D I M A C S

Center for Discrete Mathematics and Theoretical Computer Science

DIMACS SPECIAL FOCUS ON DISCRETE PROBABILITY: 1997 ANNUAL REPORT

DIMACS Center • Rutgers University • 96 Frelinghuysen Road • Piscataway, NJ 08854-8018 • USA
TEL: 732-445-5928 • FAX: 732-445-5932 • EMAIL: center@dimacs.rutgers.edu
WWW: http://dimacs.rutgers.edu/ • TELNET: telnet info.rutgers.edu 90

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs - Research, Bell Labs, NEC Research, and Telcordia Technologies (formerly Bellcore)

Overview:

This program began during the 1996 - 1997 academic year and continued into 1997 - 1998 at a much lower level. It is jointly organized and sponsored with the Institute for Advanced Study, which is cohosting visitors and workshops with DIMACS.

Organizers:

- Enrico Bombieri, Institute for Advanced Study
- Jeff Kahn, Rutgers University
- Peter Winkler, Bell Labs

Sponsors:

The National Science Foundation
The National Security Agency
AT&T Labs - Research
Bell Labs
Bellcore
The Institute for Advanced Study
Rutgers University
Princeton University

Theme of program:

Probabilistic methods and ideas have by now become central to large parts of discrete mathematics and theoretical computer science. At the same time, recent years have seen a number of major developments in parts of probability and statistical mechanics dealing with topics of an essentially discrete nature (e.g., percolation, random walks and Markov chains, particle systems, branching processes).

A major goal of this program is to foster interaction between the DM/TCS community and the probability and statistical community. This is already happening to some extent, with both the application of increasingly sophisticated tools from probability to DM and TCS, and the solution of problems from probability by combinatorial means. Nonetheless, such interactions, and even awareness of what these communities have to offer each other, are still far more exception than rule. So the program is intended to give a substantial push to a broader and more rapid development of these connections. The program is also reaching out from the mathematical sciences to the physical sciences, especially Statistical Mechanics.

Workshops 1997 - 1998:

- **Workshop on Statistical Physics Methods in Discrete Probability, Combinatorics and Theoretical Computer Science**
  - Dates: March 23-27, 1997
  - Location: Institute for Advanced Study and Rutgers University
  - Attendance: 90
  - Organizers: Jennifer T. Chayes, UCLA and Dana Randall, Georgia Tech

  This workshop focused on new developments and directions at the interface between statistical physics and the fields of discrete probability, combinatorics, information theory and theoretical computer science. The workshop brought together 90 researchers in various fields with the intention of developing a common language and recognition parallels among the fundamental problems they are addressing. Speakers focused on topics which included: Critical percolation - finite-size scaling and conformal invariance; Potts models, random cluster models and the Tutte polynomial; dynamics of the giant component in the random graph model; hard core models and independent sets; noisy channels and network reliability; Markov sampling methods; and phase transitions in random structures and algorithms.

This workshop received additional funding from the National Science Foundation and the National Security Agency.

- **Workshop on Probabilistic Analysis of Algorithms**
  - Dates: May 11-14, 1997
  - Location: Princeton University
  - Attendance: 95
  - Organizers: Alan Frieze, CMU and Michael Molloy, Toronto

  This workshop addressed problems arising in the mathematical analysis of algorithms with emphasis on probabilistic techniques. The theory of algorithms has undergone an extraordinarily vigorous development over the last 20 years or so, and probability theory has emerged as one of its most vital partners. By its nature probabilistic analysis cuts across the boundaries of Computer Science, Discrete Mathematics, Operations Research and Probability Theory.

  This workshop focused on the analysis of algorithms in the average case, assuming a probabilistic distribution on input instances. The goal is to obtain good estimates of various performance measures of algorithms such as solution quality and running time.

  The main objective was to explain why many simple algorithms, often used in practice, perform much better than as predicted by worst-case analysis. Pathological cases are generally rare and simple algorithms can often be shown to perform well in the average case. This is especially true in the case of NP-complete problems where it seems that the probabilistic approach may be the best way and indeed may be the only way to understand the success of heuristic combinatorial algorithms.

