|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Monday (23/05)** | **Tuesday (24/05)** | **Wednesday (25/05)** | **Thursday (26/05)** | **Friday (27/05)** |
| **9:00** Opening remarks**9:10** **Dana Ron** [Survey: Sublinear Algorithms for Approximating Graph Parameters](#_Dana_Ron_(Tel)**10:00****Krzysztof Onak** [A Near-Optimal Sublinear-Time Algorithm for Approximating the Minimum Vertex Cover Size](#_Krzysztof_Onak_(CMU))**10:30** Break**11:00****Justin Romberg** [Survey: Compressive Sensing in Signal Processing](#_Justin_Romberg_(Georgia)**11:50****Piotr Indyk** [On the Power of Adaptivity in Sparse Recovery](#_Piotr_Indyk_(MIT)) | **9:00****Paul Valiant** [Survey: Estimating the Unseen: Sublinear Statistics](#_Paul_Valiant_(Berkeley))**9:50** **Rocco Servedio** [Learning and Testing k-Modal Distributions](#_Rocco_Servedio_(Columbia)**10:20** Break**10:50** **Madhu Sudan**[Survey: Invariance in Property Testing](#_Madhu_Sudan_(Microsoft)**11:40****Arnab Bhattacharyya**[A Unified Framework for Testing Linear-Invariant Properties](#_Arnab_Bhattacharyya_(MIT))**12:10****Tali Kaufman**[Locally Testable Codes and Expanders](#_Tali_Kaufman) | **9:00****Ben Recht** [Survey: From Compressed Sensing to Matrix Completion and Beyond](#_Ben_Recht_(University)**9:50** **Michael Mahoney** [Fast Approximation of Matrix Coherence](#_Michael_W._Mahoney)**10:20** Break**10:50** **Ely Porat** [Multi-pattern Search in the Streaming Model](#_Ely_Porat_(Bar-Ilan)**11:20****Alexandr Andoni** [Sublinear Algorithms via Precision Sampling](#_Alexandr_Andoni_(Microsoft)**11:50** Break**12:00****Nir Halman** [Sublinear FPTASs for Stochastic Optimization Problems](#_Nir_Halman_(Hebrew)**12:30****Asaf Shapira** [Testing Odd-Cycle-Freeness in Boolean Functions](#_Asaf_Shapira_(Georgia) | **9:00****Roger Wattenhofer** [Tutorial: Distributed Algorithms](#_Roger_Wattenhofer_(ETH)**9:50****Pierre Fraigniaud** [Local Distributed Decision](#_Pierre_Fraigniaud_(LIAFA,)**10:20** Break**11:00****Ning Xie** [Local Computation Algorithms](#_Ning_Xie_(MIT))**11:30****Yuichi Yoshida** [Optimal Constant-Time Approximation Algorithms and (Unconditional) Inapproximability Results for Every Bounded-Degree CSP](#_Yuichi_Yoshida_(Kyoto) | **9:00****Oded Lachish** [Testing a Language Accepted by a Fixed Boolean Formula](#_Oded_Lachish_(Birkbeck)**9:30****Gilad Tsur**[On Approximating the Number of Relevant Variables in a Function](#_Gilad_Tsur)**10:00** Break**10:10****Andrew McGregor**[Stream Periodicity](#_Andrew_McGregor_(University)**10:40****Robert Krauthgamer** [Polylogarithmic Approximation for Edit Distance and the Asymmetric Query Complexity](#_Robert_Krauthgamer_(Weizmann)**11:10** Break**11:40****Artur Czumaj** [Planar Graphs:  Random Walks and Bipartiteness Testing](#_Artur_Czumaj_(University)**12:10****Ronitt Rubinfeld** [Testing Properties of Collections of Distributions](#_Rocco_Servedio_(Columbia)**12:25**Closing remarks |
| **1:00** Lunch | **1:00** Lunch | **1:00** Lunch | **1:00** Lunch | **1:00** Lunch |
| **3:00****David Woodruff**[Survey: Valutazione delle Norme con le Applicazioni](#_David_Woodruff_(IBM)**3:50** Break**4:30****Amit Chakrabarti** [Gap-Hamming-Distance: the Journey to an Optimal Lower Bound](#_Amit_Chakrabarti_(Dartmouth))**5:00****Graham Cormode** [Mergeable Summaries](#_Graham_Cormode_(AT&T) | **3:00****Nir Ailon** [Survey: Johnson-Lindenstrauss Transform(s)](#_Nir_Ailon_(Technion))**3:50****Jelani Nelson** [Sparse Johnson-Lindenstrauss Transforms](#_Jelani_Nelson_(MIT))**4:20** Break**4:50****Sofya Raskhodnikova** [Testing and Reconstruction of Lipschitz Functions](#_Sofya_Raskhodnikova_(Pennsylvania)**5:20****C. Seshadhri** [Is Submodularity Testable?](#_C._Seshadhri_(Sandia) | **2:00** (???)Excursion to Ravenna (with dinner) | **3:00****Christian Sohler** [Every Property of Hyperfinite Graphs is Testable](#_Christian_Sohler_(TU)**3:30****Oded Goldreich** [Finding Cycles and Trees in Sublinear Time](#_Oded_Goldreich_(Weizmann))**4:00** Break**4:30**Open Problem Session |  |
| **7:30** Dinner | **7:30** Dinner |  | **7:30** Dinner |  |

