Abstracts Collection
Search Methodologies
— Dagstuhl Seminar —

Rudolf Ahlswede[1], Ferdinando Cicalese[2] and Ugo Vaccaro[3]

[1] Universität Bielefeld, D
ahlswede@mathematik.uni-bielefeld.de
[2] University of Salerno, I
cicalese@dia.unisa.it
[3] University of Salerno, I
uv@dia.unisa.it

Abstract. From 05.07.09 to 10.07.09, the Dagstuhl Seminar 09281 on "Search Methodologies " was held in Schloss Dagstuhl – Leibniz Center for Informatics. Abstracts of the presentations given during the seminar are put together in this paper. The first section describes the seminar topics and goals in general. We also briefly comment on how the topics were addressed in the talks. Links to extended abstracts or full papers are provided, if available.

09281 Seminar topics and goals – Search Methodologies

The main purpose of this seminar was to provide a common forum for researchers interested in the mathematical, algorithmic, and practical aspects of the problem of *efficient searching*, as seen in its polymorphic incarnation in the areas of computer science, communication, bioinformatics, information theory, and related fields of the applied sciences. We believe that only the on site collaboration of a variety of established and young researchers engaged in different aspects of *search theory* might provide the necessary humus for the identification of the basic search problems at the conceptual underpinnings of the new scientific issues in the above mentioned areas. We aim at uncovering common themes and structures among these problems, by analyzing them through interdisciplinary lens, and tools from a variety of areas, ranging from Algorithmics to Computational Complexity, from Information Theory to Combinatorics. The more recent challenges provided by the areas of Communications and Molecular Biology call for more attention at the application side of the problems. Therefore, together with the conceptual understanding and the efficient algorithmic solutions, we shall focus also on the studies of new heuristics and experimental methods as well as the theoretical understanding of the well established ones.

We carefully chose a group of outstanding researchers, of different expertise but nonetheless fluent in diverse languages of sciences. They brought their different views of the themes of the original proposal of this seminar. Through the

Dagstuhl Seminar Proceedings 09281
Search Methodologies
http://drops.dagstuhl.de/opus/volltexte/2009/2245

several discussions and the two open problem sessions, we aimed at laying the basis for new perspectives, and solutions to arise.

We shall now briefly describe some of the main areas of research and the problems addressed in the talks and in the common discussions.

The ubiquitous nature of group testing makes it a gold mine for investigators in Search Theory. Group testing has been proved to find applications in a surprising variety of situations, including quality control in product testing searching for files in storage systems, screening for experimental variables, data compression, computation of statistics in the data stream model, and testing for concentration of chemical and pathogenic contaminants. Group testing has been recently applied to Computational Molecular Biology, where it is used for screening library of clones with hybridization probes, and sequencing by hybridization. The contributions by P. Damaschke, G.O.H. Katona, A.J. Macula, and E. Triesch reported on some recent development in this area. In the presentation by A. Zhigljavsky, the case when tests can be affected by noise is also considered. Fault-tolerant search strategies were also considered in C. Deppe's talk. He reported on the equivalence between combinatorial channels with feedback and combinatorial search with adaptive strategies, giving new constructive bounds, when the error is proportional to the blocklength/the number of tests.

The study of gene expression, protein structure, and cell differentiation has produced huge databases which are heterogeneous, distributed, and semi-structured. We are interested in the problem of processing queries that involve specialized approximate pattern matching and complex geometric relations. See, e.g., the contribution by E. Porat for application of group testing to problems of approximate pattern matching.

In multi-access communication one has to coordinate the access of a set of stations to a shared communication medium. It is known that this problem and probabilistic group testing are strongly tied. We focussed on the fascinating relations among the combinatorial structures that are at the conceptual bottom of deterministic multi-access communication and non-adaptive group testing, namely superimposed codes and their many variants. The importance of these structures, that appear in an astonishing variety of problems cannot be overestimated. C. Colbourn, H. Aydinian, E. Porat and G. Wiener, presented some new combinatorial constructions for selection by intersection and superimposed encoding.

A new area of research where group testing techniques are finding fertile ground for new developments is the one of compressed sensing. The presentation by C. Colbourn and O. Milenkovic focussed on some aspects of this new fascinating area of investigation.

