Program

Monday May 25, 2015

9:30 - 10:00 *Registration and Coffee*

10:00 - 10:45 **Opening Remarks, Greetings**

10:45 - 11:00 *Coffee break*

11:00 - 11:30 **Michal Feldman -** *Combinatorial Auctions via Posted Prices*

We study anonymous posted price mechanisms for combinatorial auctions in a Bayesian framework. In a posted price mechanism, item prices are posted, then the consumers approach the seller sequentially in an arbitrary order, each purchasing her favorite bundle from among the unsold items at the posted prices. These mechanisms are simple, transparent and trivially dominant strategy incentive compatible (DSIC). We show that when agent preferences are fractionally subadditive (which includes all sub-modular functions), there always exist prices that, in expectation, obtain at least half of the optimal welfare. Our result is constructive and of a black-box nature: given black-box access to a combinatorial auction algorithm A, sample access to the prior distribution, and appropriate query access to the sampled valuations, we can compute, in polytime, prices that guarantee at least half of the expected welfare of A. As a corollary, we obtain the first poly-time (in n and m) constant-factor DSIC mechanism for Bayesian submodular combinatorial auctions, given access to demand query oracles. Our results also extend to valuations with complements, where the approximation factor degrades linearly with the level of complementarity.

Joint work with Nick Gravin and Brendan Lucier.

11:30 - 12:00 **Miklos Santha -** *On the complexity of trial and error for constraint satisfaction problems*

In this talk I describe some results in a trial and error model of computing that was introduced by Bei, Chen and Zhang in STOC 2013,Êand applied to some constraint satisfaction problems. In this model the input is hidden by an oracle which, for a candidate assignment,Êreveals some information about a violated constraint if the assignment is not satisfying.ÊIn a systematic study of constraint satisfaction problemsÊwe define several types of revealing oracles, and develop a transfer theorem for each type of the revealing oracle, under a broad class of parameters. To any hidden CSP with a specific type of revealing oracle, the transfer theorem associates another, potentially harder CSP in the normal setting,Êsuch that their complexities are polynomial time equivalent. This in principle transfers the study of a large class of hidden CSPs, possibly with a promise on the instances, to the study of CSPs in the standard setting. The transfer theorems can be applied to get polynomial-time algorithms or hardness results for hidden CSPs, including satisfaction problems, monotone graph properties, isomorphism problems, and the exact version of the Unique Games problem.Ê

Joint work with G. Ivanyos, R. Kulkarni, Y. Qiao and A. Sundaram.

12:00 - 12:30 **Irit Dinur -** *Old and new PCP constructions*

The PCP theorem (AS,ALMSS 1991) guarantees that every NP language has a Probabilistically Checkable Proof (PCP) system allowing a verifier to check a witness very efficiently using randomness, and allowing for small error. Most of the talk will not assume prior knowledge.

I will talk about some recent work in which we make (some) progress towards proving the so-called "sliding-scale conjecture". This is a conjecture of BGLR from 1993 about the tradeoff between the number of bits read from the PCP proof and the error of the verifier. Our work revisits older constructions and analyzes them using the more modern "modular-composition" approach.

Based on joint work with Prahladh Harsha and Guy Kindler.

12:30 - 14:00 *Lunch (To be found on Campus. Please see our list of food courts.)*

14:00 - 14:30 **Serge Abiteboul -** *Turning your digital self into a knowledge base*

A Web user today has his/her data and information distributed in a number of services that operate in silos. Computer wizards already know how to control their personal data to some extent. It is now becoming possible for everyone to do the same, and there are many advantages to doing so. Everyone should now be in a position to manage his/her personal information. Furthermore, we will argue that we should move towards personal knowledge bases and discuss advantages to do so. We will consider technical issues that are raised.

14:30 - 15:00 **Tova Milo -** *Managing General and Individual Knowledge in Crowdsourcing Applications*

Crowd mining frameworks combine general knowledge, which can refer to an ontology or information in a database, with individual knowledge obtained from the crowd, which captures habits and preferences. To account for such mixed knowledge, along with user interaction and optimization issues, such frameworks must employ a complex process of reasoning, automatic crowd task generation and result analysis. In this talk, we describe a generic architecture for crowd mining applications. This architecture allows us to examine and compare the components of existing crowdsourcing systems and point out extensions required by crowd mining. It also highlights new research challenges and potential reuse of existing techniques/components. We exemplify this for the OASSIS project and for other prominent crowdsourcing frameworks.

