Talk titles and abstracts for the 2nd FILOFOCS workshop

Invited Speakers

Noga Alon - Easily testable graph properties

A graph on n vertices is epsilon-far from a property P if one has to add or delete from it at least $\epsilon \cdot n^2$ edges to get a graph satisfying P. A graph property is easily testable if it is possible to distinguish between graphs satisfying P and ones that are ϵ-far from P by inspecting the induced subgraph on a random subset of $poly(1/\epsilon)$ vertices. I will consider the problem of characterizing the easily testable graph properties, which is wide open, and report on some (modest) progress in its study.

Oded Goldreich - Finding Cycles and Trees in Sublinear Time

We present sublinear-time (randomized) algorithms for finding simple cycles of length at least $k \geq 3$ and tree-minors in bounded-degree graphs. The complexity of these algorithms is related to the distance of the graph from being C_k-minor free (resp., free from having the corresponding tree-minor). In particular, if the graph is far (i.e., $\Omega(1)$-far) from being cycle-free, then the algorithm finds a cycle of polylogarithmic length in time $\tilde{O}(\sqrt{N})$, where N denotes the number of vertices. This time complexity is optimal up to polylogarithmic factors.

The foregoing results are the outcome of our study of the complexity of *one-sided error* testers in the bounded-degree graphs model. For example, we show that cycle-freeness of N-vertex graphs can be tested with one-sided error within time complexity $\tilde{O}(poly(1/\varepsilon) \cdot \sqrt{N})$. This matches the known $\Omega(\sqrt{N})$ query lower bound, and contrasts with the fact that any minor-free property admits a *two-sided error* tester of query complexity that only depends on the proximity parameter ε. For any constant $k \geq 3$, we extend this result to testing whether the input graph has a simple cycle of length at least k. On the other hand, for any fixed tree T, we show that T-minor freeness has a one-sided error tester of query complexity that only depends on the proximity parameter ε.

Our algorithm for finding cycles in bounded-degree graphs extends to general graphs, where distances are measured with respect to the actual number of edges. Such an extension is not possible with respect to finding tree-minors in $o(\sqrt{N})$ complexity.

Joint work with Artur Czumaj, Dana Ron, C. Seshadhri, Asaf Shapira, and Christian Sohler.

Haim Kaplan - Submatrix maximum queries in Monge matrices and Monge partial matrices, and their applications

We describe a data structure for submatrix maximum queries in Monge matrices or partial Monge matrices, where a query seeks the maximum element in a contiguous submatrix of the given matrix. The structure, for an $n \times n$ Monge matrix, takes $O(n \log n)$ space, $O(n \log^2 n)$ preprocessing time, and answers queries in $O(\log^2 n)$ time. For partial Monge matrices the space and preprocessing grow by $\alpha(n)$ (the inverse Ackermann function), and the query remains $O(\log^2 n)$. Our design exploits an interpretation of the column maxima in a Monge (resp., partial Monge) matrix as an upper envelope of pseudo-lines (resp., pseudo-segments). We will also describe few applications of this data structure.

Nati Linial - Local combinatorics and why computer scientists should care

In many application areas we are dealing with very large graphs. Sometimes the graph under consideration is so large that we are not even able to store it on a computer. Computing hard graph parameters is completely out of question. What can we do in such situations? Which graph parameters should we extract and what should we do with them? Furthermore, is there any systematic way to model things so as to view the graph as coming from some simply defined distribution? At present I am unable to answer these questions, but in this talk I will report on some relevant ongoing work. This work revolves around studying the local structure of the graph. Specifically, for the practical questions mentioned above, the suggested approach is to look at the distribution of small subgraphs of the big graph. In particular I will mention the theory of graph limits due to Lovasz, Szegedy and collaborators. I will also mention Razborov's flag algebra. If time permits I will explain how the notion of local view extends to various other combinatorial objects.

Talks

Dana Ron - Exponentially improved algorithms and lower bounds for testing signed majorities

A signed majority function is a linear threshold function $f : \{+1, 1\}^n \to \{+1, 1\}$ of the form $f(x) = sign(\sum_{i=1}^n \sigma_i x_i)$ where each $\sigma_i \in \{+1, -1\}$. Signed majority functions are a highly symmetrical subclass of the class of all linear threshold functions, which are functions of the form $sign(\sum_{i=1}^n w_i x_i - \theta)$ for arbitrary real w_i, θ.

