Mathematisches Forschungsinstitut Oberwolfach

Report No. 44/2001

Combinatorics, Probability and Computing

September 23rd – September 29th, 2001

Organisers:
B. Bollobás (Cambridge and Memphis)
I. Wegener (Dortmund)

Over the past few decades, the interplay between the three areas of combinatorics, probability and computer science has considerably enriched all three subjects; new techniques in probability theory have been developed in order to tackle combinatorial problems, which are often inspired in turn by important applications in computing. This meeting brought together researchers in all three areas to explore recent developments and new applications. Among the themes treated in the talks and discussions were new developments in: many different aspects of the satisfiability problem for Boolean formulae, the theory of pseudo-random graphs, stochastic processes on infinite graphs, randomised algorithms for approximating the permanent of a matrix, conditions for normal approximation, random lifts of graphs, allocation and assignment problems, abstract combinatorial programming, Ramsey properties of random graphs, probabilistic models for RNA structures, complexity analysis for various graph problems, and a unified probabilistic analysis for a class of parameters connected with search-trees.

It is planned that a selection of papers from the meeting will be published as a special issue of the journal *Combinatorics, Probability and Computing.*

Abstracts

Hashing, random forests and Brownian motion

Svante Janson

We study hashing with linear probing. We derive asymptotic distributions for the total displacement and for related quantities such as the maximal individual displacement.

For a wide range of the parameters, including the case of hash tables filled to a certain fraction, the total displacement is asymptotically normal. This can be explained by the fact that there are many blocks with almost independent contributions, and that no single block alone gives a significant contribution. The technical tool is a simple conditional central limit theorem.

For very dense hash tables, the total displacement is dominated by the contribution from one or a few blocks, and the limit is no longer normal.

There is a one to one correspondence between hash tables and random labelled rooted forests with a given number of components, where block lengths correspond to tree sizes. Both hash tables and the random rooted forests can be described by a random walk, conditioned on a certain event. In the dense case, this conditioned random walk converges after rescaling, as the size of the hash table or forest tends to infinity, to a Brownian excursion or a related process, which implies the non-normal limit law in this case.

Random Lifts of Graphs

Nathan Linial

In this talk I surveyed a new class of random graphs that we have been investigating for several years now. Given a (finite connected) graph G and an integer n, we consider a class of random graphs $L_n(G)$ (the n-lifts of G). A graph H in this class has a vertex set that equals $V \times [n]$. For every edge $xy \in E(G)$ we select a permutation $\pi = \pi_{xy}$ that is chosen uniformly at random from S_n, the symmetric group of order n. We connect, for every i, the vertex (x, i) in H with $(y, \pi(i))$.

These graphs have an interesting "split personality". On the one hand they behave like random graphs of bounded degrees, but at the same time they reflect some of the properties of the *base graph* G. Among the results I mentioned:

- **Amit and Linial:** If $\delta \geq 3$ is the smallest vertex degree in G, then no n-lift of G has connectivity larger than δ (this is trivial). An n-lift of G is δ-connected with probability $1 - o(1)$.
- **Amit, Linial and Matousek:** Every n-lift of G is $\chi(G)$-colorable (trivial). Almost every n-lift has chromatic number $> \Omega(\sqrt{\frac{\chi(G)}{\log(\chi(G))}})$.
- **Linial and Rozenman:** A zero-one law for perfect matchings: For every base graph G, either almost every lift of G has a perfect matching, ot almost none have a perfect matching.

Many open problems and variations on the basic theme were mentioned.

Sparse Pseudorandom Graphs

Yoshiharu Kohayakawa

(joint work with Vojtěch Rödl, Papa Sissokho and Endre Szemerédi)

The subject of quasi-random graphs was introduced in the eighties by Thomason and Chung, Graham and Wilson. They realized the surprising fact that several important properties shared by almost all graphs are asymptotically equivalent in a deterministic sense (related work in this area was also published in the eighties by Alon, Frankl, and Rödl).

Suppose $(G^n)_{n\geq 1}$ is a sequence of graphs with $|V(G^n)| = n$, and let

$$p = p(n) = |E(G^n)|\binom{n}{2}^{-1}.$$

Three basic properties are as follows:

NSUB(k): For any graph H on k vertices, the number of labelled copies of H in G^n (not necessarily induced) is

$$N(H, G^n) = (1 + o(1))n^k p^e,$$

where e is the number of edges in H.

DISC: For all X, $Y \subset V(G^n)$ with $X \cap Y = \emptyset$, if $e(X, Y)$ denotes the number of edges between X and Y, then

$$\Big|e(X, Y) - p|X||Y|\Big| = o(pn^2).$$

EIG: Let A denote the 0–1 adjacency matrix of G^n, with 1s denoting edges. Let λ_i $(1 \leq i \leq n)$ be the eigenvalues of A and adjust the notation so that $\lambda_1 \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Then

$$\lambda_1 = (1 + o(1))pn \quad \text{and} \quad |\lambda_2| = o(pn).$$

TUPLE(2): For all but at most $o(n^2)$ pairs $\{x_1, x_2\} \subset V(G^n)$, we have

$$\Big||\Gamma(x_1) \cap \Gamma(x_2)| - p^2 n\Big| = o(p^2 n).$$

A classical result in this area is that the properties above (with any fixed integer $k \geq 4$ in NSUB) are asymptotically equivalent, if p is a constant. If $p = p(n) \to 0$ as $n \to \infty$, this equivalence breaks down. Recently, Chung and Graham investigated how one can generalize this equivalence to the case of vanishing density. In particular, they observed that, if $p = p(n) \gg n^{-1/2}$, then the properties NSUB(C^4) and TUPLE(2) are equivalent, and imply properties EIG and DISC. (By NSUB(C^4) above, we mean that $N(C^4, G^n) = (1+o(1))(np)^4$, where C^4 is the cycle of length 4.) Moreover, they identified the importance of the property BDD($C, 2$), defined in a slightly more general form below.

BDD(C, Δ): For all $1 \leq r \leq \Delta$ and for all $\{x_1, \ldots, x_r\} \subset V(G)$, we have

$$\Big||\Gamma(x_1) \cap \cdots \cap \Gamma(x_r)| - np^r\Big| \leq Cp^r n.$$

An example of Alon (1994) shows that the conditions above, even taken together (except for NSUB), *do not even imply that G^n contains a triangle*, even when $p = p(n)$ is of order as large as $n^{-1/3}$.

In joint work with V. Rödl and P. Sissokho, we recently proved that if (i) H is triangle-free, (ii) $np^{\Delta} \gg 1$, where $\Delta = \Delta(H)$ is the maximal degree in H, and (iii) TUPLE(2) and BDD(C, Δ) hold for G^n for some fixed constant $C > 1$, then G^n contains H as a subgraph, as long as n is large enough. In fact, we show that NSUB(H) holds for H.

In recent work with V. Rödl and E. Szemerédi, we managed to prove some results concerning the embedding of "large", bounded-degree graphs in "positive-density, pseudorandom subgraphs" of sparse random graphs.

Some probabilistic algorithms for k-SAT

Uwe Schöning

We present several probabilistic algorithms for k-SAT, especially for the (NP-complete) case $k = 3$. These algorithms are based on the local search paradigm. That is, starting from some arbitrary (random or special) assignment a deterministic procedure `test`(a, m) systematically searches for a satisfying assignment which is m-close to the initial assignment a with respect to Hamming distance. This basis procedure is modified in several ways: first, the initial assignments can be chosen either in a systematic way (like a covering code) or in a random fashion. Second, the procedure `test` can be substituted by some random walk. This last variant obtains the bound $(4/3)^n$ in the case of 3-SAT. It can be improved somewhat by a more sophisticated choice of the initial probability distribution on so called independent clauses.