Evaluating a function by probing the smallest possible set of variables is at the core of studying the decision tree model for Boolean functions. Function evaluation algorithms play also a central role in automatic diagnosis and more generally in computer aided decision making systems. Relevant to this area of research was the presentation by M. Milanič who reported on game tree evaluation in the priced information model. Game tree search was also dealt with in I. Althoefer presentation which focussed on Monte Carlo techniques.

Data compression is another area of investigation, which is tightly connected to search. An easy example of such connection is given by the Huffman trees which provide optimal prefix free compression and, equivalently, search strategy with optimal average number of questions. Variants of the Huffman coding problem are also important in problems of information transmission and storing. M. Golin presented new dynamic programming based approach for variants of the Huffman coding problem. T. Gagie reported on constructing minimax trees.

The presentation by E. Kranakis reported on a different model of search, the one of *randezvous* problems. Here, several agents living in a common domain, want to find each other at a common place and time. The question is what strategies they should choose to maximize their probability of meeting. Such problems have applications in the fields of synchronization, operating system design, operations research, and even search and rescue operations planning.

There were also some contributions that extended beyond the set of main topics: K. Kobayashi reported on new results on the capacity formula of finite state channels, and S. Riis, presentation introduced the (private) entropy of a directed graph in a new network coding sense, and related it to the concepts of the guessing number of a graph.

Monte Carlo Search in Trees and Game Trees

Ingo Althoefer (Universität Jena, DE)

Seminal papers on game tree search were written by Zermelo (1912; very theoretical) and Shannon (1950; more practical). They were starting points in a process which led to computer dominance in the game of chess.

Other games, especially also the classic Asian game of Go, turned out to be too hard for the Shannon-based approach of pruning and artificial evaluating.

A breakthrough came only within the last 4 years, when Monte Carlo techniques were used in appropriate forms in game tree search. An early predecessor in this direction was Bernd Bruegmanns technical report "Monte Carlo Go" from 1993.

In the talk we give an overview on Monte Carlo techniques in tree and game tree search, including a description of the UCB- and UCT-heuristics for multi-armed bandit problems, and present several "strange" phenomena.

Keywords: Tree search, game tree search, uct, monte carlo go, monte carlo tree search;

On error-tolerant pooling designs based on vector spaces

Harout Aydinian (Bielefeld University, DE)

In the classical group testing model we have a set $[n] = \{1, \ldots, n\}$ of n items containing at most d defective items. The basic problem of group testing is to identify the set of all defective items with a small number of group tests. Each *group test*, also called a *pool*, is a subset of items. It is assumed that there is a testing mechanism that for each subset $A \subset [n]$ gives one of two possible outcomes: *negative* or *positive*. The outcome is positive if A contains at least one defective and is negative otherwise. A group testing algorithm is called *nonadaptive* or a *pooling design* if all tests are specified without knowledge of the outcomes of other tests. Pooling designs have many applications in molecular biology, such as DNA screening, nonunique probe selection, gene detection, etc. A pooling design is associated with a $(0,1)-$ inclusion matrix $M = (m_{ij})$, where the rows are indexed by tests $A_1, \ldots, A_t \subset [n]$, the columns are indexed by items $1, \ldots, n$, and $m_{ij} = 1$ if and only if $j \in A_i$. The major tool used for construction of pooling designs are $d-$*disjunct* matrices. Let M be a binary $t \times n$ matrix where the columns $C_1, \ldots, C_n$ are viewed as subsets of $\{1, \ldots, t\}$ represented by their characteristic vectors. M is called d–disjunct if no column is contained in the union of d others. A d–disjunct matrix of size $t \times n$ can identify up to d defective items using t tests, given number of items n. Moreover, it has an advantage of a simple decoding. A pooling design is called *error-tolerant* if it can detect/correct some errors in test outcomes. Biological experiments are known to be unreliable , which, in fact, is a practical motivation for constructing efficient error-tolerant pooling designs. For error correction in tests the notion of a d^z–*disjunct matrix* was introduced. A d^z–disjunct matrix can detect $z-1$ errors and correct $\lfloor (z-1)/2 \rfloor$ errors. Let $t(d,n)$ denote the minimum number of rows for a d–disjunct matrix with n columns. Asymptotically $t(d,n)$ corresponds to the minimum number of tests needed for identification of at most d defectives. In the literature known are the asymptotic bounds $\Omega(d^2 \log n / \log d) \leq t(d,n) \leq O(d^2 \log n)$.
We present a new construction of error-tolerant pooling designs associated with vector spaces over finite fields. Our construction, by means of d^z–disjunct matrices, is based on packings in finite projective spaces. Given integers $1 \leq s < k \leq \frac{1}{2}m$ and a prime power q we construct a d–disjunct $t \times n$ matrix $P(m,k,s)_q$, where $t = q^m - 1$, $n = q^{(s+1)(m-k)}$, and $d = q^{k-s}$. We show that for any $d \leq q^{k-s}$ the matrix $P(m,k,s)_q$ is d^z–disjunct with $z = q^k - 1 - d(q^s - 1)$. The construction can tolerate many errors and has good performance for practical applications. In particular, for $m = 2k$ and a fixed s the lower bound for $t(d,n)$ is attained.

