Discrete Algorithms Seminar

List of Presentations

Spring 2008

January 30, 2008

Title: "Fast and Reliable Reconstruction of Phylogenetic Trees with Very Short Edges."

Authors: Ilan Gronau, Shlomo Moran, and Sagi Snir.

Journal or Conference: SODA 2008: 379–388.

Abstract: Phylogenetic reconstruction is the problem of reconstructing an evolutionary tree from sequences corresponding to leaves of that tree. A central goal in phylogenetic reconstruction is to be able to reconstruct the tree as accurately as possible from as short as possible input sequences. The sequence length required for correct topological reconstruction depends on certain properties of the tree, such as its depth and minimal edge-weight. Fast converging reconstruction algorithms are considered state-of the-art in this sense, as they require asymptotically minimal sequence length in order to guarantee (with high probability) correct topological reconstruction of the entire tree. However, when the original phylogenetic tree contains very short edges, this minimal sequence-length is still too long for practical purposes. Short edges are not only very hard to reconstruct; their presence may also prevent the correct reconstruction of long edges.

Link Paper: `http://www.cs.technion.ac.il/~ilangr/papers/short_edges_SODA08.pdf`

Speaker: Andres Varon.

February 6, 2008

Title: "Rent, Lease or Buy: Randomized Algorithms for Multislope Ski Rental."

Authors: Zvi Lotker , Boaz Patt-Shamir , and Dror Rawitz.

Journal or Conference: STACS 2008.

Abstract: In the Multislope Ski Rental problem, the user needs a certain resource for some unknown period of time. To use the resource, the user must subscribe to one of several options, each of which consists of a one-time setup cost (.buying price.), and cost proportional to the duration of the usage (.rental rate.). The larger the price, the smaller the rent. The actual usage time is determined by an adversary, and the goal of an algorithm is to minimize the cost by choosing the best option at any point in time. Multislope Ski Rental is a natural generalization of the classical Ski Rental problem (where the only options are pure rent and pure buy), which is one of the fundamental problems of online computation. The Multislope Ski Rental problem is an abstraction of many problems where online decisions cannot be modeled by just two options, e.g., power management in systems which can be shut down in parts. In this paper we study randomized algorithms for Multislope Ski Rental. Our results include the best possible online randomized strategy for any additive instance, where the cost of switching from one option to another is the difference in their buying prices; and an algorithm that produces an e-competitive randomized strategy for any (non-additive) instance.

Link Paper: `http://www.eng.tau.ac.il/~boaz/cd/stacs-proc.pdf`

Speaker: Michael Lampis.

February 13, 2008

Title: "Ranking Tournaments: Local Search and a New Algorithm."

Authors: Tom Coleman and Anthony Wirth.

Journal or Conference: ALENEX 2008: 133–141.

Abstract: Ranking data is a fundamental organizational activity. Given advice, we may wish to rank a set of items to satisfy as much of that advice as possible. In the Feedback Arc Set (FAS) problem, advice takes the form of pairwise ordering statements, 'a should be ranked before b'. Instances in which there is advice about every pair of items is known as a tournament. This task is equivalent to ordering the nodes of a given directed graph to minimize the number of arcs pointing in one direction.

In the past, much work focused on finding good, effective heuristics for solving the problem. Recently, a proof of the NP-completeness of the problem (even when restricted to tournaments) has accompanied new algorithms with approximation guarantees, culminating in the development of a PTAS (polynomial time approximation scheme) for solving FAS on tournaments.

In this paper we re-examine many of these existing algorithms and develop some new techniques for solving FAS. The algorithms are tested on both synthetic and Rank Aggregation-based datasets. We find that, in practice, local-search algorithms are very powerful, even though we prove that they do not have approximation guarantees. Our new algorithm is based on reversing arcs whose nodes have large indegree differences, eventually leading to a total ordering. Combining this with a powerful local-search technique yields an algorithm that beats existing techniques on a variety of data sets.

Link Paper: `http://www.siam.org/proceedings/alenex/2008/alx08_013colemant.pdf`

Speaker: Simon Shamoun.

February 20, 2008

Title: "Broadcasting on Networks of Workstations."

Authors: Samir Khuller, Yoo-Ah Kim, and Yung-Chun (Justin) Wan.