We study the query complexity of testing whether an unknown $f : \{+1, -1\}^n \to \{+1, -1\}$ is a signed majority function versus ϵ-far from every signed majority function. While it is known that the broader class of all linear threshold functions is testable with $poly(1/\epsilon)$ queries (independent of n), prior to our work the

best upper bound for signed majority functions was $O(\sqrt{n}) \cdot poly(1/\epsilon)$ queries (via a non-adaptive algorithm), and the best lower bound was $\Omega(\log n)$ queries for non-adaptive algorithms.

As our main results we exponentially improve both these prior bounds for testing signed majority functions:

We give a $poly(\log n, 1/\epsilon)$-query adaptive algorithm (which is computationally efficient) for this testing problem;

We show that any non-adaptive algorithm for testing the class of signed majorities to constant accuracy must make $n^{\Omega(1)}$ queries. This directly implies a lower bound of $\Omega(\log n)$ queries for any adaptive algorithm.

Our testing algorithm performs a sequence of restrictions together with consistency checks to ensure that each successive restriction is "compatible" with the function prior to restriction. This approach is used to transform the original n-variable testing problem into a testing problem over $poly(\log n, 1/\epsilon)$ variables where a simple direct method can be applied. Analysis of the degree-1 Fourier coefficients plays an important role in our proofs.

Iftach Haitner - Limits on the Usefulness of Random Oracles

In the **random oracle** model, parties are given oracle access to a random function (i.e., a uniformly chosen function from the set of all functions), and are assumed to have unbounded computational power (though they can only make a bounded number of oracle queries). This model provides powerful properties that allow proving the security of many protocols, even such that cannot be proved secure in the standard model (under any hardness assumption). The random oracle model is also used for showing that a given cryptographic primitive cannot be used in a black-box way to construct another primitive; in their seminal work, Impagliazzo and Rudich [STOC 89] showed that no key-agreement protocol exists in the random oracle model, yielding that key-agreement cannot be black-box reduced to one-way functions. Their work has a long line of followup works (Simon [EC 98], Gertner et al. [STOC 00] and Gennaro et al. [SICOMP 05], to name a few), showing that given oracle access to a certain type of function family (e.g., the family that implements public-key encryption) is not sufficient for building a given cryptographic primitive (e.g., oblivious transfer). Yet, the following question remained open:
What is the exact power of the random oracle model?

We make progress towards answering the above question, showing that, essentially, any no private input, semi-honest two-party functionality that can be securely implemented in the random oracle model, can be securely implemented information theoretically (where parties are assumed to be all powerful, and no oracle is given). We further generalize the above result to function families that provide some natural combinatorial property. Our result immediately yields that that the only no-input functionalities that can be securely realized in the random oracle model (in the sense of secure function evaluation), are the trivial ones (ones that can be securely realized information theoretically). In addition,

we use the recent information theoretic impossibility result of McGregor et al. [FOCS 10], to show the existence of functionalities (e.g., inner product) that cannot be computed both accurately and in a differentially private manner in the random oracle model; yielding that protocols for computing these functionalities cannot be black-box reduced to the existence of one-way functions.

Tova Milo - Crowd Mining

Harnessing a crowd of Web users for data collection has recently become a wide-spread phenomenon. A key challenge is that the human knowledge forms an open world and it is thus difficult to know what kind of information we should be looking for. Classic databases have addressed this problem by data mining techniques that identify interesting data patterns. These techniques, however, are not suitable for the crowd. This is mainly due to properties of the human memory, such as the tendency to remember simple trends and summaries rather than exact details. Following these observations, we develop here for the first time the foundations of crowd mining. We first define the formal settings. Based on these, we design a framework of generic components, used for choosing the best questions to ask the crowd and mining significant patterns from the answers. We suggest general implementations for these components, and test the resulting algorithm's performance on benchmarks that we designed for this purpose. Our algorithm consistently outperforms alternative baseline algorithms.

Benny Applebaum - Locally Computable Universal One-Way Hash Functions with Linear Shrinkage

We study the problem of constructing locally computable cryptographic hash functions which compress a long n-bit input string to a shorter m-bit output string. We present the first construction that achieves: (1) constant output locality, i.e., every bit of the output depends only on a constant number of bits of the input; (2) constant input locality, i.e., every bit of the input affects only on a constant number of bits of the output; and (3) linear shrinkage, i.e., $m = (1 - \epsilon)n$ for some constant $\epsilon > 0$.

Previous constructions gained only sub-linear shrinkage of $m - n = n^{1-\epsilon}$ and failed to achieve constant input locality. Our construction is based on the one-wayness of "random" local functions – a variant of an assumption made by Goldreich. As an application, we obtain a digital signature scheme with a minimal *additive* complexity overhead: signing n-bit messages with security parameter κ takes only $O(n + \kappa)$ time instead of $O(n\kappa)$ as in typical constructions. Previously, such signatures were only known to exist under an *exponential* hardness assumption.

