Internationales Begegnungs- und Forschungszentrum f ür Informatik

Schloß Dagstuhl

Seminar Report 9415

Expander Graphs, Random Graphs, and their Application in Computer Science

April 11 – 15, 1994

O v e r v i e w

The Dagstuhl Seminar on *Expander Graphs, Random Graphs, and their Application in Computer Science* was organized by Friedhelm Meyer auf der Heide (Universität Paderborn), Hans Jürgen Prömel (Universität Bonn), and Eli Upfal (IBM San José, Weizmann Institute). It brought together 26 participants from 8 countries, 5 of them came from overseas.

The 24 talks presented cover a wide range of topics including various properties of random graphs and expanders, randomized algorithms for allocation, scheduling or simulations, constructions of fault tolerant networks, local graph algorithms, probabilistic learning, and randomized Boolean circuits.

Abstracts of all talks as well as problems presented at the open problem session, chaired by Paul Erdős, are documented in this seminar report.

An interesting excursion to Trier, one of the oldest German cities, took place on Wednesday.

The outstanding environment and organization of Schloß Dagstuhl greatly contributed to the success of the seminar.

Reported by Christian Scheideler and Volker Stemann

Participants

Andrei Broder, DEC SRC - Palo Alto
Walter Deuber, Universität Bielefeld
Devdatt Dubhashi, MPI Saarbrücken
Martin Dyer, University of Leeds
Paul Erdős, Hungarian Academy of Sciences
Alan Frieze, Carnegie Mellon University Pittsburgh
Anna Karlin, DEC SRC - Palo Alto
Michal Karonski, Adam Mickiewicz University
Marek Karpinski, Universität Bonn
Bernd Kreuter,Universität Bonn
Luděk Kučera, Charles University Praha
Yuan Ma, MIT Cambridge
Colin McDiarmid, Oxford University
Friedhelm Meyer auf der Heide, Universität-GH Paderborn
Moni Naor, Weizmann Institute
Hans Jürgen Prömel, Universität Bonn
Rüdiger Reischuk, Institut für Thoeretische Informatik Lübeck
Andrzej Ruciński, Adam Mickiewicz University
Christian Scheideler, Universität-GH Paderborn
Eli Shamir, The Hebrew University of Jerusalem
Paul Spirakis, University of Patras
Anand Srivastav, Freie Universität Berlin
Angelika Steger, Universität Bonn
Volker Stemann, Universität-GH Paderborn
Andrew Thomason, University of Cambridge
Eli Upfal, Almaden Research Center San José

Contents

Anand Srivastav
Randomized Scheduling : Approximation and Non-Approximability

Angelika Steger
The Average Number of Linear Extensions of a Partial Order

Volker Stemann
Fast, Simple Dictionaries and Shared Memory Simulations on Distributed Memory Machines; Lower Bounds

Andrew Thomason
Topological Cliques and Highly Linked Graphs

Eli Upfal
Balanced Allocations

Abstracts

Optimal Construction of Edge-Disjoint Paths in Random Graphs

by Andrei Broder (joint work with A. M. Frieze, S. Suen, and E. Upfal)

Given a graph $G = (V, E)$ with n vertices, and m edges, and a set of κ pairs of vertices in V, we are interested in finding for each pair (a_i, b_i), a path connecting a_i to b_i, such that the set of κ paths so found is edge-disjoint. (For arbitrary graphs the problem is $\mathcal{NP}$-complete, although it is in $\mathcal{P}$ if κ is fixed.)

We present a polynomial time randomized algorithm for finding the optimal number of edge disjoint paths (up to constant factors) in the random graph $G_{n,m}$, for all edge densities above the connectivity threshold. (The graph is chosen first, then an adversary chooses the pairs of endpoints.) Our results give the first tight bounds for the edge disjoint paths problem for any non-trivial class of graphs.

A Variant of Ramsey's Theorem

by Walter A. Deuber (joint work with A. Stacey)

Alon, Spencer, Shelah proved the following variant of Ramsey's theorem.

Theorem : *Let σ be a permutation of m. For every $\delta \in \mathbb{N}$ there exists $A(m, \delta) \in \mathbb{N}$ with the following property : Let the edges of the complete graph on $\{1, \ldots, A(m, \delta)\}$ be colored with δ many colors. Then there exists a monochromatic complete subgraph X with $m + 1$ vertices having type σ.*

Definition : $X = \{x_0, \ldots, x_m\} \subset \mathbb{N}$ *has type* σ *iff* $\forall i, j$

$$x_i - x_{i-1} < x_j - x_{j-1} \quad \text{iff} \quad \sigma(i) < \sigma(j).$$

Here we give a proof of this theorem which gives an upper bound on the third level of the Ackermann hierarchy compared with the fifth level in the original proof. The lower bound is exponential (second level).

Vizing in Parallel

by Devdatt P. Dubashi (joint work with Alessandro Panconesi)

We give a distributed randomized algorithm that finds a proper edge-colouring of a graph. Applied to a graph with n vertices and maximum degree Δ, the algorithm :

- Given any fixed $\lambda > 0$, finds a proper colouring using at most $(1 + \lambda)\Delta$ colours in time $O(\log n)$.
- Given any fixed positive integer s, finds a proper colouring using at most $\Delta + (\log \Delta)^s = (1 + o(1))\Delta$ colours in time $O(\log n + \log \Delta \log \log \Delta)$.

These performance bounds are shown to hold with success probability arbirtarily close to 1, when applied to a class of graphs where $\Delta = \Omega(\log^{1+c} n)$ for any $c > 0$.

The algorithm is a very simple application of the Rödl Nibble, a probabilistic strategy from combinatorics, and the analysis involves a certain preaudo-random process in graphs.

Counting Contingency Tables

by Martin Dyer (joint work with Ravi Kannan, John Mount)

The problem of counting the number of contingency tables with given row and column sums has applications in Statistics. We show that the problem is #P-hard in the general case. The question of randomized approximation is considered. We show that, if the row and column sums are bounded below by small polynomials in the table size, then a modification of a random walk suggested by Diaconis may be used to generate a random table almost uniformly. It is shown how this may be used to approximately count tables.