Journal or Conference: SPAA 2005.

Abstract: Broadcasting and multicasting are fundamental operations. In this work we develop algorithms for performing broadcast and multicast in clusters of workstations. In this model, sending a message from one machine to another machine in the same cluster takes 1 time unit, and sending a message to a machine in a different cluster takes C time units. Lowekamp and Beguelin proposed heuristics for this model, but their algorithms may produce broadcast times that are arbitrarily worse than optimal. We develop the first constant factor approximation algorithms for this model. Algorithm LCF (Largest Cluster First) for the basic model is simple, efficient and has a worst case approximation guarantee of 2. We then extend these models to more complex models where we remove the assumption that an unbounded amount of simultaneous communication may happen using the global network. The algorithms for these models build on the LCF method developed for the basic problem. Finally, we develop broadcasting algorithms for the postal model where the sending processor does not block for C time units when the message is in transit. Moreover, we assume that the messages are small and so the bottleneck is the latency.

Link Paper: `www.cs.umd.edu/~samir/grant/kkw05.ps`

Speaker: Yosef Alayev.

February 27, 2008

Title: "On the approximation of optimal structures for RNA-RNA interaction."

Authors: Saad Mneimneh.

Journal or Conference: IEEE transactions on Computational Biology and Bioinformatics.

Abstract: The interaction of two RNA molecules is a common mechanism for many biological processes. Small interfering RNAs represent a simple example of such an interaction. But other more elaborate instances of RNA-RNA interaction exist. Therefore, algorithms that predict the structure of the RNA complex thus formed are of great interest. Most of the proposed algorithms are based on dynamic programming. RNA-RNA interaction is generally NP-complete; therefore, these algorithms (and other polynomial time algorithms for that matter) are not guaranteed to produce optimal structures. Our goal is to characterize this sub-optimality. We demonstrate the existence of constant factor approximation algorithms that are based on dynamic programming. In particular, we describe 1/2 and 2/3 factor approximation algorithms. These algorithms (and other algorithms in the literature) recursively divide the problem into two independent sub-problems. We define an *entangler* and prove that 2/3 is a theoretical upper bound on the approximation factor of algorithms that produce entangler-free solutions, e.g. the mentioned dynamic programming algorithms.

Link Paper:

`http://csdl.computer.org/dl/trans/tb/5555/01/ttb2007990111.pdf`
`http://engr.smu.edu/~saad/rnai.pdf`

Speaker: Valia Mitsou.

March 5, 2008

Title: "Graph Balancing: A Special Case of Scheduling Unrelated Parallel Machines."

Authors: Tomas Ebenlendr, Marek Krcal, and Jiri Sgall.

Journal or Conference: SODA 2008: 483–490.

Abstract: We design a 1.75-approximation algorithm for a special case of scheduling parallel machines to minimize the makespan, namely the case of restricted assignment where each job can be assigned to at most two machines. We also show that even for this special case it is NP-hard to compute better than 1.5 approximation. This is the first improvement of the approximation ratio 2 of Lenstra, Shmoys, and Tardos [Approximation algorithms for scheduling unrelated parallel machines, Math. Program. 46:259–271, 1990], for any special case with unbounded number of machines. Our lower bound yields the same ratio as their lower bound which works for restricted assignment, and which is still the state-of-the-art lower bound even for the most general case.

Link Paper:

`http://www.cs.dartmouth.edu/~theory/Papers/graphbalancingunrelatedproc.pdf`

Speaker: Matt Johnson.

March 12, 2008

Title: "Better Approximation of Betweenness Centrality."

Authors: Robert Geisberger, Peter Sanders, and Dominik Schultes.

Journal or Conference: ALENEX 2008.

Abstract: Estimating the importance or centrality of the nodes in large networks has recently attracted increased interest. Betweenness is one of the most important centrality indices, which basically counts the number of shortest paths going through a node. Betweenness has been used in diverse applications, e.g., social network analysis or route planning. Since exact computation is prohibitive for large networks, approximation algorithms are important. In this paper, we propose a framework for unbiased approximation of betweenness that generalizes a previous approach by Brandes. Our best new schemes yield significantly better approximation than before for many real world inputs. In particular, we also get good approximations for the betweenness of unimportant nodes.

