40th Annual Symposium on Foundations of Computer Science

40^{th} Annual Symposium on
Foundations of Computer Science

October 17–19, 1999
New York City, New York

sponsored by

IEEE Computer Society Technical Committee on
Mathematical Foundations of Computing

IEEE
COMPUTER
SOCIETY

Los Alamitos, California
Washington · Brussels · Tokyo

IEEE Catalog Number 99CB37039
ISBN: 0-7695-0409-4 (Softbound)
ISBN: 0-7803-5955-0 (Casebound)
ISBN: 0-7695-0411-6 (Microfiche)
ISSN: 0272-5428
PR00409

Additional copies may be ordered from:

IEEE Computer Society
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: + 1-714-821-8380
Fax: + 1-714-821-4641
E-mail: cs.books@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: + 1-732-981-0060
Fax: + 1-732-981-9667
mis.custserv@computer.org

IEEE Computer Society
Asia/Pacific Office
Watanabe Building,
1-4-2 Minami-Aoyama
Minato-ku, Tokyo 107-0062 JAPAN
Tel: + 81-3-3408-3118
Fax: + 81-3-3408-3553
tokyo.ofc@computer.org

Editorial production by Anne Rawlinson

Cover art production by Alex Torres

Printed in the United States of America by The Printing Company

IEEE COMPUTER SOCIETY

Proceedings

40^{th} Annual Symposium on Foundations of Computer Science

Foreword

The papers in these proceedings were presented at the 40th Annual Symposium on Foundations of Computer Science (FOCS '99), sponsored by the IEEE Computer Society Technical Committee on Mathematical Foundations of Computing. The conference was held in New York, NY, October 17-19, 1999.

The program committee consisted of Susanne Albers (Max Planck Institute Saarbrucken), Eric Allender (Rutgers University), James Aspnes (Yale University), Yair Bartal (Lucent Bell Labs), Paul Beame (University of Washington), Andrei Broder (Altavista Company), Nader Bshouty (Technion and University of Calgary), Oded Goldreich (Weizmann Institute), Ming Li (University of Waterloo), Joseph Mitchell (SUNY Stonybrook), Michael Mitzenmacher (Harvard University), Dana Randall (Georgia Tech), Satish Rao (U.C. Berkeley), David Shmoys (Cornell University), Peter Shor (AT&T Research), and Luca Trevisan (Columbia University).

The program committee met on June 27-28, 1999 and selected 67 papers from the 218 detailed abstracts submitted. The submissions were not refereed, and many of them represent reports of continuing research. It is expected that most of these papers will appear in a more complete and polished form in scientific journals in the future. In addition to the regular program, László Lovász gave his Donald E. Knuth Prize lecture at the conference.

The committee selected two papers to jointly receive the Machtey Award for the best student-authored paper. These two papers were "A $\frac{5}{2}n^2$-Lower Bound for the Rank of $n \times n$-Matrix Multiplication over Arbitrary Fields" by Markus Bläser and "Improved Bounds for Sampling Colorings" by Eric Vigoda. There were many excellent candidates for this award.

The committee wishes to thank all of those who submitted papers for consideration, as well as those who helped with the process of evaluating the submissions. A list of the latter individuals appears in these proceedings under the heading "Reviewers." The program committee chair also wishes to thank Steve Tate, Peter Shor, and Lance Fortnow for help with the electronic submission server and electronic program committee software, and Anne Rawlinson for the production of these proceedings. Finally, a special thanks to Luca Trevisan for acting as our gracious host at Columbia for our program committee meeting.