Vertex-Disjoint Paths in Random Graphs

by Alan Frieze (joint work with A.Z.Broder, S.Suen, E.Upfal)

Consider the random graph $G_{n,m}$ where $d = 2m/n = \ln n + \omega$, $\omega \to \infty$ with n. We consider the problem of finding vertex disjoint paths between prescribed sets of vertices $a_i, b_i, 1 \leq i \leq \kappa$. We prove the following theorems:

Theorem 1 *There exist positive constants α, β such that w.h.p. for all $A = \{a_1, a_2, \ldots, a_\kappa\}$, $B = \{b_1, b_2, \ldots, b_\kappa\}$ satisfying*

(i) $\kappa \leq \alpha n \ln d / \ln n$,

(ii) $|N(v) \cap (A \cup B)| \leq \beta d$,

there exist vertex disjoint paths joining a_i to b_i for $i = 1, 2, \ldots, k$. Furthermore, there is a polynomial time algorithm for finding these paths.

Theorem 2 *There exists a positive constant γ such that w.h.p. for all sets of partitions of $[n]$ into pairs $a_i, b_i, 1 \leq i \leq n/2$ there is a partition of $[n/2]$ into sets $X_1, X_2, \ldots X_r$, $r = \gamma \ln n / \ln d$, of roughly equal size, so that for each $1 \leq i \leq r$ there are vertex disjoint paths joining a_j to b_j for $j \in X_i$.*

Theorem 3 *There exists a positive constant θ such that w.h.p. for all graphs H with λn vertices and maximum degree μd, $\lambda\mu \leq \theta$, $G_{n,m}$ contains a topological copy of H.*

On the Evolution of the Butterfly

by Anna R. Karlin (joint work with Greg Nelson, Hisao Tamaki)

We study the robustness of the butterfly network against random static faults. Suppose that each edge of the butterfly is present independently of other edges with probability p. Our main result is that there is a 0-1 law on the existence of a linear-sized component. More formally, there is a critical probability p^*, $0.34 < p^* < 0.39$, such that for p above p^*, the faulted butterfly almost surely contains a linear-sized component, whereas for p below p^*, the faulted butterfly almost surely does not contain a linear-sized component.

Lower Bounds on Testing Membership to a Polygon by Randomized Computation Trees

by Marek Karpinski (joint work with D. Grigoriev)

We introduce a new method for proving lower bounds for randomized computation trees. We prove, for the first time, that the minimum depth for an arbitrary randomized computation tree testing membership to a polygon with N nodes is $\Omega(\log N)$. This gives also the first (optimal) lower bound for the deterministic algebraic computation trees for this problem. Finally,we prove that the corresponding lower bound for the randomized algebraic exp-log computation trees is $\Omega(\sqrt{\log N})$.

On the Number of Graphs of Girth at least g

by Bernd Kreuter

Let $\mathcal{H}$ be a family of graphs. Denote by $Forb_n(\mathcal{H})$ the set of all graphs not containing a graph from $\mathcal{H}$ as a weak subgraph and let $ex(n, \mathcal{H})$ be the maximum number of edges a graph in $Forb_n(\mathcal{H})$ can have. A theorem of Erdős, Frank and Rödl says that if $\mathcal{H}$ contains only graphs of chromatic number at least three then

$$\log |Forb_n(\mathcal{H})| = (1 + o(1))ex(n, \mathcal{H}).$$

They conjectured that the same would be true if $\mathcal{H}$ is also allowed to contain bipartite graphs which are not trees.

For C_4, the cycle of length four, $ex(n, C_4)$ is $(1 + o(1))n^{3/2}/2$. Kleitman and Winston were able to show that $\log |Forb_n(C_4)| \leq 108n^{3/2}$. We show as a generalization of this result that

$$\log |Forb_n(C_4, C_6, \ldots, C_{2k})| = O(n^{1+\frac{1}{k}})$$

which is also an upper bound for the number of graphs of girth at least $2k + 1$. It is known that $ex(n, C_4, C_6, \ldots, C_{2k}) = O(n^{1+1/k})$ which is sharp for $k = 3$ and $k = 5$.

Interconnecting at Random — The Diameter of a Random Regular Graph

by Luděk Kučera

The motivation for the research presented in the paper is to study how good random regular graphs are when used as interconnection networks. We concentrate to the problem of the diameter of such a graph, because this is one of the most important measures of a network and it is very small for random regular graphs compared to other interconnecting networks.

We present a method to investigate random regular graphs, which is an alternative to configurations and makes it possible to obtain not only known asymptotical results but also numerical bounds as e.g. that the probability that the diameter of a random 5-regular graph with 1024 vertices is less than 6 is at least 0.57. Some numerical results of this type are presented.

Fault-Tolerant Sorting Networks

by Yuan Ma (joint work with Tom Leighton and Greg Plaxton)

Sorting networks have intrigued and challenged computer scientists for decades, and they have proved to be very useful for a variety of applications, including circuit switching and packet routing. With the rapid advance of computer technologies, the study of the fault-tolerance properties of sorting networks has gained increasing importance since the presence of faulty elements is inevitable in any large system.

In this talk, I will present networks and parallel algorithms for sorting that work correctly even when a large number of comparators/comparisons are faulty. Both theoretical and simulation results will be presented. In particular, our networks and parallel algorithms require $o(n \log^2 n)$ comparators/comparisons. Prior to our work, all known fault-tolerant networks and parallel algorithms for sorting used more than $n \log^2 n$ comparators/comparisons.

Every 7-Uniform 7-Regular Hypergraph can be 2-Coloured : The Local Lemma Fights Back !

by Colin McDiarmid

A classical application of the Local Lemma showed that for each $k \geq 9$, every k-uniform k-regular hypergraph can be 2-coloured (that is, the edges can be coloured blue or red so that no edge is monochromatic). Alon and Bregman used a quite different method to show that this holds for every $k \geq 8$. Here, the Local Lemma fights back : we use it to show that the result holds for every $k \geq 7$.

Routing in Arbitrary Networks

by Friedhelm Meyer auf der Heide (joint work with Berthold Vöcking)

Given a network $G = (V, E)$ of degree d, and shortest paths $W_{v,v'}$ between all pairs $v, v' \in V$, the oblivious routing problem for a function $f : V \to V$ is :
For each $v \in V$, send a packet along $W_{v,f(v)}$ from v to $f(v)$. In a step, each v can choose one of the packets it currently stores and send it to the next node on its path. Which packet to choose is determined by a precedence rule.