Optimal Multiple-Choice Allocation

Peter Sanders

(joint work with Sebastian Egner and Jan Korst)

The following allocation problem has been intensively studied in the last decade: Consider n balls that shall be put into m bins. For each ball there are two possible bins picked independently and uniformly at random. The task is to decide on one of the choices for each ball such that the maximum occupancy is minimized. Korst observed that an optimal allocation can be computed in polynomial time using maximum flow computations. In the first part of the talk we answer the question "How good is optimal". It turns out that a maximum occupancy of $\lceil n/m \rceil$ or $\lceil n/m \rceil + 1$ can be achieved with high probability. This improves on results $O(n/m)$ and $n/m + O(\log \log m)$ respectiveley for simple suboptimal scheduling algorithms.

In the second part of the talk it is explained how this low maximum occupancy can be exploited for scheduling parallel disks. The bins are disks. The balls are blocks to be retrieved together. By storing two randomly placed copies of each logical block of a virtual memory, we can make two random choices available. This result is also the key ingredient in the observation that a realistic model for parallel disks (independent disks) can efficiently emulate on an unrealistic but easier to program model (parallel heads). In order to cover more aspects of realistic machines it is then explained how application oriented generalizations can be accomodated.

- Achieving maximal occupancy $\lceil n/m \rceil$ for n close to a multiple of m results in very high efficiency.
- Finding optimal schedules in time $O(n \log n)$.
- Tolerating disk failures: store copies on *different* disks and update the analysis.

- Variable length blocks: Generalize the analysis and allow pieces of blocks to be read from both copies to avoid NP-hardness.
- Model communication bottlenecks at disk controllers, I/O busses etc.
- Reduced redundancy.
- Asynchronous scheduling.

Uniform and non-uniform bounds in normal approximation under local dependence

LOUIS H. Y. CHEN

(joint work with Qi-Man Shao)

There are generally three approaches to Stein's method for normal approximation. One is by induction (Bolthausen (1984)), another by using smoother functions (Stein (1972 and (1986)), and the third through a concentration inequality (Chen and Shao (2001)). The last approach is originally due to Stein (Ho and Chen (1978)) and was developed by Chen (1986 and 2000) and Chen and Shao (2001). In the latter, the technique is also developed for non-uniform bounds.

This paper is concerned with normal approximation for locally dependent random variables using Stein's method. Both uniform and non-uniform bounds are obtained by taking the concentration inequality approach.

Local dependence is a more general notion than m-dependence in sequences of random variables. It is defined for random variables with any index set. An example of local dependence is one defined in terms of common edges for random variables indexed by the vertices of a graph. See, for example, Baldi, Rinott and Stein (1989).

Several general results are obtained under different orders of local dependence. These are then applied to various special cases, some of which have been considered by others. The uniform bounds obtained in these special cases are either similar or better than those of other authors, while the non-uniform bounds obtained did not exist in the literature. These bounds are also best possible in terms of order by comparison with the classical results for independent random variables.

Stochastic Processes on Graphs and Amenability

ALAN STACEY

A graph $G = (V, E)$ is *amenable* if

$$\inf_{\substack{\emptyset \neq W \subset V \\ W \text{ finite}}} \frac{|\partial_E W|}{|W|} = 0.$$

All the graphs we consider are infinite, connected and of bounded degree. Canonical examples of nonamenable graphs are homogeneous trees, $\mathbb{T}_d$, in which every vertex has degree $d \geq 3$, and $\mathbb{T}_d \times \mathbb{Z}$. The d-dimensional lattices $\mathbb{Z}^d$ are amenable.

We survey an area of current research in which the behaviour of a stochastic process on a graph typically depends on whether or not the graph is amenable. The simplest interesting example is the simple symmetric random walk on a graph, which has spectral radius less than 1 if and only if the graph is nonamenable (Gerl 1987).

For percolation on *transitive* (i. e. vertex-transitive) graphs, it is known that there is a threshold value, p_u, above which there is a unique infinite cluster and below which there are either zero or infinitely many infinite clusters (Häggstrom, Peres, Schonmann 1999). It is known (Burton and Keane 1989) that on transitive graphs amenability implies $p_u = p_c$. The reverse implication is conjectured to hold, but has been proven only in certain special cases.

The contact process and branching random walk on graphs are considered. For each of these processes λ_1 is the threshold for global survival and λ_2 is the threshold for local (or strong) survival. For the contact process it is known that on homogeneous trees of degree $d \geq 3$, one has $\lambda_1 < \lambda_2$ (Pemantle(1990), Liggett(1996), Stacey(1996)) and on $\mathbb{Z}^d$ one has $\lambda_1 = \lambda_2$ (Bezuidenhout and Grimmett 1990). These results also hold, and are substantially easier to prove, for the branching random walk.

Certain conjectures linking amenability with equality of λ_1 and λ_2 for the two processes were disproved by Pemantle and Stacey (2001). In particular, examples were found of trees, of bounded degree, which are amenable and for which the branching random walk has $\lambda_1 < \lambda_2$; and nonamenable examples were found where $\lambda_1 = \lambda_2$. These arose from considering the process on Galton-Watson trees. Also, a spherically symmetric nonamenable tree of bounded degree was constructed on which the contact process has $\lambda_1 = \lambda_2$. However, the equivalence of amenability and $\lambda_1 = \lambda_2$ remains open for the contact process on transitive and, more generally, *quasi-transitive* graphs (those for which the action of the automorphism group on the vertices has finitely many orbits). For the branching random walk (b.r.w.), this issue is now resolved (Stacey 2001). On a transitive graph there is a straightforward link between the behaviour of the b.r.w. and the simple random walk which establishes the equivalence. This link does not extend to quasi-transitive graphs, although if the branching random walk is modified (in a certain natural way) then the link, and thereby the equivalence, holds for all graphs of bounded degree. For the standard b.r.w. on a quasi-transitive graph, however, a new proof shows that $\lambda_1 = \lambda_2$ if and only if the graph is amenable.

Optimal myopic algorithms for random 3-SAT

Gregory Sorkin

3-SAT is a canonical NP-complete problem: satisfiable and unsatisifable instances cannot generally be distinguished in polynomial time. However, random 3-SAT formulas show a phase transition: for any large number of variables n, sparse random formulas (with $m \leq 3.145n$ clauses) are almost always satisfiable, dense ones (with $m \geq 4.596n$ clauses) are almost always unsatisfiable, and the transition occurs sharply when m/n crosses some threshold. It is believed that the limiting threshold is around 4.2, but it is not even known that a limit exists.

Proofs of the satisfiability of sparse instances have come from analyzing heuristics: the better the heuristic analyzed, the denser the instances that can be proved satisfiable with high probability. To date, the good heuristics have all been extensions of unit-clause resolution, all expressible within a common framework and analyzable in a uniform manner through the differential equation method.

Here, we determine an optimal "tuning" of any algorithm expressible in this framework. We extend the analysis via differential equations, and we make extensive use of a new optimization problem we call "maximum-density multiple-choice knapsack". The structure of optimal knapsack solutions elegantly characterizes the choices made by an optimized algorithm. We improve the known satisfiability bound from density 3.145 to 3.26.