Joint work of: Ahlswede, Rudolf; Aydinian, Harout

Locating and Detecting Arrays for Interaction Faults

Charles Colbourn (ASU - Tempe, US)

Suppose that k factors may affect the outcome of an experiment. Suppose further that among the k factors, there is one interaction among at most t of the factors that causes an abnormal behaviour or fault. One is to determine a set of $N << k$ tests, each consisting of a subset of the factors to be turned on, while the remainder are turned off. We are to find this fault nonadaptively, by determining which factors participate in the faulty interaction. We consider two variants of this question. In the first, a test yields an outcome of 1 when it contains the faulty interaction and 0 when it does not.

In the second, each factor assumes a real value, and each test results in a value which is the product of the test vector with the vector of factor values. Here the goal is to find the set of factors whose values are nonzero, assuming that there are at most t of them. The first problem is addressed by using locating arrays, a special type of covering arrays. The second is a specific case of compressive sensing of signals. The similarity between the two problems is more than cosmetic; indeed while the first appears to be purely combinatorial and the second linear algebraic, we demonstrate that locating arrays provide useful solutions to the numerical problem arising in compressive sensing.

Keywords: Covering array, locating array, compressive sensing, interaction fault

Joint work of: Colbourn, Charles J.; Syrotiuk, Violet R.; McClary, Daniel W.

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2009/2240

Competitive Group Testing and Learning Hidden Vertex Covers with Minimum Adaptivity

Peter Damaschke (Chalmers UT - Göteborg, SE)

Suppose that we are given a set of n elements d of which are "defective".

A group test can check for any subset, called a pool, whether it contains a defective. It is well known that d defectives can be found by using $O(d \log n)$ pools. This nearly optimal number of pools can be achieved in 2 stages, where tests within a stage are done in parallel. But then d must be known in advance. Here we explore group testing strategies that use a nearly optimal number of pools and a few stages although d is not known to the searcher. One easily sees that $O(\log d)$ stages are sufficient for a strategy with $O(d \log n)$ pools. Here we prove a lower bound of $\Omega(\log d / \log\log d)$ stages and a more general pools vs. stages tradeoff.

As opposed to this, we devise a randomized strategy that finds d defectives using $O(d \log(n/d))$ pools in 3 stages, with any desired probability $1 - \epsilon$.

Open questions concern the optimal constant factors and practical implications. A related problem motivated by, e.g., biological network analysis is to learn hidden vertex covers of a small size k in unknown graphs by edge group tests. (Does a given subset of vertices contain an edge?) We give a 1-stage strategy using $O(k^3 \log n)$ pools, with any FPT algorithm for vertex cover enumeration as a decoder.

Joint work of: Damaschke, Peter; Sheikh Muhammad, Azam

Adaptive coding strategies and the counting of sequences with forbidden strings

Christian Deppe (Bielefeld University, DE)

In [1] we considered the problem to find bounds for the maximal size $M(n,t,q)_f$ of a t-error correcting code with noiseless feedback over a q-ary alphabet. The code concept is suited for communication over a q-ary channel with input and output alphabets $\mathcal{X} = \{0, \ldots, q-1\}$. When a word of length n is sent by the encoder, it is changed by the channel in at most t letters. The sender always knows what the receiver got, because of the noiseless feedback.