Given f, L_f denotes the longest path among $W_{v,f(v)}$, $v \in V$, and C_f denotes the congestion, i.e. the maximum number of paths from $W_{v,f(v)}$, $v \in V$, passing through the same node. Clearly, $\Omega(C_f + L_f)$ is a lower bound, $O(C_f \cdot L_f)$ is an upper bound for routing f. Leighton, Maggs, Ranade, Rao have presented a randomized routing protocol that achieves runtime $O(C_f + L_f + \log n)$ w.h.p. in levelled networks, where routing is done from sinks to sources. In the talk, a protocol is presented that achieves the above time bound in arbitrary networks if the paths $W_{u,w}$ are shortest paths.

What Can be Computed Locally ?

by Moni Naor (joint work with Alain Mayer and Larry Stockmeyer)

We study computation that can be done locally in a distributed network. By locally we mean within time (or distance) independent of the size of the network. We consider Locally Checkable Labeling (LCL) problems, where the legality of a labeling can be checked locally (e.g., vertex coloring). Our results include the following :

- There are non-trivial LCL problems that have local algorithms.
- It is undecidable, in general, whether a given LCL has a local algorithm.
- However, it is decidable whether a given LCL has an algorithm that operates in a given time t.
- Randomization cannot make an LCL problem local; i.e., if a problem has a local randomized algorithm then it has a local deterministic algorithm.
- There is a variant of the dining philosophers problem which can be solved locally.

We also investigate the issue of single step color reduction and show that a technique of Szegedi and Vishwanatan can be made constructive.

Malign Distributions and Asymptotic Circuit Complexity

by Rüdiger Reischuk (joint work with Andreas Jakoby and Christian Schindelhauer)

A distribution μ is *malign* for a class of computational problems if for each problem in that class the average time complexity with respect to μ is almost as bad as the worst case complexity (no more than a constant factor smaller). It has been shown that uniform complexity classes, that is algorithms defined by machine programs, have malign distributions. In this talk the nonuniform computational model of Boolean circuits is considered. For the worst case it is known that almost all Boolean functions on n inputs require circuit delay $n - \log\log n + O(1)$.

We show that circuits do not have malign distributions. For any distribution μ on $\{0,1\}^n$ one can find a Boolean function f over that domain with a huge gap between its worst case time complexity, which is $n - \log n - \log\log n$, and its average delay, which is constant.

Asymptotically, however, it is shown that almost all Boolean functions of n inputs have average complexity at least $n - \log n - \log\log n$. Finally we show that for any Boolean function of worst case complexity t one can construct a distribution μ such that the average complexity with respect to μ is at least $t - \log n - \log t$.

Balanced Extensions of Graphs

by Andrzej Ruciński

For a graph G, let $v_G = |V(G)|$, $e_G = |E(G)|$, $d_G = e_G/v_G$ and $m_G = \max_{H \subseteq G} d_H$. A graph G is called *balanced* if $m_G = d_G$. A graph F is a *balanced extension* of G if $m_F = d_F = m_G$.

These notations have their origins in the theory of random graphs (small subgraphs problem), but the following problem is deterministic.

Let $ext(G) = \min\{v_F : F \text{ is a balanced extension of } G\}$. Let $a_n = \max_{G_n} ext(G_n)$, the maximum over all graphs on n vertices. Recently, A. Vince and myself proved that for $n > n_0$

$$a_n = \lfloor \frac{1}{8}(n+3)^2 \rfloor.$$

However, Luczak and myself showed that, setting $m_{G_n} = 1 + \epsilon$, $\epsilon = \epsilon(n) \geq \frac{1}{n}$, one has

$$ext(G_n) = O(\max\{n, \frac{n}{\epsilon}\})$$

whenever $e < \frac{1}{9}$ or $\epsilon > 3.25$. The proof of this result uses expander-like properties of random graphs. The above gap is due to technicalities.

Fast, Simple Dictionaries and Shared Memory Simulations on Distributed Memory Machines; Upper Bounds

by Christian Scheideler (joint work with F. Meyer auf der Heide and V. Stemann)

Assume that a set U of memory locations is distributed among n memory modules, using some number a of hash functions $h_1, \ldots, h_a$, randomly and independently drawn from a high performance universal class of hash functions. Thus each memory location has a copies. Consider the task of accessing b out of the a copies for each of given keys $x_1, \ldots, x_{\epsilon n} \in U$, $b < a$ and $0 < \epsilon \leq 1$. We present and analyse a process executing the above task on distributed memory machines (DMMs) with n processors. Efficient implementations are presented, implying

- a simulation of an n-processor PRAM on an n-processor optical crossbar DMM with delay $O(\log \log n)$,
- a simulation as above on an arbitrary DMM with delay $O(\frac{\log \log n}{\log \log \log n})$, the fastest known PRAM simulation,
- a static dictionary with parallel access time $O(\log^* n + \frac{\log \log n}{\log a})$, if a hash functions are used. In particular, an access time of $O(\log^* n)$ can be reached if $(\log n)^{1/\log^* n}$ hash functions are used.

Learning by Statistical Queries and by Noisy Examples

by Eli Shamir (joint work with Clara Schwartzman)

The general topic is algorithms for learning a suitable approximation f^{approx} of a target function f, out of a (known) class Φ of binary functions over a common domain X. Customarily, the learning is based on getting true values ("labels") $f(x_i)$ at several sample points $x_i \in X$. These labels are supplied by an "Oracle" EX.

We consider a *factorization* of the learning process, so that the examples $f(x_i)$ are used (only) to evaluate "statistical queries" EF_j, which in turn evaluate the derived output f^{approx}. The original formulation of statistical queries by M. Kearns (1993) is extended to include non-linear functionals F_j (i.e. defined on product of independent copies of $Y = X \times \mathbb{R}$) and allow arbitrary range. A major advantage of the statistical queries model is that the learning algorithm stays robust under certain level of isotropic classification noise. We show that the Fourier-transform-based Kushilevitz-Mansour algorithm can be done by second order statistical queries, hence it is robust under classification noise. As a consequence one gets an efficient learning of noisy parity functions.