15:00 - 15:30 **Michel de Rougemont -** *Streaming Property Testing of Visibly Pushdown Languages*

In the context of language recognition, we demonstrate the superiority of streaming property testers against streaming algorithms and property testers, when they are not combined. Initiated by Feigenbaum a streaming property tester is a streaming algorithm recognizing a language under the property testing approximation: it must distinguish inputs of the language from those that are ϵ-far from it, while using the smallest possible memory (rather than limiting its number of input queries).

Our main result is a streaming ϵ-property tester for visibly pushdown languages with one-sided error using memory space $\text{poly}((\log n)/\epsilon)$, important for the recognition of regular unranked ordered trees.

This construction relies on a new (non-streaming) property tester for weighted regular languages based on a previous tester by Alon *et al.* We provide a simple application of this tester for streaming testing special cases of instances of visibly pushdown languages that are already hard for both streaming algorithms and property testers.

15:30 - 16:00 *Coffee break*

16:00 - 16:30 **Marc Renault -** *Paid Exchanges are Worth the Price*

We consider the list update problem as defined in the seminal work on competitive analysis by Sleator and Tarjan [1985]. In this problem, a sequence of requests, consisting of items to access in a linked list, is given. After an item is accessed it can be moved to any position forward in the list (towards the head of the list) at no cost (free exchange), and, at any time, any two adjacent items can be swapped at a cost of 1 (paid exchange). The cost to access an item is equal to its current position in the list. The goal is to dynamically rearrange the list so as to minimize the total cost (accrued from accesses and exchanges) over the request sequence.

We show a lower bound of 12/11 on the worst-case ratio between the performance of an (offline) optimal algorithm that can only perform free exchanges and that of an (offline) optimal algorithm that can perform both paid and free exchanges. This answers an outstanding question that has been open since 1996 [Reingold and Westbrook 1996].

This is a joint work with Alejandro López-Ortiz and Adi Rosén.

16:30 - 17:00 **Uri Zwick -** *An improved version of the Random-Facet pivoting rule for the simplex algorithm*

Random-Facet is a randomized pivoting rule for the simplex method used to solve linear programming (LP) problems. It was devised independently by Kalai and by Matousek, Sharir and Welzl. It solves any LP problem using a sub-exponential number of pivoting steps, making it the fastest known pivoting rule for the simplex algorithm. We present a slightly improved version of this pivoting rule. Using the new pivoting rule, a linear program comprised of $O(d)$ constraints in d variables can be solved using an expected number of $2^{O(\sqrt{d})}$ pivoting steps, improved from the previous bound of $2^{O(\sqrt{d \log d})}$.

Joint work with Thomas Dueholm Hansen.

Tuesday May 26, 2015

9:00 - 9:30 **Shiri Chechik -** *Approximate Distance Oracles with Improved Bounds*

The design of algorithms for fast shortest path retrieval has attracted much attention in recent years, and has been studied extensively both from the theoretical and practical perspectives. While classic shortest path algorithms, e.g., Dijkstra's algorithm, enable one to easily compute a shortest path in a given graph, such algorithms suffer from two main drawbacks. First, they require that the entire graph needs to be stored, resulting in high memory usage when the input graph is very dense. Second, the time required by such algorithms to answer a single query is linear in the number of nodes, much too slow for many applications.

One approach to overcome these drawbacks, and achieve a sub-linear query time, is to pre-process the input graph and construct an appropriate data structure. A distance oracle is a data structure that allows fast retrieval of a distance estimate for any pair of nodes.

In this talk I will discuss a new construction for distance oracles in general undirected weighted graphs, which obtains improved bounds over previous works (both in size and query time).