**Workshop on Sublinear Algorithms**

Bertinoro International Center for informatics

May 22 - 27, 2011

Abstracts

# [Nir Ailon](http://www.cs.technion.ac.il/~nailon/) (Technion)

## Johnson-Lindenstrauss Transform(s)

The Johnson-Lindenstrauss technique has been for many years used as an algorithmic black-box for trading off dimensionality (and its curse) with approximation. More recently there have been attempts to open this black box and improve it as an algorithm in its own right. Much work has been devoted to reducing both time and randomness. In this tutorial I will survey work on the time complexity aspects, and connect it with other important fields such as compressed sensing.

# [Alexandr Andoni](http://www.mit.edu/~andoni/) (Microsoft Research Silicon Valley)

## Sublinear Algorithms via Precision Sampling

Suppose we want to estimate a sum of bounded reals a1+a2+…+an with access to only some "limited information" about ai's. A classical setting is where we estimate the entire sum by knowing only a random subset of ai's. Naturally, there is a trade-off between the size of the subset and the resulting approximation.

Motivated by applications where this trade-off is not good enough, we introduce Precision Sampling, which is an estimation technique that uses a more general kind of "limited information" about ai's.  Instead of obtaining a subset of ai's as above, here we obtain a rough estimate for each of the n quantities, up to various prescribed degrees of "precision" (approximation). The trade-off then becomes between the precision of the estimates of ai's and the resulting approximation to the total sum. We show one can obtain a trade-off that is qualitatively better in the precision sampling setting than in the aforementioned (vanilla sampling) setting.

Our resulting tool leads to new sublinear algorithms. In one instance, it gives a simplified algorithm for a class of streaming problems. In another instance, the tool plays an instrumental role in a query-efficient algorithm for estimating edit distance.

Joint work with Robi Krauthgamer and Krzysztof Onak.

# [Arnab Bhattacharyya](http://web.mit.edu/~abhatt/www/) (MIT)

## A Unified Framework for Testing Linear-Invariant Properties

In a sequence of recent papers, Sudan and coauthors have investigated the relation between testability of properties of Boolean functions and the invariance of the properties with respect to transformations of the domain. Linear-invariance is arguably the most common such symmetry for natural properties of Boolean functions on the hypercube. Hence, it is an important goal to find necessary and sufficient conditions for testability of linear-invariant properties. We obtain the following results:

1. We show that every linear-invariant property that can be characterized by forbidding induced solutions to a (possibly infinite) set of linear equations can be tested with one-sided error.

2. We show that every linear-invariant property that can be tested with one-sided error can be characterized by forbidding induced solutions to a (possibly infinite) set of systems of linear equations.

We conjecture that our result from item (1) can be extended to cover systems of linear equations. We further show that the validity of this conjecture would have the following implications:

1. It would imply that every linear-invariant property that is closed under restrictions to linear subspaces is testable with one-sided error. Such a result would unify several previous results on testing Boolean functions, such as the results on testing low-degree polynomials and results on testing Fourier dimensionality.

2. It would imply that a linear-invariant property P is testable with one-sided error \*if and only if\* P is closed under restrictions to linear subspaces, thus resolving Sudan's problem.