Short Vertex-Disjoint Paths, Multiconnectivity and Expander Properties in Random Graphs

by Paul G. Spirakis (joint work with S. Nikoletseas, K. Palem and M. Yung)

Let H be an undirected graph. A random graph of type-H is obtained by selecting edges of H independently and with probability p. We can thus represent a communication network H in which the links fail independently and with probability $f = 1-p$. We study here two simple but fundamental types of H : the clique of n nodes (leading to the well known random graph $G_{n,p}$) and a random member of the set of all regular graphs of degree r (leading to a new type of random graphs, of the class $G_{n,p,r}$. Note that $G_{n,p} = G_{n,p,n-1}$). Exact or asymptotic information about the remaining (with high probability) structure of type-H random graphs is of interest to applications in reliable network computing. Here we show that :

1. If p is at least $\max(p_1, p_2)$ where $p_1 = \Omega(\frac{x}{n-xl})$ and $p_2 =$ the x-connectivity threshold of $G_{n,p}$ and $l \geq 2$, then $G_{n,p}$ has for vertex pairs u, v at least x disjoint paths of length at most $2 \cdot l$ each and this holds for all u, v with probability tending to 1 as n tends to infinity.
2. $G_{n,p,r}$ is r-connected except for O(1) vertices whp for all failure probabilities $f = 1-p \leq n^{-\epsilon}$ where $\epsilon > 0$ is a fixed constant.
3. $G_{n,p,r}$ is highly disconnected a.e. for constant f and any $r < \frac{1}{2}\sqrt{logn}$ but is r-connected whp when $r \geq a \cdot \log n$, $a \geq 2$.
4. Even when $G_{n,p,r}$ is disconnected , it still has a giant connected component of small diameter even when $r = O(1)$. An $O(n \log n)$ algorithm is given to construct the giant component.
5. The second eigenvalue of the adjacency matrix of members of $G_{n,p,r}$ is concentrated around its mean and that mean is $O((r \cdot p)^{3/4})$ given that $r \cdot p > 160$. Thus a random member of $G_{n,p,r}$ remains whp a certifiable efficient expander provided that $r \cdot p > 160$.

Randomized Scheduling : Approximation and Non-Approximability

by Anand Srivastav (joint work with Peter Stangier)

We investigate the following resource constrained scheduling problem. Given m identical processors, s renouvable, but *limited* resources, n independent tasks each of one *unit length*, where each task needs one processor and a task dependent amount of every resource in order to be processed. In addition each teask possesses an integer ready-time, which means it cannot be processed before its ready-time. The optimization problem is to assign all tasks to discrete times in $\{1, \dots, n\}$, minimizing the latest completion time C_{max} of any task subject to the processor, resource and ready-time constraints.

The problem is NP-hard even under much simpler assumptions. The best previously known polynomial-time approximation is due to Röck and G. Schmidt, who gave in 1983 in the case of zero ready-times an $O(m)$-factor algorithm. For the analysis of their algorithm, the assumption of identical (zero-) ready-times is essential.

The main contribution of this paper is to break the $O(m)$ barrier (even for non zero ready-times) and to show the first constant factor approximations, which are best possible, unless $P = NP$.

Let C_{opt} be the integer minimum of our scheduling problem and let the integer C denote the size of the minimal schedule, if we consider the LP relaxation, where fractional assignments of the tasks to scheduling times are allowed. We briefly call solutions to the LP relaxation fractional schedules and to the original integer problem, integral schedules.

On the one hand we find for every $\epsilon = \frac{1}{k}$, $k \in \mathbb{N}$, with a new randomized rounding technique a schedule of size at most $\lceil(1+\epsilon)C\rceil$, provided that all resource bounds are $\Omega(\frac{1}{\epsilon^2}\log(Cs))$. Since C is a lower bound for C_{opt}, this gives a schedule of size at most $\rceil(1+\epsilon)C_{opt}\rceil$. For $\epsilon = \frac{1}{C}$ we surprisingly get a schedule of size at most $C_{opt}+1$, provided that the resource bounds are $\Omega(C^2\log(Cs))$. All these algorithms can be derandomized achieving the same approximation guarantee.

On the other hand we can prove for the apparently simpler problem with zero ready-times under the assumption that the resource bounds are $\Omega(C^2\log(Cs))$ (as required in the positive approximation result) the NP-completeness of the problem of deciding whether there exists an integral schedule of size C, even if a fractional schedule of size C and an integral schedule of size $C+1$ are known. The proof is based on an interesting reduction to the chromatic index problem.

The Average Number of Linear Extensions of a Partial Order

by Angelika Steger (joint work with G. Brightwell and H.J. Prömel)

In a seminal paper Kleitman and Rothschild (1975) gave an asymptotic formula for the number of partial orders with ground-set $[n]$. We give a shorter proof of this result and extend it to count the number of pairs $(P, \prec)$, where P is a partial order on $[n]$ and $\prec$ is a linear extension of P. This gives us an asymptotic formula for (a) the average number of linear extensions of an n-element partial order and (b) the number of suborders of an n-element linear order.

Fast, Simple Dictionaries and Shared Memory Simulations on Distributed Memory Machines; Lower Bounds

by Volker Stemann (joint work with F. Meyer auf der Heide and C. Scheideler)

We prove a lower bound for executing the process mentioned in "Fast, simple dictionaries and shared memory simulations on Distributed Memory Machines; upper bounds" , showing that our implementations of this process are optimal.

Topological Cliques and Highly Linked Graphs

by Andrew Thomason (joint work with Béla Bollobás)

Mader showed the existence of a function $f(p)$ such that every graph G satisfying $e(G) \geq f(p)|G|$ contains a topological clique of order p. He, and also Erdős and Hajnal, conjectured that $f(p)$ is of order p^2.

Recently Komlos and Szemeredi proved that $f(p) = O(p^2\log^{14}p)$ and Alon and Seymour proved that $f(p) = O(p^2\log^{1/2}p)$. We prove the conjecture by showing that $f(p) \leq 22p^2$. We also show that if G is $22k$-connected then it is k-linked; that is, for any $2k$ vertices $s_1, \ldots, s_k, t_1, \ldots, t_k$ there exist vertex disjoint $s_i - t_i$ paths, $1 \leq i \leq k$.

