07391 Abstracts Collection
Probabilistic Methods in the Design and Analysis of Algorithms
— Dagstuhl Seminar —

Martin Dietzfelbinger[1], Shang-Hua Teng[2], Eli Upfal[3], and Berthold Vöcking[4]

[1] TU Ilmenau, DE
martin.dietzfelbinger@tu-ilmenau.de
[2] Boston Univ., US
steng@cs.bu.edu
[3] Brown Univ. - Providence, US
eli@cs.brown.edu
[4] RWTH Aachen, DE
voecking@cs.rwth-aachen.de

Abstract. From 23.09.2007 to 28.09.2007, the Dagstuhl Seminar 07391 "Probabilistic Methods in the Design and Analysis of Algorithms" was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. The seminar brought together leading researchers in probabilistic methods to strengthen and foster collaborations among various areas of Theoretical Computer Science. The interaction between researchers using randomization in algorithm design and researchers studying known algorithms and heuristics in probabilistic models enhanced the research of both groups in developing new complexity frameworks and in obtaining new algorithmic results. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Links to extended abstracts or full papers are provided, if available.

Keywords. Algorithms, Randomization, Probabilistic analysis, Complexity

It is difficult to overstate the importance of probabilistic methods in Theoretical Computer Science. They belong to the most powerful and widely used tools, for example in designing efficient randomized algorithms for tackling hard optimization problems; in establishing various lower bounds in complexity theory; in the proofs of many useful discrete properties in extremal combinatorics; in providing frameworks such as the average-case and smoothed analysis for measuring the performance of algorithms; in the theory of interactive proofs. The body of work using probabilistic methods has experienced an impressive growth in the recent years. The following topics attracted enormous attention both from theorists as well as practitioners during the recent years.

In the area of randomized algorithms, several new probabilistic techniques were developed. For example, there are several exciting recent developments in the probabilistic metric embedding with tree metrics. Because various optimization problems can be solved optimally on trees (e.g., by the dynamic programming approach), quality approximations of arbitrary metrics by tree metrics provide a systematic approach for designing approximation algorithms for general metrics. Further, new techniques for designing randomized data structures were developed that draw on methods from the theory of random graphs and random walks in graphs. A core issue here is the efficient simulation of high-degree randomness without the assumption of the inputs being random.

Impressive progress has also been obtained regarding the probabilistic analysis of algorithms. In practice, scientists and engineers often use heuristic algorithms for optimization problems ranging from network design to industrial optimizations. Most of these algorithms, after years of improvements, work well in practice. However, their worst-case complexity might still be very poor, for example, exponential in the input size. It is an old observation in quite a few application areas that the worst-case instances of an algorithm might not be "typical" and might never occur in practice. So worst-case analysis can improperly suggest that the performance of the algorithm is poor. Trying to rigorously understand and model the practical performance for such heuristic algorithms is a major challenge in Theoretical Computer Science.

Probabilistic methods have played an active role in developing analysis frameworks that provide "practical enough" measures, yet one can still conduct rigorous analyses using these frameworks. For example, the recently developed *smoothed analysis* uses small random perturbations for defining performance measures. This framework applies to algorithms whose inputs are subject to slight random noises. The smoothed complexity of an algorithm is then the maximum over its inputs of the expected running time of the algorithm under slight perturbations of that input. Smoothed complexity is measured in terms of the size of the input and the magnitude of the perturbation.

Another area in which random inputs play an important role is stochastic optimization. Here uncertainty in the data is modeled by probability distributions. Stochastic optimization has a wide range of applications in various areas, including logistics, transportation, financial instruments, and network design. In recent years, there has been significant progress in analyzing important algorithms and heuristics used in this field. For example, the sample average approximation (SAA) method solves stochastic programs by sampling from the distribution of input scenarios. Recent theoretical results show that the SAA method has the properties of a fully randomized approximation scheme for a large class of multistage stochastic optimization problems.

The workshop covered recent progress in randomized algorithms and probabilistic measures of algorithms including the smoothed analysis, average-case analysis, semi-random analysis, and stochastic optimization. The presentations covered a large range of optimization problems such as linear programming, integer programming, random games, computational geometry, and scheduling. The most important contribution of the seminar is the exchange of new ideas between researchers using probabilistic methods in different contexts. In addition of providing an opportunity for information sharing and collaborations, the workshop exposed young researchers, students, and postdocs to recent developments and outstanding issues in probabilistic methods.