Link Paper:

`http://www.siam.org/proceedings/alenex/2008/alx08_09geisbergerr.pdf`
`http://algo2.iti.uni-karlsruhe.de/schultes/hwy/betweennessFull.pdf`

Speaker: Marcin Szczodrak.

March 19, 2008

Title I: "Bounds on Multiprocessing Timing Anomalies."

Authors: Ron Graham.

Journal or Conference: SIAM Journal on Applied Mathematics, Vol. 17, No. 2. (Mar., 1969), pp. 416–429.

Abstract: It is well known to workers in the field of parallel computation that a multiprocessing system consisting of many identical processors acting in parallel may exhibit certain somewhat unexpected "anomalies," even though the system operates under a rather natural set of rules; e.g., it can happen that increasing the number of processors can increase the length of time required to execute a given set of tasks. In this paper we study a typical model of such a multiprocessing system, and we determine the precise extent by which the execution time for a set of tasks can be influenced because of these timing anomalies. A special case of this model will be shown to generate an interesting number theoretic question, partial answers to which are given in the latter half of the paper.

Link Paper:

`http://www.sci.brooklyn.cuny.edu/~philaris/scheduling/graham1969LPT.pdf`

Title II: "Worst-Case Analysis of a Placement Algorithm Related to Storage."

Authors: Ashok K. Chandra and C. K. Wong.

Journal or Conference: SIAM Journal of Computing, pp. 249–263, 1975.

Abstract: In this paper, a discrete minimization problem arising from storage allocation considerations is studied. Owing to the complexity of finding an optimum solution, a heuristic is proposed and its performance is analyzed. The worst-case ratio of the cost by this algorithm to that by the optimum algorithm is shown to lie between 1.03 and 1.04, implying that this algorithm produces a solution within 4 per cent of the optimum. A generalization of this problem to a class of cost functions is also considered. The worst-case ratios for these functions tend, in the limit, to that of the cost function studied by Graham in his classical paper [1].

Link Paper:

`http://naughtybag.wordpress.com/files/2008/03/chandrawong_loadbalancing_1975.pdf`

Speaker: Yi Feng.

April 2, 2008

Title: "Geometric Algorithms for Optimal Airspace Design and Air Traffic Controller Workload Balancing."

Authors: Amitabh Basuani, Joseph S. B. Mitchell, and Girishkumar Sabhnani.

Journal or Conference: ALENEX 2008: 75–89.

Abstract: The National Airspace System (NAS) is designed to accommodate a large number of flights over North America. For purposes of workload limitations for air traffic controllers, the airspace is partitioned into approximately 600 sectors; each sector is observed by one or more controllers. In order to satisfy workload limitations for controllers, it is important that sectors be designed carefully according to the traffic patterns of flights, so that no sector becomes overloaded. We formulate and study the airspace sectorization problem from an algorithmic point of view, modeling the problem of optimal sectorization as a geometric partition problem with constraints. The novelty of the problem is that it partitions data consisting of trajectories of moving points, rather than static point set partitioning that is commonly studied. First, we formulate and solve the 1d version of the problem, showing how to partition a line into "sectors" (intervals) according to historical trajectory data. Then, we apply the 1D solution framework to design a 2D sectorization heuristic based on binary space partitions. We also devise partitions based on balanced "pie partitions" of a convex polygon. We evaluate our 2D algorithms experimentally. We conduct experiments using actual historical flight track data for the NAS as the basis of our partitioning. We compare the workload balance of our methods to that of the existing set of sectors for the NAS and find that our resectorization yields competitive and improved workload balancing. In particular, our methods yield an improvement by a factor between 2 and 3 over the current sectorization in terms of the time-average and the worst-case workloads of the maximum workload sector. An even better improvement is seen in the standard deviations (over all sectors) of both time-average and worst-case workloads.

Link Paper: `http://www.siam.org/proceedings/alenex/2008/alx08_08basua.pdf`

Speaker: Deniz Sarioz.

April 9, 2008

Title: "Fault-tolerant relay node placement in wireless sensor networks: problems and algorithms."

Authors: Weiyi Zhang, Guoliang Xue, and Satyajayant Misra.

Journal or Conference: Infocom 2007.

