Program

Wednesday May 2, 2012

9:00 - 9:15 **Opening Remarks**

9:15 - 10:00 **Guy Even -** *Analysis of LP-decoding and message passing decoding using local optimality*

We present a unified analysis framework that captures recent advances in the study of local-optimality characterizations for codes on graphs. These local-optimality characterizations are based on combinatorial structures embedded in the Tanner graph of the code. Local-optimality implies both maximum-likelihood (ML) optimality and linear-programming (LP) decoding optimality. Also, an iterative message-passing decoding algorithm is guaranteed to find the unique locally-optimal codeword, if one exists.

This proof technique is demonstrated by considering a definition of local optimality that is based on the simplest combinatorial structures in Tanner graphs, namely, paths of length h. Guarantees for successful decoding are obtained even when h exceeds the girth of the Tanner graph.

We will also discuss more advanced characterizations of local optimality that are based on skinny trees and d-trees in the computation tree.

Joint work with Nissim Halabi.

10:00 - 10:30 **Coffee break**

10:30 - 11:15 **Benny Chor -** *Genetic Code Symmetry and Efficient Design of GC-Constrained Coding Sequences*

Cloning of DNA sequences (40-80 bases) into phage display libraries using polymerase chain reaction (PCR) is a process whose efficiency is strongly affected by the distribution of G-C bases in the DNA sequence. As the genetic code is not one-to-one, there is some flexibility in choosing the DNA sequence coding for the given peptide (short protein sequence). The number of possible DNA sequences is exponential in the peptide length, so a computational problem naturally arises: Finding the DNA sequence whose parameters are optimal. We develop an efficient, linear time 'one pass' algorithm for this search problem. Our algorithm strongly relies on an interesting symmetry, which we observed in the standard genetic code. Most non-standard genetic codes examined possess this symmetry as well, yet some do not.

Joint work with Matan Gavish (TAU/HU/Stanford) and Amnon Peled (Hadassah)

11:15 - 12:00 **Fabio Pardi -** *Reconstructing evolutionary trees from distances*

Several popular methods for inferring evolutionary trees (or for hierarchical clustering) are based on a matrix of pairwise distances between species (or any kind of objects): the objective is to construct a tree with edge lengths so that the pairwise distances between the leaves in that tree are as close as possible to the input distances. In evolutionary biology, each of these distances is typically an estimate of the amount of change separating two species and is estimated from molecular sequences using probabilistic models of sequence evolution.

The fundamental step of distance-based tree reconstruction is to fit the edge lengths of a tree of fixed structure to the given distance estimates. This step implicitly depends on the variances assumed for these estimates. In this talk, I will discuss a number of tree reconstruction methods, showing my work on them and showing in particular how their properties

(such as their robustness to noisy data) are affected by the variance model they assume. I am currently investigating variance assumptions leading to objective functions that can be optimized very rapidly. This has the potential to lead to very fast and accurate tree reconstruction algorithms.

12:00 - 14:00 **Lunch (to be found in restaurants around...)**

14:00 - 14:45 **Adi Rosén -** *Online Computation with Advice*

We consider the model of online computation with advice whose purpose is to model and analyze online algorithms which have access to some partial information about the future. The model allows one to define a spectrum of algorithms going from classical online algorithms (no information about the future) to offline optimal ones (information defining the optimal action at any time). The aim is that of analyzing the interplay between the amount of information about the future available to the online algorithm, and the achievable competitive ratio.

We define the model and show its applicability by giving several results for some classical online problems such as Metrical Task Systems and the k-server problem. We also consider some connections to other models of online computation, and discuss several open problems.

Based on joint work with Yuval Emek, Pierre Fraigniaud and Amos Korman and with Marc Renault.

14:45 - 15:30 **Christian Konrad -** *Language and Graph Problems in the Streaming Model*

This talk is an overview talk about some recent results in streaming algorithms. Streaming algorithms process the input date via passes while using sublinear memory. We consider language membership problems as well as graph problems.

The talk starts out with the problem of recognizing well-parenthesized expressions in the streaming model [Magniez et al. STOC 2010]. Here, we discuss also the bridge to communication complexity that serves as a device for proving space lower bounds for streaming algorithms. Recognizing well-parenthesized expressions allows to check well-formedness of XML documents. This link gave rise to a recent work on the problem of validating XML documents in the streaming setting [Konrad, Magniez ICDT 2012] that we discuss subsequently.