For $q = 2$ this model was considered by Berlekamp [2], who derived striking results for triples of performance $(M, n, t)_f$, that is, the number of messages M, block length n and the number of errors t. It is convenient to use the notation of relative error $\tau = t/n$ and rate $R = n^{-1} \log M$. We start in [1] with the investigation in the q-ary case. The Hamming bound $H_q(\tau)$ for $C_q^f(\tau)$, the supremum of the rates achievable for τ and all large n, is a central concept.

We introduced in [1] a new strategy (the rubber method), where we used the "0" as a rubber symbol. Very important here is that the "0" is used only as "protocol" information. In the r-rubber method we used $r+1$ consecutive symbols as a rubber sequence. We showed in [1] for $q \geq 3$

(i)

$$C_q^f(\tau) \begin{cases} \leq H_q(\tau) & \text{if } 0 \leq \tau \leq \frac{1}{q} \\ = (1-2\tau)\log_q(q-1) & \text{if } \frac{1}{q} \leq \tau \leq \frac{1}{2} \\ = 0 & \text{if } \frac{1}{2} \leq \tau \leq 1 \end{cases}$$

(ii) The rate functions F_r obtained by the r-rubber method are tangents to $H_q(\tau)$ going through $(\frac{1}{r+1}, 0)$.

Here we improve the lower bound for $0 \leq \tau \leq \frac{1}{q}$ by using more than one rubber sequence. $\mathcal{R} = \{(0,0,\ldots,0),(1,1,\ldots,1),\ldots,(q_1,q_1,\ldots q_1)\} \subset \{0,1,\ldots,q-1\}^r$ denotes the set of rubbers and $\mathcal{S}_{\mathcal{R}}^n$ denote the skeleton, that are all sequences of length n without a rubber as a subblock. We encode with the function $b : \mathcal{M} \to \mathcal{S}_{\mathcal{R}}^{n-(r+1)t} \times (\mathcal{R}\backslash(/,/,\ldots,/))^t$ our messages. After the transmission the receiver can decode the skeleton and the order in which we use the rubbers. Therefore the rubbers are not only used like in [1] for "protocol" information. We get $R = \frac{1}{n}\log_q(|\mathcal{S}_{IR}^{n-(r+1)t}|q_1^t)$. To establish the rate we have to calculate the number of remaining sequences (without the rubber sequences as subblocks).

We give formulas for the cardinality of a set of sequences of length n with forbidden subblocks. For $q_1 = q-1$ and $r = 2$ we get the rate $(1-2\tau)\log_q(q-1)$, which is the same like in [1]. If we choose $1 < q_1 < q-1$ and $q > 3$ the rates in some intervals are bigger than the rates given in [1].

References

1. R. Ahlswede, C. Deppe, and V. Lebedev, Nonbinary error correcting codes with noiseless feedback, localized errors or both, Annals of European Academy of Sciences, No. 1, 285 - 309, 2005.
2. C.R. Berlekamp, Block coding for the binary symmetric channel with noiseless delayless feedback, Proc. Symposium on Error Correcting Codes, Univ. Visconsin, 1968.

Joint work of: Ahlswede, Rudolf; Deppe, Christian, Lebedev, Vladimir

Minimax Trees in Linear Time with Applications

Travis Gagie (University of Toronto, CA)

A minimax tree is similar to a Huffman tree except that, instead of minimizing the weighted average of the leaves' depths, it minimizes the maximum of any leaf's weight plus its depth. Golumbic (1976) introduced minimax trees and gave a Huffman-like, $O(n \log n)$-time algorithm for building them. Drmota and Szpankowski (2002) gave another $O(n \log n)$-time algorithm, which takes linear time when the weights are already sorted by their fractional parts. In this paper we give the first linear-time algorithm for building minimax trees for unsorted real weights.

Keywords: Data structures, data compression, prefix-free coding

Joint work of: Gawrychowski, Pawel; Gagie, Travis

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2009/2242

See also: Proceedings of the 20th International Workshop on Combinatorial Algorithms; Springer LNCS 5874.

A Generic Top-Down Dynamic-Programming Approach to Prefix-Free Coding

Mordecai Golin (HKUST - Kowloon, HK)

Given a probability distribution over a set of n words to be transmitted, the "Huffman Coding" problem is to find a minimal-cost prefix free code for transmitting or storing those words. The basic Huffman coding problem can be solved in O(n log n) time but variations are more difficult.