Epidemiology and Graph Cuts

Elliot Anshelevich (Rensselaer Polytechnic, USA)

The connection between epidemiology and graph cuts leads to many new problems that are both important and interesting. Specifically, we consider the question of immunizing a few individuals in a social network in order to stop or slow the process of an epidemic. We give a short overview of known results in this area, and point out fundamental problems that are still open. For example, given a graph, a set of infected nodes, and transmission probabilities on edges, no good algorithms currently exist for finding the nodes to "immunize" in order to save as many lives as possible.

Joint work of: Anshelevich, Elliot; Kempe, David

Computing Local Statistics on Massive Graphs

Luca Becchetti (University of Rome "La Sapienza", I)

Web applications and monitoring tools often generate large or huge amounts of data. Many data collections are stored as large scale graphs (e.g., Web crawls, social networks etc.).

We propose efficient techniques to compute link-based statistical indices that measure the local degree of clustering in the neighborhoods of nodes of large graphs. In particular, we consider the number of triangles and the related local clustering coefficient. Previous work concentrates on the total number of triangles.

Our algorithms use or refine known probabilistic techniques to compute estimations of such indices for all nodes at the same time, using an amount of time and space that scale gracefully with the network size.

The statistics we are interested in can be used later for ranking, Web spam detection and other tasks, or they can be used in the study of the Web or other complex networks that exhibit small-world properties.

Keywords: Massive data sets, Minwise random permutations, Semi-streaming

Joint work of: Becchetti, Luca; Boldi, Paolo; Castillo, Carlos; Gionis, Aristides

Balanced Allocations: Balls-into-Bins Revisited and Chains-into-Bins

Petra Berenbrink (Simon Fraser University, CA)

The study of balls-into-bins games or occupancy problems has a long history since these processes can be used to translate realistic problems into mathematical ones in a natural way. In general, the goal of a balls-into-bins game is to allocate a set of independent objects (tasks, jobs, balls) to a set of resources (servers, bins, urns) and, thereby, to minimize the maximum load.

In this paper we analyse the maximum load for the chains-into-bins problem where we have n bins and the balls are connected in n/ℓ chains of length ℓ. In this process, the balls of one chain have to be allocated to ℓ consecutive bins. We allow each chain d i.u.r. bin choices. The chain is allocated using the rule that the maximum load of any bin receiving a ball of that chain is minimized. We show that, for $d > 1$, the maximum load is $\ln\ln(n/\ell)/\ln d + O(1)$ with probability $1-O(1/\ln\ln(n/\ell))$. This shows that the maximum load is decreasing with increasing chain length.

Keywords: Balls-into-bins, Chains-into-bins

Joint work of: Batu, Tugkan; Berenbrink, Petra; Cooper, Colin

Single-Source Stochastic Routing

Shuchi Chawla (University of Wisconsin - Madison, USA)

We introduce and study the following model for routing uncertain demands through a network. We are given a capacitated multicommodity flow network with a single source and multiple sinks, and demands that have known values but unknown sizes. We assume that the sizes of demands are governed by independent distributions, and that we know only the means of these distributions and an upper bound on the maximum-possible size. Demands are irrevocably routed one-by-one, and the size of a demand is unveiled only after it is routed. A routing policy is a function that selects an unrouted demand and a path for it, as a function of the residual capacity in the network. Our objective is to maximize the expected value of the demands successfully routed by our routing policy. We distinguish between safe routing policies, which never violate capacity constraints, and unsafe policies, which can attempt to route a demand on any path with strictly positive residual capacity. We design safe routing policies that obtain expected value close to that of an optimal unsafe policy in planar graphs. Unlike most previous work on similar stochastic optimization problems, our routing policies are fundamentally adaptive. Our policies iteratively solve a sequence of linear programs to guide the selection of both demands and routes.

Keywords: Stochastic optimization, Routing, Iterative rounding

Full Paper:
http://www.cs.wisc.edu/~shuchi/papers/stoch-routing.pdf

The Solution Space Geometry of Random Constraint Satisfaction Problems

Amin Coja-Oghlan (HU Berlin, D)

Random instances of constraint satisfaction problems such as k-SAT or k-Coloring are notoriously hard for many algorithmic techniques, in particular if the density of the instance is close to the threshold for the existence of a solution. Recent non-rigorous work from the statistical physics community has led to new algorithms (e.g., "Survey Propagation") that perform emprically very well on random instances. These algorithms are based on hypotheses about the "solution space geometry" of random constraint satifaction problems, i.e., the number and relative location of the solutions. In this talk I will discuss these hypotheses as well as their algorithmic implications, and present some recent mathematically rigorous work on proving some of the hypotheses.