  The workshop focused on five major sub-areas: Algorithmic Theory of Random Graphs and Related Combinatorial Problems; Analysis of Euclidean Functionals; Linear Programming and Integer Programming Problems; Analysis of Fundamental Algorithms; and Average Case Complexity.

- **Workshop on Microsurveys in Discrete Probability**
  - Dates: June 2-6, 1997
  - Location: Institute for Advanced Study, Princeton, New Jersey
  - Attendance: 76
  - Organizers: David Aldous, Berkeley, and Jim Propp, MIT

  The goal of this workshop was to focus on "microsurveys" of recently active areas within discrete probability. The five day program included three sessions for open problems and scheduled topics which included: A simple family of tree-valued Markov processes; connections between hyperplane arrangements and card-shuffling; exact sampling from Markov chains; random walks in scenery: every vertex has a color; the emergence of the giant component in the random graph process; the scale-invariant Poisson process; mixing times for Markov chains; and quadratic dynamical systems.

Workshops Planned for the Rest of 1997-98

- **Workshop on Randomization Methods in Algorithm Design**
  - Dates: December 12 - 14, 1997
  - Location: Princeton University
  - Organizers: Panos Pardalos, University of Florida; Sanguthevar Rajasekaran, University of Florida; Jose Rolim, University of Geneva

Selected Research Results

Solution of a Question of Shannon. The *Shannon capacity* of an undirected graph G is the limit, as n tends to infinity, of the n-th root of the independence number of the nth normal power of G. Motivated by a question in information theory, Shannon conjectured in 1956 that the capacity of a disjoint union of two graphs is always equal to the sum of their capacities (he showed that it is always at least that big, and that equality holds when at least one of the graphs is perfect). Noga Alon (V) showed that this is false in a very strong sense: the capacity of a union can be bigger than any fixed power of the sum of the capacities. This has the following somewhat counter-intuitive information theoretic interpretation. By using two channels alternately, it is sometimes possible to obtain capacity that exceeds considerably the sum of the capacities of each of them.

Statistical Mechanics, the Bethe Lattice, and Random 3-Colorings of a Binary Tree Motivated by problems from statistical mechanics where we try to model the behavior of physical phenomena (like magnetism) using the random behavior of small particles, Graham Brightwell (V) and Peter Winkler (PM, BL) have done extensive work on the Bethe lattice (Cayley tree). They have obtained two theorems classifying systems which exhibit phase transitions, and another classification involving ergodicity of graph homomorphisms, plus results on memory in branching random walks. The first classification theorem specifies the constraint graphs for which there are node activities with multiple Gibbs distributions. An illustration from the paper by Brightwell and Winkler appears on the cover of the 1997-98 catalog of the Graduate School of Rutgers - New Brunswick. The illustration shows a random 3-coloring of a binary tree subject to the condition that no red point may be adjacent to a green one. Brightwell and Winkler have used these random colorings as an abstract mathematical model of the phenomena studied in statistical mechanics and have shown that such colorings can exhibit phase transitions akin to the phase transitions exhibited by physical systems. (The illustration shows a "phase" in which green is favored over red, even though the local conditions are the same for those two colors.) The picture was chosen for the cover of the catalog in part because it symbolized the collaborations between mathematicians and computer scientists and researchers in other fields of science, such as physics, chemistry, biology, or the social sciences, collaborations in which graduate students participated through DIMACS.

A Startling Example in Statistical Mechanics Graham Brightwell (V), Olle Haggstrom (V), and Peter Winkler (PM, BL) have used Brightwell-Winkler theory (above), together with an idea of Schonmann and Tanaka, to devise a startling example in statistical mechanics. Random green-yellow-red colorings, with no red node adjacent to a green one and equal conditional probability x for red and green, constitute the Widom-Rowlinson model of statistical mechanics. On a cubic lattice it is known that when x is high one gets multiple phases, i.e., spontaneous symmetry-breaking, and when x is low one doesn't. Physicists believe that there is just one critical value of x where the change takes place, but this remains open (as it does for various other models as well). Applying the calculations on the tree version of this model to the k-branching tree and modifying the tree led Brightwell, Haggstrom, and Winkler to an example where there were three critical points. They proved that there is a unique critical point on the k-branching tree, namely $x = ((k+1)/k)^k/(k-1)$; but if one takes k=40 and adds 7 dangling edges to each node of the tree, bizarre things happen: symmetry is broken for x in the open interval (.217935, .401318) and again after $x = 4.55188$, but elsewhere there is only one (symmetric) phase. There are other known examples of triple critical points but they involve external fields; we believe this is the first case of multiple critical points in a symmetric system.

