introduced by Papadimitriou et al. Our experimental analysis has shown that the algorithm might work quite well in practice. Results in this area appeared in [10, 14].

Distributed and network algorithms Next generation 3G/4G wireless data networks allow multiple codes (or channels) to be allocated to a single user, where each code can support multiple data rates. Providing fine-grained QoS to users in such networks poses the two dimensional challenge of assigning *both* power (rate) and codes for every user. This gives

rise to a new class of parallel scheduling problems. We abstract general downlink scheduling problems suitable for proposed next generation wireless data systems, devise a communication-theoretic model for multirate wireless channels. Our focus is on optimizing the maximum response time of jobs, which is suitable measure for streams of user requests. We have presented provable results on the algorithmic complexity of these scheduling problems, devising very simple, online algorithms for approximating the optimal maximum response time.

In the Hose Model for Network Design problems we are not given a node-to-node traffic matrix, but just upper bounds on the total volume of traffic entering or leaving some set of terminals. The goal is to install capacity at a minimum cost as to support every traffic matrix satisfying the upper bound constraints. In this setting, we have considered problem of designing a minimum cost tree-network in the hose model. This problem is known to be NP-hard when we are given different upper bounds on the amount of traffic leaving or entering a terminal but polynomially solvable if the two bounds are equal for every specific vertex. We have shown that also the case where the total outgoing traffic is equal to the total incoming traffic is polynomially time solvable. Moreover, in this case, the cost of an optimal tree reservation is within a factor of three of the cost of an optimal reservation that is allowed to use arbitrary subgraphs.

In an another paper we have studied stochastic graph models of the WebGraph, presenting a new model that describes the WebGraph as an ensemble of different regions generated by independent stochastic processes. Models such as the Copying Model of Kumar et al. [FOCS 2001] and the Evolving Network Model by Barabasi are simulated and compared on several relevant measures such as degree and clique distribution.

Caching and prefetching have often been studied as separate tools for enhancing the access to the World Wide Web. In this setting, we have proposed integrated Caching and Prefetching Algorithms for improving the performances of web navigation, devising a new prefetching algorithm that uses a limited form of user cooperation to establish which documents to prefetch in the local cache at the client side. We have shown that our prefetching technique is highly beneficial only if integrated with a suitable caching algorithm. Works in this area appeared in [2, 3, 12, 13].

Approximate and on-line algorithms We have studied the multicast routing and admission control problem on unit-capacity tree and mesh topologies in the throughput-model. The problem is a generalization of the edge-

disjoint paths problem and is NP-hard both on trees and meshes. We have addressed both the offline and the online version of the problem: In the offline setting, we have given the first constant-factor approximation algorithm for trees, and an $O((\log \log n)^2)$-factor approximation algorithm for meshes. In the online setting, we have provided the first polylogarithmic competitive online algorithm for tree and mesh topologies. We have aldo studied the problem of scheduling parallel machines online, allowing preemptions while disallowing migration of jobs that have been scheduled on one machine to another. We have measured the quality of service provided by an algorithm by the average stretch, defined as the average ratio of the amount of time the job spends in the system (the response time) to its processing time. This problem is of relevance in many applications, e.g., wireless data servers and distributed server systems in wired networks. We have proved a first costant competitive ratio for this problem for the algorithm introduced in STOC99 by Awerbuch et al. for minimizing average flow time without migration. Finally, we have studied the on-line traveling salesman problem where the objective is to minimize the maximum flow time. This objective is particularly interesting for applications. However, it is possible to prove that there can be no competitive algorithm, neither deterministic nor randomized. We have introduced a natural restriction on the adversary for this problem on the real line. Our main result is an algorithm that achieves a costant competitive ratio against the non-abusive adversary. Results in this area appeared in [1, 11, 15]. In the framework of on-line algorithms, the study of the on line version of the so-called Quota Traveling Salesman Problem has also been started.

Experimentation and visualization We have surveyed different aspects in algorithm engineering as a field that provides methodologies and tools for developing and engineering efficient algorithmic codes. In this context, we have considered the role of algorithm engineering in integrating and reinforcing the traditional theoretical approaches for the design and analysis of algorithms, addressing issues related to design, analysis, implementation, tuning, debugging and experimental evaluation of computer programs for solving algorithmic problems. We have also applied different experimental methodologies to the analysis of agorithms for the two-layer crossings minimization problem, which arises in many applications such as diagram drawing and computer-aided circuit layout design. Based on those experiments, which yielded useful clues to the structure of the problem, we have proposed and analysed a new effective algorithm, based on solving feedback

problems on directed graphs. Works in this area appeared in [4, 5, 6].

Bibliography

[1] L. Becchetti, S. Leonardi and S. Muthukrishnan. Scheduling to minimize Average Stretch without Migration. Accepted for publication on *Journal of Computer and System Sciences*.

[2] L. Becchetti, S. Diggavi, S. Leonardi, A. Marchetti-Spaccamela, S. Muthukrishnan, T. Nandagopal and A. Vitaletti. Parallel Scheduling Problems in Next Generation Wireless Networks Proceedings *Fourteenth Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA 02)*, pp. 238-247, 2002.

[3] M. Curcio, S. Leonardi and A. Vitaletti. An Experimental Study of Prefetching and Caching Algorithms for the World Wide Web. To appear in *Proceedings of the 4th Workshop on Algorithm Engineering and Experiments (ALENEX'02)*, San Francisco (CA), January 2002.

[4] C. Demetrescu and I. Finocchi. *Local-Ratio Approximation Algorithms for Feedback Problems on Directed Graphs*. To appear in *Information Processing Letters*.

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

[6] C. Demetrescu and I. Finocchi and G.F. Italiano. *Algorithm Engineering*. To appear in *Bulletin of the EATCS*.

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

[8] C. Demetrescu and G.F. Italiano. Improved Bounds and New Trade-Offs for Dynamic All Pairs Shortest Paths. In *Proceedings of the 29-th International Colloquium on Automata, Languages, and Programming (ICALP'02), Málaga, Spain*, July 2002.

[9] C. Demetrescu and M. Thorup. Oracles for Distances Avoiding a Link-failure. In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'02), San Francisco, CA*, January 2002.

[10] D.Dubashi, L.Laura and A.Panconesi. Analysis and Experimental evaluation of a Simple Algorithm for Collaborative Filtering in the Planted Partition Model. DIS technical report 23-02.

[11] M.R. Henzinger and S. Leonardi. Scheduling multicasts on unit capacity trees and meshes. Accepted for publication in *Journal of Computer and System Sciences*.

[12] G.F. Italiano, S. Leonardi, G. Oriolo. Design of Networks in the Hose Model. To appear in Proc. of the *2nd International Workshop on Approximation and Randomization Algorithms in Communication Networks*, pp. 65-76, Carleton Scientific Press, 2002.

[13] L. Laura, S. Leonardi, G. Caldarelli and P. De Los Rios. A Multi-Layer Model of the WebGraph. *2nd International Workshop on Web Dynamics*, Honolulu, Hawaii, May 2002.

[14] L.Laura, U.Nanni and F.Sarracco. Query expansion through Automatic Generated LSI-Similarity Thesaurus. DIS technical report 24-02.

[15] M. Lipmann, S.O. Krumke, L. Laura, A. Marchetti-Spaccamela, W.E. de Paepe, D. Poensgen, L. Stougie. Non-abusiveness helps: An O(1)-competitive algorithm for minimizing the maximum flow time in the online traveling salesman problem. In Proceedings of the 5th International Workshop on Approximation Algorithms for Combinatorial Optimization, Lecture Notes in Computer Science, Springer, 2002.