D I M A C S

Center for Discrete Mathematics and Theoretical Computer Science

DIMACS SPECIAL FOCUS ON DISCRETE PROBABILITY: 1999 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 - 1999 at a much lower level. While there are currently no plans for workshops under this theme, there remains great interest in it among DIMACS members, visitors, postdocs, and students, as is reflected in the research described in this section. This special focus has been jointly organized and sponsored with the Institute for Advanced Study, which has cohosted visitors and workshops with DIMACS.

Organizers:

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

Sponsors and Collaborating Institutions:

The National Science Foundation
The National Security Agency
AT&T Labs - Research
Bell Labs
Georgia Tech
The Institute for Advanced Study
NEC Research Institute
National Institute for Standards and Technology (NIST)
Princeton University
Rutgers University
Telcordia Technologies

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 has been to foster interaction between the dm/tcs community and the probability and statistics community. This was already happening to some extent before the special focus, 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, were still far more exception than rule. So the program has been intended to give a substantial push to a broader and more rapid development of these connections. The program also reaches out from the mathematical sciences to the physical sciences, especially statistical mechanics, and has been opening up and/or enhancing lines of communication and collaboration between mathematical and physical scientists.

Workshops Feb. 1999 - Jan. 2000:

- **Workshop: Combinatorial Methods for Statistical Physics Models**
  - Dates: April 29 - May 1, 1999
  - Location: Georgia Institute of Technology

– Organizers: Dana Randall and Prasad Tetali, Georgia Tech

– Co-Sponsors: Georgia Institute of Technology, The Southern Applied Analysis Center (SAAC), The Algorithms, Combinatorics and Optimization Program (ACO)

This workshop focused on recent developments at the interface among combinatorics, statistical physics and theoretical computer science. Topics included Gibbs measures and phase transitions in various models (such as the Potts model, hardcore lattice gases and dimer systems), percolation theory, and mixing rates of finite Markov chains. It was a follow-up to the earlier activities of the Special Focus on Discrete Probability, during which we explored the interface among discrete mathematics, discrete probability, and statistical mechanics. The aim of the workshop was to address some of the fascinating questions in combinatorics and discrete probability theory that have arisen with somewhat independent motivations (from the above topics) and which seem to require techniques from several disciplines for solutions. There were introductory lectures intended for a wide audience as well as more focused talks highlighting recent research trends; beginners to the topic with a background in basic combinatorics and probability were encouraged to attend. Peter Winkler opened the workshop by speaking about percolation and collision. Jennifer Chayes followed with a discussion of the scaling window for 2SAT. Roman Kotecky presented an overview of the $q = 3$ Potts model, Sorin Istrail talked about the impossibility of exact solubility of the general 3-dimensional Ising model, David Wilson discussed scaling limits for minimal and random spanning trees in two dimensions, and Umesh Vazirani discussed group theory, Fourier transforms, and quantum computation. This was followed by a presentation on independence properties and combinatorial uss of monomer-dimer systems by Jeff Kahn, one on mixing properties of the monomer-dimer model, with applications to random sampling, by Rob van den Berg, and one on slow mixing for the independent set model and critical Swendsen-Wang by Christian Borgs. Mark Huber spoke about sampling independent sets, Gregory Sorkin about the random assignment problem, Rick Kenyon about long-range properties of spanning trees in Z^2, and Jeff Steif about amenability and nonamenability of graphs and connections with Ising. Finally, Claire Kenyon discussed mixing rates of block dynamics in one dimension, Eric Vigoda talked about improved bounds for sampling colorings, Stefan Boettcher described the virtues of large fluctuations in extremal optimization, and Kari Eloranta discussed bounded ice on planar lattices.

Selected Research Results

Testability of Graph Properties We have obtained several results in the field of *Combinatorial Property Testing.* Property testing deals with the question whether a given input x satisfies a prescribed property P or is "far" from any input satisfying it. A challenge here is to develop efficient randomized algorithms which will be able to answer the above question extremely quickly and yet reliably. Different choices of the property P lead to very diverse algorithmic problems, whose solution is based frequently on deep combinatorial results and techniques. Noga Alon (V), Eldar Fischer, Michael Krivelevich (PD) and Mario Szegedy (PM, ATTL) studied testability of graph properties. Let P be a property of graphs. An ϵ-test for P is a randomized algorithm which, given the ability to make queries whether a desired pair of vertices of an input graph G with n vertices are adjacent or not, distinguishes with high probability between the case of G satisfying P and the case that it has to be modified by adding and removing more than ϵn^2 edges to make it satisfy P. The property P is called *testable* if for every ϵ there exists an ϵ-test for P whose query complexity is independent of the size of the input graph. Alon, et al. studied testability of graph properties defined by first order statements, and showed that properties whose logical description does not contain a quantifier alternation of type $\forall\exists$ are always testable, while some properties containing this alternation are not. The proof is based on a quite involved application of the Szemerédi Regularity Lemma.