Keywords: Constraint satisfaction problems, Random structures, Survey propagation

Joint work of: Achlioptas, Dimitris; Coja-Oghlan, Amin; Frieze, Alan

Testing Expansion in Bounded-Degree Graphs

Artur Czumaj (University of Warwick, GB)

We consider the problem of testing expansion in bounded degree graphs. We focus on the notion of vertex-expansion: an a-expander is a graph $G = (V, E)$ in which every subset $U \subseteq V$ of at most $|V|/2$ vertices has a neighborhood of size at least $\alpha|U|$. Our main result is that one can distinguish good expanders from graphs that are far from being weak expanders in time $\widetilde{O}(\sqrt{n})$. We prove that the property testing algorithm proposed by Goldreich and Ron (2000) with appropriately set parameters accepts every α-expander with probability at least $2/3$ and rejects every graph that is ε-far from an α^*-expander with probability at least $2/3$, where $\alpha^* = \Theta(\frac{\alpha^2}{d^2 \log(n/\varepsilon)})$ and d is the maximum degree of the graphs. The algorithm assumes the bounded-degree graphs model with adjacency list graph representation and its running time is $O(\frac{d^2\sqrt{n}\log(n/\varepsilon)}{\alpha^2\varepsilon^3})$.

Keywords: Property testing, Expansion, Randomized algorithms

Joint work of: Czumaj, Artur; Sohler, Christian

Quasirandomness

Benjamin Doerr (MPI für Informatik - Saarbrücken, D)

An object is called quasirandom if it imitates a particular property of a random object. Jim Propp's "rotor router model" is a quasirandom analogue of a random walk.

A rotor on each vertex ensures that each vertex serves its neighbors in an as balanced manner as possible. In this talk, we review several famous results and findings on quasirandom walks (using the rotor router mechanism instead of random decisions). We shall then present two very recent results, which will both appear in the proceedings of SODA 2008. The first is that infinite trees may display an unbounded discrepancy between the Propp and the random model. The second is the first application of the rotor router idea to an algorithmic problem. We show that most results for the classical rumor spreading protocol still hold if we use a mostly quasirandom protocol (only the initial rotor directions are chosen at random). In fact, for several network topologies the broadcast time improves in the quasirandom model.

Keywords: Quasirandomness, Rotor router model, Propp machine, Broadcast

Joint work of: Cooper, Joshua; Doerr, Benjamin; Friedrich, Tobias; Sauerwald, Thomas; Spencer, Joel

How Small Worlds Emerge

Devdatt Dubhashi (Chalmers UT - Göteborg, S)

We discuss some models of evolving random graphs that explain the mergence of small worlds and connections with the celebrated model of Jon Kleinberg.

An Efficient Randomized Controller for Dynamic Networks

Yuval Emek (Weizmann Inst. - Rehovot, IL)

Consider the setting in which events occur in an online fashion at arbitrary nodes of a distributed network modeled as an undirected tree.

Events may be *topological*, that is, insertion or removal of nodes, or *non-topological*.

It is required for the process to halt when the number t of events that occurred throughout the network is within some given range.

We introduce the notion of an (M, W, q)-*randomized controller* which signals termination when t satisfies

1. $t \geq M - W$; and
2. with probability at least $1 - q$, $t \leq M$.

The efficiency of a randomized controller is measured by the expected number of messages it sends.

We design an (M, W, q)-randomized controller that sends $O\big(N \log^2 \log(1/q) \cdot \log(M/(W+1))\big)$ messages on expectation, where N is the total number of nodes that ever existed in the network.

By fixing $q = 2^{-\log^c N}$ for any constant $c > 1$, we obtain a randomized controller with a (one-sided) negligible failure probability and expected message complexity $O\left(N \log^2 \log N \log(M/(W+1))\right)$.

This work should be put in contrast with the seminal work of Afek, Awerbuch, Plotkin, and Saks and with the recent work of Korman and Kutten which investigates the deterministic version of the problem.

They construct an (M, W)-*controller*, which is essentially an $(M, W, 0)$-randomized controller, that sends $O\left(N \log^2 N \log(M/(W+1))\right)$ messages.

Due to the establishment of a simple $\Omega\left(N \log(M/(W+1))\right)$ lower bound on the message complexity required to implement an (M, W, q)-randomized controller for any constant $q < 1$, our construction provides an exponential improvement (with respect to the gap from the lower bound) at the cost of admitting a one-sided negligible failure probability.