In the second part of the talk, we consider graph problems in the semi-streaming model. Here, the input stream is a sequence of the edges of an input graph. We focus on the matching problem, and we briefly touch on the semi-matching problem. The difficulty of these problems relies strongly on the arrival order of the input stream. We place particular emphasis on the arrival order, and we discuss worst-case order as well as random order. Concerning bipartite graphs, we also consider vertex-based orderings of the edges wrt. the vertices of one bipartition.

We conclude with open problems in this research area.

15:30 - 16:00 **Coffee break**

16:00 - 16:45 **Michel de Rougemont -** *Approximate verification and Enumeration problems*

We consider Verification problems, such as the Equivalence of Automata, Probabilistic Automata, and MDP's (Markov Decision Processes). The enumeration problem consists in enumerating words which distinguish two Automata and strategies which distinguish two MDPs. As these problems are hard in some sense, we present probabilistic methods in the case of Probabilistic Automata and Approximate methods (in the sense of Property Testing) for Automata and MDP's.

Thursday May 3, 2012

9:15 - 10:00 **Boaz Patt-Shamir -** *Low-congestion distributed algorithms*

The traditional model for computing over a communication network (called "LOCAL") allows sending a message of arbitrary size in a single time step. This way, the time complexity is a measure of the locality of algorithms: saying that an algorithm runs in time T is equivalent, under the LOCAL model, to saying that the problem can be solved if each node learns about all its neighbors to distance at most T. Therefore, in this model any problem can be solved in time linear in the network diameter. While work on the LOCAL model has produced many interesting results, it is widely accepted that this model does not capture the true complexity of distributed computing. A better approximation of reality is the CONGEST model, where a link can carry only a bounded number of bits in a time step. Usually, it is assumed that message size is $O(\log n)$ bits, so that each message can carry a constant number of node IDs and values of polynomial magnitude. It turned out that in this model, many problems cannot be solved in $o(\mathrm{sqrt}(n))$ time, even in networks of diameter, say, $O(\log n)$. In this talk we review some known results in the CONGEST model, as well as some new progress directions.

10:00 - 10:30 **Coffee break**

10:30 - 11:15 **Amos Korman -** *Ant Colonies and Distributed Computing*

In this talk, I will discuss an ongoing research project whose goal is to initiate a systematic interdisciplinary methodology connecting distributed computing and biological ensembles. The high level goal is to demonstrate that applying results and ideas from the field of distributed computing can be useful for obtaining a better understanding of the behavior of large biological ensembles. This approach will be demonstrated by focusing on one of the most beautiful of biological cooperation systems - the ant colony.

From the perspective of theoretical distributed computing, I shall introduce models and concepts inspired by ants (and other biological ensembles) which will find interest from the purely theoretical point of view. Indeed, most research in distributed computing is motivated by computer networks applications. Unfortunately, even though distributed computing is extremely common in nature, there are very few works in distributed computing that study such phenomena. This project, which is motivated by a more scientific approach than an engineering one, aims to progress our knowledge in this direction. Indeed, the models I shall introduce respect the motivations of biological ensembles which may be quite different than computer networks.

From the perspective of biology, the approach will demonstrate that fundamental properties and concepts such as memory of ants and their communications can be viewed from an information theoretical point of view, and that tools from computer science can be used for their understandings. Indeed, the goal would be to accompany some of the theoretical results by suitable experiments on living ants. By doing so, we shall demonstrate that theoretical results in distributed computing can be used to make predictions in biology and possibly even provide evidence for the existence of structural properties of a biological organism (e.g., its memory capacity, its communication bandwidth etc.). This latter goal is ambitious and extremely intriguing.

11:15 - 12:00 **Pierre Fraigniaud -** *Distributed Decision*

This talk will survey our recent results on distributed decision, in various settings, including LOCAL computing (a.k.a. Linial model), shared memory wait-free computing, and computing

with mobile entities (e.g., robots, software agents). In all these setting, the processes or mobile entities have to collectively decide distributed languages, with the constraints that if the input configuration is in the language then all of them should output "yes", otherwise one of them should output "no". We shall discuss of the impact of non-determinism, of oracles, and of randomization. The surveyed results are recent results obtained in collaboration with Amos Korman, Andrzej Pelc, David Peleg, Sergio Rajsbaum, and Corentin Travers.

12:00 - 14:00 **Lunch (to be found in restaurants around...)**

14:00 - 14:45 **Amos Fiat -** *Envy, Greed, and some other Deadly Sins*

We will review recent work on envy free mechanisms and issues of price discrimination in particular. Price discrimination refers to any nonuniform pricing policy used by a firm with market power to maximize its profits. Price discrimination is profitable because consumers who value the good more are willing to pay more.