A Tutte Polynomial for Colored Graphs Bela Bollobas (V) and Oliver Riordan defined a polynomial W on graphs with colors on the edges, by generalizing the spanning tree expansion of the Tutte polynomial as far as possible. In particular, they gave necessary and sufficient conditions on the edge weights for this expansion not to depend on the order used and gave a contraction-deletion formula for W analogous to that for the Tutte polynomial. They showed that generalizations of the Tutte polynomial obtained from its rank generating function formulation, or from a random cluster model, can be obtained from W. Finally, they found the most general conditions under which W gives rise to a link invariant.

Dependent Random Graphs and Spatial Epidemics Jacob van den Berg (V), G.R. Grimmett, and R.B. Schinazi have extended certain exponential decay results of subcritical percolation to a class of locally dependent random graphs, introduced by Kuulasmaa as models for d-dimensional spatial epidemics. In these models, infected individuals eventually die (are removed) and are not replaced. They combined these results with certain continuity and re-scaling arguments in order to improve our knowledge of the phase diagram of a modified epidemic model in which new susceptibles are born at some positive rate. In particular, they showed that, throughout an intermediate phase where the infection rate lies between two certain critical values, no coexistence is possible for sufficiently small positive values of the recovery rate. This result provides a converse to results of Durrett–Neuhauser and Andjel–Schinazi. They also showed that such an intermediate phase indeed exists for every $d \geq 1$ (i.e., that the two critical values mentioned above are distinct). An important technique used is the general version of the BK inequality for disjoint occurrence, proved in 1994 by David Reimer (GS, RU).

The Existence of Finitary Codings for a Class of Random Fields. Jeff Steif (V) and Jacob van den Berg (V) studied the existence of finitary codings from an i.i.d. process to Markov random fields. They showed that the presence of a phase transition is an obstruction for the existence of the above coding; this yields a large class of Bernoulli shifts for which no such coding exists. Conversely, they showed that for the stationary distribution of a monotone exponentially ergodic probabilistic cellular automata such a coding exists. These constructions were inspired by the Propp-Wilson algorithm for exact simulation. In particular, combining the above with a result of Martinelli and Olivieri, this yielded the fact that for the plus state for a d-dimensional ferromagnetic Ising model, $d \geq 2$, there is (not) such a map when the interaction parameter is below (above) its critical value.

Threshold of Broadcast in Random Graphs A broadcasting process involves sending a piece of information which resides at one vertex in a graph to all the remaining vertices. At each time step, a vertex knowing the information can send it to one of its neighbors. The broadcast problem is to find the minimum number of time steps needed. The problem is NP hard in general. It is not hard to show that the edges used by the optimal broadcast algorithm form a tree. The broadcast problem is to find the tree and the minimum broadcast time. Hui Chen (PD) considered the problem, for a random graph $G_{n,p}$, of finding the threshold for p such that with high probability (whp), a piece of information can be broadcasted in the graph in exactly $\lceil \log_2 n \rceil$ steps. (An event E_n is said to occur whp if $\mathbf{Pr}[E_n] = 1 - o(1)$ as $n \to \infty$.) The best upperbound to date is due to Frieze and Molloy who showed that if $p \geq 18 \ln n/n$ then whp, there exists a rooted broadcasting tree on every vertex. Chen showed that a trivial lower bound is also sufficient for the existence of a rooted optimal broadcast tree on each vertex of a random graph $G_{n,p}$.