Abstract: Two fundamental functions of the sensor nodes in a wireless sensor network are to sense its environment and to transmit sensed information to a base station. One approach to prolong sensor network lifetime is to deploy some relay nodes whose main function is to communicate with the sensor nodes, other relay nodes, and the base stations. It is desirable to deploy a minimum number of relay nodes to achieve certain connectivity requirement. In this paper, we study four related fault-tolerant relay node placement problems, each of which has been previously studied only in some restricted form. For each of them, we discuss its computational complexity and present a polynomial time O(1)-approximation algorithm with a small approximation ratio. When the problem reduces to a previously studied form, our algorithm either improves the previous best algorithm or reduces to the previous best algorithm.

Link Paper: `http://optimization.asu.edu/~xue/papers/INFOCOM2007-RelayNode.pdf`

Speaker: Rui Zhang.

April 16, 2008

Title: "Better bounds for online load balancing on unrelated machines."

Authors: Ioannis Caragiannis.

Journal or Conference: SODA 2008: 972–981.

Abstract: We study the problem of scheduling permanent jobs on unrelated machines when the objective is to minimize the Lp norm of the machine loads. The problem is known as load balancing under the Lp norm. We present an improved upper bound for the greedy algorithm through simple analysis; this bound is also shown to be best possible within the class of deterministic online algorithms for the problem. We also address the question whether randomization helps online load balancing under Lp norms on unrelated machines; this is a challenging question which is open for more than a decade even for the L2 norm. We provide a positive answer to this question by presenting the first randomized online algorithms which outperform deterministic ones under any (integral) Lp norm for p = 2,...,137. Our algorithms essentially compute in an online manner a fractional solution to the problem and use the fractional values to make random choices. The local optimization criterion used at each step is novel and rather counterintuitive: the values of the fractional variables for each job correspond to flows at an approximate Wardrop equilibrium for an appropriately defined non-atomic congestion game. As corollaries of our analysis and by exploiting the relation between the Lp norm and the makespan of machine loads, we obtain new competitive algorithms for online makespan minimization, making progress in another longstanding open problem.

Link Paper: `http://www.ceid.upatras.gr/caragian/soda08.pdf`

Speaker: Panos Hilaris.

April 30, 2008

Title: "Yet another algorithm for dense max-cut: Go Greedy."

Authors: Claire Mathieu and Warren Schudy.

Journal or Conference: SODA 2008.

Abstract: We study dense instances of MaxCut and its generalizations. Following a long list of existing, diverse and often sophisticated approximation schemes, we propose taking the nave greedy approach; we prove that when the vertices are considered in random order, our algorithms are still approximation schemes. Our algorithms may be simple, but the analysis is not. It relies on smoothing the vertices defining the partial cuts and on proving certain martingale properties. We also give a simpler proof of the result from Alon, Fernandez de la Vega, Kannan, and Karpinski [1] that dense problems have sample complexity $\tilde{O}(1/\epsilon^4)$. Like previous work, our results generalize to dense maximum constraint satisfaction problems.

Link Paper: `http://www.cs.brown.edu/~ws/papers/maxcut.pdf`

Speaker: Ou Liu.

May 07, 2008

Title: "An Experimental Study of Recent Hotlink Assignment Algorithms."

Authors: Tobias Jacobs.

Journal or Conference: ALENEX 2008.

Abstract: The concept of *hotlink assignment *aims at enhancing the structure of web sites such that the user's expected navigation effort is minimized. We concentrate on sites that are representable by trees and assume that each leaf carries a weight representing its popularity. The problem of optimally adding at most one additional outgoing edge ("hotlink") to each inner node has been widely studied. A considerable number of approximation algorithms have been proposed and worst-case bounds for the quality of the computed solutions have been given. However, only little is known about the practical behaviour of most of these algorithms yet. This paper contributes to close this gap by evaluating all recent strategies experimentally. Our experiments are based on trees extracted from real websites as well as on synthetic instances. The latter are generated by a new method that simulates the growth of a web site over time. We also propose a memory-efficient way to implement an optimal hotlink assignment algorithm, making it possible to compute optimal solutions for larger instances than before. Finally, we present a new approximation algorithm that is easy to implement and exhibits an excellent behaviour in practice.

Link Paper: `http://www.siam.org/proceedings/alenex/2008/alx08_014jacobst.pdf`

Speaker: Selcuk Cevher.