Price discrimination may have some drawbacks:

(1) Customer resentment: In a 28 month study, covering 50,000 customers, Anderson and Simester [2010] found that customers who felt cheated due to price discrimination "...react by making fewer subsequent purchases from the firm. The effect is largest among the firm's most valuable customers: those whose prior purchases were most recent and at the highest prices." Similar sentiment has been observed in queues, where people prefer longer waits and no queue jumping over shorter, but unfair, queues [Avi-Itzhak et al. 2007].

(2) It may violate price discrimination laws: The Robinson-Patman Act of 1936 is a US federal law that prohibits certain forms of price discrimination by requiring that the seller offer the same prices to customers at a given level of trade. The European Community Competition Law also forbids some forms of price discrimination [Geradin and Petit 2006].

Based on joint papers with Edith Cohen, Michal Feldman, Haim Kaplan, Stefano Leonardi, Svetlana Olonetsky, and Piotr Sankowski.

14:45 - 15:30 **Raghav Kulkarni -** *Fourier spectrum of Boolean functions, Sparsity-granularity Theorem, and its application*

The "sparsity" of a Boolean function is the number of its non-zero Fourier coefficients. The "granularity" of a Boolean function is the smallest integer k such that all its Fourier coefficients can be expressed as an integer multiple of $1/2^k$. Recently (2009) Gopalan, O'Donnell, Servedio, Shpilka, and Wimmer proved a somewhat surprising theorem about the Fourier spectrum of Boolean functions, which we call the "Sparsity-granularity Theorem". The theorem asserts that for any Boolean function, the logarithm of sparsity is roughly of the same order of magnitude as its granularity.

We will present some applications of the Sparsity-granularity Theorem. In particular, we will present:

1. the (old) application for property testing upper bounds (due to Gopalan et al.)
2. two (new) applications: a) lower bounding parity tree complexity of matroids b) log-rank conjecture for AC^0-XOR functions.

We will conclude with a bunch of open questions in search of more applications of the Sparsity-granularity Theorem.

The new results are joint work with Miklos Santha.

15:30 - 16:00 **Coffee break**

16:00 - 16:45 **Tova Milo -** *Mob Data Sourcing*

Crowd-based data sourcing is a new and powerful *data procurement* paradigm that engages Web users to collectively contribute data, analyze information and share opinions. Crowd-based data sourcing democratizes data-collection, cutting companies' and researchers' reliance on stagnant, overused datasets and bears great potential for revolutionizing our information world. Yet, triumph has so far been limited to only a handful of successful projects such as Wikipedia or IMDb. This comes notably from the difficulty of managing huge volumes of data and users of questionable quality and reliability. Every single initiative had to battle, almost from scratch, the same non-trivial challenges. The ad hoc solutions, even when successful, are application specific and rarely sharable. In this talk we consider the development of solid scientific foundations for Web-scale data sourcing. We believe that such a principled approach is essential to obtain knowledge of superior quality, to realize the task more effectively and automatically, be able to reuse solutions, and thereby to accelerate the pace of practical adoption of this new technology that is revolutionizing our life. We will consider the logical, algorithmic, and methodological foundations for the management of large scale crowd-sourced data as well as the the development of applications over such information.

16:45 - 17:30 **Serge Abiteboul -** *Viewing the Web as a Distributed Knowledge Base*

Information of interest may be found on the Web in a variety of forms, in many systems, and with different access protocols. A typical user may have information on many devices (smartphone, laptop, TV box, etc.), many systems (mailers, blogs, Web sites, etc.), many social networks (Facebook, Picasa, etc.). This same user may have access to more information from family, friends, associations, companies, and organizations. Today, the control and management of the diversity of data and tasks in this setting are beyond the skills of casual users. Facing similar issues, companies see the cost of managing and integrating information skyrocketing.

We are interested here in the management of such data. Our focus is not on harvesting all the data of a particular user or a group of users and then managing it in a centralized manner. Instead, we are concerned with the management of Web data in place in a distributed manner, with a possibly large number of autonomous, heterogeneous systems collaborating to support certain tasks.

Our thesis is that managing the richness and diversity of user-centric data residing on the Web can be tamed using a holistic approach based on a distributed knowledge base. All Web informations are represented as logical facts, and Web data management tasks as logical rules. We discuss Webdamlog, a variant of datalog for distributed data management that we use for this purpose. The automatic reasoning provided by its inference engine, operating over the Web knowledge base, greatly benefits a variety of complex data management tasks that currently require intense work and deep expertise.