One of the standard techniques for solving these variations utilizes a top-down dynamic programming approach.

In this talk we show that this approach is amenable to dynamic programming speedup techniques, permitting a speedup of an order of magnitude for many algorithms in the literature for such variations as mixed radix, reserved length and one-ended coding. These speedups are immediate implications of a general structural property that permits batching together the calculation of many DP entries.

This is joint work with Xiaoming XU and Jiajin YU. A preliminary version appeared in SODA'09.

Keywords: Dynamic Programming, Huffman Coding

Finding at least one defective element in two rounds

Gyula O.H. Katona (Alfréd Rényi Institute of Mathematics - Budapest, HU)

Suppose that some elements of the set $[n] = \{1, 2, \ldots, n\}$ are defective. Their number and positions are unknown. Subsets of $[n]$ can be used for tests. If this test is $A \subset [n]$ then the result of the test is YES if at least one of the defectives is in A. Otherwise the answer is NO. The goal of the search is to find at least one of the defectives, or to claim that there is none. We prove that the number of tests is at least n in the non-adaptive case. On the other hand, if two rounds can be used, then the number of tests in the worst case is at least $(2 + o(1))\sqrt{n}$, and this is sharp.

Keywords: Search of at least one, two rounds

Some Aspects of Finite State Channel related to Hidden Markov Process

Kingo Kobayashi (National Inst. of Inform. & Communications Techn., JP)

We have no satisfactory capacity formula for most channels with finite states.

Here, we consider some interesting examples of finite state channels, such as Gilbert-Elliot channel, trapdoor channel, etc., to reveal special characters of problems and difficulties to determine the capacities.

Meanwhile, we give a simple expression of the capacity formula for Gilbert-Elliot channel by using a hidden Markov source for the optimal input process. This idea should be extended to other finite state channels.

Keywords: Finite state channel, Hidden Markov source, Gilbert-Elliot channel, Trapdoor Channel

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2009/2243

Memory/Time Tradeoffs for Rendezvous on a Ring

Evangelos Kranakis (Carleton Univ. - Ottawa, CA)

The *rendezvous* problem (gathering of agents widely dispersed in some domain at a common place and time) has been studied under many guises and in many settings. We present deterministic and randomized algorithms for rendezvous on synchronous, anonymous, oriented rings. We give tight time (as well as expected time)/memory tradeoffs for two identical agents to rendezvous.

Keywords: Mobile agent, rendezvous, deterministic, randomized

Covert Combinatorial DNA Taggant Signatures for Authentication, Tracking and Trace-Back

Anthony J. Macula (SUNY - Geneseo, US)

This effort attempts to develop a new biomolecular covert taggant method that marks objects with synthetic DNA nano-scale biosensors, called ComDTags, for verification of authenticity and for forensic track and trace-back. Pharmaceutical companies, brand owners and governments need thousands of covert and unique signatures to protect their drugs, products, documents and citizens. This effort hopes to demonstrate that the ComDTag DNA taggant system can construct millions of unique, information-rich and covert ComDTag signatures and that these signatures can be detected and decoded only by authorized users.

The ComDTag system "sandwiches" a piece of molecular biology between two "slices" of mathematics. First, it uses mathematics to design the synthetic DNA that makes the storage of information in ComDTags possible. Then it uses the specificity of DNA strand recognition and the wet laboratory method of polymerase chain reaction (PCR) to store information and to generate a signal. Finally, it uses mathematical group testing to decode the PCR signal and identify the ComDTag signatures and the information they contain.

Counterfeiting costs the U.S. economy 200 billion per year and is responsible for the loss of 750,000 American jobs. Since 2000, the number of U.S. counterfeit drug cases has increased 800%. In developing countries, 20% of medicines are thought to be counterfeit, costing legitimate businesses 75 billion worldwide. The counterfeiting of the drug heparin has been linked to the deaths of over a hundred Americans with hundreds more having severe allergic reactions. These disturbing facts indicate that there is both an increasing commercial and national security need for safe and covert means to prevent, track and trace-back counterfeit products. Covert ComDTag signatures can feasibly help to meet these needs.