Joint work with Yoni Moses.

Nathanael Francois - Streaming Complexity of Checking Priority Queues

Given a sequential access to the insert/extract operations on a data structure, one would like to decide, a posteriori only, if it corresponds to the evolution of a reliable structure. In a context of massive data, one would like to minimize both the amount of reliable memory of the checker and the number of passes on the sequence of operations. Chu, Kannan and McGregor initiated the study of checking priority queues in this setting. They showed that use of timestamps allows to check a priority queue with a single pass and memory space square root of N (up to a logarithmic factor). Later, Chakrabarti, Cormode, Kondapally and McGregor removed the use of timestamps, and proved that more passes do not help. We show that, even in the presence of timestamps, more passes do not help, solving an open problem of those previous works. On the other hand, we show that a second pass, but in reverse direction, allows polylogarithmic memory space, extending a phenomenon the first time observed by Magniez, Mathieu and Nayak for checking well-parenthesized expressions.

Amnon Ta-Shma - Inverting well-conditioned matrices in Quantum Logspace

We show that quantum computers improve on the the best known classical algorithms for matrix inversion (and singular value decomposition) as far as space is concerned. This adds to the (still short) list of important problems where quantum computers are of help.

Specifically, we show that the inverse of a well conditioned matrix can be approximated in quantum logspace with intermediate measurements. This should be compared with the best known classical algorithm for the problem that requires $\Omega(\log^2 n)$ space.

We also show how to approximate the spectrum of a normal matrix, or the singular values of an arbitrary matrix, with ε additive accuracy, and how to approximate the singular value decomposition of a matrix whose singular values are well separated.

The technique builds on ideas from several previous works, includingsimulating a Hamiltonian in small quantum space (building on [AT04] and [vMW12]), treating a Hermitian matrix as a Hamiltonian and running the quantum phase estimation procedure on it (building on [HHL09]) and making small space probabilistic (and quantum) computation consistent through the use of offline randomness and the shift and truncate method (building on [SZ99]).

Title: What can be decided locally without identifiers?

Abstract: The issue of identifiers plays a crucial role in distributed computing. In the context of local decision (i.e., when nodes of a network need to decide the legality of a distributed instance based on information they can gather only from nodes at bounded distances), symmetry breaking does not seem to play any role, and therefore the role of identities may seem to be less crucial than in local construction. Let LD be the class of all distributed languages that can be

decided in a constant number of rounds (in the LOCAL model). We study the question of whether LD* = LD, where LD* is the class of all distributed languages that can be decided in a constant number of rounds, by an ID-oblivious algorithm (i.e., an algorithm where the output of nodes does not depend on the identity assignment). We show that equality holds under some reasonable assumptions, but does not hold in the general case. That is, we show that in the general case, $LD^* \subset LD$. Interestingly, the proof of this result relies on a reduction from a problem in classical (sequential) computability theory. To the best of our knowledge, this is the first time that a result from computability theory is used in the context of network computing.

Amos Korman - What can be decided locally without identifiers?

The issue of identifiers plays a crucial role in distributed computing. In the context of local decision (i.e., when nodes of a network need to decide the legality of a distributed instance based on information they can gather only from nodes at bounded distances), symmetry breaking does not seem to play any role, and therefore the role of identities may seem to be less crucial than in local construction. Let LD be the class of all distributed languages that can be decided in a constant number of rounds (in the LOCAL model). We study the question of whether LD* = LD, where LD* is the class of all distributed languages that can be decided in a constant number of rounds, by an ID-oblivious algorithm (i.e., an algorithm where the output of nodes does not depend on the identity assignment). We show that equality holds under some reasonable assumptions, but does not hold in the general case. That is, we show that in the general case, $LD^* \subset LD$. Interestingly, the proof of this result relies on a reduction from a problem in classical (sequential) computability theory. To the best of our knowledge, this is the first time that a result from computability theory is used in the context of network computing.

Uri Zwick - Improved upper bounds for Random-Edge and Random-Jump on abstract cubes

We obtain improved exponential upper bounds for two very natural randomized algorithms for finding the sink of an Acyclic Unique Sink Orientation (AUSO) of the Boolean hypercube. AUSOs form an appealing combinatorial abstraction of linear programming. For Random-Edge, we obtain an upper bound of about 1.8^n, improving a result of Gaertner and Kaibel. For Random-Jump, we obtain an upper bound of about 1.5^n, improving a result of Mansour and Singh.