Joint work with Elena Grigorescu and Asaf Shapira.

# [Amit Chakrabarti](http://www.cs.dartmouth.edu/~ac/) (Dartmouth)

## Gap-Hamming-Distance: the Journey to an Optimal Lower Bound

In the Gap-Hamming-Distance problem (GHD), Alice and Bob receive n-bit strings and must decide whether they are "close" (Hamming distance \leq n/2-n1/2) or "far" (distance \geq n/2+n1/2). How many bits must they communicate to solve this problem, with error at most 1/3? Since being proposed in the early 2000s by Indyk and Woodruff, for its applications to space lower bounds for data stream algorithms, the problem has been extensively studied.

In this talk, I shall describe the journey from the early results of Indyk and Woodruff, to stronger intermediate results obtained through a deeper understanding of the problem, and finally to a recent optimal lower bound due to myself and Oded Regev. We now know that any GHD protocol requires (n) communication, which completely settles the problem.

Two simplifications of our proof have appeared within the last two months, and I shall touch upon these as well.

# [Graham Cormode](http://dimacs.rutgers.edu/~graham/) (AT&T Research)

## Mergeable Summaries

In dealing with massive distributed data, exact computations may be possible but can still be very expensive to perform. Instead, it is often desirable to look to more lightweight computations that give (guaranteed) approximations. A central approach is the idea of the mergeable summary: a compact data structure that summarizes a data set, which can be merged with others to summarize the union of the data without significant growth in size. This framework enables parallel computation.

Samples and sketches are two examples of well-known, effective mergeable summaries. I'll present recent results on new, compact mergeable summaries for various tasks, such as approximating frequent items, order statistics, and range counts.

Joint work with Pankaj Agarwal, Zengfeng Huang, Jeff Phillips, Zhewei Wei and Ke Yi.

# [Artur Czumaj](http://www.dcs.warwick.ac.uk/~czumaj/) (University of Warwick)

## Planar Graphs:  Random Walks and Bipartiteness Testing

We study the testability of hereditary properties in arbitrary planar graphs. Our main result is a constant time testing of bipartiteness in arbitrary planar graphs. The previous bound for this class of graphs was O(n1/2), and the constant-time testability was only known for planar graphs with **bounded degree**. Previously used transformations of unbounded-degree sparse graphs into bounded-degree sparse graphs cannot be used to reduce the problem to the testability of bounded-degree planar graphs. Our approach extends to arbitrary minor-free graphs.

Our algorithm is based on random walks. The challenge here is to analyze random walks for a class of graphs that has good separators, i.e., bad expansion. Standard techniques that use a fast convergence to a uniform distribution do not work in this case. Informally, our analysis technique self-reduces the problem of finding an odd length cycle in a multigraph G induced by a collection of cycles to another multigraph G’ induced by a set of shorter odd-length cycles, in such a way that when a random walks finds a cycle in G’ with probability p, then it does so with probability f(p) in G. This reduction is applied until the cycles collapse to selfloops that can be easily detected.

Joint work with Morteza Monemizadeh, Krzysztof Onak, and Christian Sohler.

# [Pierre Fraigniaud](http://www.liafa.jussieu.fr/~pierref/) (LIAFA, Paris)

## Local Distributed Decision

Inspired by sequential complexity theory, we focus on a complexity theory for distributed decision problems. A central theme in distributed network algorithms concerns understanding and coping with the issue of locality. In this context, solving a decision problem requires the processors to independently inspect their local neighborhoods and then collectively decide whether a given global input instance belongs to some specified language. We consider the standard LOCAL model of computation and define LD(t) (for local decision) as the class of decision problems that can be solved in t communication rounds. We first study the question of whether randomization helps in local distributed computing, and to what extent. Specifically, we define the corresponding randomized class BPLD(t,p,q), containing all languages for which there exists a randomized algorithm that runs in t rounds, accepts correct instances with probability at least p and rejects incorrect ones with probability at least q. We show that p2+q=1 is a threshold for the containment of LD(t) in BPLD(t,p,q). More precisely, we show that there exists a language that does not belong to LD(t) for any t=o(n) but does belong to BPLD(0,p,q) for any p,q in (0,1] such that p2+q \leq1. On the other hand, we show that, restricted to hereditary languages, BPLD(t,p,q) = LD(O(t)), for any function t and any p, q in (0,1] such that p2+q>1. In addition, we investigate the impact of non-determinism on local decision, and establish some structural results inspired by classical computational complexity theory. Specifically, we show that non-determinism does help, but that this help is limited, as there exist languages that cannot be decided non-deterministically. Perhaps surprisingly, it turns out that it is the combination of randomization with non-determinism that enables to decide all languages in constant time. Finally, we introduce the notion of local reduction, and establish some completeness results.

