Discrete Algorithms Seminar

List of Presentations

Fall 2009

September 2, 2009

Title: "Combinatorial Optimization on Graphs of Bounded Treewidth"

Authors: Hans L. Bodlaender and Arie M. C. A. Koster

Journal or Conference: Computer Journal, 51(3):255–269, 2008

Abstract: There are many graph problems that can be solved in linear or polynomial time with a dynamic programming algorithm when the input graph has bounded treewidth. For combinatorial optimization problems, this is a useful approach for obtaining fixed-parameter tractable algorithms. Starting from trees and series-parallel graphs, we introduce the concepts of treewidth and tree decompositions, and illustrate the technique with the Weighted Independent Set problem as an example. The paper surveys some of the latest developments, putting an emphasis on applicability, on algorithms that exploit tree decompositions, and on algorithms that determine or approximate treewidth and find tree decompositions with optimal or close to optimal treewidth. Directions for further research and suggestions for further reading are also given.

Internet: `http://citeseerx.ist.psu.edu /viewdoc/download?doi=10.1.1.107.2561&rep=rep1&type=pdf`

Speaker: Michael Lampis

September 9, 2009

Title: "Logic, Graphs, and Algorithms"

Authors: Martin Grohe

Journal or Conference: Electronic Colloquium on Computational Complexity (ECCC), 14(091), 2007

Abstract: Algorithmic meta theorems are algorithmic results that apply to whole families of combinatorial problems, instead of just specific problems. These families are usually defined in terms of logic and graph theory. An archetypal algorithmic meta theorem is Courcelle's Theorem, which states that all graph properties definable in monadic second-order logic can be decided in linear time on graphs of bounded tree width. This article is an introduction into the theory underlying such meta theorems and a survey of the most important results in this area.

Internet: `http://eccc.hpi-web.de/eccc-reports/2007/TR07-091/index.html`

Speaker: Michael Lampis

September 16, 2009

Title: "Minimum Manhattan Network is NP-Complete"

Authors: Francis Y. L. Chin, Zeyu Guo, and He Sun

Journal or Conference: SoCG 2009: 393–402

Abstract: A rectilinear path between two points $p, q \in R^2$ is a path connecting p and q with all its line segments horizontal or vertical segments. Furthermore, a Manhattan path between p and q is a rectilinear path with its length exactly $dist(p, q) := |p.x - q.x| + |p.y - q.y|$. Given a set T of n points in R^2, a network G is said to be a Manhattan network on T, if for all $p, q \in T$ there exists a Manhattan path between p and q with all its line segments in G. For the given point set T, the Minimum Manhattan Network (MMN) Problem is to find a Manhattan network G on T with the minimum network length. In this paper, we shall prove that the decision version of MMN is strongly NP-complete, using the reduction from the well-known 3-SAT problem, which requires a number of gadgets. The gadgets have similar structures, but play different roles in simulating the 3-SAT formula. The reduction has been implemented with a computer program.

Internet: `http://i.cs.hku.hk/~chin/paper/mmn-scg.pdf`

Speaker: Yi Feng

September 23, 2009

Title: "Computation of Perfect DCJ Rearrangement Scenarios with Linear and Circular Chromosomes"

Authors: Sèverine Bérard, Annie Chateau, Cedric Chauve, Christophe Paul, and Eric Tannier

Journal or Conference: RECOMB-CG 2008: 158–169

Abstract: We study the problem of transforming a multichromosomal genome into another using Double Cut-and-Join (DCJ) operations, which simulates several types of rearrangements, as reversals, translocations, and block-interchanges. We introduce the notion of a DCJ scenario that does not break families of common intervals (groups of genes colocalized in both genomes). Such scenarios are called perfect, and their properties are well known when the only considered rearrangements are reversals. We show that computing the minimum perfect DCJ rearrangement scenario is NP-hard, and describe an exact algorithm which exponential running time is bounded in terms of a specific pattern used in the NP-completeness proof. The study of perfect DCJ rearrangement leads to some surprising properties. The DCJ model has often yielded algorithmic problems which complexities are comparable to the reversal-only model. In the perfect rearrangement framework, however, while perfect sorting by reversals is NP-hard if the family of common intervals to be preserved is nested, we show that finding a shortest perfect DCJ scenario can be answered in polynomial time in this case. Conversely, while perfect sorting by reversals is tractable when the family of common intervals is weakly separable, we show that the corresponding problem is still NP-hard in the DCJ case. This shows that despite the similarity of the two operations, easy patterns for revervals are hard ones for DCJ, and vice versa.