9:30 - 10:00 **Dana Ron -** *On Testing Dynamic Environments*

We initiate a study of testing dynamic environments, focusing on environment that evolve according to a given local rule. The testing task consists of checking whether the environment has indeed evolved from some initial configuration according to the known evolution rule. We focus on the temporal aspect of this computational problem, which is reflected in two requirements: (1) It is not possible to go Òback to the pastÓ and make a query concerning the environment at time t after making a query at time tÕ $> t$; (2) Only a small portion of the environment is inspected in each time slot.

We present some general observations, an extensive study of two special cases, two separation results, and a host of open problems. In this talk I will focus on one of the special cases: Evolution according to a one-dimensional movement, and show that it can be tested within a total number of queries that is independent of the size of the environment.

Joint work with Oded Goldreich.

10:00 - 10:30 **Ronitt Rubinfeld -** *Sampling Correctors*

In many situations, sample data is obtained from a noisy or imperfect source. In order to address such corruptions, we propose the methodology of sampling correctors. Such algorithms use the structure that the distribution is purported to have, in order to allow one to make "on-the-fly" corrections to samples drawn from probability distributions. These algorithms then act as filters between the noisy data and the end user. We show connections between sampling correctors, distribution learning algorithms, and distribution property testing algorithms. We show that these connections can be utilized to expand the applicability of known distribution learning and property testing algorithms as well as to achieve improved algorithms for those tasks.

Joint work with Clément Canonne and Themis Gouleakis.

10:30 - 11:00 *Coffee break*

11:00 - 11:30 **Benny Chor -** *Three Strikes at Protein Conservation*

We propose three different similarity measures for protein sequences and their underlying DNA sequences. We examine these measures in homologous proteins of five mammalian species: Human, rat, mouse, cow, and dog. We study functional enrichments with respect to these measures, and the mutual relations between them.

The talk will be self contained, and no prior biological knowledge is required.

Joint work with Eyal Cohen.

11:30 - 12:00 **Gregory Kucherov -** *Efficient index-based filtering for approximate sequence search*

In this talk, we first give a short introduction to filtering techniques for sequence search. We then briefly introduce FM-index, the most prominent representative of the family of succinct data structures supporting exact string search, which is largely used in modern bioinformatics applications. We further focus on FM-index-based filtering techniques, and illustrate them through the problem of computing "suffix-prefix overlaps" motivated by genome assembly problem.

Joint work with D.Tsur (Ben-Gurion University).

12:00 - 12:30 **Fabio Pardi -** *Distance-based methods in phylogenomics*

With the continuous growth of genome sequencing capabilities, phylogenetic tree inference is increasingly based on large collections of genomic alignments Ñ each alignment derived from the comparison of orthologous genomic regions across different species. We have entered the era of 'phylogenomics', the study of evolution at the genomic scale. Classical methods in phylogenomics are commonly classified in two categories: super-tree and super-alignment methods. Super-tree approaches infer a phylogenetic tree per input alignment and then try and combine the trees obtained into a larger phylogeny, while super-alignment methods first concatenate the input alignments, and then infer a tree from the resulting super-alignment, using standard methodology. Super-tree methods are usually considered to be more efficient, while super-alignment methods are more accurate.

In this talk I will introduce a third category Ñ distance-based phylogenomic methods Ñ where tree inference is based on a collection of distance matrices, one matrix of pairwise distances estimated per input alignment. This approach inherits the advantage in efficiency of supertree methods, without one of their main pitfalls: that of only focusing on the topology of the source trees, thus ignoring branch lengths (both in input and output). As an example of this approach, I will present a novel phylogenomic distance-based method, named ERaBLE (Evolutionary Rates and Branch Length Estimation), to estimate the branch lengths of a given reference topology, and the relative evolutionary rates of the genes employed in the analysis. Our experiments show that ERaBLE is very fast and fairly accurate, e.g., allowing to deal with the OrthoMaM database, composed of $6,953$ exons from up to 40 mammals.