Combined with more overt taggants, ComDTag signatures can yield more robust, secure and useful covert track and trace-back solutions for commercial and national security interests. There are feasibly millions of unique ComDTag signatures available and they have unique layering (superimposable) capabilities so that they can be repeatedly applied to products manufactured or assembled at both primary and secondary production facilities. This gives ComDTag the additional capability to protect processes, e.g., drug synthesis and supply chains.

Keywords: Taggants, Group testing; Polymerase chain reaction, Combinatorial designs, Biomolecular computing, DNA codes, DNA sequence design, Real-time PCR, Graph theory, Networks

Competitive Evaluation of Threshold Functions and Game Trees in the Priced Information Model

Martin Milanic (University of Primorska, SI)

Consider the following *function evaluation problem*: A function $f : V \to \mathbb{R}$ has to be evaluated for a fixed but unknown *assignment* σ, i.e., a choice of the values for the set of variables $V = \{x_1, x_2, \ldots, x_n\}$.

Each variable x_i has an associated non-negative cost $c(x_i)$ which is the cost incurred to probe x_i, i.e., to read its value $x_i(\sigma)$. For each $i = 1, \ldots, n$, the cost $c(x_i)$ is fixed and known beforehand.

The goal is to *adaptively* identify and probe a minimum cost set of variables $U \subseteq V$ whose values uniquely determine the value of f for the given assignment, regardless of the value of the variables not probed. The cost of U is the sum of the costs of the variables it contains, i.e., $c(U) = \sum_{x \in U} c(x)$. We use $f(\sigma)$ to denote the value of f w.r.t. σ, i.e., $f(\sigma) = f(x_1(\sigma), \ldots, x_n(\sigma))$.

A set of variables $U \subseteq V$ is a *proof* with respect to a given assignment σ for the variables of V if the value $f(\sigma)$ is determined by the restriction $\sigma_{|U}$ of σ to U. An evaluation algorithm $\mathbb{A}$ for f is a rule to adaptively read the variables in V until the set of variables read so far is a proof for the value of f. The cost of algorithm $\mathbb{A}$ for an assignment σ is the total cost incurred by $\mathbb{A}$ to evaluate f under the assignment σ.

Given a cost function $c(\cdot)$, we let $c^f_{\mathbb{A}}(\sigma)$ denote the cost of the algorithm $\mathbb{A}$ for an assignment σ and $c^f(\sigma)$ the cost of the cheapest proof for f under the assignment σ. We say that $\mathbb{A}$ is ρ-competitive if $c^f_{\mathbb{A}}(\sigma) \leq \rho c^f(\sigma)$, for every possible assignment σ. We use $\gamma_c^{\mathbb{A}}(f)$ to denote the competitive ratio of $\mathbb{A}$, that is, the infimum over all values of ρ for which $\mathbb{A}$ is ρ-competitive. The best possible competitive ratio for any deterministic algorithm, then, is

$$\gamma_c^f = \inf_{\mathbb{A}} \gamma_c^{\mathbb{A}}(f),$$

where the infimum is computed over all possible deterministic algorithms $\mathbb{A}$.

With the aim of evaluating the dependence of the competitive ratio on the structure of f, one defines the extremal competitive ratio $\gamma^{\mathbb{A}}(f)$ of an algorithm $\mathbb{A}$ as

$$\gamma^{\mathbb{A}}(f) = \sup_c \gamma_c^{\mathbb{A}}(f).$$

The best possible extremal competitive ratio for any deterministic algorithm, then, is

$$\gamma(f) = \inf_{\mathbb{A}} \gamma^{\mathbb{A}}(f).$$

This last measure is meant to capture the structural complexity of f independent of a particular cost assignment and algorithm.

The above framework for studying the function evaluation problem is due to Charikar *et al.* [1].

In [2], Cicalese and Laber introduced a new linear-programming-based approach (the $\mathcal{LPA}$) for designing competitive algorithms for the function evaluation problem. The $\mathcal{LPA}$ is based on the solution of a linear program defined on the set of minimal proofs of the function in question. Cicalese and Laber proved that for any monotone Boolean function the $\mathcal{LPA}$ provides algorithms with the best extremal competitive ratio.

Moreover, for some proper subclasses of the monotone Boolean functions, this competitiveness is achievable with efficient (polynomial time) implementation of the $\mathcal{LPA}$.