Joint work with Amos Korman, and David Peleg.

# [Oded Goldreich](http://www.wisdom.weizmann.ac.il/~oded/) (Weizmann)

## 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 Ck-minor free (resp., free from having the corresponding tree-minor). In particular, if the graph is far (i.e., (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 $\tilde{O}(poly(1/)\sqrt{N})$. This matches the known (\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 . 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 .

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.

This is joint work with Artur Czumaj, Dana Ron, C. Seshadhri, Asaf Shapira, and Christian Sohler.

# [Nir Halman](http://pluto.huji.ac.il/~halman/) (Hebrew University of Jerusalem and MIT)

## FPTASs for Stochastic Optimization Problems

We present a framework for obtaining Fully Polynomial Time Approximation Schemes (FPTASs) for stochastic optimization problems that run in sublinear time. The functions in our model are either convex or monotone and are over discrete domains.

The framework is based on approximating univariate functions by piecewise linear functions and monitoring the propagation of errors.

This talk is based on several joint works with (subsets of) Diego Klabjan (Northwestern), Chung-Lun Li (The Hong Kong Polytechnic University), Jim Orlin (MIT) and David Simchi-Levi (MIT).

# [Piotr Indyk](http://people.csail.mit.edu/indyk/) (MIT)

## On the Power of Adaptivity in Sparse Recovery

Abstract: ???

# Tali Kaufman

## Locally Testable Codes and Expanders

Expanders and Error correcting codes are two celebrated notions, coupled together, e.g. in the seminal work on Expander Codes by Sipser and Spielman that yields some of the best codes. In recent years there is a growing interest in codes admitting algorithms that run in \*constant\* time. E.g. codes for which it is possible to distinguish in constant time between codewords and vectors far from the code. Such codes are known as locally testable codes (LTCs) and they are intimately related to the PCP Theorem.

It is known that Expander Codes based on \*best\* expanders can NOT yield LTCs, so it seems that expansion is somewhat contradicting to the notion of LTCs. We show that, nevertheless, the graph associated with an LTC \*must\* be essentially an expander. Namely, every LTC can be decomposed into a constant number of "basic" codes whose associated graph is an expander.

Based on a joint work with Irit Dinur

# [Robert Krauthgamer](http://www.wisdom.weizmann.ac.il/~robi/) (Weizmann Institute)

## Polylogarithmic Approximation for Edit Distance and the Asymmetric Query Complexity

We present a near-linear time algorithm that approximates the edit distance between two strings within a significantly better factor than previously known. This result emerges in our investigation of edit distance from a new perspective, namely a model of asymmetric queries, for which we present near-tight bounds.

Another consequence of this new model is the first rigorous separation between edit distance and Ulam distance, by tracing the hardness of edit distance to phenomena that were not used by previous analyses.

Joint work with Alexandr Andoni and Krzysztof Onak.

# [Oded Lachish](http://www.dcs.bbk.ac.uk/~oded/) (Birkbeck University of London)

## Testing a Language Accepted by a Fixed Boolean Formula

For a given formula  we define SAT() to be the set of all assignments that satisfy . We study the query complexity of testing and estimating SAT(), in a model where the testers have absolute knowledge of . Our main results are.