Balanced Allocations

by Eli Upfal (joint work with Y. Azar, A.Z. Broder and A.R. Karlin)

Suppose that we sequentially place balls into n boxes by putting each ball into a randomly chosen box. It is well known that when we are done, the fullest box has with high probability $(1 + o(1)) \ln n / \ln \ln n$ balls in it. Suppose instead, that for each ball we choose two boxes at random and place the ball into the one which is less full at the time of placement. We show that with high probability, the fullest box contains only $\ln \ln n / \ln 2 + O(1)$ balls – exponentially less than before. Furthermore, we show that a similar gap exists in the infinite process, where at each step one ball, chosen uniformly at random, is deleted, and one ball is added in the manner above. We discuss consequences of this and related theorems for dynamic resource allocation, hashing, and on-line load balancing.

Problems

Problems in Probabilistic Recurrences

by Devdatt Dubhashi

1. Probabilistic Recurrences

The behaviour of many randomised divide–and–conquer algorithms is a stochastic process which is captured succinctly by a *probabilistic recurrence relation*

$$T(x) = a(x) + T(H(x)). \tag{1}$$

For instance $T(x)$ might stand for the running time of an algorithm that on input x, expends an amount of work $a(x)$ and generates a sub–problem of size $H(x)$. Here $a(x)$ is a fixed function, but $H(x)$ is a random variable in the range $[0, x]$ with a distribution determined by the algorithm in question.

In the literature, the analysis of many randomised algorithms fit this framework. However, their analyses are frequently carried out by disparate *ad hoc* techniques. Karp, [1] recognised that all these algorithms can be analysed uniformly in the above framework and gave general theorems which could be applied in the fashion of a "cook-book" substitution to give the desired performance guarantees on the algorithms (see [2] for some typical examples, or numerous ones exhibited in [1]) To state the hypothesis and results of Karp, we introduce some notations and definitions.

We make only a weak assumption on the distribution of the random variable $H(x)$, namely that $E[H(x)] \leq m(x)$, for a fixed deterministic function, $m(x)$, satisfying $0 \leq m(x) \leq x$. Also, $a(x)$ and $m(x)$ are non-decreasing functions. The equation

$$\tau(x) = a(x) + \tau(m(x)) \tag{2}$$

can be regarded as the deterministic counterpart of the probabilistic recurrence or intuitively, as an equation governing the expected values. Whenever this equation has a solution, it has a unique least non-negative solution $u(x)$, given by $u(x) = \sum_{i \geq 0} a(m^{(i)}(x))$, where inductively, we define $m^{(0)}(x) := x$ and $m^{(i+1)}(x) := m(m^{(i)}(x))$ for $i \geq 0$.

We can now state a *Master Theorem* for the solution to the probabilistic recurrence equation (1) above. The first part is from the seminal work of Karp, [1], while the second part from [2] extends Karp's Theorem by removing some technical restrictions (while in the presence of these restrictions, it reduces essentially to the statement in the first part).

Theorem :

- *If $m(x)$ and $a(x)$ are continuous functions satisfying (1) $m(x)/x$ is nondecreasing and (2) $a(x)$ is strictly increasing on $\{x \mid a(x) > 0\}$. Then for every positive real x and every positive integer w,*

$$\Pr[T(x) \geq u(x) + wa(x)] \leq (m(x)/x)^w.$$

- *Let* $\Delta = \Delta(x) := \max_{b \le y \le x}(m(y)/y)$, *where* b *is the terminating point of the recurrence. Then for every positive real* x *and for sufficiently large positive integer* w,

$$\Pr[T(x) \ge u(x) + wa(x)] \le \Delta(x)^{k/10}.$$

Normally, one would like to see a large deviation result of the form $\Pr[T(x) > E[T(x)] + \cdots] < \cdots$. So the natural question is: how is the solution to the deterministic equation (2) related to $E[T(x)]$. We can give the following partial answer:

Proposition : *Let* a *and* m *both be* **convex** *functions. Then* $E[T(x)] \le u(x)$.

Proof. First observe that the stochastic process described by the probabilistic recurence (1), determines a sequence of non–increasing random variables

$$x =: X_0, X_1, \cdots, X_i, \cdots$$

such that

$$E[X_{i+1} \mid X_i] \le m(X_i) \tag{3}$$

for each $i \ge 0$. Hence we have

$$\begin{aligned} E[X_{i+1}] &= E[E[X_{i+1} \mid X_i]] \\ &\le E[m(X_i)], \text{ using (3)} \\ &\le m(E[X_i]), \text{ since } m \text{ is convex.} \end{aligned}$$

By induction then,

$$E[X_i] \le m^{[i]}(x) \tag{4}$$

for each $i \ge 0$. Finally then, since

$$T(x) = \sum_{i \ge 0} a(X_i)$$

we have

$$\begin{aligned} E[T(x)] &= \sum_{i \ge 0} E[a(X_i)] \\ &\le \sum_{i \ge 0} a(E[X_i]), \text{ since } a \text{ is convex} \\ &\le \sum_{i \ge 0} a(m^{[i]}(x)), \text{ using (4)} \\ &= u(x) \end{aligned}$$

∎

Hence in this situation (a, m convex), the Master Theorem yields the large deviation bounds in the usual form. Moreover in many applications, the conditions on a, m do indeed hold. However it would be nice to replace these conditions on a, m by more natural ones or perhaps to remove them altogether.

2. Lower Bounds in Distributed Computation and Communication Complexity

In the simplest formulation of the *communication complexity* problem, there are two agents X and Y having pieces of data x and y respectively and they wish to collectively compute a function $f(x,y)$. The basic question is: how much information must they exchange in order to do this? A truly distributed version of this question would involve a setting where the answer to be computed is a pair $(f_X(x,y), f_Y(x,y))$, X wishing to know f_X and Y, f_Y. In general, one could have n agents, $X_1, \cdots, X_n$, agent X_i having in his possession data x_i, for $1 \leq i \leq n$ and the answer to be computed is a vector–valued function $F(x_1, \cdots, x_n) = (f_1(x_1, \cdots x_n), \cdots, f_n(x_1, \cdots, x_n))$, agent X_i desiring to know f_i.