Paul Beame
Program Committee Chair

Committees

General Conference and Technical Committee Chair

Alok Aggarwal, IBM Research

Local Arrangements and Finance Chair

Piotr Berman, Pennsylvania State University

Program Committee Chair

Paul Beame, University of Washington

Program Committee

Susanne Albers
Eric Allender
Jim Aspnes
Yair Bartal
Paul Beame
Andrei Broder
Nader Bshouty
Oded Goldreich
Ming Li
Joseph Mitchell
Michael Mitzenmacher
Dana Randall
Satish Rao
David Shmoys
Peter Shor
Luca Trevisan

Machtey Award

"A $\frac{5}{2}n^2$-Lower Bound for the Rank of $n \times n$-Matrix Multiplication over Arbitrary Fields"
Markus Bläser

"Improved Bounds for Sampling Colorings"
Eric Vigoda

Reviewers

Martin Abadi
Bill Aiello
Noga Alon
Rajeev Alur
Andris Ambainis
Arne Andersson
Dana Angluin
Esther Arkin
Yossi Azar
Richard Beigel
Mihir Bellare
Josh Benaloh
Shai Ben-David
Michael Bender
Michael Ben-Or
Johannes Bl¨ omer
Graham Brightwell
Adam Buchsbaum
Harry Buhrman
Jonathan Buss
Jin-Yi Cai
William Chan
Paul Chew
Fan Chung
Richard Cleve
Edith Cohen
Claude Crepeau
Philipe Cucker
Drew Dean
Tamal Dey
Wayne Eberly
Alon Efrat
Nadav Eiron
Ran El-Yaniv
Jeff Erickson
Sandor Fekete
Paolo Ferragina
Amos Fiat
Faith Fich
James Fill
Guang Gao
Juan Garay
Joachim von zur Gathen
Phil Gibbons
Rafi Hassin
Monika Henzinger
Mark Huber
Piotr Indyk
Sandy Irani
Jeff Jackson
Mark Jerrum
Tao Jiang
David Johnson
Michael Kaminski
Paul Kearney
Mike Kearns
Jon Kleinberg
Hugo Krawczyk
Eyal Kushilevitz
Richard Ladner
Jeff Lagarias
Leslie Lamport
Friedhelm Meyer auf der Heide
Peter Bro Miltersen
Ramamohan Paturi
Yuval Rabani
Sivaram Rajagopalan
Rajeev Raman
Ramamoothi Ravi
Klaus Reinhardt
Dana Ron
Amin Shokrollahi
Amit Sahai
Raimund Seidel
Steve Seiden
Rocco Servidio
Nir Shavit
Carl H. Smith
Jack Snoeyink
Madhu Sudan
´ Eva Tardos
David Toman
Martin Tompa
Salil Vadhan
Moshe Vardi
Ramarathnam Venkatesan
Eric Vigoda
Paul Vitanyi
John Watrous
Jeff Westbrook
Avi Wigderson
Ronald de Wolf
Mihalis Yannakakis
Francis Zane

Corporate Sponsors

IBM
AT&T
Microsoft

Table of Contents

Session 1

Chair: David Shmoys

Session 2

Chair: Paul Beame

Session 3

Chair: Joseph Mitchell

Session 4

Chair: Michael Mitsenmacher

Session 5A

Chair: Andrei Broder

Session 5B

Chair: Eric Allender

Session 6A

Chair: Dana Randall

Session 6B

Chair: James Aspnes

Session 7A

Chair: Satish Rao

Session 7B

Chair: Peter Shor

Session 8A

Chair: Joseph Mitchell

Session 8B

Chair: Paul Beame

Session 9A

Chair: Yair Bartal

Session 9B

Chair: Luca Trevisan

Session 10A

Chair: Ming Li

Session 10B

Chair: Luca Trevisan

Session 11A

Chair: Susanne Albers

Session 11B

Chair: Nader Bshouty

Session 1

Chair: David Shmoys

Primal-Dual Approximation Algorithms for Metric Facility Location and k-Median Problems

*Kamal Jain**   *Vijay V. Vazirani**

Abstract

We present approximation algorithms for the metric uncapacitated facility location problem and the metric k-median problem achieving guarantees of 3 and 6 respectively. The distinguishing feature of our algorithms is their low running time: $O(m \log m)$ and $O(m \log m(L+\log(n)))$ respectively, where n and m are the total number of vertices and edges in the underlying graph. The main algorithmic idea is a new extension of the primal-dual schema.