Internet: `http://www.cecm.sfu.ca/~cchauve/Publications/JCB_RCG08_1.pdf`

Speaker: Andres Varon

September 30, 2009

Title: "The Alcuin Number of a Graph"

Authors: Péter Csorba, Cor A. J. Hurkens, and Gerhard J. Woeginger

Journal or Conference: ESA 2008: 320–331

Abstract: We consider a planning problem that generalizes Alcuin's river crossing problem (also known as: The wolf, goat, and cabbage puzzle) to scenarios with arbitrary conflict graphs. We derive a variety of combinatorial, structural, algorithmical, and complexity theoretical results around this problem.

Internet: `http://www.renyi.hu/~csp/alcuin.pdf`

Speaker: Valia Mitsou

October 14, 2009

Title: "Matching Point Sets with Respect to the Earth Movers Distance"

Authors: Sergio Cabello, Panos Giannopoulos, Christian Knauer, and Günter Rote

Journal or Conference: Computational Geometry: 39 (2008) 118-133

Abstract: The Earth Mover's Distance (EMD) between two weighted point sets (point distributions) is a distance measure commonly used in computer vision for color-based image retrieval and shape matching. It measures the minimum amount of work needed to transform one set into the other one by weight transportation. We study the following shape matching problem: Given two weighted point sets A and B in the plane, compute a rigid motion of A that minimizes its Earth Mover's Distance to B. No algorithm is known that computes an exact solution to this problem. We present simple FPTASs and polynomial-time $(2+\epsilon)$-approximation algorithms for the minimum Euclidean EMD between A and B under translations and rigid motions.

Internet: `http://www.imfm.si/preprinti/PDF/01003.pdf`

Speaker: Simon Shamoun

October 21, 2009

Title: "Selfish Bin Packing"

Authors: Leah Epstein and Elena Kleiman

Journal or Conference: ESA 2008: 368–380

Abstract: Following recent interest in the study of computer science problems in a game theoretic setting, we consider the well known bin packing problem where the items are controlled by selfish agents. Each agent is charged with a cost according to the fraction of the used bin space its item requires. That is, the cost of the bin is split among the agents, proportionally to their sizes. Thus, the selfish agents prefer their items to be packed in a bin that is as full as possible. The social goal is to minimize the number of the bins used. The social cost in this case is therefore the number of bins used in the packing.

A pure Nash equilibrium is a packing where no agent can obtain a smaller cost by unilaterally moving his item to a different bin, while other items remain in their original positions. A Strong Nash equilibrium is a packing where there exists no subset of agents, all agents in which can profit from jointly moving their items to different bins. We say that all agents in a subset profit from moving their items to different bins if all of them have a strictly smaller cost as a result of moving, while the other items remain in their positions.

We measure the quality of the equilibria using the standard measures PoA and PoS that are defined as the worst case worst/best asymptotic ratio between the social cost of a (pure) Nash equilibrium and the cost of an optimal packing, respectively. We also consider the recently introduced measures SPoA and SPoS, that are defined similarly to the PoA and the PoS, but consider only Strong Nash equilibria.

We give nearly tight lower and upper bounds of 1.6416 and 1.6428, respectively, on the PoA of the bin packing game, improving upon previous result by Bilò, and establish the fact that PoS = 1. We show that the bin packing game admits a Strong Nash equilibrium, and that SPoA=SPoS. We prove that this value is equal to the approximation ratio of a natural greedy algorithm for bin packing.

Internet: `http://math.haifa.ac.il/lea/sbp.pdf`

Speaker: George Rabanca

October 28, 2009

Title: "Bounded Delay Packet Scheduling in a Bounded Buffer"

Authors: Stanley P. Y. Fung

Journal or Conference: CoRR abs/0907.2741: (2009)

Abstract: We study the problem of buffer management in QoS-enabled network switches in the bounded delay model where each packet is associated with a weight and a deadline. We consider the more realistic situation where the network switch has a finite buffer size. A 9.82-competitive algorithm is known for the case of multiple buffers (Azar and Levy, SWAT'06). Recently, for the case of a single buffer, a 3-competitive deterministic algorithm and a 2.618-competitive randomized algorithm was known (Li, INFOCOM'09). In this paper we give a simple deterministic 2-competitive algorithm for the case of a single buffer.