12:30 - 14:00 *Lunch (To be found on Campus. Please see our list of food courts.)*

14:00 - 14:30 **Nati Linial -** *Random simplicial complexes and why we care*

Graphs are the perfect modeling tool in an unusually broad range of application domains. This is because so many complex systems are determined by pairwise interactions of the system's constituents. But there are so many complex systems where the underlying interactions are k-way for $k > 2$. What do we do then? In principle one could appeal to hypergraph theory. Unfortunately hypergraph theory is not nearly as developed as graph theory. There is a special kind of hypergraphs called simplicial complexes. Such hypergraphs have a very strong geometric and topological nature and offer therefore an access to a variety of geometric tools and ideas. About 10 years ago Roy Meshulam and I introduced a model of random d-dimensional simplicial complexes which for $d = 1$ coincides with $G(n, p)$ random graphs. In my talk I will provide a taste of this beautiful emerging theory.

I will mention joint papers with R. Meshulam, T. Luczak, L. Aronshtam and Y. Peled.

14:30 - 15:00 **Allan Borodin -** *Weakly submodular functions (or whatever you want to call it)*

We introduce a variant of submodular functions which we initially have called "weakly submodular functions". This class includes all monotone submodular (and monotone meta-submodular) functions as well as some "mildly" supermodular functions. We show that several natural functions belong to this class and relate our class to some other recent submodular function extensions. In particular, this class includes the max sum dispersion problem. We consider the optimization problem of maximizing a weakly submodular function subject to uniform and general matroid constraints. For a uniform matroid constraint, the "standard greedy algorithm" achieves a constant approximation ratio. More specifically, for any cardinality constraint p, the greedy algorithm has a constant approximation ratio bounded by $\alpha(p)$ that experimentally appears to be converging (from below) to 5.95 as p increases. For a general matroid constraint with with rank s, we prove that the local search algorithm has constant approximation ratio bounded by $\rho(s)$ which analytically is converging (from above) to 10.22 as s increases.

Joint work with Dai Tri Man Le and Yuli Ye.

15:00 - 15:30 **Yossi Azar -** *Speed Scaling in the Non-clairvoyant Model*

In recent years, there has been a growing interest in speed scaling algorithms, where a set of jobs need to be scheduled on a machine with variable speed so as to optimize the flow-times of the jobs and the energy consumed by the machine. A series of results have culminated in constant-competitive algorithms for this problem in the clairvoyant model, i.e., when job parameters are revealed on releasing a job (Bansal, Pruhs, and Stein, SODA 2007; Bansal, Chan, and Pruhs, SODA 2009). Our main contribution in this paper is the first constant-competitive speed scaling algorithm in the non clairvoyant model, which is typically used in the scheduling literature to model practical settings where job volume is revealed only after the job has been completely processed. Unlike in the clairvoyant model, the speed scaling problem in the non-clairvoyant model is non-trivial *even for a single job*. Our non-clairvoyant algorithm is defined by using the existing clairvoyant algorithm in a novel inductive way, which then leads to an inductive analytic tool that may be of independent interest for other online optimization problems. We also give additional algorithmic results and lower bounds for speed scaling on multiple identical parallel machines.

Joint work with N. Devanur, Z. Huang and D. Panigrahi.

15:30 - 16:00 *Coffee break*

16:00 - 16:30 **Micha Sharir -** *Incidence Geometry: The New Algebraic Revolution*

I will survey the immense progress in combinatorial and computational geometry in the past seven years, triggered by the infusion of techniques from algebraic geometry and algebra, as introduced by Guth and Katz and further developed by the community at large. This has led to solutions of very hard problems, with the crown jewel being a nearly complete solution to Erdos's 1946 problem on distinct distances in the plane.

In the short time that I will have I will only highlight some of the key developments, and demonstrate the new approach by a few examples.

16:30 - 17:00 **Doron Shaharabani -** *The Offset Filtration of Convex Objects*

We consider offsets of a union of convex objects. We aim for a filtration, a sequence of nested cell complexes, that captures the topological evolution of the offsets for increasing radii. We describe methods to compute a filtration based on the Voronoi diagram of the given convex objects. We prove that, in two and three dimensions, the size of the filtration is proportional to the size of the Voronoi diagram. Our algorithm runs in $\Theta(n \log n)$ in the 2-dimensional case and in expected time $O(n^{3+\epsilon})$, for any $\epsilon > 0$, in the 3-dimensional case. Our approach is inspired by alpha-complexes for point sets, but requires more involved machinery and analysis primarily since Voronoi regions of general convex objects do not form a good cover. We show by experiments that our approach results in a similarly fast and topologically more stable method compared to approximating the input by point samples.