1. Introduction

Given costs for opening facilities and costs for connecting cities to facilities, the uncapacitated facility location problem seeks a minimum cost solution that connects each city to an open facility. Clearly, this problem is applicable to a number of industrial situations. For this reason it has occupied a central place in operations research since the early 60's [3, 25, 33, 34], and has been studied from the perspectives of worst case analysis, probabilistic analysis, polyhedral combinatorics and empirical heuristics (see [10, 28, 29]). In the last few years, there has been renewed interest in tackling this problem, this time from the perspective of approximation algorithms [12, 23, 24, 27, 32]. In this paper, we carry this further by developing an approximation algorithm based on the primal-dual schema. We further use this algorithm as a subroutine to solve a related problem, the k-median problem. The latter problem differs in that there are no costs for opening facilities, instead a number k is specified, which is an upper bound on the number of facilities that can be opened. The two algorithms achieve approximation guarantees of 3 and 6 respectively.

Both of our algorithms work only for the metric case, i.e., when the connecting costs satisfy the triangle inequality; both problems are **NP**-hard for this case as well. If the connection costs are unrestricted, approximating either problem is as hard as approximating set cover, and therefore cannot be done better than $O(\log n)$ factor, unless $\mathbf{NP} \subseteq \tilde{\mathbf{P}}$. For the first problem, this is straightforward to see, and for the second, this is established by Lin and Vitter [26].

The distinguishing feature of our algorithms is their low running time: $O(m \log m)$ and $O(m \log m(L + \log(n)))$ respectively, where n and m are the total number of vertices and edges in the underlying graph ($n = n_c + n_f$ and $m = n_c \times n_f$, where n_c and n_f are the number of cities and facilities) and L is the number of bits needed to represent a connecting cost. In particular, the running time of the first algorithm is dominated by the time taken to sort the connecting costs of edges. It is worth pointing out that our facility location algorithm is also suitable for distributed computation.

The first constant factor algorithm for the metric uncapacitated facility location problem was given by Shmoys, Tardos and Aardal [32], improving on Hochbaum's bound of $O(\log n)$ [20] (see [27] for another $O(\log n)$ factor algorithm). Their approximation guarantee was 3.16. After some improvements [23, 11], the current best factor is $(1 + 2/\epsilon)$, due to Chudak and Shmoys [12]. The drawback of these algorithms, based on LP-rounding, is that they need to solve large linear programs, and so have prohibitive running times for most applications. A different approach was recently used by Korupolu, Plaxton and Rajaraman [24] (see also [14]). They showed that a well known local search heuristic achieves an approximation guarantee of $(5 + \epsilon)$, for any $\epsilon > 0$. However, even this algorithm has a high running time of $(n^6 \log n/\epsilon)$. Regarding hardness results, the work of [19, 35] establishes that a better factor than 1.463 is not possible, unless $\mathbf{NP} \subseteq \tilde{\mathbf{P}}$.

Researchers have felt that the primal-dual schema should be adaptable in interesting ways to the combinatorial structure of individual problems, and that its full potential has not yet been realized in the area of approximation algorithms. Our work substantiates this belief. We extend the scope of this schema in the following way: All primal-dual approximation algorithms obtained so far [6, 18, 36, 17, 31, 30, 21] work with a pair of covering and packing linear programs, i.e., a primal-dual pair of LP's such that all components of the constraint matrix, objective function vector and right hand side vector are non-negative. This includes, for instance, [36, 17]

*Address: College of Computing, Georgia Institute of Technology, Atlanta, GA 30332–0280, E-mail: {kjain,vazirani}@cc.gatech.edu. Research supported by NSF Grant CCR-9820896.

0-7695-0409-4/99 $10.00 © 1999 IEEE

in which the overall LP-relaxation does have negative coefficients; however, the problem is decomposed into phases, and the relaxation used in each phase is a covering program. On the other hand, our algorithm works with primal and dual programs that do have negative coefficients.

Despite this added complexity, our algorithm has a particularly simple description: Each city j keeps raising its dual variable, α_j, until it gets connected to an open facility. All other primal and dual variables simply respond to this change, trying to maintain feasibility or satisfying complementary slackness conditions. For the latter, we give a new mechanism as well.