* Testing
	+ If SAT() is binary and read-once, then the query complexity is at most quasipolynomial in 1/
	+ If SAT() is binary, monotone and read-k-times then the query complexity is at most quasipolynomial in 1/ and k
* Estimating
	+ If SAT() is binary, monotone, read-once, then the query complexity is at most quasipolynomial in 1/

# [Michael W. Mahoney](http://cs.stanford.edu/people/mmahoney/) (Stanford University)

## Fast Approximation of Matrix Coherence

The concept of matrix coherence measures the extent to which the singular vectors of a matrix are correlated with the standard basis, and as such it is of interest in the recently-popular problem of matrix completion. A more refined concept is that of statistical leverage. Statistical leverage is usually quantified by the diagonal elements of the projection matrix onto the best rank-k approximation to a matrix. Thus, it is useful in large-scale diagnostic data analysis for identifying outliers in large data matrices and data graphs. Moreover, it is the key structural nonuniformity that must be dealt with (i.e., either rapidly approximated or rapidly uniformized at the preprocessing step) in developing fast random sampling and random projection algorithms for matrix problems such as least-squares regression and low-rank matrix approximation.

As a corollary of our main result, we obtain an algorithm to approximate the coherence of an arbitrary matrix in time qualitatively faster than the naive algorithm. Our main result is a randomized algorithm that takes as input an arbitrary m x n matrix A and a rank parameter k; it returns as out output a relative-error approximation to every diagonal element of the projection matrix onto the best rank-k approximation to A; and it runs in time qualitatively faster than the time to compute a basis for that space.
In particular, given an m x n matrix A, with m>>n, the algorithm returns relative-error approximations to all the diagonal elements of the projection matrix onto the left singular subspace in roughly O(m n log n) time, as opposed to O(mn2) time required by the naive algorithm. In addition, numerical implementations run faster than traditional deterministic algorithms for matrices as small as hundreds by thousands.
Our analysis boils down to computing a relative-error approximation to an underconstrained least-squares approximation, or relatedly it can be viewed as an application of Johnson-Lindenstrauss ideas.

# [Andrew McGregor](http://www.cs.umass.edu/~mcgregor/) (University of Massachusetts, Amherst)

## Stream Periodicity

Abstract: ???

# [Jelani Nelson](http://web.mit.edu/minilek/www/) (MIT)

## Sparse Johnson-Lindenstrauss Transforms

The Johnson-Lindenstrauss (JL) lemma (1984) states that any n points in d-dimensional Euclidean space can be embedded into k=O(log n/2) dimensions so that all pairwise distances are preserved up to 1+/-. Furthermore, this embedding can be achieved via a linear mapping. The JL lemma is a useful tool for speeding up solutions to several high-dimensional problems: closest pair, nearest neighbor, diameter, minimum spanning tree, etc. It also speeds up some clustering and string processing algorithms, reduces the amount of storage required to store a dataset, and can be used to reduce memory required for numerical linear algebra problems such as linear regression and low rank approximation.

The original proofs of the JL lemma let the linear mapping be specified by a random dense k x d matrix (e.g. i.i.d. Gaussian entries). Thus, performing an embedding requires dense matrix-vector multiplication. There has been much recent work on developing distributions that allow for embedding vectors quickly, begun by the work of [Ailon, Chazelle, STOC 2006]. Unfortunately, these works cannot take advantage of sparsity of the vector to be embedded, and they take (d) time to embed a vector even with only one non-zero entry. This feature is particularly unfortunate in streaming applications, where updates can be seen as 1-sparse vectors.

One way to speed up embeddings for sparse vectors is to develop distributions over linear mappings whose corresponding matrices themselves are sparse. In this talk, we give two JL distributions with the sparsest known matrices. In fact, these are the first distributions where every matrix in their support has only o(1) of its entries non-zero for all settings of  and n (specifically, only an O()-fraction of entries in each column are non-zero).

This is based on joint work with Daniel Kane (Harvard University).

# [Krzysztof Onak](http://people.csail.mit.edu/konak/) (CMU)

## A Near-Optimal Sublinear-Time Algorithm for Approximating the Minimum Vertex Cover Size

I will present a near-optimal sublinear-time algorithm for approximating the size of a minimum vertex cover. Let VC denote this quantity for the input graph. Our algorithm computes an estimate alpha such that VC \leq  \leq 2VC + n, where  is a parameter to the algorithm. The query complexity and running time of the algorithm are \tilde{O}(d’ poly(1/)), where d’ denotes the average vertex degree in the graph.

Joint work with Dana Ron, Michal Rosen, and Ronitt Rubinfeld.

# [Ely Porat](http://u.cs.biu.ac.il/~porately/) (Bar-Ilan University)

## Multi-pattern Search in the Streaming Model

Abstract: ???