Internet: `http://arxiv.org/PS_cache/arxiv/pdf/0907/0907.2741v1.pdf`

Speaker: Yosef Alayev

November 4, 2009

Title: "Finding, Counting and Listing all Triangles in Large Graphs, An Experimental Study"

Authors: Thomas Schank and Dorothea Wagner

Journal or Conference: WEA 2005: 606–609

Abstract: A rigorous experimental study of algorithms for counting and listing all triangles in large existing networks as well as generated graphs is presented. In the past, this fundamental graph problem has been studied intensively from a theoretical point of view. Recently, triangle counting has also become a widely used tool in network analysis. Due to the very large size of networks like the Internet, WWW or social networks, the efficiency of algorithms for triangle counting and listing is an important issue. The main intention of this work is to evaluate the practicability of triangle counting and listing in very large graphs with various degree distributions. We give a surprisingly simple enhancement of a well known algorithm that performs best, and makes triangle listing and counting in huge networks feasible.

Internet: `http://i11www.iti.uni-karlsruhe.de/extra/publications/sw-fclt-05_t.pdf`

Speaker: Shu-Yuan Wu

November 11, 2009

Title: "Non-Adaptive Fault Diagnosis for All-Optical Networks via Combinatorial Group Testing on Graphs"

Authors: Nicholas J. A. Harvey, Mihai Patrascu, Yonggang Wen, Sergey Yekhanin, and Vincent W. S. Chan

Journal or Conference: INFOCOM 2007: 697–705

Abstract: We consider the fault diagnosis problem in all-optical networks, focusing on probing schemes to detect faults. Our work concentrates on non-adaptive probing schemes, in order to meet the stringent time requirements for fault recovery. This fault diagnosis problem motivates a new technical framework that we introduce: group testing with graph-based constraints. Using this framework, we develop several new probing schemes to detect network faults. The efficiency of our schemes often depends on the network topology; in many cases we can show that our schemes are near-optimal by providing tight lower bounds.

Internet: `http://research.microsoft.com/en-us/um/people/yekhanin/papers/nonadapt.pdf`

Speaker: Jie Chu

November 18, 2009

Title: “An Improved Construction of Progression-Free Sets”

Authors: Michael Elkin

Journal or Conference: SODA 2010: 886–905

Abstract: The problem of constructing dense subsets S of $\{1, 2, \ldots, n\}$ that contain no arithmetic triple was introduced by Erdős and Turán in 1936. They have presented a construction with $|S| = \Omega(n^{\log_3 2})$ elements. Their construction was improved by Salem and Spencer, and further improved by Behrend in 1946. The lower bound of Behrend is

$$|S| = \Omega\left(\frac{n}{2^{2\sqrt{2}\sqrt{\log_2 n}} \cdot \log^{1/4} n}\right) .$$

Since then the problem became one of the most central, most fundamental, and most intensively studied problems in additive number theory. Nevertheless, no improvement of the lower bound of Behrend was reported since 1946.

In this paper we present a construction that improves the result of Behrend by a factor of $\Theta(\sqrt{\log n})$, and shows that

$$|S| = \Omega\left(\frac{n}{2^{2\sqrt{2}\sqrt{\log_2 n}}} \cdot \log^{1/4} n\right) .$$

In particular, our result implies that the construction of Behrend is not optimal. Our construction and proof are elementary and self-contained.

Internet: `http://www.siam.org/proceedings/soda/2010/SODA10_072_elkinm.pdf`

Speaker: Ben Baumer

November 25, 2009

Title: "Multi-Channel Scheduling Algorithms for Fast Aggregated Convergecast in Sensor Networks"

Authors: Amitabha Ghosh, Özlem Durmaz Incel, V.S. Anil Kumar, and Bhaskar Krishnamachari

Journal or Conference: MASS 2009:

Abstract: AbstractFast and periodic collection of aggregated data is of considerable interest for mission-critical and continuous monitoring applications in sensor networks. In the many-to-one communication paradigm, known as convergecast, we consider applications wherein data packets are aggregated at each hop along a tree-based routing topology and focus on maximizing the data collection rate at a sink node. The primary hindrance in maximizing the data collection rate is the presence of interfering links. In this work, we employ TDMA scheduling and exploit the benefits of multiple frequency channels by assigning them appropriately to the parents of the routing tree in order to mitigate the effects of interference. Our key result in the paper lies in proving that minimizing the schedule length for an arbitrarily deployed network (possibly even worst case) in the presence of multiple frequencies is NP-hard, and in designing approximation algorithms with provable performance guarantee for geometric networks. In particular, we design a constant factor approximation algorithm for networks modeled as unit disk graphs, where every node has a uniform transmission range, and a $\Delta(T) \log n$ approximation algorithm for general disk graphs, where nodes could have different transmission ranges; n is the number of nodes in the network and $\Delta(T)$ is the maximum node degree on a given routing tree T. We also prove that a constant factor approximation is achievable on unit disk graphs even for unknown routing topologies so long as the maximum degree of any node in the tree is bounded by a constant. To the best of our knowledge, these are the first approximation results on the optimal schedule length under multiple frequencies. Our other contributions are in showing that finding the minimum number of frequencies required to remove all the interfering links in an arbitrary network in NP-hard. We give an upper bound on the maximum number of such frequencies required and propose a polynomial time algorithm that minimizes the schedule length under this scenario. Finally, we evaluate our algorithms through simulations and show various trends in performance for different network parameters.

Internet: `http://anrg.usc.edu/~amitabhg/MultiChannelScheduling_MASS2009.pdf`

Speaker: Matt Johnson

December 2, 2009

Title: "Quincy: Fair Scheduling for Distributed Computing Clusters"

Authors: Michael Isard, Vijayan Prabhakaran, Jon Currey, Udi Wieder, Kunal Talwar and Andrew Goldberg

Journal or Conference: SOSP 2009: 261–276

Abstract: This paper addresses the problem of scheduling concurrent jobs on clusters where application data is stored on the computing nodes. This setting, in which scheduling computations close to their data is crucial for performance, is increasingly common and arises in systems such as MapReduce, Hadoop, and Dryad as well as many grid-computing environments. We argue that data-intensive computation benefits from a fine-grain resource sharing model that differs from the coarser semi-static resource allocations implemented by most existing cluster computing architectures. The problem of scheduling with locality and fairness constraints has not previously been extensively studied under this resource-sharing model.

We introduce a powerful and flexible new framework for scheduling concurrent distributed jobs with fine-grain resource sharing. The scheduling problem is mapped to a graph datastructure, where edge weights and capacities encode the competing demands of data locality, fairness, and starvation-freedom, and a standard solver computes the optimal online schedule according to a global cost model. We evaluate our implementation of this framework, which we call Quincy, on a cluster of a few hundred computers using a varied workload of data-and CPU-intensive jobs. We evaluate Quincy against an existing queue-based algorithm and implement several policies for each scheduler, with and without fairness constraints. Quincy gets better fairness when fairness is requested, while substantially improving data locality. The volume of data transferred across the cluster is reduced by up to a factor of 3.9 in our experiments, leading to a throughput increase of up to 40

Internet: `http://www.sigops.org/sosp/sosp09/papers/isard-sosp09.pdf`

Speaker: Juan Liu

December 9, 2009

Title: "Vector Bin Packing with Multiple-Choice"

Authors: Boaz Patt-Shamir and Dror Rawitz

Journal or Conference: CoRR abs/0910.5599: (2009)

Abstract: We consider a variant of bin packing called multiple-choice vector bin packing. In this problem we are given a set of items, where each item can be selected in one of several D-dimensional incarnations. We are also given T bin types, each with its own cost and D-dimensional size. Our goal is to pack the items in a set of bins of minimum overall cost. The problem is motivated by scheduling in networks with guaranteed quality of service (QoS), but due to its general formulation it has many other applications as well. We present an approximation algorithm that is guaranteed to produce a solution whose cost is about $\ln D$ times the optimum. For the running time to be polynomial we require $D = O(1)$ and $T = O(\log n)$. This extends previous results for vector bin packing, in which each item has a single incarnation and there is only one bin type. To obtain our result we also present a PTAS for the multiple-choice version of multidimensional knapsack, where we are given only one bin and the goal is to pack a maximum weight set of (incarnations of) items in that bin.

Internet: `http://arxiv.org/PS_cache/arxiv/pdf/0910/0910.5599v1.pdf`

Speaker: Ali Assarpour