Among such classes is the class of functions admitting a compact circuit representation with k-out-of-n gates (aka threshold trees). This class also includes the classic read once functions (aka AND-OR trees).

The existence of an efficient implementation of the $\mathcal{LPA}$ for general monotone Boolean functions remains a major open problem.

In one recent and one upcoming publication [3,4], we obtain some further developments in this line of research:

- We extend the result by Cicalese and Laber to the case when the cost of reading a variable depends on the value of the variable.
- We study the class of threshold functions, which generalize k-out-of-n functions and have applications in several contexts. We show an interesting connection between the separating structures of threshold functions and the solution of the LP used by the $\mathcal{LPA}$: A direct consequence of this result is the existence of a polynomial implementation of the $\mathcal{LPA}$ with the best competitiveness against the worst case costs for threshold functions given via a separating structure. We also show that a pseudo-polynomial implementation of the $\mathcal{LPA}$ exists for the class of functions that are representable

by read once formulas whose connectives are threshold functions given by their separating structure. In the case the threshold functions are provided via their complete DNF our algorithm runs in polynomial time.
– We characterize the best possible extremal competitive ratio for the class of game tree functions, which are a generalization of AND/OR tree functions where AND and OR are replaced by MIN and MAX and variables are not restricted to be Boolean valued.

References

1. M. Charikar, R. Fagin, V. Guruswami, J. M. Kleinberg, P. Raghavan, and A. Sahai. Query strategies for priced information. *Journal of Computer and System Sciences*, 64(4):785–819, 2002.
2. F. Cicalese and E. S. Laber. Function evaluation via linear programming in the priced information model. In *Proc. of ICALP 2008*, LNCS **5125**, pp. 173–185. 2008.
3. F. Cicalese and M. Milanič. Computing with priced information: when the value makes the price. In *Proc. of ISAAC 2008*, LNCS **5369**, pp. 378–389. 2008.
4. F. Cicalese and M. Milanič. Competitive evaluation of threshold functions in the priced information model. Accepted for publication in *Annals of Operations Research*.

Keywords: Function evaluation, priced information, competitive analysis, threshold function, game tree

Joint work of: Cicalese, Ferdinando; Milanic, Martin

Iterative algorithms for low-rank matrix completion

Olgica Milenkovic (Univ. of Illinois - Urbana, US)

We present an overview of algorithms for low-rank matrix completion problems, both for the error-free and noisy reconstruction regime. We then proceed to describe a new class of methods inspired by compressive sensing algorithms suitable for both sparse and dense matrices of low-rank.

Keywords: Compressive sensing, low-rank matrix completion

Joint work of: Dai, Wei; Milenkovic, Olgica

Explicit Non-Adaptive Combinatorial Group Testing Schemes

Ely Porat (Bar-Ilan University - Ramat-Gan, IL)

Group testing is a long studied problem in combinatorics: A small set of r ill people should be identified out of the whole (n people) by using only queries (tests) of the form "Does set X contain an ill human?".

In this paper we provide an explicit construction of a testing scheme which is better (smaller) than any known explicit construction. This scheme has $\Theta(min[r^2 \log n, n])$ tests which is as many as the best non-explicit schemes have. In our construction we use a fact that may have a value by its own right: Linear error-correction codes with parameters $[m, k, \delta m]q$ meeting the Gilbert-Varshamov bound may be constructed quite efficiently, in $\Theta[q^k m)$ time.

Keywords: Prime Numbers, Group Testing, Streaming, Pattern Matching

Joint work of: Porat, Ely; Rotschild, Amir

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2009/2241

Pattern matching with don't cares and few errors

Ely Porat (Bar-Ilan University - Ramat-Gan, IL)

We present solutions for the k-mismatch pattern matching problem with don't cares. Given a text t of length n and a pattern p of length m with don't care symbols and a bound k, our algorithms find all the places that the pattern matches the text with at most k mismatches. We first give an $\Theta(n(k + \log m \log k) \log n)$ time randomised algorithm which finds the correct answer with high probability. We then present a new deter- ministic $\Theta(nk^2 \log^2 m)$ time solution that uses tools originally developed for group testing. Taking our derandomisation approach further we develop an approach based on k-selectors that runs in $\Theta(nk\text{polylog}m)$ time. Further, in each case the location of the mismatches at each alignment is also given at no extra cost.