Joint work with Dan Halperin and Michael Kerber.

20:00 - WORKSHOP DINNER

Wednesday May 27, 2015

9:00 - 9:30 **Amnon Ta-Shma -** *Fully Explicit Hitting Set Generators for Low-Degree Polynomials*

We give a fully explicit construction of hitting set generators for low-degree polynomials, with close to optimal parameters. Unlike the fully explicit construction by Lu (CCC'12), that achieves slightly better parameters, our construction is purely algebraic and does not rely on previous constructions. Our analysis relies on the Riemann-Roch theorem and the isolation lemma for sparse families.

9:30 - 10:00 **Moti Medina -** *Online Path Computation and Function Placement in SDNs*

We consider a task of serving requests that arrive in an online fashion in Software-Defined Networks (SDNs). Each request specifies a "routing and processing plan" for a stream (e.g., the stream arrives at a node s, then it needs to pass through a firewall, then undergo encryption, and leave through a node t). Each processing stage can be performed by a designated subset of servers in the system. The algorithm needs to either reject the request or to admit it into the system along with a detailed routing (a.k.a. "path computation") and processing assignment ("function placement"). Each request also specifies the communication bandwidth and the processing load it requires, and each component in the system (link or processor) has a specified capacity; a solution may not violate the capacity constraints.

In this work we first formalize the problem, and propose a new service model that allows us to cope with requests with *unknown duration*. The new service model extends the traditional accept/reject schemes with a new possible response of "stand by." Our main result is an online algorithm for path computation and function placement that guarantees, *in each time step*, goodput of at least $\Omega\left(\frac{\text{OPT}^*}{\log n}\right)$, where n is the system size and OPT^* is an upper bound on the maximal possible goodput. The guarantee holds assuming that requests ask for at most an $O(1/\log n)$-fraction of the capacity of any component in the system. However, the guarantee holds even though our algorithm serves requests in an all-or-nothing fashion using a single path without preemption, while OPT^* may serve request fractions, may split the allocation over multiple paths, and may arbitrarily preempt and resume service of requests.

This is a joint work with Guy Even and Boaz Patt-Shamir.

10:00 - 10:30 **Boaz Patt-Shamir -** *Distributed Algorithms for Approximate Weighted Shortest Paths*

We study small-message distributed algorithms (the CONGEST model) for the all-pairs shortest-paths problem. It is known that the problem cannot be solved with approximation ratio 1.5 in $o(n)$ time, and not with any non-trivial ratio in time $o(\sqrt{n})$. We present a deterministic algorithm for $(1+\epsilon)$-approximation in time $O((n \log n)/\epsilon^2)$. We also consider routing schemes that involve node relabeling. We show how to compute an $O(k)$-approximation, for any integer $k > 1$, in time $O((n^{1/2+1/k} + D) \cdot \text{polylog}(n))$, where D is the hop diameter of the network. This algorithm uses labels of $O(\log n)$ bits, which is asymptotically optimal.

Joint work with Christoph Lenzen.

10:30 - 11:00 *Coffee break*

11:00 - 11:30 **Pierre Fraigniaud -** *Randomized Local Network Computing*

We carry on investigating the line of research questioning the power of randomization for the design of distributed algorithms. In their seminal paper, Naor and Stockmeyer [STOC 1993] established that, in the context of network computing, in which all nodes execute the same algorithm in parallel, any construction task that can be solved locally by a randomized Monte-Carlo algorithm can also be solved locally by a deterministic algorithm. This result however holds in a specific context. In particular, it holds only for distributed tasks whose solutions can be locally checked by a deterministic algorithm. In this paper, we extend the result of Naor and Stockmeyer to a wider class of tasks. Specifically, we prove that the same derandomization result holds for every task whose solutions can be locally checked using a 2-sided error randomized Monte-Carlo algorithm. This extension finds applications to, e.g., the design of lower bounds for construction tasks which tolerate that some nodes compute incorrect values. In a nutshell, we show that randomization does not help for solving such resilient tasks.

Joint work with Laurent Feuilloley.