Testability of Formal Languages Noga Alon (V), Michael Krivelevich (PD), Ilan Newman and Mario Szegedy (PM, ATTL) studied testability of formal languages. Their main result is as follows. For a regular language $L \in \{0,1\}^*$ and an integer n there exists a randomized algorithm which always accepts a word w of length n if $w \in L$, and rejects it with high probability if w has to be modified in at least ϵn positions to create a word in L. The algorithm queries $\tilde{O}(1/\epsilon)$ bits of w. This result shows that all regular languages are quickly testable. They also studied testability of more complex languages and showed, in particular, that the query complexity of testing context-free languages cannot be bounded by any function of ϵ.

Monotonicity Testing Consider a problem of *testing whether a given Boolean function is monotone.* A function $f : \{0,1\}^n \to \{0,1\}$ is said to be *monotone* if $f(x) \leq f(y)$ for every $x \preceq y$, where $\preceq$ denotes the natural partial order among strings: $x \preceq y$ iff $x_i \leq y_i$ for all $1 \leq i \leq n$. The testing algorithm can request the value of the function on arguments of its choice, and is required to distinguish monotone functions from functions which are far from being monotone. More precisely, the testing algorithm is given a *distance* parameter $\epsilon > 0$, and oracle access to an unknown function f mapping $\{0,1\}^n$ to $\{0,1\}$. If f is monotone then the algorithm should accept it with probability at least $2/3$, and if f is at distance greater than ϵ from any monotone function then the algorithm should reject it with probability at least $2/3$. Distance between functions is measured in terms of the fraction of the domain on which the functions differ. The complexity measures that are usually focused on are the *query complexity* and the *running time* of the testing algorithm. Monotonicity testing is an example of combinatorial property testing. Improving on an earlier algorithm, O. Goldreich, S. Goldwasser, E. Lehman, D. Ron, and A. Samorodnitsky (PD) showed that there is an algorithm for Boolean monotonicity testing with query complexity and running time of order $O(n/\epsilon)$. A key step in the improved analysis is application of *shifting* of Boolean functions. This technique was extended further by Goldreich, et al. and then by Goldreich, E. Dodis, Lehman, D. Ron, S. Raskhodnikova, and Samorodnitsky (PD) to obtain monotonicity testers for wider classes of functions.

Combinatorial Aspects of Testing k-colorability of Graphs. Graph coloring is an old subject with many critical modern applications. Noga Alon (V) and Michael Krivelevich (PD) considered-combinatorial aspects of testing k-colorability of graphs and proved the following result. Let G be a graph on n vertices and suppose that at least ϵn^2 edges have to be deleted from it to make it k-colorable. Then most induced subgraphs of G on $O(k \ln k/\epsilon^2)$ vertices are not k-colorable. If G is as above for $k = 2$, then most induced graphs on $\tilde{O}(1/\epsilon)$ vertices are non-bipartite, and this is tight up to the poly-logarithmic factor. These combinatorial results lead immediately to algorithms for testing k-colorability whose query complexity is $O(1/\epsilon^4)$ for every fixed $k \geq 3$ and $\tilde{O}(1/\epsilon^2)$ for the case $k = 2$, thus improving results of Goldreich, Goldwasser and Ron.

Choosability Properties of Random Hypergraphs Michael Krivelevich (PD) and Van Vu studied choosability properties of random hypergraphs. Choosability problems arise when we specify in advance the set of allowable colors for an edge. A hypergraph $H = (V, E)$ is called k-choosable, if for every family of color lists $\mathcal{S} = \{S(v) \subset Z : v \in V\}$ satisfying $|S(v)| = k$ for all $v \in V$, there exists a choice function $f : V \to Z$ such that $f(v) \in S(v)$ for all $v \in V$, and no edge of H is monochromatic under f. The choice number of H, denoted by $ch(H)$, is the minimal k for which H is k-choosable. A random hypergraph $H(k, n, p)$ is a k-uniform hypergraph on the set V of n labeled vertices, where each k-tuple of V is chosen to be an edge of H independently and with probability $p = p(n)$. The main result of Krivelevich and Vu states that for every $k \geq 2$ and for all values of p down to $p(n) = O(n^{-k+1})$ the ratio between the choice number and the chromatic number of $H(k, n, p)$ does not exceed $k^{1/(k-1)}$ asymptotically. Moreover, for large values of p, namely, when $p \geq n^{-(k-1)^2/(2k)+\epsilon}$ for an arbitrary positive constant ϵ, the choice number and the chromatic number of $H(k, n, p)$ have almost surely the same asymptotic value.