Mixing Times of Lozenge Tiling and Card Shuffling Markov Chains David Wilson (V) introduced a simple technique that can be applied to a number of classes of Markov chains to obtain upper bounds and lower bounds on their mixing times, the number of steps the Markov chains take to approach their equilibrium distribution. One application is to a class of Markov chains introduced by Luby, Randall, and Sinclair to generate random tilings of regions by lozenges. The mixing time bounds derived by Wilson improve substantially upon the previous results. In another application he resolved a few questions raised by Diaconis and Saloff-Coste, by lower bounding the mixing time of various card shuffling Markov chains. His lower bounds are within a constant factor of their upper bounds. This resolved a question that had been open for about a decade. When he applied this technique to modify an analysis of Bubley and Dyer, he obtained an $O(n^3 \log n)$ upper bound on the mixing time of the Karzanov-Khachiyan Markov chain for linear extensions.

Exact Sampling Algorithms for Storage and Network Models Robert Lund and David Wilson (V) developed algorithms for simulating random samples from the invariant measure of storage and network models without relying upon the analytic form of the invariant measure. Since the invariant measure of many storage and network models is intractable, the proposed algorithms can be used to study the invariant measure of such systems through simulation techniques. Backwards coupling methods were used along with the sample path monotonicities inherent in these models to develop an efficient algorithm for simulating from the limiting distribution. Previously, some sort of truncation was required.

Sampling Points in a Convex Body Russ Bubley, Martin Dyer and Mark Jerrum (V) have described a new method for proving the polynomial-time convergence of an algorithm for sampling (almost) uniformly at random from a convex body in high dimension. Previous approaches had been based on estimating conductance via isoperimetric inequalities. Bubley, Dyer, and Jerrum showed that a more elementary coupling argument can be used to give a similar result.

Set Partitions with Unequal Block Sizes A. Knopfmacher, A. M. Odlyzko (PM, ATTL), B. Pittel (V), L. B. Richmond, D. Stark, G. Szekeres, and N.C. Wormald have determined the asymptotic behavior of the number of set partitions of an n-element set into blocks of distinct sizes. The behavior is more complicated than is typical for the set partition problems. Although there is a simple generating function, the usual analytic methods for estimating coefficients fail in the direct approach.

Confirming two Conjectures about Integer Partitions Boris Pittel (V) has settled two old conjectures about integer partitions. For a given integer n, let Λ_n denote the set of all integer partitions $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m > 0$, $(m \geq 1)$, of n. For the dominance order "$\preceq$"on Λ_n, Pittel showed that if two partitions λ, μ are both chosen from Λ_n uniformly at random, and independently of each other, then $\Pr(\lambda \preceq \mu) \to 0$ as $n \to \infty$. This statement answers affirmatively a question posed by Macdonald in 1979. The proof is based on the limit joint distribution of the largest parts counts found by Fristedt. A slight modification of the argument confirms a conjecture made by Wilf in 1982, namely that, for n even, the probability of a random partition being graphical is zero in the limit.

Matchings and Hamiltonian Cycles in Sums of Random Trees A *random tree* on the set $[n] = \{1, \ldots, n\}$ is a tree on this set chosen uniformly at random from the family of all trees on the set $[n]$. Let k be a positive integer. For trees $T_1, \ldots, T_k$, all of them on the set $[n]$, we define their sum as the graph on the vertex set $[n]$ and edge set being the union of edge sets of the trees $T_1, \ldots, T_k$, where the parallel edges coalesce. We denote by $\mathbf{ST}_n(k)$ the sum of k random trees on $[n]$. Let f be a mapping from $[n] \to [n]$. Let $D(f)$ be its associated functional digraph, i.e., the graph with vertex set $[n]$ and edges $(i, f(i))$, $i \in [n]$. Similarly to the sum of trees we define the sum of mappings. We use $\mathbf{SM}_n(k)$ for the sum of k random mappings. Alan Frieze showed in 1996 that $\mathbf{SM}_n(2)$ almost surely has a perfect matching and Cooper and Frieze showed in 1996 that $\mathbf{SM}_n(4)$ almost surely has a hamiltonian cycle. The problem of whether or not $\mathbf{SM}_n(3)$ almost surely has a hamiltonian cycle is one of the most important open questions in the theory of random graphs. Alan Frieze (V), Michel Karonski, and Lubos Thoma (PD) have proved that

$$\lim_{\substack{n\to\infty \\ n \text{ even}}} Prob(\mathbf{ST}_n(1) \text{ has a perfect matching }) = 0$$

and

$$\lim_{\substack{n\to\infty \\ n \text{ even}}} Prob(\mathbf{ST}_n(2) \text{ has a perfect matching }) = 1$$

and that

$$\lim_{n\to\infty} \text{Prob}(\mathbf{ST}_n(5) \text{ has a hamiltonian cycle }) = 1.$$

