Discrete Algorithms Seminar

List of Presentations

Spring 2009

January 28, 2009

Title: "Overhang"

Authors: Mike Paterson and Uri Zwick

Journal or Conference: SODA 2006: 231–240

Abstract: How far off the edge of the table can we reach by stacking n identical blocks of length 1? A classical solution achieves an overhang of $\frac{1}{2}H_n$, where $H_n = \sum_{i=1}^{n} \frac{1}{i} \approx \ln n$ is the nth harmonic number, by stacking all the blocks one on top of another with the ith block from the top displaced by $\frac{1}{2i}$ beyond the block below. This solution is widely believed to be optimal. We show that it is exponentially far from optimal by giving explicit constructions with an overhang of $\Omega(n^{1/3})$. We also prove some upper bounds on the overhang that can be achieved. The stability of a given stack of blocks corresponds to the feasibility of a linear program and so can be efficiently determined.

Internet:

`http://arxiv.org/PS_cache/arxiv/pdf/0710/0710.2357v1.pdf`
`http://www.cs.tau.ac.il/~zwick/papers/overhang-SODA.pdf`

Speaker: Amotz Bar-Noy

February 4, 2009

Title: "A Cubic Kernel for Feedback Vertex Set"

Authors: Hans L. Bodlaender

Journal or Conference: STACS 2007: 320–331

Abstract: The Feedback Vertex Set problem on unweighted, undirected graphs is considered. Improving upon a result by Burrage et al. [7], we show that this problem has a kernel with $O(k^3)$ vertices, i.e., there is a polynomial time algorithm, that given a graph G and an integer k, finds a graph G' and integer k', such that G has a feedback vertex set of size at most k, if and only if G' has a feedback vertex set of size at most k', and G' has $O(k^3)$ vertices. Moreover, if G' has a feedback vertex set of size at most k', then a minimum size feedback vertex set of G' directly gives a minimum size feedback vertex set of size k'. This kernelization algorithm can be used as the first step of an FPT algorithm for Feedback Vertex Set, but also as a preprocessing heuristic for Feedback Vertex Set.

Internet: `http://webdoc.sub.gwdg.de/ebook/serien/ah/UU-CS/2006-042.pdf`

Speaker: Michail Lampis

February 11, 2009

Title: "How to Guard a Graph?"

Authors: Fedor V. Fomin, Petr A. Golovach, Alex Hall, Matus Mihalak, Elias Vicari, and Peter Widmayer

Journal or Conference: ISAAC 2008: 318–329

Abstract: We initiate the study of the algorithmic foundations of games in which a set of cops has to guard a region in a graph (or digraph) against a robber. The robber and the cops are placed on vertices of the graph; they take turns in moving to adjacent vertices (or staying). The goal of the robber is to enter the guarded region at a vertex with no cop on it. The problem is to find the minimum number of cops needed to prevent the robber from entering the guarded region. The problem is highly non-trivial even if the robber's or the cops' regions are restricted to very simple graphs. The computational complexity of the problem depends heavily on the chosen restriction. In particular, if the robber's region is only a path, then the problem can be solved in polynomial time. When the robber moves in a tree, then the decision version of the problem is NP-complete. Furthermore, if the robber is moving in a DAG, the problem becomes PSPACE-complete.

Internet: //

`http://www.springerlink.com/content/v04337631v088127/fulltext.pdf`
`ftp://ftp.inf.ethz.ch/pub/publications/tech-reports/6xx/605.pdf`

Speaker: Valia Mitsou

February 18, 2009

Title: "On the Approximability of Dodgson and Young Elections"

Authors: Ioannis Caragiannis, Jason A. Covey, Michal Feldman, Christopher M. Homan, Christos Kaklamanis, Nikos Karanikolas, Ariel D. Procaccia, and Jeffrey S. Rosenschein

Journal or Conference: SODA 2009: 1058–1067

Abstract: The voting rules proposed by Dodgson and Young are both designed to ?nd the alternative closest to being a Condorcet winner, according to two di?erent notions of proximity; the score of a given alternative is known to be hard to compute under either rule. In this paper, we put forward two algorithms for ap- proximating the Dodgson score: an LP-based randomized rounding algorithm and a deterministic greedy algorithm, both of which yield an O(logm) approximation ratio, where m is the number of alternatives; we observe that this result is asymptotically optimal, and further prove that our greedy algorithm is optimal up to a factor of 2, unless problems in NP have quasi-polynomial time algorithms. Although the greedy algorithm is computationally superior, we argue that the randomized rounding algorithm has an advantage from a social choice point of view. Further, we demonstrate that computing any reasonable approximation of the ranking produced by Dodgson's rule is NP-hard. This result provides a complexity-theoretic explanation of sharp discrepancies that have been observed in the Social Choice Theory literature when comparing Dodgson elections with simpler voting rules. Finally, we show that the problem of calculating the Young score is NP-hard to approximate by any factor. This leads to an inapproximability result for the Young ranking.