Keywords: Online algorithms, Distributed algorithms, Dynamic networks, Controller, Size estimation

Worst Case and Probabilistic Analysis of the 2-Opt Algorithm for the TSP

Matthias Englert (RWTH Aachen, D)

2-Opt is probably the most basic and widely used local search heuristic for the TSP. This heuristic achieves amazingly good results on “real world” Euclidean instances both with respect to running time and approximation ratio. There are numerous experimental studies on the performance of 2-Opt. However, the theoretical knowledge about this heuristic is still very limited. Not even its worst case running time on Euclidean instances was known so far. We clarify this issue by presenting a family of Euclidean instances on which 2-Opt can take an exponential number of steps.

Previous probabilistic analyses were restricted to instances in which n points are placed uniformly at random in the unit square where it was shown that the expected number of steps is bounded by $O(n^{10})$ for Euclidean instances. We consider a more advanced model of probabilistic instances in which the points can be placed according to general distributions on the unit square. In particular, we allow different distributions for different points. We study the expected running time in terms of the number n of points and the maximal density phi of the probability distributions. We show an upper bound on the expected length of any 2-Opt improvement path of $O(n^{4+1/3}\phi^{8/3})$. In addition, we prove an upper bound of $O(\sqrt{\phi})$ on the expected approximation factor.

Joint work of: Englert, Matthias; Röglin, Heiko; Vöcking, Berthold

Fluctuating Frugality: VCG Path Auction in the Giant Component

Abraham Flaxman (Microsoft Research - Redmond, USA)

Packet routing in the Internet is a natural application area for algorithmic mechanism design. It appeared as an example application in the seminal work [N. Nissan and A. Ronen, Algorithmic mechanism design, *Proc. of the 31st ACM Symposium on Theory of Computing* (1999) 129–140], and has been the subject of on-going study ever since. The present paper investigates a simple game-theoretic model of routing in a graph, where a selfish agent controls each edge, and the goal is to route a packet between vertices s and t along the shortest path. The Vickery-Clarke-Groves (VCG) mechanism is a well-known auction protocol which obtains minimum-length (s, t)-path by inducing truthful behavior in selfish agents. Previous work has shown that the frugality ratio (the ratio of the VCG payment to the shortest path length) may be arbitrarily large. In real-world graphs, however, the observed frugality ratio has not been too large (around 1.3). This was partially explained in [M. Mihail, C. Papadimitriou, and A. Saberi, On certain connectivity properties of the Internet topology, *Proc. of the 44th IEEE Symposium on Foundations of Computer Science* (2003) 28–35], which provided bounds on frugality ratio in the Erdős-Rényi graph $G_{n,d/n}$. If F denotes the expected frugality ratio, then the previous work shows that $1 + \Omega\left(d^{-1}\right) \leq F \leq O(1)$. In this paper, we obtain a more precise bound on the frugality ratio in $G_{n,d/n}$, and show that, as a function of d, $G_{n,d/n}$ has $\liminf_{n\to\infty}(F - 1) = \Theta\left(d^{-1}\right)$ and $\limsup_{n\to\infty}(F - 1) = \Theta(1)$. This requires deriving very precise bounds on the average length of the shortest path and the "second shortest" path between vertices in a sparse random graph.

Keywords: Random graphs, VCG path auction

Joint work of: Flaxman, Abraham; Reed, Bruce

Selfish Load Balancing

Tom Friedetzky (University of Durham, GB)

We investigate the convergence to Nash equilibria of different (selfish) reallocation protocols. Our models include uniform and weighted tasks and uniform resources. We show upper and lower bounds on the convergence time to exact and approximate equilibria.

Keywords: Load balancing, Equilibria

Joint work of: Berenbrink, Petra; Friedetzky, Tom; Goldberg, Leslie Ann; Goldberg, Paul; Hajirasouliha, Iman; Hu, Zengjian; Martin, Russel

See also: Distributed Selfish Load Balancing (SODA'06) and Convergence to Equilibria in Distributed, Selfish Reallocation Processes with Weighted Tasks (ESA'07)

Weak Sparse Regularity and Combinatorial Optimization

Alan M. Frieze (CMU - Pittsburgh, USA)

Szemerédi's regularity lemma has proved to be of fundamental importance in combinatorics. It is also useful in the context of finding approximation algorithms for many *dense* problems. For this purpose, the full strength of the lemma is not needed and Frieze and Kannan introduced the notion of a *weak regularity lemma.* Kohayakawa and Rödl introduced a sparse version of the regularity lemma that is applicable to graphs without "dense" spots. In this paper we show how to construct a sparse weak regular partition and how it can sometimes be used in approximation schemes.