Many open problems remain. It is non-trivial to extend the methods to 4-SAT and beyond. If results are to be applicable to “real-world” 3-SAT instances, then the theory should be extended to formulas that need not be uniformly random, but obey some weaker conditions. Also, there is theoretical evidence that in the unsatisfiable regime it is difficult to prove the unsatisfiability of a given formula, while in the known region of satisfiability, linear-time algorithms produce satisfying assignments with high probability. Is the unsatisfiable regime truly hard, and is the whole of the satisfiable regime truly easy? In particular, as the scope of myopic, local algorithms is expanded so that they examine more and more variables, can such algorithms solve random instances arbitrarily close to the threshold density?

Abstract combinatorial programming and efficient property testers

Christian Sohler

(joint work with Artur Czumaj)

The goal of property testing is to distinguish between the case whether a given object has a certain property or is ‘far away’ from any object having this property. In the first part of the talk we prove that k-colorability of graphs can be tested in time independent of the size of the graph (this was first proven by Goldreich, Goldwasser, and Ron). We present a testing algorithm that examines only $\tilde{O}(k^4/\epsilon^4)$ entries in the adjacency matrix of the input graph, where ϵ is a distance parameter independent of the size of the graph.

In the second part of the talk we present a general proof technique that can be used to show that certain properties that are closed under restrictions (if an object has a property then any ‘subobject’ also has the property; e. g., if the object is a graph then colorability is closed under restrictions while connectivity is not). We introduce *abstract combinatorial programming* which can be roughly described as linear programming where you forget about the geometry and allow that a set of constraints defines multiple bases. Then we show that a property that is closed under restrictions can be tested, if there is a gap and feasibility preserving reduction that maps any object to an abstract combinatorial program of small dimension. We illustrate our approach with three examples: testing low degree uni-variate polynomials, radius clustering, and graph coloring.

BDD-based Cryptanalysis of Stream Ciphers

Matthias Krause

Many stream ciphers occuring in practice produce their output bit stream according to a rule $y = C(L(x))$, where $L(x)$ denotes an internal linear bit-stream produced by a fixed number of linear feedback shift registers starting from a secret initial state $x = (x_0, \ldots, x_{n-1})$, and C denotes some nonlinear compression function. We present an algorithm for computing the secret key x from a given output bitstream y of length $\approx n$, which uses Free Binary Decision Diagrams (FBDDs), a data structure for minimizing and manipulating Boolean functions. We show that if the decision whether $C(z) = y$ can be performed by polynomial size FBDDs (this is usually the case), the effective key-length of the cipher is bounded by $\frac{1-\alpha}{1+\alpha}n$, where α denotes the information rate (per bit) which y reveals about the internal bit-stream z. This yields the best known upper bounds on the effective key length for several stream ciphers of practical use, for instance a $0.656n$ upper bound on the effective key length of the self-shrinking generator, a $0.6364n$ upper bound on the effective key length of the A5-generator, used in the GSM-standard, a $0.6n$ upper bound on the effective key length of E_0 encryption standard in the one level mode, and a $0.8823n$ upper bound on the effective key length of E_0 in the two level mode, as it is used in the Bluetooth wireless LAN system.

Approximating the permanent

Mark Jerrum

(joint work with Alistair Sinclair and Eric Vigoda)

The permanent $\operatorname{per} X$ of an $n \times n$ matrix $X = (x_{ij})$ is a multivariate polynomial akin to the more familiar determinant, except that all monomials are given positive sign. If X is interpreted as the adjacency matrix of a bipartite graph G (with x_{ij} corresponding to the edge from vertex i on the left to j on the right) then $\operatorname{per} X$ is the generating (or partition) function of perfect matchings in G. In contrast to the determinant, which can be evaluated efficiently using Gaussian elimination, the permanent is known to be #P-complete, even when X is restricted to being a 0,1-matrix. This classical result of Valiant almost certainly rules out a polynomial-time algorithm for computing the permanent of such a matrix exactly.

The proof of #P-completeness of the permanent uses polynomial interpolation in an essential way, and interpolation is numerically unstable: it may be necessary to evaluate a polynomial to very high accuracy in order to know the coefficients even roughly. Therefore the completeness result does not rule out the possibility that the permanent of a 0,1-matrix can be computed to within relative error $1 \pm \varepsilon$ in time polynomial in n and ε^{-1}. (An approximation algorithm with this property is technically a "fully polynomial randomised approximation scheme" or FPRAS.) The question of the existence of an FPRAS, which had been open for some time, has recently been resolved affirmatively.

The full solution builds on a partial solution of Jerrum and Sinclair from the late eighties. (It is convenient at this juncture to switch to the perfect matching formulation.) They analysed a Markov chain, proposed by Broder, on perfect and near-perfect matchings (i. e., matchings leaving two vertices uncovered) of a bipartite graph G. They showed that the Markov chain is "rapidly mixing" – i. e., that it converges to near-stationarity in time polynomial in n – provided G satisfies a certain graph-theoretic condition. Given samples of perfect matchings from a near-uniform distribution it is possible to estimate the number of perfect matchings to within small relative error in polynomial time.

The new approach removes the restriction on the graph G. The idea is to modify the Markov chain by assigning to each near-perfect matching a weight that is a function of the position of the pair of holes. If the weight is made inversely proportional to the total number of near-perfect matchings with that hole pattern, then the modified Markov chain is unconditionally rapidly mixing. Although we don't know at the outset what these ideal weights are (indeed they are related to the very quantities we are trying to estimate), we are able to converge to them through an iterative procedure. As a consequence, we obtain an FPRAS for $\operatorname{per} X$ when X is a 0,1-matrix. The method generalises to arbitrarily non-negative real matrices. No further extension is possible, since the presence of even one negative entry in X is enough to render approximate evaluation of $\operatorname{per} X$ computationally intractable (under some reasonable complexity-theoretic assumption).

Efficient recognition of random unsatisfiable k-SAT instances.

Andreas Goerdt
(joint work with Joel Friedman and Michael Krivelevich)

We consider the family of probability spaces of random instances of the k-SAT problem, that is formulas in conjunctive normal form where each clause consists of exactly k literals. We paramaterize these probability spaces by n, the number of underlying variables, and $m = m(n)$, the number of random clauses.

The following threshold behaviour is well known and proved by a theorem of Friedgut: There exist constants c_k such that fomulas with $c_k(1-\varepsilon)n$ random clauses are satisfiable with high probability whereas formulas with $c_k(1+\varepsilon)n$ random clauses are unsatisfiable with high probability. From the algorithmic point of view it is interesting to note that formulas at the threshold seem to be hard instances. Therefore it is an obvious project to get deterministic polynomial time algorithms working with high probability for random formulas as close as possible to the threshold. The behaviour of such algorithms on the satisfiable side of the threshold is conceptually clear: Try to find a satisfying assignment. On the unsatisfiable side of the threshold this would read: Try to find a witness of unsatisfiability in deterministic polynomial time. And this is exactly the question we address. Previous work of Beame, Pitassi, Karp, and Saks shows that random k-SAT instances with $n^{k-1}/(\log n)^{k-2}$ clauses have polynomial size backtracking trees and thus can be certified efficiently unsatisfiable.