11:30 - 12:00 **Mor Baruch -** *Randomized Proof-Labeling Schemes*

Proof-labeling schemes, introduced by Korman, Kutten and Peleg [PODC 2005], are a mechanism to certify that a network configuration satisfies a given boolean predicate. Such mechanisms find applications in many contexts, for example, the design of fault-tolerant distributed algorithms. In a proof-labeling scheme, predicate verification consists of neighbors exchanging labels, whose contents depends on the predicate. In this paper, we introduce the notion of randomized proof-labeling schemes where messages are randomized and correctness is probabilistic. We show that randomization reduces label size exponentially while guaranteeing probability of correctness arbitrarily close to one. In addition, we present a novel label-size lower bound technique that applies to both deterministic and randomized proof-labeling schemes. Using this technique, we establish several tight bounds on the verification complexity of MST, acyclicity, connectivity, and longest cycle size.

This is a joint work with Pierre Fraigniaud and Boaz Patt-Shamir.

12:00 - 12:30 **Ran Canetti -** *Cryptographic Software Obfuscation and Applications*

Software obfuscation, namely making software unintelligible while preserving its functionality, has long been considered a futile and inherently ineffective concept. Yet, recent results in cryptography have demonstrated that, under appropriate assumptions on the hardness of certain algebraic problems, software can indeed be meaningfully and effectively obfuscated. This new prospect holds both great promise and great danger to cyberspace, and thus also to our society at large.

The talk will present the basic ideas behind software obfuscation, recent developments towards more efficient and more secure obfuscation, and a number of salient applications.

12:30 - 14:00 *Lunch (To be found on Campus. Please see our list of food courts.)*

14:00 - 14:30 **Noga Alon -** *Signrank and VC-dimension*

The signrank of an N by N matrix A with $\{1, -1\}$ entries is the minimum possible rank of a real matrix B in which every entry has the same sign as the corresponding entry of A. The VC-dimension of A is the maximum cardinality of a set of columns I of A so that for every subset J of I there is a row i of A so that $A_{i,j} = +1$ for all j in J and $A_{i,j} = -1$ for all j in $I - J$.

I will study the problem of estimating the maximum possible signrank of an N by N matrix with VC-dimension d and will mention its applications in Communication Complexity and Learning.

Joint work with Shay Moran and Amir Yehudayoff.

14:30 - 15:00 **Sophie Laplante -** *Relative Discrepancy Does Not Separate Information and Communication Complexity*

Does the information complexity of a function equal its communication complexity? We examine whether any currently known techniques might be used to show a separation between the two notions. Recently, Ganor et al. [2014] provided such a separation in the distributional setting for a specific input distribution μ. We show that in the nondistributional setting, the relative discrepancy bound they defined is, in fact, smaller than the information complexity, and hence it cannot be used to separate information and communication complexity. In addition, in the distributional case, we provide an equivalent linear program formulation for relative discrepancy and relate it to variants of the partition bound, resolving also an open question regarding the relation of the partition bound and information complexity. Last, we prove the equivalence between the adaptive relative discrepancy and the public-coin partition bound, which implies that the logarithm of the adaptive relative discrepancy bound is quadratically tight with respect to communication.

15:00 - 15:30 **Rotem Oshman -** *Using Information Theory to Understand Multi-Party Communication*

The advent of massively parallel computation on very large data sets ("big data") raises many interesting and fundamental questions in distributed computing and multi-party communication complexity: the cost of computation in systems like MapReduce and OpenMPI is largely dominated by communication and synchronization between the machines, so understanding these complexity measures can guide us in the development of better algorithms for parallel and cloud computing. In this talk I will discuss recent applications of tools and ideas from information theory to understand the time and communication complexity of various problems in synchronous parallel computation.

15:30 - 16:30 *Cakes and Farewell*

We are thankful to the sponsors of the 4th FILOFOCS workshop:

- The Israeli Ministry of Science, Technology and Space
- The CNRS (French National Center for Scientific Research)
- Universié Paris Diderot
- LIAFA
- Tel-Aviv University
- I-CORE ALGO: Israeli Center of Research Excellence in Algorithms
- Modas Project (an ERC project)
- ACC: Advanced Communications Center, Tel-Aviv University