Keywords: Prime Numbers, Group Testing, Streaming, Pattern Matching

Joint work of: Porat, Ely; Rotschild, Amir; Clifford, Raphael; Efremo, Klim

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2009/2244

Graph Entropy, Network Coding, and Guessing Games

Sören Riis (Queen Mary, University of London, UK)

We introduce the (private) entropy of a directed graph (in a new network coding sense) as well as a number of related concepts. We show that the entropy of a directed graph is identical to its guessing number and can be bounded from below with the number of vertices minus the size of the graph's shortest index code. We show that the Network Coding solvability of each specific multiple unicast network is completely determined by the entropy (as well as by the shortest index code) of the directed graph that occur by identifying each source node with each corresponding target node. Shannon information inequalities can be used to calculate upper bounds on a graph entropy as well as calculating the size of the minimal index code. Recently, a number of new families of so-called non-shannon-type information inequalities have been discovered. It has been shown that there exist communication networks with a capacity strictly ess than required for solvability, but where this fact cannot be derived using Shannon's classical information inequalities. Based on this result we show that there exist graphs with an entropy that cannot be calculated using only Shannon's classical information inequalities, and show that better estimate can be obtained by use of certain non-shannon-type information inequalities.

Searching for defective edges in graphs and hypergraphs

Eberhard Triesch (RWTH Aachen, DE)

Suppose a hypergraph of rank r is given and d of its hyperedges are defective.

Find all defective hyperedges by using the following tests:

For each subset W of the vertex set, we may test whether at least one defective hyperedge is a subset of W.

We are looking for the worst case behaviour of sequential algorithms.

(The case d=1 corresponds to group testing.) The talk surveys the known results, particularly if the hypergraph is a graph or a 3-uniform hypergraph. A new algorihm for the 3-uniform hypergraphs is presented.

Keywords: Graph, hypergraph, group testing

Joint work of: Triesch, Eberhard; Korneffel, Torsten

Rounds in Combinatorial Search

Gabor Wiener (Budapest Univ. of Technology & Economics, HU)

The search complexity of a separating system $\mathcal{H} \subseteq 2^{[m]}$ is the minimum number of questions of type "$x \in H$?" (where $H \in \mathcal{H}$) needed in the worst case to determine a hidden element $x \in [m]$.

If we are allowed to ask the questions in at most k batches then we speak of the *k-round* (or *k-stage*) complexity of $\mathcal{H}$, denoted by $\mathrm{c}_k(\mathcal{H})$. While 1-round and m-round complexities (called non-adaptive and adaptive complexities, respectively) are widely studied (see for example Aigner [**?**]), much less is known about other possible values of k, though the cases with small values of k (tipically $k = 2$) attracted significant attention recently, due to their applications in DNA library screening.

It is clear that $|\mathcal{H}| \geq \mathrm{c}_1(\mathcal{H}) \geq \mathrm{c}_2(\mathcal{H}) \geq \ldots \geq \mathrm{c}_m(\mathcal{H})$.

A group of problems raised by G. O. H. Katona is to characterize those separating systems for which some of these inequalities are tight. In this paper we are discussing set systems $\mathcal{H}$ with the property $|\mathcal{H}| = \mathrm{c}_k(\mathcal{H})$ for any $k \geq 3$. We give a necessary condition for this property by proving a theorem about traces of hypergraphs which also has its own interest.

Keywords: Search, adaptive, non-adaptive, hypergraph, trace

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2009/2239

Existence theorems for search problems with lies

Anatoly Zhigljavsky (Cardiff University, GB)

For a wide range of search problems with lies we derive upper bounds for the length of optimal nonadaptive algorithms. The method of deriving these bounds is probabilistic and is therefore not constructive; it does not provide ways to construct the optimal algorithms. The search problems we consider include many known combinatorial group testing problems such as testing in binary, additive and multiaccess channel models.

For several general models we show that the leading asymptotic term for the number of tests does not depend on the number of lies and is thus the same as for the zero-lie case. However, the other terms in the asymptotic upper bounds depend on the number of lies and make a serious impact on the upper bounds in the nonasymptotic situation.

Keywords: Discret search, search with lies, Probabilistic method, Existence theorems, Group testing