Our results improve on these bounds: For even k we show that formulas with $n^{k/2+\varepsilon}$ clauses can be efficiently certified unsatisfiable with high probability, for odd k we get by a simple reduction $(k+1)/2$ in the exponent. For 3-SAT instances we get an efficient algorithm for $n^{3/2+\varepsilon}$ random clauses. Our algorithms work by assigning two graphs to a formula and computing the eigenvalue spectrum of the adjacency matrix of these graphs. Then we show that the eigenvalue spectrum is such that it certifies unsatisfiabilty. Note that the eigenvalue spectrum of an adjacency matrix can be approximated to any degree of precision in polynomial time.

At the heart of our algorithms lies the following observation, which we state for $k = 4$: If a 4-SAT instance is satisfiable then there is a set of $n/2$ variables such that all all-positive clauses have at least one variable not from this set or all all-negative clauses have the analogous property. We assign the following two graphs to a formula: The vertices are the $\binom{n}{2}$ unordered pairs of variables. In the first graph two such vertices are connected by an edge if the corresponding variables are in one all-positive clause. In the second graph the edges are induced in the same way by the all-negative clauses. Note that we do not consider the mixed clauses of the underlying formula. Now the following holds: If the underlying formula is satisfiable then at least one of these graphs has an independent set of vertices having at least $\binom{n}{2}/2$ vertices.

Our algorithms rely on bounding the size of independent sets by the eigenvalues of the adjacency matrix of the underlying graph. We call a graph ε-regular if the degree of each vertex is between $d(1-\varepsilon)$ and $d(1+\varepsilon)$. We call a graph ν-separated if for $i \geq 2$ $|\lambda_i| \leq \nu\lambda_1$ where the eigenvalues of the adjacency matrix of the graph are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. One of our results is the following bound on the size of any independent set: If a graph is ν-separated and ε-balanced then it has no independent set larger than $n/5 + nf(\nu,\varepsilon)$ where $f(\nu,\varepsilon)$ goes to 0 if ν and ε do. This result allows to show the absence of independent sets with at least $\binom{n}{2}/2$ vertices in our graph.

Ramsey properties of random graphs: sharp threshold and online coloring

Andrzej Ruciński

Consider the graph Ramsey property $\mathcal{R}$ consisting of all graphs for which every 2-coloring of their edges results in a monochromatic copy of a triangle K_3. It has been known that there exist constants c and C such that

$$\lim_{n\to\infty} \Pr[G(n,p) \in \mathcal{R}] = \begin{cases} 1 & \text{if } p > C/\sqrt{n} \\ 0 & \text{if } p < c/\sqrt{n} \end{cases}$$

Using a recent criterion of Friedgut for the existence of a sharp threshold, Friedgut, Rödl, Ruciński and Tetali have shown that there exists a function $b = b(n)$ such that for all $\epsilon > 0$

$$\lim_{n\to\infty} \Pr[G(n,p) \in \mathcal{R}] = \begin{cases} 1 & \text{if } p > (1+\epsilon)b(n)/\sqrt{n} \\ 0 & \text{if } p < (1-\epsilon)b(n)/\sqrt{n} \end{cases}$$

As one of the main ingredients of the proof, the following lemma has been established. For a graph F, let $Base(F)$ be the set of all pairs of vertices in $V(F)$ which together with any two edges of F form a triangle.

Lemma *For all $\lambda > 0$ and $c > 0$ there exists $\alpha > 0$ such that with probability tending to 1 as $n \to \infty$, for any subgraph F of $G(n, cn^{-1/2})$ with at least $\lambda n^{3/2}$ edges, the set $Base(F)$ contains at least αn^3 triangles.*

This lemma has also some interesting consequences with respect to online coloring of the edges of $G(n,M)$. Consider the random graph process revealing its edges one by one: $e_1, \dots, e_{\binom{n}{2}}$, $n \geq 6$, and the following one-person game related to it. The Player's goal is to 2-color the edges as they come and not to create a monochromatic triangle for as long as possible. The game is over when a monochromatic triangle is created.

While it is not hard to prove that the expected length of the game is about $n^{4/3}$, the proof of the upper bound suggests the following relaxation: instead of coloring online, the random graph is generated in only two rounds, and the Player colors the edges first after round one and then at the end. Given the size of the first round, how large a second round can the Player survive?

In the most extreme case, when round one consists of a random graph with $cn^{3/2}$ edges, it follows from the Lemma above that asymptotically almost surely the addition of ωn random edges makes the graph have the Ramsey property $\mathcal{R}$, where $\omega = \omega(n)$ is any function which tends to ∞ as $n \to \infty$.

During my talk, I conjectured that, in fact, only ω random edges should suffice. This was confirmed during the workshop in collaboration with Yoshi Kohayakawa.

Point Distributions and Large Tetrahedra

Hanno Lefmann

An old conjecture of Heilbronn states that for every distribution of n points in the 2-dimensional unit square $[0,1]^2$ (or unit disc) there exist three distinct points which form a triangle of area at most c/n^2 for some fixed constant $c > 0$. Erdös observed that this conjecture, if true, would be best possible, as the points $(i \bmod n, i^2 \bmod n)_{i=0,\dots,n-1}$ on the moment-curve in the $n \times n$ grid would show after rescaling. However, Komlós, Pintz and Szemerédi in 1982 disproved Heilbronn's conjecture by proving that for every n there exists a configuration of n points in the unit square $[0,1]^2$ with every three points forming a triangle of area at least $c' \cdot \log n/n^2$, where $c' > 0$ is constant. Using techniques from derandomization, this existence argument was made constructive in the sequel, namely a polynomial time algorithm was given, which finds n points in $[0,1]^2$ achieving the lower bound $\Omega(\log n/n^2)$ on the minimum triangle area. Upper bounds on Heilbronn's triangle problem were given by Roth and Schmidt in a series of papers and the currently best upper bound is due to Komlós, Pintz and Szemerédi and is of the order $n^{-8/7+\varepsilon}$ for every fixed $\varepsilon > 0$.

Recently, Barequet considered a k-dimensional version of Heilbronn's problem by looking at the minimum volume of simplices arising from distributions of n points in $[0,1]^k$. For given dimension $k \geq 3$ he showed, that for every n there exist n points in the k-dimensional unit cube $[0,1]^k$ such that the minimum volume of every simplex spanned by any $(k+1)$ of these n points is at least $\Omega(1/n^k)$. Barequet gave three approaches for proving his lower bound. One uses a random argument; another is similar to Erdös' construction (and was probably known to him), namely taking the points $P_l = (l^j \bmod n/n)_{j=1,\ldots,k}$ for $l = 0, 1, \ldots, n-1$ on the moment-curve.

Barequet's lower bound was improved by this author by a factor $\Theta(\log n)$ for dimensions $k \geq 3$, using a probabilistic existence argument based on a variant of a result of Ajtai, Komlós, Pintz, Spencer and Szemerédi. For the corresponding arguments the continuous structure of $[0,1]^k$ was crucial. With Niels Schmitt we investigated, how to find a deterministic polynomial time algorithm for this problem. For the case of dimension $k = 3$ we obtained the following result:

Theorem *For every positive integer n one can find deterministically in polynomial time a configuration of n points in the unit cube $[0,1]^3$ such that the volume of any tetrahedron spanned by any four of these points is at least $\Omega(\log n/n^3)$.*

The proof of this result is based on techniques from combinatorics and number theory and some of the main ideas were presented in the talk.

Quasi-periodicity and quasi-randomness

Vera T. Sós
(joint work with Miklós Simonovits)

Considering different structures (graphs, numbers, permutations, etc.), an important question is when deterministic objects can be considered as randomlike ones and how randomlike objects can be generated in non-random ways. Depending on the specific problem, these questions have different aspects and may have several nonequivalent answers.