Internet: `http://www.siam.org/proceedings/soda/2009/SODA09_115_caragiannisi.pdf`

Speaker: Matt Johnson

February 25, 2009

Title: "Revisiting the TTL-based Controlled Flooding Search: Optimality and Randomization"

Authors: Nicholas Chang and Mingyan Liu

Journal or Conference: MOBICOM 2004: 85–99

Abstract: In this paper we consider the problem of searching for a node or an object (i.e., piece of data, file, etc.) in a large network. Applications of this problem include searching for a destination node in a mobile ad hoc network, querying for a piece of desired data in a wireless sensor network, and searching for a shared file in an unstructured peer-to-peer network. We limit our attention in this study to the class of controlled flooding search strategies where query/search packets are broadcast and propagated in the network until a preset TTL (time-to-live) value carried in the packet expires. Every unsuccessful search attempt results in an increased TTL value (i.e., larger search area) and the same process is repeated. The primary goal of this study is to derive search strategies (i.e., sequences of TTL values) that will minimize the cost of such searches associated with packet transmissions. The main results of this paper are as follows. When the probability distribution of the location of the object is known a priori, we present a dynamic programming formulation with which optimal search strategies can be derived that minimize the expected search cost. We also derive the necessary and sufficient conditions for two very commonly used search strategies to be optimal. When the probability distribution of the location of the object is not known a priori and the object is to minimize the worst-case search cost, we show that the best strategies are randomized strategies, i.e., successive TTL values are chosen from certain probability distributions rather than deterministic values. We show that given any deterministic TTL sequence, there exists a randomized version that has a lower worst-case expected search cost. We also derive an asymptotically (as the network size increases) optimal strategy within a class of randomized strategies.

Internet: `http://www.eecs.umich.edu/~mingyan/pub/mobicom04.pdf`

Speaker: Yi Feng

March 4, 2009

Title: "Four-Dimensional Hilbert Curves for R-Trees"

Authors: Herman Haverkort and Freek van Walderveen

Journal or Conference: ALENEX 2009: 63–73

Abstract: Two dimensional R-trees are a class of spatial index structures in which objects are arranged to enable fast window queries: report all objects that intersect a given query window. One of the most successful methods of arranging the objects in the index structure is based on sorting the objects according to the positions of their centers along a two-dimensional Hilbert space filling curve. Alternatively one may use the coordinates of the objects' bounding boxes to represent each object by a four-dimensional Hilbert-type curve. In experiments by Kamel and Faloutsos and by Arge et al. the first solution consistently outperformed the latter when applied to point data, while the latter solution clearly outperformed the first on certain artificial rectangle data. These authors did not specify which four-dimensional Hilbert-type curve was used; many exist.

In this paper we show that the results of the previous papers can be explained by the choice of the four-dimensional Hilbert-type curve that was used and by the way it was rotated in four-dimensional space. By selecting a curve that has certain properties and choosing the right rotation one can combine the strengths of the two-dimensional and the four-dimensional approach into one, while avoiding their apparent weaknesses. The effectiveness of our approach is demonstrated with experiments on various data sets. For real data taken from VLSI design, our new curve yields R-trees with query times that are better than those of R-trees that were obtained with previously used curves.