# [Sofya Raskhodnikova](http://www.cse.psu.edu/~sofya/) (Pennsylvania State University)

## Testing and Reconstruction of Lipschitz Functions

A function f : D -> R has Lipschitz constant c if dR(f(x), f(y)) <= c dD(x, y) for all x, y in D, where dR and dD denote the distance functions on the range and domain of f, respectively. We say a function is Lipschitz if it has Lipschitz constant 1. (Note that rescaling by a factor of 1/c converts a function with a Lipschitz constant c into a Lipschitz function.) In other words, Lipschitz functions are not very sensitive to small changes in the input. We initiate the study of testing and local reconstruction of the Lipschitz property of functions.

We consider functions over domains {0,1}d, {1,...,n} and {1,...,n}d, equipped with L1 distance. We design efficient testers of the Lipschitz property for functions of the form f:{0,1}d ->  Z, where  is in (0,1] and Z is the set of multiples of , and also of the form f: {1,...,n} -> R, where R is (discretely) metrically convex. In the first case, the tester runs in time O(d min{d,r}/), where r is the diameter of the image of f; in the second, in time O((\log n)/). We give corresponding lower bounds of (d) and (log n) on the query complexity (in the second case, only for nonadaptive 1-sided error testers). Our lower bound for functions over {0,1}d is tight for the case of the {0,1,2} range and constant .

We also present a local filter of the Lipschitz property for functions of the form f:{1,...,n}d -> R with lookup complexity O((log n+1)d). For functions of the form {0,1}d, we show that every nonadaptive local filter has lookup complexity exponential in d. The testers that we developed have applications to programs analysis. The reconstructors have applications to data privacy.

Joint work with Madhav Jha.

# [Ben Recht](http://pages.cs.wisc.edu/~brecht/) (University of Wisconsin)

## From Compressed Sensing to Matrix Completion and Beyond

Matrix completion—where one seeks to recover a low rank matrix from an incomplete set of observations— is a recurring problem in collaborative filtering, dimensionality reduction, and systems and control theory. While the general problem of finding the lowest rank matrix satisfying a set of equality constraints is NP-hard, a wave of recent results have provided very general settings under which one can perfectly recover all of the missing entries of a low-rank matrix by solving convex optimization problems. In this talk, I will review this emerging body of work, discussing both theoretical foundations and practical applications and drawing connections with fast Monte Carlo algorithms for matrix approximation pioneered by the TCS community. I will conclude the talk with some recent work on further extending the catalog of objects and structures that can be recovered from partial information using convex optimization.

# [Justin Romberg](http://users.ece.gatech.edu/justin/Justin_Romberg.html) (Georgia Tech)

## Compressive Sensing in Signal Processing

We will overview the fundamental results and recent theoretical trends in compressive sensing, and review some recent hardware designs which operate within the CS framework. The theoretical portion of the talk will include reconstruction guarantees for structured random matrices (e.g. sampling using random convolution or random modulation) and recent progress on sampling ensembles of correlated signals using low-rank recovery algorithms. The applications will include single-pixel imaging, radar imaging, acoustic source localization, and accelerated forward modeling for seismic imaging.

# [Dana Ron](http://www.eng.tau.ac.il/~danar/) (Tel Aviv University)

## Sublinear Algorithms for Approximating Graph Parameters

Abstract: ???

# [Ronitt Rubinfeld](http://people.csail.mit.edu/ronitt/) (Tel Aviv University and MIT)

## Testing Properties of Collections of Distributions

Abstract: ???

# [Rocco Servedio](http://www.cs.columbia.edu/~rocco/) (Columbia University)

## Learning and Testing k-Modal Distributions

A k-modal probability distribution over the domain {1,…,N} is one whose histogram has at most k "peaks" and "valleys". Such distributions are a natural generalization of the well-studied class of monotone increasing (or monotone decreasing) probability distributions.

We study the problem of learning an unknown k-modal distribution from samples. We also study a related problem in property testing: given access to samples from an unknown k-modal distribution p, determine whether p is identical to q or p is "far" from q. Here q is a k-modal distribution which may be given explicitly or may be available via sample access. These problems were previously studied in the case k=0 (monotone distributions) and k=1 (unimodal distributions).

We give algorithms for these problems that are provably close to the best possible in terms of sample and time complexity. An interesting feature of our approach is that our learning algorithms use ingredients from property testing and vice versa.