This framework is not merely a mathematical nicety, but in fact has a natural computational interpretation in the *distributed* model of computing, [3,4]. One has an underlying graph whose nodes correspond to processors and edges to bidirectional communication links. The computation is *synchronised* and proceeds in a number of rounds. We are interested in *feasible* or *efficient* algorithms, namely those that terminate in a number of rounds *polylogarithmic* in the number of vertices of the underlying graph. The information available at each processor is its own *local* neighbourhood (i.e. its set of neighbours, or equivalently, its row in the adjacency matrix of the underlying graph). The function desired to be computed is a *global* function of the whole underlying graph, each processor desiring to know only the localisation of the function restricted to its neighbourhood.

For example, one may wish to compute an edge colouring of the graph, each processor wishing to know only the colours of edges incident upon it. In [5], such a distributed algorithm is presented which works in a polylogarithmic number of rounds and produces, for any fixed positve integer s, a $\Delta + \Delta/(\log\Delta)^s = (1+o(1))\Delta$ edge colouring of the graph (where Δ is the maximum degree of the graph). We conjecture that this is optimal. Specifically, that any algorithm running in time $O(\log^{O(1)} n)$ cannot use less than $\Delta + \Delta/\log^s\Delta$ colours for all $s > 0$ and some evidence towards this is offered in [5]. We propose that the appropriate framework in which to obtain such lower bounds in the distributed model of computing is to view them as generalised communication complexity problems as suggested above.

References :

[1] Richard M. Karp, "Probabilistic Recurrence Relations", *Proc. 23 ACM Symp. on The Theory of Computing*, pp. 190–197, 1991.

[2] S. Choudhuri, D. Dubhashi, A. Panconesi, "Probabilistic Recurrence Relations Revisited", unpublished, 1994

[3] N. Linial, "Local–Global Phenomena in Graphs", *Combinatorics, Probability and Computing*, 2:4, pp.491–503, 1993.

[4] A. Panconesi, *Locality in Distributed Computing*, Ph.D. Thesis, Cornell University, 1993.

[5] D. Dubhashi and A. Panconesi, "Near Optimal Distributed Edge Colouring", Schloss Dagstuhl 1994.

Probability Methods in Combinatorial Number Theory and some Unsolved Problems

by Paul Erdős

Denote by $r(u,v)$ the smallest integer for which if we color the edges of a complete graph of $r(u,v)$ vertices by two colors there is always a complete monochromatic graph of u vertices of color 1 or a complete monochromatic graph of v vertices of color 2. It is known that

$$cn \cdot 2^{n/2} < r(n,n) < \frac{\binom{2n}{n}}{n^{1/2}}$$

The upper bound is due to Scekeres and Thomason, the lower bound which is probabilistic is due to me. I offer 100 dollars for a proof that $\lim_{n\to\infty} r(n,n)^{1/n}$ exists and 250 for its value.

Several other combinatorial problems were discussed. Denote by $f(n)$ the number of solutions of $n = a_i + a_j$, $a_1 < a_2 < ...$ is a sequence of integers. I proved by probabilistic methods that there is a sequence $a_1 < a_2 < \ldots$ for which for every n it holds

$$c_1 \log n < f(n) < c_2 \log n \qquad (1)$$

I offer 100 dollars for a constructive proof of (1) or even for a constructive proof of the existence of a sequence satisfying

$$0 < f(n) < n^{\epsilon}.$$

I offer 500 dollars for a proof (or disproof) that there is no sequence for which

$$\lim_{n\to\infty} \frac{f(n)}{\log n} = c, \quad 0 < c < \infty$$

holds, also 500 dollars for a proof or disproof of our old conjecture with Turan : If $f(n) > 0$ for all $n > n_G$ then $\lim_{n\to\infty} f(n) = \infty$.

Here is a problem of V.T.Sós, Sárközy and myself : Is it true that for every $\epsilon > 0$ there is a sequence of integers

$$1 \le a_1 < \ldots < a_k \le n, \quad n > n_0(\epsilon), \quad a_i + a_j \text{ all distinct}$$

for which if $s_1 < s_2 < \ldots < s_t \le n$ are the integers of the form $a_i + a_j$ then

$$\max\{s_{i+1} - s_i\} < c \cdot n^{1/2} \ ? \qquad (2)$$

(2) holds for some Sidon sequence.

Let there be given $2^{n-2}+1$ points in the plane no three on a line. Scekeres conjectured more than 60 years ago that one can always find n of them which form the vertices of a convex n-gon. If $f(n)$ is the smallest such integer Scekeres and I proved

$$2^{n-2} + 1 \le f(n) \le \binom{2n-4}{n-2}.$$

$f(6) = 17$ is open (100 marks for a proof or disproof).

Problems for the Hypercube

by Alan Frieze

Let C_k be the k-dimensional Hypercube.

Question 1 : How many perfect matchings m_k does C_k have ? Known :

$$2^{2^k} \leq m_k \leq k^{2^{k-1}}, \quad k \geq 3$$

(a) $m_k =?$, (b) $m_k \approx ?$, (c) $\log m_k \approx ?$

Question 2 : How many cycles does a random 2-factor of C_k have ?

Question 3 : Find a p such that

$$\lim_{k\to\infty} \Pr\{C_k \subseteq G_{n,p}\} = 1, \quad n = 2^k$$

A Complexity Problem for Hypergraphs

by Marek Karpinski

Given a random hypergraph $H = (V, E)$. What is the parallel complexity of computing a maximal independent set (MIS) in H ? In the case of graphs (hypergraphs of dimension 2) a parallel greedy algorithm running in expected $O(\log n)$ time is known (Calkin, Frieze, 1990). It is also known that in a nonexplicit model of computation in which a hypergraph is given by the independence system oracle, the lower bound for any randomized parallel algorithm for the MIS problem working with p processors is $\Omega((n/\log(np))^{1/3})$ (Karp, Upfal, Wigderson, 1984).