Internet: `http://www.siam.org/proceedings/alenex/2009/alx09_007_haverkorth.pdf`

Speaker: Andi Toce

March 11, 2009

Title: "Approximating the Throughput of Multiple Machines in Real-Time Scheduling"

Authors: Amotz bar-Noy, Sudipto Goha, Joseph (Seffi) Naor, and Baruch Scieber

Journal or Conference: SIAM Journal on Computing (SICOMP), 31(2):331–352, 2001

Abstract: We consider the following fundamental scheduling problem. The input to the problem consists of n jobs and k machines. Each of the jobs is associated with a release time, a deadline, a weight, and a processing time on each of the machines. The goal is to find a non preemptive schedule that maximizes the weight of jobs that meet their respective deadlines. We give constant factor approximation algorithms for four variants of the problem, depending on the type of the machines (identical vs. unrelated) and the weight of the jobs (identical vs. arbitrary). All these variants are known to be NP-hard, and the two variants involving unrelated machines are also MAX-SNP hard. The specific results obtained are as follows:

- For identical job weights and unrelated machines: a greedy 2-approximation algorithm.
- For identical job weights and k identical machines: the same greedy algorithm achieves a tight $\frac{(1+1/k)^k}{(1+1/k)^k-1}$ approximation factor.
- For arbitrary job weights and a single machine: an LP formulation achieves a 2-approximation for polynomially bounded integral input and a 3-approximation for arbitrary input. For unrelated machines, the factors are 3 and 4, respectively.
- For arbitrary job weights and k identical machines: the LP-based algorithm applied repeatedly achieves a $\frac{(1+1/k)^k}{(1+1/k)^k-1}$ approximation factor for polynomially bounded integral input and a $\frac{(1+1/2k)^k}{(1+1/2k)^k-1}$ approximation factor for arbitrary input.
- For arbitrary job weights and unrelated machines: a combinatorial $(3 + 2\sqrt{2} \approx 5.828)$-approximation algorithm.

Internet: `http://www.sci.brooklyn.cuny.edu/~amotz/PUBLICATIONS/schedule.pdf`

Speaker: Yosef Alayev

March 18, 2009

Title: "Finding Approximate Repetitions Under Hamming Distance"

Authors: Roman Kolpakov and Gregory Kucherov

Journal or Conference: Theoretical Computer Science (TCS), 303(1):135–156, 2003

Abstract: The problem of computing periodicities with K possible mismatches is studied. Two main definitions are considered, and for both of them an $O(nK \log K + S)$ algorithm is proposed (n the word length and S the size of the output). This improves, in particular, the bound obtained by G. Landau and J. Schmidt in 1993 (Proceedings of the Fourth Annual Symposium on Combinatorial Pattern Matching, Lecture Notes in Computer Science, Vol. 684, Springer, Berlin, Padova, Italy, pp. 120133). Finally, other possible definitions are briefly analyzed.

Internet: `http://www2.lifl.fr/~kucherov/PAPERS/KolpakovKucherovTCS03.pdf`

Speaker: Justin Tojeira

March 25, 2009

Title: "The Fragment Assembly String Graph"

Authors: Eugene W. Myers

Journal or Conference: European Conference on Computational Biology (Madrid, Spain, 2005), 79–85

Abstract: We present a concept and formalism, the string graph, that represents all that is inferable about a DNA sequence from a collection of shotgun sequencing reads collected from it. We give time and space efficient algorithms for constructing a string graph given the collection of overlaps between the reads and in particular, present a novel linear expected time algorithm for transitive reduction in this context. The result demonstrates that the decomposition of reads into k-mers employed in the de Bruijn graph approach of Pevzner et al. is not essential and in fact creates both efficiency problems and unnecessary conceptual complexities. The current paper is the first in a series and presents the basic algorithm and preliminary results that demonstrate the efficiency and scalability of the method. The result is a step toward a next-generation whole genome shotgun assembler that will easily scale to mammalian genomes.

Internet: `http://research.janelia.org/myers/Papers/string.graph.pdf`

Speaker: Yupu Liang

April 1, 2009

Title: "Loopless Generation of Multiset Permutations using a Constant Number of Variables by Prefix Shifts"

Authors: Aaron Williams

Journal or Conference: SODA 2009: 987–996

Abstract: This paper answers the following mathematical question: Can multiset permutations be ordered so that each permutation is a prefix shift of the previous permutation? Previously, the answer was known for the permutations of any set, and the permutations of any multiset whose corresponding set contains only two elements. This paper also answers the following algorithmic question: Can multiset permutations be generated by a loopless algorithm that uses sublinear additional storage? Previously, the best loopless algorithm used a linear amount of additional storage. The answers to these questions are both yes.