Keywords: Weak sparse regularity, Approximation schemes

Joint work of: Coja-Oghlan, Amin; Cooper, Colin; Frieze, Alan

On Random Betweenness Constraints

Andreas Goerdt (TU Chemnitz, D)

Despite of their wide applicability, in bioinformatics and AI for example, betweenness constraints have (to the authors' knowledge) not been considered from the random structures point of view. They are essentially different from classical constraints like satisfiability, graph colouring,...in that the underlying domain for each variable is unbounded. In case of betweenness constraints assignments presenting possible solutions are total orderings of the variables. And a betweenness constraint is a set of triples (x, y, z) of three distinct variables meaning semantically that y must lie between x and z in the ordering considered.

Thus we get the obvious satisfiability problem. This problem is known to be NP-complete as shown by Opatrny already in the seventies.

Picking betwenness constraints with cn randomly chosen triples, n being the number of variables considered, one can see that typical instances start to get unsatisfiable at $c \approx 0.8$ and are unsatisfiable for $c \approx 1.3$. We considered n up to 30 and observed that the transition satisfiable-unsatisfiable gets steeper with increasing n. Thus in line with analogous problems, in particular random 3-SAT being heavily investigated recently, we might think that the following "betweenness constraint threshold conjecture" holds: There exists a constant C threshold value such that instances become unsatisfible at this threshold C. (This can be made more precise in a quite natural way.) It is obvious to try to prove relevant results.

Often in analogous situations like 3-SAT a theorem of Friedgut is applicable. This theorem shows that the transition satisfiable-unsatisfiable is sharp, but does not imply the existence of a *fixed* constant C at which this transition happens. Interesting enough this theorem does not seem to be easily applicable here and the sharpness of the transition does not seem to be known by now.

We prove that betweenness constraints for $c \approx 2.55$ typically are unsatisfiable. On the other hand we show that constraints for $c < 1/4$ are satisfiable with probability $> \varepsilon > 0$. This says at least that they are not unsatisfiable with high probabilty. Note that the giant component threshold for 3-uniform hypergraphs occurs for $(1/6) \times n$ many hyperedges.

Thus the second result does not hold on simple grounds that the underlying hypergraph is of a very simple structure.

Instead we prove the absence of a certain specialized kind of cycles in the underlying hypergraph of a random betweenness constraint.

With the absence of these cycles a quite natural heuristics to construct a satisfying ordering can be shown to succeed.

Keywords: Random structures, Betwenness constraints

Joint work of: Goerdt, Andreas; Lanka, André

Matrix Norms and Rapid Mixing for Spin Systems

Leslie Ann Goldberg (University of Liverpool, GB)

We give a systematic development of the application of matrix norms to rapid mixing in spin systems. We show that rapid mixing of both random update Glauber dynamics and systematic scan Glauber dynamics occurs if any matrix norm of the associated dependency matrix is less than 1. We give improved analysis for the case in which the diagonal of the dependency matrix is 0 (as in heat bath dynamics). We apply the matrix norm methods to random update and systematic scan Glauber dynamics for colouring various classes of graphs. We give a general method for estimating a norm of a symmetric non-regular matrix. This leads to improved mixing times for any class of graphs which is hereditary and sufficiently sparse including several classes of degree-bounded graphs such as non-regular graphs, trees, planar graphs and graphs with given tree-width and genus. This paper is joint work with Martin Dyer and Mark Jerrum.

Joint work of: Dyer, Martin; Goldberg, Leslie Ann; Jerrum, Mark

Full Paper:
http://arxiv.org/abs/math/0702744

Recent Progress in Computing Approximate Nash Equilibria

Paul Goldberg (University of Liverpool, GB)

It is known that it is "hard" to compute exact Nash equilibria, in particular this is the case for two-player games in normal form. In this talk I give an overview of some recent results on the complexity of computing approximate Nash equilibria of two-player games.

Keywords: Algorithms, Economics, Theory

Stochastic Analysis of Online Steiner Tree

Anupam Gupta (CMU - Pittsburgh, USA)

In the online Steiner tree problem, vertices come one-by-one and have to be connected to the root. The greedy algorithm is an $O(\log n)$ competitive algorithm, and this is the best possible. One can obtain better guarantees when the input sequence is not chosen by an adversary, but instead are draws from some distribution over the vertices of the graph.

Keywords: Online algorithms, Stochastic analysis

See also: To Appear in SODA 2008

On Embedding Edit Distance into L_1

Robert Krauthgamer (Weizmann Inst. - Rehovot, IL)

The edit distance (aka Levenshtein distance) between two strings is the number of character insertions, deletions and substitutions required to transform one string to the other. A very powerful paradigm for solving computational problems on the metric space induced by the edit distance is to embed this metric into L_1, using a low-distortion map (if possible).

