DIMACS Center
Rutgers University

Discrete Random Systems

Annual Report

July 2007

Participants who spent 160 hours or more

Fred Roberts, Rutgers University, Principal Investigator
Dana Randall, Georgia Institute of Technology, Member of Organizing Committee
Gregory Sorkin, IBM Research, Member of Organizing Committee

Other Participants:

Christian Borgs, Microsoft Research, Member of Organizing Committee
Eric Vigoda, Georgia Institute of Technology, Member of Organizing Committee
Peter Winkler, Dartmouth, Member of Organizing Committee

Conference on Probabilistic Combinatorics & Algorithms: A Conference in Honor of Joel Spencer's 60th Birthday
Dates: April 24 - 25, 2006 (This was a "pre-special focus" kickoff event, funded separately.)

Organizers:
Noga Alon, Tel Aviv University and the Institute for Advanced Study
Janos Pach, NYU and Renyi Institute, Budapest, Hungary
Aravind Srinivasan, University of Maryland
Prasad Tetali, Georgia Institute of Technology

Workshop: Properties of Large Graphs: From Combinatorics to Statistical Physics and Back
Dates: October 16 - 20, 2006

Organizers:
László Lovász, Microsoft Research
Eötvös Loránd University
Benny Sudakov, Princeton University

Georgia Tech Kickoff for the Special Focus on Discrete Random Systems
Dates: November 20, 2006

Organizer:
Dana Randall, Georgia Institute of Technology

This was attended by members of the general theory and ACO (algorithms, combinatorics and optimization) communities at large in Atlanta, GA.

Workshop: Complex Networks and their Applications
Dates: January 22 - 24, 2007

Organizers:
Fan Chung Graham, UCSD
Ashish Goel, Stanford University
Milena Mihail, Georgia Institute of Technology
Chris Wiggins, Columbia University

Workshop: Phase Transitions in Random Structures and Algorithms
Dates: March 19 - 23, 2007

Organizers:
Gregory Sorkin, IBM Research
Eric Vigoda, Georgia Institute of Technology

Working Group: Current Topics in Markov Chains and Phase Transitions
Dates: March 26 - 30, 2007

Organizers:
Dana Randall, Georgia Institute of Technology
Eric Vigoda, Georgia Institute of Technology

Workshop: Markov Chain Monte Carlo: Synthesizing Theory and Practice
Dates: June 4 - 7, 2007

Organizers:
Jim Fill, Johns Hopkins University
Jim Hobert, University of Florida
Antonietta Mira, University of Insubria (Italy)
Luke Tierney, University of Iowa
Peter Winkler, Dartmouth

Workshop on Puzzling Mathematics and Mathematical Puzzles: a Gathering in Honor of Peter Winkler's 60th Birthday
Dates: June 8 - 9, 2007

Organizers:
Graham Brightwell, London School of Economics
Dana Randall and Tom Trotter, Georgia Institute of Technology

Partner Organizations

Telcordia Technologies: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

AT&T Labs - Research: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

NEC Laboratories America: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Alcatel – Lucent Bell Labs: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Princeton University: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Avaya Labs: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Georgia Institute of Technology: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning and all aspects of the project. Also financial support.

HP Labs: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

IBM Research: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Microsoft Research: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Rensselaer Polytechnic Institute: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Stevens Institute of Technology: Collaborative Research
Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Activities

During the past decade there has been tremendous interplay between discrete mathematics, theoretical computer science, and statistical physics. This project brings together researchers from these various fields to explore topics at their interface. The focus is on probabilistic algorithms and models that arise in the study of physical systems and combinatorial structures. Strong themes running through these interactions include: phase transitions; probabilistic combinatorics; Markov Chain Monte Carlo and other random walks; and random structures and randomized algorithms.

The active participation of scientists across disciplines in this area of Discrete Random Systems is indicative of the rich potential for further groundbreaking cross-fertilization. There has already been significant impact from these recent interactions. Physicists are gaining insights into physical systems, and in particular the simulations of these systems, through complexity theory and new algorithmic paradigms developed by computer scientists and discrete mathematicians. Computer scientists are gaining a much richer understanding of computational possibilities, and computational limitations, by considering the physical implications of their inquiries. Discrete mathematicians are developing new tools for studying large combinatorial systems by looking at infinite analogues studied in physics and applying randomized computational methods.