Joint work with Costis Daskalakis and Ilias Diakonikolas.

# [C. Seshadhri](http://www.cs.princeton.edu/~csesha/) (Sandia National Laboratories)

## Is Submodularity Testable?

This talk is really more about initiating discussion about a new property testing problem. I will talk about some results, but they seem to indicate our ignorance more than anything else (so I'm going to be a little vague in this abstract). We want to property test submodularity over the boolean hypercube. This is a well-studied property that commonly appears when modelling utility functions and has lots of applications in combinatorial optimization. Property testing also makes a lot of "sense" for submodularity. This is because many algorithms assume access to a truly submodular function (whose description is actually 2n, where n is the dimension of the hypercube), but only query a very small portion (poly(n) points) of the domain.

Verifying (exactly) the property of submodularity can be reduced to verifying the monotonicity of n functions (over the boolean hypercube). Property testing, as we painfully learned, seems to be a completely different prospect. Indeed, we prove results showing that from a sublinear testing perspective, submodularity behaves very differently from monotonicity. For example, we give examples of far from submodular functions that have surprisingly few local violations.

Joint work with Jan Vondrak (appeared in ICS 2011).

# [Asaf Shapira](http://people.math.gatech.edu/~asafico/) (Georgia Tech)

## Testing Odd-Cycle-Freeness in Boolean Functions

Call a boolean function f Odd-Cycle-Free if for every odd integer k, and for every k points x1,…,xk in the support of f the sum x1+…+xk does not vanish. We show that the property of being odd-cycle-free is testable with poly(1/) queries. We prove this by analyzing the eigenvalues of the Cayley graph associated with a Boolean function.

We also prove that there is a canonical tester for odd-cycle-freeness making poly(1/) queries, meaning that the testing algorithm operates by picking a random linear subspace of dimension O(log(1/)) and then checking if the restriction of the function to the subspace is odd-cycle-free or not. The test is analyzed by studying the effect of random subspace restriction on the Fourier coefficients of a function. Our work implies that using a canonical tester, instead of an arbitrary tester, to test odd-cycle-freeness results in no more than a polynomial blowup in the query complexity. The question of whether a canonical tester with polynomial blowup exists for all linear-invariant properties remains an open problem.

Joint work with A. Bhattacharyya, E. Grigorescuy and P. Raghavendra.

# [Christian Sohler](http://ls2-www.cs.uni-dortmund.de/~sohler/) (TU Dortmund)

## Every Property of Hyperfinite Graphs is Testable

A property testing algorithm for a property  in the bounded degree graph model [GR97] is an algorithm that, given access to the adjacency list representation of a graph G=(V,E) with maximum degree at most d, accepts G with probability at least 2/3 if G has property , and rejects G with probability at least 2/3, if it differs on more than dn edges from every d-degree bounded graph with property . A property is *testable*, if for every , d, and n, there is a property testing algorithm A,n,d that makes at most q(,d) queries to an input graph of n vertices, that is, a non-uniform algorithm that makes a number of queries that is independent of the graph size.

A k-disc around a vertex v of a graph G=(V,E) the subgraph induced by all vertices of distance at most k from v. We show that the structure of a planar graph on large enough number of vertices, n, and with constant maximum degree d, is determined, up to the modification (insertion or deletion) of at most dn edges, by the frequency of k-discs for certain k=k(,d), that is independent of the size of the graph. We can replace planar graphs by any hyperfinite class of graphs, which includes, for example, every graph class that does not contain a set of forbidden minors.

We use this result to obtain new results and improve upon existing results in the area of property testing. In particular, we prove that

* graph isomorphism is testable for every class of hyperfinite graphs,
* every graph property is testable for every class of hyperfinite graphs,
* every property of hyperfinite graphs is testable in the bounded degree
* graph model,
* a large class of graph parameters is approximable for hyperfinite graphs.

Our results also give a partial explanation of the success of motifs in the analysis of complex networks.

Joint work with Ilan Newman.

# [Madhu Sudan](http://research.microsoft.com/en-us/um/people/madhu/) (Microsoft Research New England)

## Invariance in Property Testing

Abstract: ???