A Distribution Problem for DMMs

by Friedhelm Meyer auf der Heide

Let $P_1, \ldots, P_n$ be processors, $M_1, \ldots, M_n$ memory modules, $U = \{1, \ldots, p\}$ for some (large) integer p. Let $h : U \to \{1, \ldots, n\}$ be some function, $t \in \mathbb{N}$. Consider the following process :

Initially, each P_i has keys $x_{i,1}, \ldots, x_{i,t} \in U$.
Each P_i tries to deliver $x_{i,1}, \ldots, x_{i,t}$ to $M_{h(x_{i,1})}, \ldots, M_{h(x_{i,t})}$ in this order, i.e. the following process is executed.

$a(i) = 1$ for all $i \in \{1, \ldots, n\}$
While not all P_i are finished do :
  $\forall i$: P_i tries to deliver $y_i = x_{i,a(i)}$ to $M_{h(y_i)}$.
  $\forall j$: Each M_j accepts one arbitrary of the incoming y_i's.
    If it accepts y_i, then also all $y_{i'}$ with $y_{i'} = y_i$ are accepted.
  $\forall i$: If y_i is accepted, then $a(i) = a(i) + 1$.
    If $a(i) = t + 1$ then P_i is finished and quits.

It is known that, for $t \geq \log n$ and h a random function, the above process needs $O(\log n)$ rounds with high probability, *provided that all $x_{i,j}$'s are distinct.* On the other hand, there are easy examples showing that time $\Omega((\log n)^2)$ can show up if some $x_{i,j}$'s may be identical.

The Problem : Assume that for each i, $x_{i,1}, \ldots, x_{i,t}$ are distinct. How many rounds are sufficient, with high probability, if h is a random function ?

The Carpool Problem

by Moni Naor

Consider the following problem: on a set of n nodes labeled $\{1, \ldots, n\}$ there is a (possibly infinite) sequence of edges, i.e. pairs of nodes. Given a an undirected edge it should be oriented. The goal is devise a method of orienting the edges so that in every node at every point in time the difference between the indegree and outdegree is as small as possible. (Note that it can be shown that for every sequence there exists an orientation such that at any point in along the sequence the max difference is 1.) The problem comes in three flavors:

1. Find a *deterministic* rule of orienting the edges and analyze it on the the worst input sequence.
2. Suggest a rule and analyse it under the some assumption on the distribution of the sequence, in particular that each edge in the sequence is chosen uniformly from all possible edges and independently of the rest of the sequence.
3. Suggest a *randomized* rule for orienting the edges and analyze its expected performance on the worst sequence.

The greedy algorithm is the one where an edge is oriented from the node with the smaller difference between the outdegree and indegree to the one with the larger difference. In the deterministic version of the rule ties are broken according to the lexicographic order. In the randomized version of the rule ties are broken at random. Ajtai, Aspnes, Naor, Rabani, Schulman and Waarts have investigated all three flavors of the problem and obtained the following:

1. The deterministic performance is $n/2$: there is a method (the greedy algorithm) that achieves this bound and for any deterministic rule there is a sequence where this difference will occur.
2. The expected difference of the greedy algorithm on the uniform distribution on the edges is $\Theta(\log \log n)$.
3. There is a rule (local greedy) with expected performance $O(\sqrt{n \log n})$ on any sequence. The lower bound is $\Omega(\sqrt[3]{\log n})$.

The main open problem therefore is to close the gap in flavor (3). In particular, analyse the randomized greedy algorithm on the worst input.

Conjecture: The performance of the randomized greedy algorithm is $\Theta(\log n)$.

The motivation (and source of the name) of the problem is the following: a set of n persons forms a carpool. On each day a subset of the persons arrive and one of them is designated as the driver. A scheduling rule is required so that the driver will be determined in a 'fair' way. The definition of fairness, suggested by Fagin and Williams, is that the number of time a participants drives (the actual load) will be as close as possible to the total driving load this participant would have driven if it were possible to equally partition a drive between all the people who arrived on a certain day (the desired load). Ajtai et. al showed how to reduce the carpool problem to the edge orientation problem while losing at most a factor of 2 in the maximum difference between the desired load and the actual load.

A Problem for Convex Hulls of Regular Graphs

by A. Ruciński

This problem comes from the analysis of the exponent in Janson's inequality. For a given graph G, let

$$\Omega_G = \{(|V(H)|, |E(H)|) : H \subseteq G, E(H) \neq \emptyset\}.$$

Let $c(G)$ be the number of straight line segments forming the boundary of the convex hull of Ω_G less two. Let $c(\mathcal{F}) = max_{F \in \mathcal{F}} c(F)$.

It is known that $c(\mathcal{G}_n) \sim \frac{2}{5}n$, where $\mathcal{G}_n$ is the family of all graphs on n vertices, and that $c(\mathcal{B}_n) = \Theta(n^{2/3})$, where $\mathcal{B}_n$ is the family of all balanced graphs. Every regular graph is balanced.

Problem : What is $c(\mathcal{R}_n)$, where $\mathcal{R}_n$ is the family of all regular graphs on n vertices ?

Literature :

A. Ruciński, *On convex hulls of graphs*, Ars Combinatorica 1991

T. Łuczak, A. Ruciński, *Convex hulls of dense balanced graphs*, J. Comput. Appl. Math. 41 (1992) 205-213

Random Allocations in Bunches — Is there a Sharp Threshold ?

by Eli Shamir

Let Q be a collection of bins and $\mathcal{B}$ be a collection of bunches — each bunch is a subset of Q. Randomly pick a bunch $B \in \mathcal{B}$, put a ball in each bin of B. Let the random variable T be the cover time of Q, that is the first time each bin contains a ball. The problem is to study T.

Interesting cases :

- $Q_n = \{0,1\}^n = 2^n$ vertices of the hypercube.
- $\mathcal{B}_k^n$ = all subspaces of codimension k (each has 2^{n-k} bins)
- $\tilde{\mathcal{B}}_k^n$ = all Hamming balls (centered at $x \in Q^n$) of 2^{n-k} vertices

It is known that $\sigma(T) =_{n\to\infty} O(E(T))$, $E(T) = c_k^n n$ (σ : standard variation, E : expectation). But is $\sigma =_{n\to\infty} o(E(T))$ true, i.e. is there a sharp threshold of T ?