This work is part of the Webdam ERC Project.

18:30 - 21:00 **Reception (cocktail dinatoire)** - on the ground floor of the IHP

Friday May 4, 2012

9:15 - 10:00 **Yossi Azar -** *Fast approximation algorithms for submodular optimization problems*

We consider three submodular optimization problems. For each of these problems we provide a different fast, combinatorial approximation algorithm. For the first problem, 'ranking with submodular valuations' our algorithm provides the best possible approximation. For the second problem, 'maximization submodular function under linear packing constraints', we match the best approximation provided by non-combinatorial algorithm. For the third problem, 'submodular Max-SAT', we get a non-trivial approximation in linear time.

Based on joint work with Iftah Gamzu and Ran Roth.

10:00 - 10:30 **Coffee break**

10:30 - 11:15 **Iordanis Kerenidis -** *Quantum and Classical Communication Complexity*

We will review some important results in quantum communication complexity, including the notion of quantum fingerprints, that provide an exponential separation between classical and quantum communication complexity in the simultaneous message model without public coins, and the Hidden Matching problem, that provides a separation in the bounded-error one-way communication model. We will also describe some of the main open questions in the field. No prior knowledge of quantum computing is necessary.

11:15 -12:00 **Miklos Santha -** *Quantum walks and learning graphs*

In this talk I will survey two generic methods to design quantum query algorithms which require only minimal knowledge of quantum computing. I will give an intuitive treatment of the discrete time quantization of classical Markov chains, and will describe the relatively new idea of learning graphs. With several examples I will try to illustrate classical successes and recent progresses.

12:00 - 14:00 **Lunch (to be found in restaurants around...)**

14:00 - 14:45 **Amnon Ta-Shma -** *Algebraic constructions of randomness extractors*

Extractors are boolean functions that allow, in some precise sense, extraction of randomness from somewhat random distributions, using only a small amount of truly random bits. Extractors, and the closely related dispersers, exhibit some of the most "random-like" properties of explicitly constructed combinatorial structures. Consequently, extractors have found many applications in CS.

The search for *explicit* constructions of good dispersers and extractors is at least 30 years old. The first constructions used hash functions. About 15 years ago Luca Trevisan showed a close connection between extractors and error correcting codes with good list decoding properties. Following that, many constructions that exploit this connection appeared, and many of them heavily used algebraic techniques that previously appeared in error correcting code constructions.

In the talk, I will present the problem and some of the main ideas underlying current constructions. I will also briefly present a new construction by Chris Umans and myself that employs "two levels" of the main component of Parvaresh Vardy codes, achieving the same "world-record" parameters as Dvir, Kopparty, Saraf and Sudan [DKSS] with a somewhat more direct construction.

14:45 - 15:30 **David Xiao -** *NP-hardness and cryptography*

Cryptography is one of the most successful contributions of theoretical computer science. It underlies much of modern telecommunication networks, providing secrecy and authentication on channels that are intrinsically insecure. By developing a rigorous theory of security, theoretical computer science has allowed people to use these insecure networks with the confidence that their information will be protected from eavesdroppers and hackers. Nevertheless, there is uncertainty regarding the computational hardness assumptions necessary to build cryptosystems. Many assumptions such as the hardness of factoring integers or the hardness of the RSA problem are vulnerable to sub-exponential time attacks as well as quantum attacks. One goal of theoretical cryptographers is to base the hardness of breaking cryptosystems on minimal assumptions. We already know that P different from NP is necessary to build secure cryptography. Ideally one would be able to show that this assumption is also sufficient, namely that if P does not equal NP, then one can build secure cryptosystems. In this talk, we will give an introduction to this area of research and present the state of the art in understanding this question. Roughly, this line of research indicates that basing cryptography on NP-hardness is not possible, and that cryptographic hardness is "significantly harder" than NP-hardness.

15:30 - 16:00 **Coffee break**

16:00 - 16:45 **Uri Zwick -** *Randomized pivoting rules for the simplex algorithm*

The simplex algorithm is one of the most widely used algorithms for solving linear programs in practice. Its worst case complexity, however, with essentially all known deterministic pivoting rules is exponential. There is, however, a randomized pivoting rule, called Random-Facet, discovered independently by Kalai and by Matousek, Sharir and Welzl, under which the expected running time of the simplex algorithm is subexponential. We obtain subexponential lower bounds for Random-Facet and two other randomized pivoting rules.

Joint work with Thomas Dueholm Hansen and Oliver Friedmann.