Probability that a Random Structure has a Given Property Joel Spencer (V) and Lubos Thoma (PD) have developed a procedure characterizing limit values of probabilities for a random structure having a property from a given language L. The method is based on completions of almost sure theories. Let L be a (first or second order) language. For each $n \in \mathbf{N}$ we are given a probability space (Ω_n, μ_n), where Ω_n is a finite set of models of L and μ_n is a probabilistic measure on Ω_n. (In the particular application by Spencer and Thoma, we set (Ω_n, μ_n) to be the random graph $G(n, p)$.) We assume all subsets of Ω_n are μ_n-measurable. For $\psi \in L$ we write $\mu_n(\psi)$ for $\mu_n(\{M \in \Omega_n : M \models \psi\})$, the probability ψ holds in a "random" model. In 1991, T. Luczak and J. Spencer classified probabilities p with $p = \Theta(\log n/n)$ at which the $0-1$ law does not hold, i.e., thresholds in the neighborhood of the point where the random graph $G(n, p)$ becomes connected. Namely, the 0-1 law does not hold for the probabilities of the form $p_{k,l} = p_{k,l}(n) := (\ln n + l \ln\ln n + c + o(1))/(kn)$ where $l \geq k-1 \geq 0$. Spencer and Thoma have obtained a full characterization of limit probabilities for the random graph $G(n, p_{k,l})$ and first order properties of graphs. Due to the generality of the results, one can expect the procedure developed by Spencer and Thoma to play a role in characterizing limiting probabilities for other random structures.

DIMACS special foci cross-fertilize each other. There was considerable cross-fertilization between the special focus on discrete probability and that on molecular biology. The following results are examples of the impact of that cross-fertilization, and were already mentioned in Section 5.2.

Thresholds for Ordering Libraries of Variable Length Clones One of the many biological problems to which mathematicians have given attention in recent years is the Physical Mapping Problem from genetics. To attain a physical map, many copies of a genome are digested by various enzymes, creating a library of overlapping clones. One technique involves further digestion of the clones to get sets of digestion lengths. Hui Chen (PD) has considered the technique of testing the clones for the existence of various DNA subsequences called probes. If the probes are (relatively) long, they will appear at most once in the genome, and they are called unique probes. If the probes are (relatively) short, their appearances throughout the genome will be according to a Poisson process (assuming randomness of the base pairs in the genome), and these are called Poisson probes. In either case, clone overlaps can be detected by commonality of probes, and this information can be used to reconstruct the actual overlap structure of the clone library, and the ordering of the probes on the genome. This information constitutes the physical map, which can be used by biologists for further sequencing, or as a guide to other genetic research. Chen showed that a gapless distribution of clones, with distinct endpoints, has a unique ordering of the atomic intervals (up to complete reversal) satisfying the so-called consecutive 1's property if and only if there are no two atomic intervals the same. The consecutive 1's property provides a simple test of whether or not a graph is an interval graph, a central concept in the history of molecular genetics and in particular in physical mapping.

Sequence Reconstruction Problems of sequence reconstruction led to some interesting results by Richard Arratia (V), Bela Bollobas (V), Don Coppersmith, and Greg Sorkin. The original question they considered arises in SBH (Sequencing by Hybridization), which is the attempt to reconstruct a long sequence from the knowlege of which q-tuples it contains, without being given the knowledge of the order in which those q-tuples occur. For random sequences, through deBruijn graphs, this reduces to a question of enumerating Euler circuits. Arratia, et al. showed that the interlacement graph, which gives only the relation between pairs of repeated pairs of (q-1)-tuples, determines the number of Euler circuits. The tool for this leads to a new graph polynomial, which may be of independent interest.