I will first present a low-distortion embedding of edit distance on permutations (aka the Ulam metric) . I will then discuss lower bounds on the distortion required to embed 0-1 strings and permutations.

Joint work of: Andoni, Alex; Charikar, Moses; Gopalan, Parikshit; Jayram, T.S.; Krauthgamer, Robert; Rabani, Yuval

On the Random Satisfiable 3CNF Process

Michael Krivelevich (Tel Aviv University, IL)

We suggest a new model for generating random satisfiable 3CNF formulas. To generate a random formula, first permute at random all $8\binom{n}{3}$ possible clauses over variables $x_1, \ldots, x_n$.

Then, start from the empty formula, go over all clauses one by one according to the random permutation, and add a new clause to the current formula only if the formula stays satisfiable after the addition, otherwise proceed to the next clause. We study the evolution of this process and the distribution over formulas obtained after scanning the first m clauses (in the random permutation's order). We show that for $m = cn$, where $c > 0$ is a sufficiently large constant, with high probabiltiy essentially all satisfying assignments are gathered in one cluster, and all but $e^{-\Omega(m/n)}n$ of the variables take the same value in all satisfying assignments, i.e. are "frozen". We also describe a polynomial time algorithm that almost surely finds a satisfying assignment in such randomly generated formula.

Keywords: Random satisfiable SAT formulas

Joint work of: Krivelevich, Michael; Sudakov, Benny; Vilenchik, Dan

Smoothed Analysis of Binary Search Trees and Quicksort Under Additive Noise

Bodo Manthey (Saarland University, D)

While the height of binary search trees is linear in the worst case, their average height is logarithmic. We investigate what happens in between, i.e., when the randomness is limited, by analyzing the smoothed height of binary search trees: Randomly perturb a given (adversarial) sequence and then take the expected height of the binary search tree generated by the resulting sequence.

As perturbation models, we consider partial permutations, where some elements are randomly permuted, and additive noise, where random numbers are added to the adversarial sequence. We prove tight bounds for the smoothed height of binary search trees under these models. We also obtain tight bounds for smoothed number of left-to-right maxima. Furthermore, we exploit the results obtained to get bounds for the smoothed number of comparisons that quicksort needs.

Keywords: Smoothed analysis, Binary search trees, Quicksort, Left-to-right maxima

Joint work of: Manthey, Bodo; Tantau, Till

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2007/1289

See also: Bodo Manthey, Till Tantau. Smoothed Analysis of Binary Search Trees and Quicksort Under Additive Noise. Electronic Colloquium on Computational Complexity (ECCC), Report 07-039, 2007

Consistent Weighted Sampling

Frank McSherry (Microsoft - Mountain View, USA)

We describe an efficient procedure for sampling representatives from a weighted set such that the probability that for any sets of weights S and T, the probability that the two choose the same sample is the Jacard similarity:

$$Pr[sample(S) = sample(T)] = \frac{\sum_x \min(S(x), T(x))}{\sum_x \max(S(x), T(x))} . \tag{1}$$

The sampling process takes expected time linear in the number of non-zero weights.

Joint work of: Manasse, Mark; McSherry, Frank; Talwar, Kunal

Probabilistic Analysis of Game Tree Evaluation

Ralph Neininger (Universität Frankfurt, D)

In the analysis of game searching methods a basic problem is to determine the value of the root of a minimax tree with certain given number at its leaves (the input). Various models for the input and related evaluation algorithms have been proposed and analyzed.

We review some of these models, in particular models with input from the set $\{0, 1\}$ and probabilistic models such as Pearl's model and the (random) incremental model. We discuss the complexity of evaluation algorithms under these models: a new tail bound for the complexity of Snir's randomized evaluation algorithm is given improving upon a Gaussian tail bound due to Karp and Zhang.

Also a limit law for the root's value in Pearl's model is given leading to a conjecture on the asymptotic distribution of the complexity of $\alpha - \beta$ pruning.

Keywords: Gametree; Minimax trees; Randomized algorithm; Tail bound

Joint work of: Ali Khan, Tämur; Devroye, Luc; Neininger, Ralph

How to Skip

Alessandro Panconesi (University of Rome "La Sapienza", I)

We study the problem of optimal skip placement in an inverted list. Assuming the query distribution to be known in advance, we formally prove that an optimal skip placement can be computed quite efficiently. Our best algorithm runs in time $O(n \log n)$, n being the length of the list.