Joint work with Thomas Dueholm Hansen and Mike Paterson.

David Xiao - Lower bounds on Information Complexity via Zero-Communication Protocols

Information complexity is an important tool for proving lower bounds in communication complexity. It is a measure of the amount of information revealed by the transcript of a communication protocol about the inputs. It has been used to prove lower bounds about problems such as Disjointness, as well as composition results such as direct sum theorems. It seems to be quite powerful and previously the relationship between information complexity and other lower bound techniques such as the corruption bound was not well understood.

We will show that information complexity subsumes essentially all known lower bound techniques in communication complexity. Namely, we prove that information complexity is at least the relaxed partition bound, which subsumes the rectangle/corruption, smooth rectangle, discrepancy, and γ_2 bounds. This gives support to the conjecture that information complexity is equal to communication complexity.

As applications of our results, we give lower bounds on the information complexity of the Gap Hamming Distance problem and the Vector in Subspace problem, which in turn gives an exponential separation between quantum communication complexity and classical information complexity.

This is joint work with Iordanis Kerenidis, Sophie Laplante, Virginie Leray, and Jeremie Roland.

Yishay Mansour - Learning and Domain Adaptation

Domain adaptation is a fundamental learning problem where one wishes to use labeled data from one or several source domains to learn a hypothesis performing well on a different, yet related, domain for which no labeled data is available. This generalization across domains is a very significant challenge for many machine learning applications and arises in a variety of natural settings, including NLP tasks (document classification, sentiment analysis, etc.), speech recognition (speakers and noise or environment adaptation) and Face recognition (different lighting conditions, different population composition).

The learning theory community has only recently started to analyze domain adaptation problems. In the talk, I will overview some recent theoretical models and results regarding domain adaptation.

This is based on joint works with Mehryar Mohri and Afshin Rostamizadeh.

Marc Renault - Online Scheduling and Bin Packing with Advice

We consider the setting of online computation with advice, and study the bin packing problem and a number of scheduling problems. We show that it is possible, for any of these problems, to arbitrarily approach a competitive ratio of 1 with only a constant number of bits of advice per request. For the bin packing problem, we give an online algorithm with advice that is $(1 + \varepsilon)$-competitive

and uses $O(\frac{1}{\varepsilon}\log\left(\frac{1}{\varepsilon^2}\right)$ bits of advice per request. For scheduling on m identical machines, with the objective function of any of makespan, machine covering and the minimization of the ℓ_p norm, $p > 1$, we present similar results. We give online algorithms with advice which are $(1+\varepsilon)$-competitive ($(1/(1-\varepsilon))$-competitive for machine covering) and use $O\left(\frac{1}{\varepsilon}\log\left(\log_{1+\varepsilon}\frac{1}{\varepsilon}\right)\right)$ bits of advice per request. We complement our results by giving a lower bound showing that for any online algorithm with advice to be optimal, for any of the above scheduling problems, a non-constant number (namely, at least $\left(1-\frac{2m}{n}\right)\log m$, where n is the number of jobs and m is the number of machines) of bits of advice per request is needed. This is a joint work with Adi Rosen and Rob van Stee.

Yossi Azar - Unrelated Machine Scheduling with Startup Costs

Motivated by applications in energy-efficient scheduling in data centers, Khuller, Li, and Saha introduced the *machine activation* problem as a generalization of the classical optimization problems of minimum set cover and minimum makespan scheduling on parallel machines. In this problem, a set of n jobs have to be distributed among a set of m (unrelated) machines, given the processing time of each job on each machine. Additionally, each machine incurs a startup cost if at least one job is assigned to it. The goal is to produce a schedule of minimum total startup cost subject to a constraint L on its makespan. While Khuller et al considered the offline version of this problem, a typical scenario in scheduling is one where jobs arrive online and have to be assigned to a machine immediately on arrival.

We give an $(O(\log(mn)\log m), O(\log m))$-competitive randomized online algorithm for this problem, i.e. the schedule produced by our algorithm has a makespan of $O(L\log m)$ with high probability, and a total expected startup cost of $O(\log(mn)\log m)$ times that of an optimal offline schedule with makespan L. Our algorithm is almost optimal since it follows from previous results that the two approximation factors cannot be improved to $o(\log m\log n)$ (under standard complexity assumptions) and $o(\log m)$ respectively.