This area is of increasing commercial, economic, and societal relevance as large-scale real-world systems are successfully modeled stochastically. Examples of these diverse applications include discrete models of biological epidemics, structural analysis of the worldwide web, and stochastic optimization methods for operations research problems in transportation and scheduling.

The DIMACS special focus on Discrete Random Systems brings together world-class researchers working at the interface between discrete probability, statistical physics, and computer science, graduate students in these different disciplines, and practitioners working in various application domains. The program is exploring a broad spectrum of topics in the emerging field of Discrete Random Systems through a series of focused workshops, research working groups, and a visitor program, all intended to facilitate new interactions. More specifically, throughout the program we are addressing open problems and applications for each of the following:

- Randomized algorithms for combinatorial, decision, optimization, and enumeration problems.
- Random structure such as random graphs and random CNF Formulas, including questions of existence, optimization, and connections between phase transitions and computational intractability.
- Probabilistic combinatorics including extremal combinatorics and the new field of quasi-random graphs.
- Phase transitions in discrete physical systems and percolation models from statistical physics.
- Rapid and torpid mixing of Markov Chain Monte Carlo (MCMC) processes in finite-sized physical systems and combinatorial models.
- Technical tools used in the above, including mathematical methods (e.g., martingales, large-deviation inequalities, conductance and related measures) and physics tools (e.g., the replica and cavity methods).
- Optimization and analysis for real-world problems that may be modeled by random structures, including the Internet and biological structures.

This special focus is a natural follow-up to a program at the MSRI on Probability, Algorithms and Statistical Physics, January 3, 2005 to May 13, 2005, with organizers David Aldous, Claire Kenyon, Harry Kesten, Jon Kleinberg, Fabio Martinelli, Yuval Peres (co-chair), Alistair Sinclair (co-chair), Alan Sokal, Peter Winkler, and Uri Zwick. The timing of this special focus was chosen so as to provide a breathing space between the activities at the MSRI and DIMACS while maintaining momentum and enabling further collaborations in this growing research area, and to connect to activities at DIMACS in Biology, Epidemiology, the Social Sciences, and homeland security that will drive the field in new directions.

This special focus differs from the MSRI program by emphasizing small working groups, formal and informal, that delve in detail into some of the topics that came up during the MSRI semester. Workshops lay the groundwork for these intense, collaborative working groups. The DIMACS program has a broader computer science focus than the MSRI program. During the semester at MSRI, the general theme was connections between theoretical compute science and mathematical physics, including phase transitions in the web graph. At DIMACS we are exploring other topics in computer science where discrete random systems are currently playing a central role, including computational applications in the biological sciences. One of this year’s workshops fostered collaborations between theoreticians and practitioners using Markov chain sampling methods to study large data sets. Interactions with other DIMACS activities such as the multi-year special foci on Computational and Mathematical Epidemiology, Information Processing in Biology, and Computation and the Socio-economic Sciences are providing additional directions for collaborations involving connections between discrete random systems and many applications.

The tutorials, workshops, and working group meetings that were held during the first year of this project are as follows:

Conference on Probabilistic Combinatorics & Algorithms: a Conference in Honor of Joel Spencer's 60th Birthday

Dates: April 24 - 25, 2006

Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Noga Alon, Tel Aviv University and the Institute for Advanced Study; Janos Pach, NYU and Renyi Institute, Budapest, Hungary; Aravind Srinivasan, University of Maryland; and Prasad Tetali, Georgia Institute of Technology
Attendance: 113

This was a “pre-special focus” kickoff event funded separately. However, we mention it here because it laid the groundwork for the special focus activities and achievements reported on here. It is clear from comments cited later in the report that it has had an important influence on the project. Professor Joel Spencer has been a leading figure in the research and exposition of probabilistic combinatorics. His research in various areas including Ramsey theory, discrepancy theory, random graphs (and connections to logic), randomized greedy methods, etc. have set the tone for the field for more than three decades. Professor Spencer has also influenced a whole generation of researchers in discrete math and computer science through his very well known books, his lectures, and inspiring personality. This conference celebrated Professor Spencer's 60th birthday, through a number of talks that covered several of Professor Spencer's research areas. There was ample time for discussions and opportunities for the junior researchers and students who attended to meet senior researchers.