Internet: `http://www.siam.org/proceedings/soda/2009/SODA09_107_williamsa.pdf`

Speaker: Benjamin Baumer

April 22, 2009

Title I: "AdWords and Generalized Online Matching"

Authors I: Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani

Journal or Conference I: JACM 54(5), 2007 (also FOCS 2005)

Abstract I: How does a search engine company decide what ads to display with each query so as to maximize its revenue? This turns out to be a generalization of the online bipartite matching problem. We introduce the notion of a tradeoff revealing LP and use it to derive an optimal algorithm achieving a competitive ratio of $1 - 1/e$ for this problem.

Internet I: `http://www.stanford.edu/~saberi/adwords.pdf`

Title II: "Online Budgeted Matching in Random Input Models with Applications to Adwords"

Authors II: Gagan Goel and Aranyak Mehta

Journal or Conference II: SODA 2008: 982–991

Abstract II: We study an online assignment problem, motivated by Adwords Allocation, in which queries are to be assigned to bidders with budget constraints. We analyze the performance of the Greedy algorithm (which assigns each query to the highest bidder) in a randomized input model with queries arriving in a random permutation. Our main result is a tight analysis of Greedy in this model showing that it has a competitive ratio of 1 - 1/e for maximizing the value of the assignment. We also consider the more standard i.i.d. model of input, and show that our analysis holds there as well. This is to be contrasted with the worst case analysis of [MSVV05] which shows that Greedy has a ratio of 1/2, and that the optimal algorithm presented there has a ratio of 1 - 1/e. The analysis of Greedy is important in the Adwords setting because it is the natural allocation algorithm for an auction-style process.

From a theoretical perspective, our result simplifies and generalizes the classic algorithm of Karp, Vazirani and Vazirani for online bipartite matching. Our results include a new proof to show that the Ranking algorithm of [KVV90] has a ratio of 1 - 1/e in the worst case. It has been recently discovered [KV07] (independent of our results) that one of the crucial lemmas in [KVV90], related to a certain reduction, is incorrect. Our proof is direct, in that it does not go via such a reduction, which also enables us to generalize the analysis to our online assignment problem.

Internet II: `http://www.cc.gatech.edu/~gagang/onlineBudgetedMatching.pdf`

Speaker: Shoshana Neuburger

April 29, 2009

Title: "Physical Search Problems Applying Economic Search Models"

Authors: Yonatan Aumann, Noam Hazon, Sarit Kraus, and David Sarne

Journal or Conference: AAAI 2008: 9–16

Abstract: This paper considers the problem of an agent searching for a resource or a tangible good in a physical environment, where at each stage of its search it observes one source where this good can be found. The cost of acquiring the resource or good at a given source is uncertain (a-priori), and the agent can observe its true value only when physically arriving at the source. Sample applications involving this type of search include agents in exploration and patrol missions (e.g., an agent seeking to find the best location to deploy sensing equipment along its path). The uniqueness of these settings is that the expense of observing the source on each step of the process derives from the last source the agent explored. We analyze three variants of the problem, differing in their objective: minimizing the total expected cost, maximizing the success probability given an initial budget, and minimizing the budget necessary to obtain a given success probability. For each variant, we first introduce and analyze the problem with a single agent, either providing a polynomial solution to the problem or proving it is NP-Complete. We also introduce an innovative fully polynomial time approximation scheme algorithm for the minimum budget variant. Finally, the results for the single agent case are generalized to multi-agent settings.

Internet: `http://u.math.biu.ac.il/~sarit/Articles/aumanetalaaai08.pdf`

Speaker: Erdal Kose

May 6, 2009

Title: "Optimally Competitive List Batching"

Authors: Wolfgang W. Bein, Leah Epstein, Lawrence L. Larmore, and John Noga

Journal or Conference: SWAT 2004: 77–89

Abstract: Batching has been studied extensively in the offline case, but many applications such as manufacturing or TCP acknowledgement require online solutions. We consider online batching problems, where the order of jobs to be batched is fixed and where we seek to minimize average flow time. We present optimally competitive algorithms for s-batch (competitive ratio 2) and p-batch problems (competitive ratio of 4). We also derive results for naturally occurring special cases. In particular, we consider the case of unit processing times.

Internet:

`http://www.springerlink.com/content/gmn1can7xpr20ppc/fulltext.pdf`
`http://arxiv.org/PS_cache/arxiv/pdf/0904/0904.1002v1.pdf`

Speaker: Ali Assarpour