In the talk we focus on two structures.

1 Sequences. The simplest examples are the $(\{n\alpha\})$ (and the $(\lfloor n\alpha \rfloor)$ = Sturmian) sequences. These sequences are fundamental in diophantine approximation, in ergodic theory etc. and are often used also in combinatorics.

They have two – seemingly contradictory – features:

- These sequences have a very strict structure, are close to periodic; are quasi-periodic in several well defined senses, (e. g. three-distance or three difference property), among non periodic sequences the characteristic sequence of an $(\lfloor n\alpha \rfloor)$ has the smallest number of different blocks of fixed size, etc.

- They are randomlike in some sense (e. g., they are uniformly distributed sequences) and therefore in some areas can be used as quasi-random (deterministic) sequences.

In the last two decades quasi-periodic (almost periodic) structures, including also those in higher dimension (e. g. the Penrose tilings in the plane), became important in many other fields, such as operation research, computer science, game theory, but also in physics, in the theory of quasi-crystals etc.

In this lecture we indicate some of this large variety of links and connections, but we concentrate on results and problems with number-theoretical and combinatorial features.

As an example we mention also Weyl trees (motivated also by computer science, introduced and investigated by Luc Devroye and A. Goudjil) which are search trees where the input sequences are $(\{n\alpha\})$ sequences. These search trees are randomlike in some respects but not in all. With Simonovits we studied certain aspects of these search trees.

2 Graphs. Thomason and Chung-Graham-Wilson gave a class of graph properties, all possessed by random graphs and at the same time equivalent to each other. With Simonovits we proved that this class (i. e., the class of quasi-random graphs) can be characterized also by the Szemerédi-partition of graphs. Using this approach we proved that some properties, which do not imply quasi-randomness on their own, do imply it if we consider the corresponding hereditarily extended properties. (Large subgraphs of random-like graphs must also be random-like.) E. g., we proved the following Theorem:

Let $\nu = v(L)$, $E = e(L)$. Denote by $\mathbf{N}(G \subseteq L)$ and $\mathbf{N}^*(G \subseteq L)$ the number of not necessarily induced and induced copies of L in G, respectively. Further, denote by $\beta_L(p)$ and $\gamma_L(p)$ the "densities" of **labelled induced** and **labelled not necessarily induced** copies of L in a p–random graph:

$$\beta_L(p) = p^E(1-p)^{\binom{\nu}{2}-E} \text{ and } \gamma_L(p) = p^E.$$

Theorem *Let L_ν be a fixed sample-graph, $p \in (0,1)$ be fixed. Let (G_n) be a sequence of graphs. If (for every sufficiently large n) for every induced $F_h \subseteq G_n$,*

$$\mathbf{N}(L_\nu \subseteq F_h) = \gamma_L(p)h^\nu + o(n^\nu),$$

then (G_n) is p–quasi–random.

For the case of induced copies the situation is more complicated. We have the following

Conjecture *Let L_ν be a fixed sample-graph, $p \in (0,1)$ be fixed. Let (G_n) be a sequence of graphs. If (for every sufficiently large n) for every induced $F_h \subseteq G_n$,*

$$\mathbf{N}^*(L_\nu \subseteq F_h) = \beta_L(p)h^\nu + o(n^\nu),$$

then (G_n) is p–quasi–random.

We can prove this for several cases, among others:

Theorem *The above conjecture holds for regular graphs L_ν.*

The results may contribute to study the problem: how to create a scale between periodic and random sequences, between "non-random" and random graphs, etc. It is worth observing that some of these questions are much better understood for sequences and others for graphs.

Randomized Rumor Spreading

Christian Schindelhauer

(joint work with Richard Karp, Scott Shenker and Berthold Vöcking)

We investigate the class of so-called epidemic algorithms that are commonly used for the lazy transmission of updates to distributed copies of a database. These algorithms use a simple randomized communication mechanism to ensure robustness. Suppose n players communicate in parallel rounds in each of which every player calls a randomly selected communication partner. In every round, players can generate rumors (updates) that are to be distributed among all players. Whenever communication is established between two players, each one must decide which of the rumors to transmit. The major problem (arising due to the randomization) is that players might not know which rumors their partners have already received. For example, a standard algorithm forwarding each rumor from the calling to the called players for $\Theta(\log n)$ rounds needs to transmit the rumor $\Theta(n \log n)$ times in order to ensure that every player finally receives the rumor with high probability.

We investigate whether such a large communication overhead is inherent to epidemic algorithms. On the positive side, we show that the communication overhead can be reduced significantly. We give an algorithm using only $O(n \log\log n)$ transmissions and $O(\log n)$ rounds. In addition, we prove the robustness of this algorithm, e. g., against adversarial failures. On the negative side, we show that any address-oblivious algorithm (i. e., an algorithm that does not use the addresses of communication partners) needs to send $\Omega(n \log\log n)$ messages for each rumor regardless of the number of rounds. Furthermore, we give a general lower bound showing that time- and communication-optimality cannot be achieved simultaneously using random phone calls, that is, every algorithm that distributes a rumor in $O(\log n)$ rounds needs $\omega(n)$ transmissions.

Exploring Graphs with Little Memory

Rüdiger Reischuk

(joint work with Andreas Jakoby and Maciej Liskiewicz)

For graph problems like reachability – decide whether there exists a path between two nodes – or planarity testing very efficient linear time algorithms have been known for long. In this talk we discuss the space complexity of such problems. For directed graphs reachability is the canonical problem (often called GAP for graph accesibility problem in this context) that is complete for the complexity class nondeterministic logarithmic space $\mathcal{NL}$. In case of undirected graphs, the problem UGAP, one can exploit random walks to find a connecting path – getting the complexity down to probabilistic logarithmic space $\mathcal{RL}$. Or one can use symmectric nondeterministic machines, that can always reverse a computational step, to put the problem into the complexity class $\mathcal{SL}$. For the complexity class $\mathcal{L}$ of problems that can be solved in deterministic logarithmic space – the minimal amount of space necessary simply to identify a node – reachability in undirected forests, UFA, is a canonical complete problem.

Our goal is to extend the class of graphs that have very space efficient algorithms, that is fall into $\mathcal{L}$. With respect to the notion of tree width, trees or forests are the simplest graphs having width 1. Increasing the width to 2 we get the class of series-parallel graphs, SP-graphs for short, that have many applications, for example in programming analysis, and have been studied extensively. SP-graphs can be defined constructively by starting with single edges and performing a sequence of operations, series and parallel composition. Alternatively, one could give a characterization by forbidden minors, which is the diamond ◇ in the directed case and the K_4 in the undirected case.

We prove that both the reachability problem and the recognition problem for directed SP-graphs are $\mathcal{L}$-complete. The lower bounds follow from simple reductions, whereas the upper bounds require an extensive study of special properties of SP-graphs. It is shown that the successor relation can be tested deterministically by selecting special combinations of paths. A new characterization of directed SP-graphs is given by a forbidden induced homeomorphic subgraph, the Zig-Zag graph Z, that can be tested space efficiently. Furthermore, we show that the series-parallel decomposition of such graphs can be computed in logarithmic space.

Finally, we discuss the situation for undirected SP-graphs, which contrary to the general case seem to be more difficult than their directed counterparts, and related open problems, for example whether for arbitrary fixed ℓ checking for K_ℓ as a minor can be done space efficiently.