Until the work of [30] (which relaxed the dual program itself), all approximation algorithms based on the primal-dual schema used the mechanism formalized in [36]: In the first phase, an integral primal solution is found, satisfying the primal complementary slackness conditions; however, this solution may have redundancies. In the second phase, a minimal solution is extracted, typically via a reverse delete procedure, and in the process, dual complementary slackness conditions get satisfied with a relaxation factor. The final algorithm has this factor as its approximation guarantee.

Our first phase is similar. For the second phase, we introduce the new procedure of *forward include* for removing redundancies. After this procedure is done, *all* complementary slackness conditions are satisfied; however, the primal solution may be infeasible. The solution is augmented – this time the primal conditions need to be relaxed by a factor of 3, which is also the approximation guarantee of the algorithm.

The k-median problem also has numerous applications, especially in the context of clustering, and has also been extensively studied. In recent years, ths problem has found new clustering applications in the area of data mining (see [7]).

A non-trivial approximation algorithm for this problem eluded researchers for many years. The breakthrough was made by Bartal, who gave a factor $O(\log n \log\log n)$ algorithm. After an improvement [8], a constant factor algorithm, using a different approach, was recently obtained by Charikar, Guha, Tardos and Shmoys [9]. Their algorithm has an approximation guarantee of $6\frac{2}{3}$; however, it has the same drawback since it uses LP-rounding. Their algorithm uses several ideas from the constant factor algorithms obtained for the facility location problem, thus making one wonder if there is a deeper connection between the two problems.

In this paper, we establish such a connection: between the LP-relaxations for the two problems. This enables us to use our algorithm for the facility location problem as a subroutine to solve the k-median problem. The idea for this lies in the following principle from economics: taxation is an effective way of controlling the amount of goods coming across the border – raising tariffs will reduce in-flow and vice versa. Given an instance of the k-median problem, we remove the restriction that at most k facilities be opened, and instead assign a cost of z for opening each of the facilities, thus obtaining an instance of the facility location problem. By changing z, we can control the number of facilities opened by our facility location algorithm. Ideally, at this point, we would like to find a value of z for which the algorithm opens exactly k facilities. We do not know how to do this. Instead, we find two solutions for "close" values of z, one opening more than k facilities, and the other opening less. An appropriate convex combination of these solutions is found that opens, fractionally, exactly k facilities. Finally, using a randomized rounding procedure, this is converted into an integral solution, sacrificing a small multiplicative factor in the process. A derandomization of this procedure is also provided.

These ideas also help solve a common generalization of the two problems – in which facilities have costs, and in addition, there is an upper bound on the number of facilities that can be opened. We give a factor 6 approximation algorithm for this problem as well; the previous bound was 9.8 [9].

The *capacitated* facility location problem, in which each facility i can serve at most u_i cities, has no non-trivial approximation algorithms. Part of the problem is that all LP-relaxations known for this problem have unbounded integrality gap (see [32]). In Section 5 we give a factor 4 approximation algorithm for the variant in which each facility can be opened an unbounded number of times; if facility i is opened y_i times, it can serve at most $u_i y_i$ cities. A special case of this version, in which the capacities of all the facilities are assumed to be equal, is solved with factor 3 in [13], again using LP-rounding.

2. The metric uncapacitated facility location problem

The *uncapacitated facility location problem* seeks a minimum cost way of connecting cities to open facilities. It can be stated formally as follows: Let G be a bipartite graph with bipartition (F, C), where F is the set of *facilities* and C is the set of *cities*. Let f_i be the cost of opening facility i, and c_{ij} be the cost of connecting city j to (opened) facility i. The problem is to find a subset $I \subseteq F$ of facilities that should be opened, and a function $\phi : C \rightarrow I$ assigning cities to open facilities in such a way that the total cost of opening facilities and connecting cities to open facilities is minimized. We will consider the *metric* version of this problem, i.e., the c_{ij}'s satisfy the triangle inequality.

We will adopt the following notation: $|C| = n_c$ and $|F| = n_f$. The total number of vertices $n_c + n_f = n$ and the total number of edges $n_c \times n_f = m$.

Following is an integer program for this problem. In this program, y_i is an indicator variable denoting whether facility i is open, and x_{ij} is an indicator variable denoting whether city j is connected to the facility i. The first constraint ensures