The placement is optimal in the sense that it minimizes the expected time to process a query.

Our theoretical results are matched by experiments with a real corpus, showing that substantial savings can be obtained with respect to the traditional skip placement strategy, that of placing consecutive skips, each spanning $\sqrt{n}$ many locations.

Joint work of: Chierichetti, Flavio; Lattanzi, Silvio; Mari, Federico; Panconesi, Alessandro

Uncoordinated Two-Sided Markets

Heiko Röglin (RWTH Aachen, D)

Various economic interactions can be modeled as two-sided markets. A central solution concept to these markets are stable matchings. It is well known that stable matchings can be computed using a centralized polynomial-time algorithm. Many markets, however, do not have any centralized matching mechanism to match agents. In those markets, matchings are formed by actions of self-interested agents. Knuth introduced uncoordinated two-sided markets and showed that the uncoordinated better response dynamics may cycle. Roth and Vande Vate showed that the random better response dynamics converges to a stable matching with probability one, but did not address the problem of convergence time.

We give an exponential lower bound for the convergence time of the random better response dynamics in two-sided markets. We also extend these results to the best response dynamics, i.e., we present a cycle of best responses, and prove that the random best response dynamics converges to a stable matching with probability one, but its convergence time is exponential. Additionally, we identify the special class of correlated two-sided markets for which we prove that no better response cycle exists and that the random best response dynamics converges in expected polynomial time.

Joint work of: Ackermann, Heiner; Goldberg, Paul; Mirrokni, Vahab; Röglin, Heiko; Vöcking, Berthold

Cover Time, Mixing Time and Randomized Rumor Spreading

Thomas Sauerwald (Universität Paderborn, D)

In this talk we study randomized rumor spreading (a.k.a. randomized broadcast or push algorithm). Here, a rumor placed at one of the vertices of a graph should be spread to all other vertices. This is done iteratively by letting each vertex who already knows the rumor to forward it to some neighbor selected independently and uniformly at random. The question is how fast the rumor propagates to all vertices.

We derive lower and upper bounds on the runtime of this procedure based on random walk parameters like mixing rate (convergence speed towards stationary distribution) and cover time (expected number of steps after all vertices have been visited).

Keywords: Random walk, Randomized algorithms

Joint work of: Elsässer, Robert; Sauerwald, Thomas

A Constant Approximation Algorithm for the a priori TSP

David Shmoys (Cornell University, USA)

In the traveling salesman problem, one is given N, a set of points, and for each pair of points in N, one is also given the distance between them, where we assume that these satisfy the triangle inequality; the aim is to find a tour τ through all points in N that minimizes the total length $c(\tau)$ of the tour. In the *a priori* TSP, one is also given a probability distribution Π over the subsets $A \subseteq N$ of so-called active sets. For each subset A, each tour τ induces a tour τ_A by "shortcutting" those points not in A; we let $c(\tau_A)$ denote the length of the resulting tour of the points in A. In the *a priori* TSP, we measure the quality of a tour τ by computing the expected length with respect to a random choice of A drawn according to Π, $E_A[c(\tau_A)]$; the aim is to compute a tour that minimizes this expectation. Let τ^* denote an optimal *a priori* tour.

We consider the case in which Π is specified by giving independent activation probabilities for each point in N. We give a simple 4-approximation algorithm for this problem, that is, in polynomial time, we compute a tour τ such that $E_A[c(\tau_A)] \leq 4E_A[c(\tau_A^*)]$.

Joint work of: Shmoys, David; Talwar, Kunal

Clustering for Metric and Non-Metric Distance Measures

Christian Sohler (Universität Paderborn, D)

Given a set P of objects with a distance measure $d(.,.)$, the k-median clustering problem is to find a set of k objects (centers) such that the sum over the objects in P of distances to the nearest center is minimized. We show that the k-median problem with distance measure $d(.,.)$ can be approximated in linear time within a factor of $(1+\varepsilon)$, if the 1-median can be solved by computing the 1-median of a random sample of a constant number of point (we assume ε and k to be constant). Based on this result we show that there is a k-median algorithm for metrics with bounded doubling dimension, the Kullback-Leibler divergence (under some restrictions), and some Bregman divergences.

Keywords: Algorithms, Clustering, Kullback-Leibler divergence

Joint work of: Ackermann, Marcel; Blömer, Johannes; Sohler, Christian

Sampling-based Approximation Algorithms for Multi-stage Stochastic Optimization

Chaitanya Swamy (University of Waterloo, CA)

Stochastic optimization problems provide a means to model uncertainty in the input data where the uncertainty is modeled by a probability distribution over the possible realizations of the data. We consider a broad class of these problems, called *multi-stage stochastic programming problems with recourse*, where the uncertainty evolves through a series of stages and one take decisions in each stage in response to the new information learned. These problems are often computationally quite difficult with even very specialized (sub)problems being $\#P$-complete.