The Complexity of Computing the MCD-Estimator

Paul Fischer

(joint work with Thorsten Bernholt)

In modern mathematical statistics and data analysis, one fundamental problem is that of constructing statistical methods which are *robust* against model deviations. For example, it is well known that the standard estimates of location and scatter – sample mean and sample variance – are not robust. A single data point which is moved far out will change these quantities arbitrarily. In general one assumes that the observed data is mainly generated by some process or distribution which one would like to analyse. We shall call the part of the data coming from the distribution of interest the data from the *true population*. The rest of the data, however, might come from other sources or is altered by noise; we call this the *outliers*. The goal is to nevertheless estimate statistical quantities of the true population.

More precisely, given N observations, MCD is the problem to select a subset of size h, for some $h > N/2$, for which the determinant of the empirical covariance matrix is minimal over all subsets of size h. For a set of points $\mathcal{X}' = \mathbf{x}_1, \dots, \mathbf{x}_h$ in $\mathbb{R}^d$ the *(empirical) covariance matrix* is defined as

$$C(\mathcal{X}') = \frac{1}{h} \sum_{i=1}^{h} (\mathbf{x_i} - \mathbf{t})\ (\mathbf{x_i} - \mathbf{t})^T \text{ with } \quad \mathbf{t} = \frac{1}{h} \sum_{i=1}^{h} \mathbf{x_i} \ .$$

There is a nice geometric interpretation of the MCD. The inverse $C^{-1}(\mathcal{X}')$ of the minimum covariance matrix $C(\mathcal{X}')$ and the mean $\mathbf{t}(\mathcal{X}')$ define an ellipsoid in $\mathbb{R}^d$. This ellipsoid nicely matches the points $\mathcal{X}'$. The determinant is a measure of volume. Hence a small determinant corresponds to an ellipsoid of small volume. If the extensions of the ellipsoid in all dimensions are small then the set $\mathcal{X}'$ is quite compact. Another way to get a small volume is that the ellipsoid is somewhat "flat", i. e., it might have a large extension in some directions but only small ones in others. This indicates that the set $\mathcal{X}'$ is "essentially lower dimensional".

In this talk we address the complexity of computing the MCD-estimator. Obviously, computing $\det(\mathcal{X}')$ for all $\binom{N}{h}$ subsets $\mathcal{X}'$ of $\mathcal{X}$ of size h solves the problem, though it might take exponential time in h. It was not clear whether the estimator itself has this complexity independent of the dimensionality d of the data. Here we show that the complexity of MCD is polynomial if the dimension is fixed. This is achieved by avoiding to consider all subsets of size h. Exploiting geometric properties of the estimator, we have been able to design an algorithm which enumerates a sequence of subsets of size h of the input data set $\mathcal{X}$ in polynomial time. We show that one of the sets enumerated has minimum covariance determinant. The running time of our algorithm is $O\left(N^{d^2}\right)$.

On the other hand it is possible to show that the decision version of the MCD problem is NP-complete if the dimension varies. This is achieved by reducing CLIQUE to MCD.

Percolation Thresholds: Bounds, Conjectures, and Counterexamples

John C. Wierman

The substitution method was used to derive several upper and lower bounds for percolation thresholds of Archimedean lattices. An Archimedean lattice is a tiling of the plane with regular polygons that is vertex-transitive. There are exactly 11 Archimedean lattices. Percolation on these lattices has been studied in the physics literature. They are related to a conjecture of Häggström, that vertex-transitive graphs either have critical probability equal to one or less than a bound $B < 1$. The new substitution method bounds establish that the $(3, 12^2)$ lattice has the largest bond percolation threshold of all Archimedean lattices, making it a prime candidate for the vertex-transitive graph with the largest critical probability less than one. The bounds are not sufficiently accurate to identify the Archimedean lattice with the largest site percolation critical probability.

The new bounds are more accurate than previous ones: For bond percolation models on the $(3, 12^2)$ and Kagomé lattices, the upper and lower bounds differ by less than 0.01. For site percolation models on the $(4, 6, 12)$ and $(4, 8^2)$ lattices, the bounds differ by less than 0.1. These are the first bounds with this level of accuracy.

The range of critical probability values for bond percolation models on fully-triangulated lattices was investigated. A sequence of such graphs was constructed which has critical probability values tending to zero. Thus, the range is at least from 0 to .3473. Based on both rigorous bounds and simulation estimates for critical probabilities, I conjecture that the standard triangular lattice has the largest bond percolation critical probability of any fully-triangulated periodic graph.

Physics have tried to formulate a "universal formula" which approximates the critical probability values of all graphs, for many years. Recent formulas have been based on only the dimension and average degree of the graph, and have claimed a maximum deviation of 0.08. The fully-triangulated graphs mentioned above all have dimension 2 and average degree 6, but show that any formula must have an error of at least .1736 on one of the fully-triangulated graphs.

Counterexamples were provided for two common beliefs that have persisted for over 40 years in percolation theory: The critical probability is not a monotone decreasing function of the average degree of the graph. For a set of graphs, the bond percolation thresholds and site percolation thresholds may not have the same order.

Random Assignment with Integer Costs

Sven Erick Alm

(joint work with Robert Parviainen)

In the random assignment problem, we are looking for a permutation π that minimizes

$$Z(n) = \sum_{i=1}^{n} c_{i,\pi(i)}\,,$$

where $C = (c_{ij})$ is the cost matrix, with entries that are i. i. d. random variables, usually $U(0,1)$ or $Exp(1)$.

Let $Z^*(n)$ be the optimal (minimal) cost, and π^* the optimal assignment.

Mzard (1988) conjectured that

$$\lim_{n\to\infty} E(Z^*(n)) = \frac{\pi^2}{6} \text{ (for } U(0,1) \text{ costs)}\,.$$

In 1992, Aldous showed that, for continuous cost distributions, there is a limit, which only depends on the density at 0. In 2000, Aldous proved Mzard's conjecture, and also gave the limiting distribution for $n \cdot c_{i,\pi^*(i)}$. He also showed that the (row) rank of $c_{i,\pi^*(i)}$ converges to a Geometric(1/2) distribution.

We use Aldous' results and a coupling argument to study four different models with integer costs.

1. The rows of C are independent random permutations of $1, \ldots, n$.
2. The costs, $c_{i,j}$, are i.i.d. uniform on $1, \ldots, n$.
3. The matrix C is a random permutation of $1, \ldots, n^2$.
4. The costs, $c_{i,j}$, are i.i.d. uniform on $1, \ldots, n^2$.

Cases 1 and 2 are scaled by $1/n$; cases 3 and 4 by $1/n^2$.

The results are:

$$\begin{aligned}
\pi^2/6 \le \lim_{n\to\infty} E(Z_1^*(n)) &\le 2\,,\\
\pi^2/6 + 1/2 \le \lim_{n\to\infty} E(Z_2^*(n)) &\le \pi^2/6 + 13/24\,,\\
\lim_{n\to\infty} E(Z_3^*(n)) &= \pi^2/6\,,\\
\lim_{n\to\infty} E(Z_4^*(n)) &= \pi^2/6\,.
\end{aligned}$$

Asymptotic normality and submap counts in random maps

Nicholas C. Wormald

It is well known that the moments, or factorial moments, determine many distributions (such as Poisson or normal). This is convenient if the central moments are computed, or if the expected value is bounded. But what if only non-central moments are known, and only asymptotically, and the expectation goes to infinity quite rapidly? Zhicheng Gao and I gave a new general result showing that the asymptotic behaviour of high factorial moments can determine the shape of asymptotically normal distributions.