# Gilad Tsur

## On Approximating the Number of Relevant Variables in a Function

In this work we consider the problem of approximating the number of relevant variables in a function given query access to the function. Since obtaining a multiplicative factor approximation is hard in general, we consider several relaxations of the problem. In particular, we consider relaxations in which we have a promise that the function belongs to a certain family of functions (e.g., linear functions), and we consider a relaxation of the property testing variant of the problem. In the latter relaxation the task is to distinguish between the case that the number of relevant variables is at most k, and the case in which it is far from any function in which the number of relevant variable is more than (1+ )k for a parameter .

We give both upper bounds and almost matching lower bounds for the relaxations we study.

This is joint work with Dana Ron.

# [Paul Valiant](http://www.cs.berkeley.edu/~pvaliant/) (Berkeley)

## Estimating the Unseen: Sublinear Statistics

Much of statistics can be described as: given samples from an unknown distribution, estimate some property of the distribution. Many natural estimators from statistics require a number of samples that is linear in a natural parameter of the problem, in particular, often, a bound on the support size of the distribution. In this talk I will survey SUBlinear approaches to these problems, starting with the Good-Turing estimator for the fraction of unseen probability mass and Fisher's estimate of the unseen number of species, both from the 1940s, and covering up through the latest results on estimating entropy, independence, and closeness of distributions. Both algorithmic results and lower bounds will be discussed, with a focus on common threads and unifying intuition.

# [Roger Wattenhofer](http://www.dcg.ethz.ch/members/roger.html) (ETH Zürich)

## Distributed Algorithms

Abstract: ???

# [David Woodruff](http://www.almaden.ibm.com/cs/people/dpwoodru/) (IBM Almaden)

## Valutazione delle Norme con le Applicazioni

Abstract: ???

# [Ning Xie](http://people.csail.mit.edu/ningxie/) (MIT)

## Local Computation Algorithms

For input x, let F(x) denote the set of outputs that are the “legal” answers for a computational problem F. Suppose x and members of F(x) are so large that there is not time to read them in their entirety. We propose a model of *local computation algorithms* which for a given input x, support queries by a user to values of specified locations yi in a legal output y in F(x).

When more than one legal output y exists for a given x, the local computation algorithm should output in a way that is consistent with at least one such y. Local computation algorithms are intended to distill the common features of several concepts that have appeared in various algorithmic subfields, including local distributed computation, local algorithms, locally decodable codes, and local reconstruction.

We develop a technique, based on known constructions of small sample spaces of k-wise independent random variables and Beck's analysis in his algorithmic approach to the Lovasz Local Lemma, which under certain conditions can be applied to construct local computation algorithms that run in *polylogarithmic* time and space. We apply this technique to maximal independent set computations, scheduling radio network broadcasts, hypergraph coloring and satisfying k-SAT formulas.

Based on joint works with Noga Alon, Ronitt Rubinfeld, Gil Tamir and Shai Vardi.

# [Yuichi Yoshida](http://www.lab2.kuis.kyoto-u.ac.jp/~yyoshida/) (Kyoto University)

## Optimal Constant-Time Approximation Algorithms and (Unconditional) Inapproximability Results for Every Bounded-Degree CSP

Raghavendra (STOC 2008) gave an elegant and surprising result: if Khot’s Unique Games Conjecture (STOC 2002) is true, then for every constraint satisfaction problem (CSP), the best approximation ratio is attained by a certain simple semidefinite programming and a rounding scheme for it.

In this talk, we show that similar results hold for constant- time approximation algorithms in the bounded-degree model. Specifically, we present the following: (i) For every CSP, we construct an oracle that serves an access, in constant time, to a nearly optimal solution to a basic LP relaxation of the CSP. (ii) Using the oracle, we give a constant-time rounding scheme that achieves an approximation ratio coincident with the integrality gap of the basic LP. (iii) Finally, we give a generic conversion from integrality gaps of basic LPs to hardness results. All of those results are unconditional. Therefore, for every bounded-degree CSP, we give the best constant-time approximation algorithm among all.

A CSP instance is called ε-far from satisfiability if we must remove at least an ε-fraction of constraints to make it satisfiable. A CSP is called testable if there is a constant- time algorithm that distinguishes satisfiable instances from ε-far instances with probability at least 2/3. Using the results above, we also derive, under a technical assumption, an equivalent condition under which a CSP is testable in the bounded-degree model.