We obtain the first fully polynomial randomized approximation scheme (FPRAS) for a broad class of multi-stage stochastic linear programming problems with any constant number of stages, without placing any restrictions on the underlying probability distribution or on the cost structure of the input. For any fixed k, for a rich class of k-stage stochastic linear programs (LPs), we show that, for any probability distribution, for any $\varepsilon > 0$, one can compute, with high probability, a solution with expected cost at most $(1 + \varepsilon)$ times the optimal expected cost, in time polynomial in the input size, $\frac{1}{\varepsilon}$, and a parameter λ that is an upper bound on the cost-inflation over successive stages. Moreover, the algorithm analyzed is a simple and intuitive algorithm that is often used in practice, the *sample average approximation* (SAA) method. In this method, one draws certain samples from the underlying distribution, constructs an approximate distribution from these samples, and solves the stochastic problem given by this approximate distribution. This is the first result establishing that the SAA method yields near-optimal solutions for (a class of) multi-stage programs with a polynomial number of samples.

As a corollary of this FPRAS, by adapting a generic rounding technique of Shmoys and Swamy, we also obtain the first approximation algorithms for the analogous class of multi-stage stochastic integer programs, which includes the multi-stage versions of the set cover, vertex cover, multicut on trees, facility location, and multicommodity flow problems.

Keywords: Stochastic optimization, Approximation algorithms, Randomized algorithms, Linear programming

Joint work of: Swamy, Chaitanya; Shmoys, David

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2007/1290

Full Paper:
http://www.math.uwaterloo.ca/∼cswamy/papers/multistage.pdf

See also: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science, pages 595-604, 2005

Constrained Random Graph Processes

Anusch Taraz (TU München, D)

In this talk we discuss various random graph processes that follow a simple common framework: given a monotone graph property A, start with the empty graph on vertex set $[n]$ and then repeat adding edges chosen uniformly at random, provided that they don't violate the property A.

We will consider different examples of properties such as cycle-free, F-free or planar. We investigate how these processes evolve and whether the final random graph will almost surely contain copies of an arbitrary fixed subgraph H or not. We conjecture that the answer is always "yes" when A is a global property, but will depend on the density of H when A is local.

Keywords: Random graph processes, Planar

Joint work of: Gerke, Stefanie; Schlatter, Dirk; Steger, Angelika; Taraz, Anusch

On the Tractability of the k-colorability Problem

Danny Vilenchik (Tel Aviv University, IL)

As part of the efforts put in understanding the intricacies of the k-colorability problem, researchers analyzed different distributions over k-colorable graphs. While the worst case is of course notoriously hard (not even reasonably approximable), the average case turns out to be "easy" in many cases. Mediating between these two extremities are semi-random models. In this work we consider semi-random variants of the planted k-colorability distributions. We are able to get a more general framework than the one suggested by Coja-Oghlan 2004, and then by Krivelevich and Vilenchik in 2006. The main contribution of this work is exploring how far can one stretch the current analytical machinery that was developed to deal with random k-colorable instances in some suitably defined semi-random model. We have both positive results, that is the current algorithmic techniques extend to a more general semi-random setting, and some hardness results.

Keywords: Computational complexity, Random structures, k-colorable graphs

Tight Bounds for Blind Search on the Integers

Philipp Wölfel (University of Calgary, CA)

We analyze a simple random process in which a token is moved in the interval $A = [0, n]$: Fix a probability distribution p over $[1, n]$. Initially, the token is placed in a position chosen uniformly at random from A.

In round t, a random value d is chosen according to our probability distribution p. If the token is in position $a \leq d$, then it is moved to position $a - d$. Otherwise it stays put. Let T be the number of rounds until the token reaches position 0. We show tight bounds for the expectation of T for the optimal distribution p, i.e., we show that the minimum of $E[T]$ over all probability distributions p is $\Theta((\log n)^2)$. For the proof, a novel potential function argument is introduced. The research is motivated by the problem of approximating the minimum of a continuous function over $[0, 1]$ with a "blind" optimization strategy. Our technique has also other applications, for example for the proof of lower bound for the greedy routing time in small-world graphs.

Joint work of: Dietzfelbinger, Martin; Rowe Jon; Wegener, Ingo; Woelfel, Philipp