Let X_n be a nonnegative integer random variable ($n \geq 1$). Provided the kth factorial moment $\mathbf{E}[X_n]_k = \mathbf{E}X_n(X_n - 1) \cdots (X_n - k + 1)$ converges to μ^k for each fixed k as $n \to \infty$, we can conclude that X_n converges in distribution to the Poisson random variable with mean μ. This is the standard "method of moments". If $\mathbf{E}[X_n]_k \sim \mu_n^k$ for each finite k but $\mu_n \to \infty$ as $n \to \infty$ there is no conclusion, and this is where the story usually ends. But the surprising fact is that if the moments behave suitably also for $k \to \infty$, we can deduce the (expected) result that X_n is asymptotically normal. What is required is that

$$\mathbf{E}[X_n]_k \sim \mu_n^k \exp\left(\frac{k^2 s_n}{2}\right)$$

where $s_n > -\mu_n^{-1}$ and a couple of other simple conditions hold. The conclusion is that $(X_n - \mu_n)/\sigma_n$ tends in distribution to the standard normal as $n \to \infty$, where $\sigma_n = \sqrt{\mu_n + \mu_n^2 s_n}$.

We gave applications to submap counts in random planar triangulations, where we use a simple argument to asymptotically determine the high moments for the number of copies of a given subtriangulation in a random 3-connected planar triangulation. Similar results were also obtained for 2-connected triangulations and quadrangulations with no multiple edges.

For these applications, the usual methods of proving asymptotic normality do not seem to apply, as they all basically rely on the variable in question being the sum of a large number of nearly independent variables. For random maps, such a framework has never been established. The new method also applies to some other situations. However, for the counts of small subgraphs in random graphs, it usually fails because the high moments are warped by the tail of the distribution.

On Random RNA Secondary Structures

Markus Nebel

A RNA molecule consists of a chain of four different types of nucleotides which only differ by the base (adenine (A), cytosine (C), guanine (G) or uracil (U)) involved. The specific sequence of bases along the chain is called the *primary structure* of the molecule. Through the creation of hydrogen bounds, the complementary bases A and U (resp. C and G) form stable base pairs. Additionally, there exists the weaker G-U pair, where the bases bind in a skewed fashion. By the creation of base pairs the primary structure is folded into a stable three-dimensional conformation called *tertiary structure* of the molecule. It is customary in sciences to study the simplified *secondary structure* by focusing ones attention just on what bases form pairs and allow the sequence to form helical regions in two dimensions. Since experimental approaches like X-ray diffraction are quite expensive much effort has been made to deduce the secondary structure from the knowledge of the primary structure. With respect to this task the notion of order of a secondary structure has been introduced by Waterman, who gave the first formal framework for secondary structures. Many authors have paid attention to enumeration problems related to the combinatorics of RNA secondary structures. Two different models have been considered. As shown by Waterman, assuming that base-pairing is possible between arbitrary pairs of nucleotides, the set of all possible structures can be modelled as a specific class of planar graphs. As pointed out by Zuker and Sankoff, a more realistic model is obtained by a stochastic approach, where we assume a Bernoulli distribution of the bases. The probability p (usually called *stickiness*) that two random bases can be paired is used to control the shape of the molecules. For $p = 1$ both models are equivalent. In all cases, parameters like the number of different structures of a given size, the number of structures of given size and order, the expected number of specific substructures but also the systematical treatment of such problems from a mathematical point of view are of interest.

In this talk we present a new approach (based on multivariate generating functions) for enumerating parameters related to secondary structures. For the first time it becomes possible to derive satisfying results for parameters which depend on the order of the molecules considered. We derive precise asymptotics (number of nucleotides $n \to \infty$) for the (expected) number of secondary structures of size n and fixed order k and for the averaged order of structures of size n, assuming the Bernoulli model with a stickiness $p > 0$. This solves an open problem which can be traced back to the original work of Waterman. Many additional results like the averaged length of *loops* (sequences of unpaired bases) or the expected number of substructures like *hairpins* and *bulges* will be presented.

The number of 2-SAT functions

Imre Leader

(joint work with Béla Bollobás and Graham R. Brightwell)

Our aim is to study the following question: of the 2^{2^n} Boolean functions on n variables, how many are expressible as 2-SAT formulae? In other words we wish to count the number of different instances of 2-SAT, counting two instances as equivalent if they have the same set of satisfying assignments. Viewed geometrically, we are asking for the number of subsets of the n-dimensional discrete cube that are unions of $(n-2)$-dimensional subcubes.

There is a trivial upper bound of $2^{4\binom{n}{2}}$, as this is the total number of 2-SAT formulae. There is also an obvious lower bound of $2^{n^2/2}$, corresponding to the monotone 2-SAT formulae. So what is the correct speed? Is there a constant c such that the number of such functions is $(c+o(1))^{n^2}$, and, if so, what can we say about the value of c?

Our main result is that, rather surprisingly, the trivial lower bound gives the correct speed: the number of 2-SAT functions is $2^{n^2/2+o(n^2)}$. We also prove some results about the number of k-SAT functions, and make a number of related conjectures.

Singularity Analysis and the Combinatorial/Probabilistic Analysis of a Class of Search-Tree Functionals

James Allen Fill

(joint work with Philippe Flajolet and Nevin Kapur)

For integer $m \geq 2$, the m-ary search tree, or multiway tree, generalizes (in order to produce quicker searches) the binary search tree, a fundamental data structure. An *m-ary tree* is a rooted tree with at most m "children" for each node (vertex), each of which is distinguished as one of m possible types. An *m-ary search tree* is an m-ary tree in which each node has the capacity to contain $m-1$ elements of the linearly ordered set $[n]$ of keys (identified here with the records themselves). There is a natural way to associate an m-ary search tree with a sequence of n distinct keys; see, e. g., Chapter 3 in Mahmoud, Hosam M., *Evolution of Random Search Trees*, Wiley, New York, 1992.

A useful probability model is the *random permutation model*, described as follows. Let π be a uniformly random permutation of $[n]$ and build the naturally associated tree. The distribution of trees under the random permutation model is the distribution induced by this construction, and we denote its probability mass function by Q.

For *binary* search trees $L(T) := -\log Q(T)$ is simply the log-product of branch sizes. The "typical" shape of a binary search tree can be described by studying the distribution of $L(T)$ when T is given the distribution Q. This was done in earlier published work of the speaker; asymptotic expressions for the expected value and variance of $L(T)$ were derived, and a limiting normal law was established. Qualitatively, the speaker's earlier work summarized the results as follows: "Thus it might be fair to say that most binary search trees have a rather 'full' shape, like the complete tree."

For m-ary search trees, the formula for Q is $Q(T) = 1 \Big/ \prod_x \binom{|T(x)|}{m-1}$, where $|T|$ is the number of nodes in a tree T, $T(x)$ is the branch of T rooted at the node x, and the product is over all nodes in T that are filled to capacity. Again we consider $L(T) := -\log Q(T)$ with $T \sim Q$. We treat the induced distribution of $L(T)$ by placing the problem in a unifying framework of additive-type functional on trees. Fix $m \geq 2$, and let f on m-ary search trees satisfy

$$f(T) = \sum_{i=1}^{m} f(T_i) + c_{|T|}, \qquad |T| \geq m-1,$$

where $(c_n)_{n \geq m-1}$ is a given sequence (often called the *toll function*) and T_i denotes the ith subtree of the root of T; the values $f(T)$ for $|T| \leq m-2$ must also be specified. Then we say that f is of *additive type*. Examples include *space requirement* ($c_n := 1$), *internal path length* ($c_n := n - (m-1)$), and our "shape functional" L ($c_n := \ln \binom{n}{m-1}$). For the first example, the small-tree ($|T| \leq m-2$) values $f(T)$ are all 1; for the others, they all vanish.