Remark : Cover time for $\mathcal{B}_k^n$ is actually the number of random k-clauses over n literals (disjunctive clauses) which have a (common) satisfying truth assignment.

E-Mail Addresses

Andrei Broder	broder@src.dec.com
Walter Deuber	
Devdatt Dubhashi	dubhashi@mpi-sb.mpg.de
Martin Dyer	dyer@scs.leeds.ac.uk
Paul Erdős	
Alan Frieze	af1p@andrew.cmu.edu
Anna Karlin	karlin@src.dec.com
Michal Karonski	karonski@plpuam11.amu.edu.pl
Marek Karpinski	marek@cs.bonn.edu
Bernd Kreuter	kreuter@unitas.or.uni-bonn.de
Luděk Kučera	ludek@kam.ms.mff.cuni.cz
Yuan Ma	yuan@theory.lcs.mit.edu
Colin McDiarmid	cmcd@uk.ac.ox.stats
Friedhelm Meyer auf der Heide	fmadh@uni-paderborn.de
Moni Naor	naor@wisdom.weizmann.ac.il
Hans Jürgen Prömel	proemel@unitas.or.uni-bonn.de
Rüdiger Reischuk	reischuk@informatik.mu-luebeck.de
Andrzej Ruciński	rucinski@plpuam11.amu.edu.pl
Christian Scheideler	chrsch@uni-paderborn.de
Eli Shamir	shamir@cs.huji.ac.il
Paul Spirakis	spirakis@cti.gr
Anand Srivastav	anands@inf.fu-berlin.de
Angelika Steger	steger@unitas.or.uni-bonn.de
Volker Stemann	vost@hni.uni-paderborn.de
Andrew Thomason	andrew@pmms.cam.ac.uk
Eli Upfal	ely@almaden.ibm.com / eli@wisdom.weizmann.ac.il

Addresses

Andrei Broder
DEC SRC - Palo alto
Digital Equipment Corporation
Systems Research Center
130 Lytton Ave.
Palo Alto CA 94301
USA
Phone: +1-415-853-2118

Walter Deuber
Universität Bielefeld
Fakultät für Mathematik
D-33615 Bielefeld
Germany

Devdatt Dubhashi
Max-Planck-Institut für Informatik
Im Stadtwald
66123 Saarbrücken
Germany
Phone: +49-681-302-5442

Martin Dyer
School of Computer Studies
University of Leeds
Leeds LS2 9IT
Great Britain
Phone: +44-532-335442

Paul Erdős
Hungarian Academy of Sciences
Mathematical Institute
Pf. 127
H-1364 Budapest
Hungary

Alan Frieze
Carnegie Mellon University
Dept. of Mathematics
Schenley Park
Pittsburgh PA 15213
USA
Phone: +1-412-268-8476

Anna Karlin
DEC SRC - Palo Alto
Digital Equipment Corporation
Systems Research Center
130 Lytton Ave.
Palo Alto CA 94301
USA
Phone: +1-415-853-2107

Michal Karonski
Adam Mickiewicz University
Dep. of Discrete Mathematics
ul. Matejki 48/49
PL-60-769 Poznan
Poland
Phone: +48-61-664713

Marek Karpinski
Universität Bonn
Institut für Informatik
Römerstr. 164
D-53117 Bonn
Germany
Phone: +49-228-550-224

Bernd Kreuter
Universität Bonn
Institut für Diskrete Mathematik
Nassestr. 2
D-53113 Bonn
Germany
Phone: +49-228-698620

Luděk Kučera
Charles University
Department of Applied Mathematics
Malostranske nam. 25
11800 Praha 1
Czech Republic
Phone: +4-22-2451-0286

Yuan Ma
MIT
Laboratory for Computer Science
545 Technologie Square
Cambridge MA 02139
USA
Phone: +1-617-621-2512

Colin McDiarmid
Oxford University
Department of Statistics
1 South Parks Road
Oxford OX1 3TG
Great Britain
Phone: +44-0865-272872

Friedhelm Meyer auf der Heide
Universität-GH Paderborn
FB 17 - Mathematik/Informatik
Warburgerstr. 100
D-33098 Paderborn
Germany
Phone: +49-5251-60-3310 /3306

Moni Naor
Weizmann Institute
Departement of Appl. Mathematics
P.O. Box 26
76100 Rehovot
Israel
Phone: +972-8-343-471

Hans Jürgen Prömel
Rhein. Friedr.-Wilhelms-Univ. Bonn
Forschungsinstitut für Diskrete Mathematik
Nassestr. 2
D-53113 Bonn
Germany
Phone: +49-228-738770

Rüdiger Reischuk
Institut für Thoeretische Informatik
Seefahrtschule
Wallstraße 40
D-23560 Lübeck
Germany
Phone: +49-451-7030-416

Andrzej Ruciński
Adam Mickiewicz University
Faculty of Mathematics and Computer Sci.
Matejki 48/49
PL-60-769 Poznan
Poland

Christian Scheideler
Universität-GH Paderborn
FB 17 - Mathematik/Informatik
Warburgerstr. 100
D-33098 Paderborn
Germany
Phone: +49-5251-60-3319

Eli Shamir
The Hebrew University of Jerusalem
Dept. of Mathematics
91904 Jerusalem
Israel

Paul Spirakis
University of Patras
Dept. of Computer Science and Engineering
Computer Technology Institute
PO Box 1122
26110 Patras
Greece
Phone: +30-61-997703 /997586

Anand Srivastav
Freie Universität Berlin
Institut für Theoretische Informatik
Takustraße 9
D-14195 Berlin
Germany
Phone: +49-030-83875100

Angelika Steger
Universität Bonn
Institut für Diskrete Mathematik
Nassestr. 2
D-53113 Bonn
Germany
Phone: +49-228-738751

Volker Stemann
Heinz Nixdorf Institut
Universität-GH Paderborn
Warburgerstr. 100
D-33098 Paderborn
Germany
Phone: +49-5251-60-3377

Andrew Thomason
University of Cambridge
Department of Pure Mathematics
16 Mill Lane
Cambridge CB2 1SB
Great Britain
Phone: +44-223-337991

Eli Upfal
Almaden Research Center
Department K 53/802
IBM Research Division
650 Harry Rd
San José CA 95120
USA
Phone: +1-408-972-1788