Packing Random Rectangles A random rectangle is the product of two independent random intervals, each being the interval between two random points drawn independently and uniformly from [0,1]. Ed Coffman (PM, BL), George S. Lueker (V), Joel Spencer (V), and Peter M. Winkler (PM, BL) proved that the number of items in a maximum cardinality disjoint subset of n random rectangles has the tight asymptotic bound $\Theta(n^{1/2})$. Although tight bounds for the problem generalized to $d > 2$ dimensions remain an open problem, Coffman, et al. showed that $\Omega(n^{1/2})$ and $O((nlog^{d-1}n)^{1/2})$ are asymptotic lower and upper bounds. In addition, they proved that $\Theta(n^{d/(d+1)})$ is a tight asymptotic bound for the case of random cubes.

Large Deviations for Ideal Quantum Systems Joel Lebowitz (PM, RU), Marco Lenci, and Herbert Spohn (V) considered a general d-dimensional quantum system of non-interacting particles in a very large (formally infinite) container. They proved that, in equilibrium, the fluctuations in the density of particles in a subdomain Ł of the container are described by a large deviation function related to the pressure of the system. That is, untypical densities occur with a probability exponentially small in the volume of Ł, with the coefficient in the exponent given by the appropriate thermodynamic potential. Furthermore, small fluctuations satisfy the central limit theorem.

An Entropy Approach to the Hard-core Model on Bipartite Graphs Jeff Kahn (PM, RU) introduced entropy as a tool for tackling the question of critical activity for the hard-core model on Z^d — a (still open) problem which was regularly under discussion during this special focus — and settled a related question of J. van den Berg (V). Specifically, Kahn used entropy ideas to study hard-core distributions on the independent sets of a finite, regular bipartite graph, specifically distributions according to which each independent set I is chosen with probability proportional to $\lambda^{|I|}$ for some fixed $\lambda > 0$. Among the results he obtained were rather precise bounds on occupation probabilities; a "phase transition" statement for Hamming cubes; and an exact upper bound on the number of independent sets in an n-regular bipartite graph on a given number of vertices.

Entropy, Independent sets and Antichains "Dedekind's Problem" of 1897 asks for the number $\psi(m)$ of elements in the free distributive lattice on m generators, or equivalently, of antichains in the Boolean algebra $\mathcal{B}_m$. In a 1969 breakthrough, Kleitman used an ingenious elaboration of the basic approach of Hansel to pinpoint the asymptotics of the logarithm of $\psi(m)$. Kleitman's argument was developed considerably further by Kleitman and Markowsky in 1975. For n-regular, N-vertex bipartite graphs with bipartition $A \cup B$, Jeff Kahn (PM, RU) gave a precise bound for the sum over independent sets I of the quantity $\mu^{|I \cap A|}\lambda^{|I \cap B|}$. (In other language, this is bounding the *partition function* for certain instances of the *hard-core model.*) He extended the result to graded partially ordered sets, which in particular provided a simple proof of the well-known bound for Dedekind's Problem given by Kleitman and Markowsky.

Proof of a Conjecture in Combinatorial Probability, with Consequences in Computational Complexity Jeff Kahn (PM, RU), Michael Saks (PM, RU) and Cliff Smyth (GS, RU) proved a well known conjecture of Steven Rudich in combinatorial probability. They formulated and proved a dual form of the well known Reimer - van den Berg - Kesten inequality, and used it to deduce Rudich's conjecture. This conjecture is known to have some interesting consequences in computational complexity. Let X be a set of Boolean variables. A *term* is a product of literals (where a *literal* is a variable or its complement). Let T be a set of terms, and consider the graph with vertex set T where two terms are neighbors if they share at least one variable. For a term t, let $N(t)$ denote the neighborhood of t in T. Consider the probability space in which we select an assignment of the variables at random. For $S \subseteq T$, let p_S denote the probability that at least one term in S is satisfied and let u_S denote the probability, that *exactly* one term in S is satisfied. Rudich's conjecture says that there is an $\epsilon > 0$ such that if $u_T > 1 - \epsilon$ then there must be a term such that $p_{N(t)} > \epsilon$. The consequence (which is due to Rudich), concerns limitations on the construction of cryptographic primitives. The theory of cryptography involves a number of "primitives" which are used as building blocks to construct protocols. These include: one-way functions, one-way permutations, trap-door functions, pseudo-random generators. Considerable research has been done to determine which primitives (or sets of primitives) suffice to construct all of the others. Of particular interest are "black box" constructions of one primitive from another, which, roughly speaking, are constructions which build the target primitive from a given primitive making use only of the properties that define the given primitive, and not the actual realization of the given primitive. Rudich showed that his conjecture implies that if there exists a black box construction of one-way permutation from one-way function then $P \neq NP$. This means that finding such a construction is as hard as separating P and NP, and thus it is unlikely that such a construction is possible by any current techniques.