We provide a general framework for the exact and asymptotic analysis of the distributions of functionals of additive type. In particular, singularity analysis (the extraction of asymptotic information about a sequence from the behavior of its generating function near singularities) can be extremely useful in the consideration of asymptotic distributions, but the tools currently available do not handle the shape functional. We expand the singularity analysis tool kit by proving that if singularity analysis can be applied to each of two generating functions, then it can also be applied to their Hadamard product. This tool allows for the handling not only of the shape functional, but also of a wide variety of asymptotic problems in combinatorics and probability.

Edited by Thorsten Bernholt

Participants

Dr. Sven Erick Alm
sea@math.uu.se
Department of Mathematics
University of Uppsala
P.O. Box 480
S-75106 Uppsala

Thorsten Bernholt
bernholt@ls2.cs.uni-dortmund.de
Lehrstuhl für Informatik II
Universität Dortmund
44221 Dortmund

Prof. Dr. Bela Bollobas
Bollobas@msci.memphis.edu
b.bollobas@dpmms.cam.ac.uk
Dept. of Mathematical Sciences
University of Memphis
Memphis, TN 38152
USA

Prof. Graham R. Brightwell
graham@tutte.lse.ac.uk
Dept. of Mathematics
London School of Economics
Houghton Street
GB-London WC2A 2AE

Prof. Dr. Louis H.Y. Chen
lhychen@math.nus.edu.sg
Department of Mathematics
National University of Singapore
Science Drive 2
Singapore 117543
SINGAPORE

Prof. James Allen Fill
jimfill@jhu.edu
Mathematical Sciences Department
Johns Hopkins University
34th and Charles Streets
Baltimore, MD 21218-2682
USA

Dr. Paul Fischer
paulf@ls2.cs.uni-dortmund.de
Lehrstuhl für Informatik II
Universität Dortmund
44221 Dortmund

Prof. Dr. Andreas Goerdt
goerdt@informatik.tu-chemnitz.de
Fakultät für Informatik und
Informationssicherheit
TU Chemnitz
Str. der Nationen 62
09107 Chemnitz

Prof. Dr. Svante Janson
svante@math.uu.se
Department of Mathematics
University of Uppsala
P.O. Box 480
S-75106 Uppsala

Prof. Dr. Mark R. Jerrum
mrj@dcs.ed.ac.uk
Division of Informatics
University of Edinburgh
King's Building
Mayfield Road
GB-Edinburgh EH9 3JZ

Dr. Yoshiharu Kohayakawa
yoshi@ime.usp.br
Instituto de Matematica e
Estatistica
Universidade de Sao Paulo
rua do Matao 1010
Sao Paulo 05508-900 - SP
BRAZIL

Prof. Dr. Matthias Krause
krause@th.informatik.uni-mannheim.de
Theoretische Informatik
Universität Mannheim
D7,27
68131 Mannheim

Prof. Dr. Imre Leader
i.leader@dpmms.cam.ac.uk
Dept. of Pure Mathematics and
Mathematical Statistics
Center for Mathematical Sciences
Wilberforce Road
GB-Cambridge CB3 OWB

Prof. Dr. Hanno Lefmann
lefmann@informatik.tu-chemnitz.de
Fakultät für Informatik und
Informationssicherheit
TU Chemnitz
Str. der Nationen 62
09107 Chemnitz

Prof. Dr. Nathan Linial
nati@cs.huji.ac.il
School of Computer Science and
Engineering
The Hebrew University
Givat-Ram
91904 Jerusalem
ISRAEL

Dr. Markus Nebel
nebel@sads.informatik.uni-frankfurt.de
FB Biologie und Informatik
Universität Frankfurt
Postfach 111932
60054 Frankfurt

Prof. Dr. Hans Jürgen Prömel
proemel@informatik.hu-berlin.de
Institut für Informatik
Humboldt-Universität zu Berlin
10099 Berlin

Prof. Dr. Rüdiger Reischuk
reischuk@tcs.mu-luebeck.de
Inst. für Theoretische Informatik
Medizinische Universität Lübeck
Wallstr. 40
23560 Lübeck

Dr. Oliver M. Riordan
O.M.Riordan@dpmms.cam.ac.uk
Trinity College
GB-Cambridge CB2 1TQ

Prof. Dr. Andrzej Rucinski
rucinski@amu.edu.pl
Department of Mathematics
UAM
ul. Matejki 48/49
60 769 Poznan
POLAND

Dr. Peter Sanders
sanders@mpi-sb.mpg.de
Max-Planck-Institut
für Informatik
Stuhlsatzenhausweg 85
66123 Saarbrücken

Dr. Christian Schindelhauer
schindel@uni-paderborn.de
Heinz-Nixdorf Institut &
FB Mathematik - Informatik
Universität Paderborn
Fürstenallee 11
33102 Paderborn

Dr. Uwe Schöning
schoenin@informatik.uni-ulm.de
Fakultät für Informatik
Universität Ulm
89069 Ulm

Prof. Dr. Alex Scott
scott@math.ucl.ac.uk
Department of Mathematics
University College London
Gower Street
GB-London WC1E 6BT

Prof. Dr. Miklos Simonovits
miki@renyi.hu
Alfred Renyi Mathematical Institute
of the Hungarian Academy of Science
Realtanoda u. 13-15
P.O.Box 127
H-1053 Budapest

Christian Sohler
csohler@uni-paderborn.de
Heinz-Nixdorf Institut &
FB Mathematik - Informatik
Universität Paderborn
Fürstenallee 11
33102 Paderborn

Dr. Gregory B. Sorkin
sorkin@watson.ibm.com
Dept. of Mathematical Sciences
IBM Thomas J. Watson Research
Center
P.O.Box 218
Yorktown Heights, NY 10598
USA

Prof. Dr. Vera T. Sos
sos@renyi.hu
Alfred Renyi Mathematical Institute
of the Hungarian Academy of Science
Realtanoda u. 13-15
P.O.Box 127
H-1053 Budapest

Dr. Alan Stacey
A.M.Stacey@dpmms.ac.uk
Centre for Mathematical Sciences
University of Cambridge
Wilberforce Road
GB-Cambridge CB3 OWB

Prof. Dr. Endre Szemeredi
szemered@cs.rutgers.edu
Department of Computer Science
Rutgers University
110 Frelinghuysen Road
Piscataway, NJ 08855
USA

Dr. Andrew Thomason
A.G.Thomason@dpmms.cam.ac.uk
Centre for Mathematical Sciences
University of Cambridge
Wilberforce Road
GB-Cambridge CB3 OWB

Prof. Dr. Ingo Wegener
wegener@ls2.informatik.uni-dortmund.de
Lehrstuhl für Informatik II
Universität Dortmund
44221 Dortmund

Prof. John C. Wierman
wierman@jhu.edu
Mathematical Sciences Department
Johns Hopkins University
34th and Charles Streets
Baltimore, MD 21218-2682
USA

Dr. Nicholas Wormald
nick@ms.unimelb.edu.au
Dept. of Mathematics and Statistics
University of Melbourne
Melbourne VIC 3010
AUSTRALIA