Based on the paper Online mixed packing and covering (SODA 2013) joint with U. Bhaskar, L. Fleischer and D. Panigrahi

Pierre Fraignaud - Rumor Spreading in Random Evolving Graphs

Randomized gossip is one of the most popular way of disseminating information in large scale networks. This method is appreciated for its simplicity, robustness, and efficiency. In this talk, we survey some of the recent results we obtained on randomized gossip in dynamic networks. We consider the edge-Markovian evolving graph model which captures natural temporal dependencies between the structure of the network at time t, and the one at time $t+1$. Precisely, a non-edge appears with probability p, while an existing edge dies with probability q. We consider both the Push protocol, in which every informed node selects, at

every time step, one of its neighboring node uniformly at random and forwards the information to this node, as well as the Flooding protocol, in which every informed node forwards the information to all its neighbors, at every time step.

Joint work with: Hervé Baumann, Andrea Clementi, Pierluigi Crescenzi, Carola Doerr, Marco Isopi, Alessandro Panconesi, Francesco Pasquale, and Riccardo Silvestri.

Boaz Patt-Shamir – On the line between buffer overflow and team formation

Consider the following two problems.

Problem 1: In the layer model of communication, messages of higher levels need to be broken into several lower-level messages (e.g., TCP streams into IP packets, IP packets into Ethernet frames etc.). When the lower level messages experience congestion, some of them may be lost, and the effect may be the loss of higher-level messages. This rather general setting leads to a concrete question: when a buffer overflows, which low-level packets should be discarded so as to maximize the number of completely-delivered high-level messages?

Problem 2: We are in charge of a large project, which requires several skills. We can cover each skill either by paying an outsourcing agency which charges some amount for each skill they provide, or by hiring workers, where each worker has her own set of skills and cost. To form a team we interview candidates, so that we know for each candidate what are her skills and what is her cost. The question is, which candidates should we hire to minimize the overall cost? The difficulty is that if we don't hire a candidate, she's lost forever (similarly to the secretary problem).

While these problems appear unrelated, it turns out that Problem 1 represents a new type of on-line set packing and that Problem 2 represents its dual, a new variant of on-line set cover. In this talk we'll discuss the problems and present competitive algorithms to solve them.

Based on work with Yuval Emek, Pierre Fraigniaud, Magnus Halldorsson, Yishay Mansour, Dror Rawitz, Adi Rosen.

Mathieu Lauriere - New lower bounds for privacy in communication protocols

Communication complexity is a central model of computation introduced by Yao in 1979, where two players, Alice and Bob, receive inputs x and y respectively and want to compute $f(x, y)$ for some fixed function f with the least amount of communication. Recently people have revisited the question of the privacy of such protocols: is it possible for Alice and Bob to compute $f(x, y)$ without revealing too much information about their inputs? There are two types of privacy for communication protocols that have been proposed: first, an information theoretic definition (see Bar-Yehuda et al in 1993 and Klauck in 2004), which for Boolean functions is equivalent to the notion of information cost introduced

by Chakrabarti et al in 2001 and that has since found many important applications; second, a combinatorial definition introduced by Feigenbaum et al in 2010 and further developed by Ada et al in 2012. We will show new results for both notions of privacy, as well as the relation between them. With new lower bound techniques both for the combinatorial and the information-theoretic definitions we can find tight bounds for the privacy of several functions, including Equality, Disjointness, Inner Product, Greater Than. Moreover we will see how it is possible to extend the definitions and some results of privacy to bounded-error randomized protocols, and will provide a relation between the two notions and the communication complexity.

Adi Rosen - Space Constrained Interval Selection

We study streaming algorithms for the interval selection problem: finding a maximum cardinality subset of disjoint intervals on the line. A deterministic 2-approximation streaming algorithm for this problem is developed, together with an algorithm for the special case of proper intervals, achieving improved approximation ratio of 3/2. We complement these upper bounds by proving that they are essentially best possible in the streaming setting: it is shown that an approximation ratio of $2 - \epsilon$ (or $3/2 - \epsilon$ for proper intervals) cannot be achieved unless the space is linear in the input size. In passing, we also answer an open question of Adler and Azar regarding the space complexity of constant-competitive randomized preemptive online algorithms for the same problem.

Joint work with Yuval Emek and Magnús Halldórsson.

Guy Even - Revisiting Online Routing with Unknown Durations

We revisit the online algorithm of Azar, Awerbuch, Plotkin, and Waarts [1994] for routing virtual circuits with unknown durations. We show that this algorithm can be cast into the primal-dual approach of Buchbinder and Naor [2005]. The main novelty is in dealing with the non-monotone changes in the variables due to rerouting of virtual circuits.