Discrete Random Systems Special Focus Kickoff Lecture
Date: October 5, 2006
Time: 11:25 am - 12:05 pm
Location: DIMACS Center, CoRE Building, Rutgers University
Speaker: Gregory Sorkin, IBM Research
Title: Phase Transitions in Optimization Problems, and Algorithmic Implications
Attendance: This was attended by members of the mathematics and statistical physics community at Rutgers and Princeton.

Phase transitions in random structures are an essential feature of statistical physics, play an important role in discrete mathematics, and have algorithmic implications. Some of the most interesting questions center around random 3-Sat (the satisfiability of randomly generated Conjunctive Normal Form formulas with 3 literals per clause), but because this is so hard to analyze, we often instead prove results for constraint satisfaction problems with clauses of length 2. For example, it is known that random 3-Sat undergoes a phase transition from almost-sure satisfiability to almost-sure unsatisfiability as the “clause density” increases, but it is not known exactly where this transition occurs, or even that the transition point tends to a constant. By contrast, the transition of random 2-Sat at density 1 is already a classical result, and more recently the transition's “scaling window” has been comprehensively analyzed. While 2-Sat is an easy decision problem, its optimization equivalent, Max 2-Sat, is NP-hard, and undergoes a similar phase transition, with the same scaling window. On the algorithmic side, it is known that sparse instances of random 3-Sat can, with high probability, be satisfied by efficient algorithms, but it is not known if this persists right up to the phase transition point. For “semi-random” 2-variable constraint satisfaction problems (a class encompassing random Max 2-Sat and random Max Cut), an algorithm running in expected linear time optimally solves instances of density up to the giant-component threshold, and indeed through the scaling window.

Workshop: Properties of Large Graphs: From Combinatorics to Statistical Physics and Back
Dates: October 16 - 20, 2006
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: László Lovász, Microsoft Research and Eötvös Loránd University; and Benny Sudakov, Princeton University
Attendance: 80

Many topics of research in mathematics, computer science and physics involve properties of very large graphs, or of graph sequences that grow asymptotically. For example, such topics arise in the study of Internet models, quasi-random graphs and other questions in graph theory, in property testing of large graphs, and in statistical mechanics. One of the areas of mathematics where these questions have been studied extensively is extremal graph theory. Given the central question in extremal graph theory concerning the existence of various small subgraphs, it might at first seem that this area has little to do with the other themes of this workshop, which are all probabilistic in nature. It turns out, however, that this is not the case. Indeed, one of the central results in this area is Szemerédi's Regularity Lemma, which has connections to, e.g., property testing and so-called quasi-random graphs. Quasi-random (also called pseudo-random) graphs were introduced by Thomason and Chung, Graham and Wilson. These graphs are deterministic, but are similar in many ways to random graphs: they have the same number of small subgraphs as a truly random graph, a partition into large subgraphs has similar cuts (number of edges between the different parts of the partitions), etc. In fact, these different characterizations have been proven to be equivalent.

Recently, a program has begun to generalize the results from quasi-random graphs to a large class of graph sequences . The work was originally motivated by an attempt to understand in which sense scale-free graphs like the Barabasi-Albert graph tend to a limit as their size goes to infinity, but turns out to have much wider scope. People working on this new program include researchers coming from many areas, including topology (Michael Freedman), group theory (Balazs Szegedy), combinatorics and computer science (Jeff Kahn, Laci Lovász, Lex Schrijver, Vera Sos, Katalin Vesztergombi) and probability theory and mathematical statistical physics (Christian Borgs, Jennifer Chayes).

One of the central notions in this program is the notion of a convergent graph sequence G_n. For dense graphs, one of the many equivalent definitions requires that for any finite graph F on k vertices, the probability that an induced subgraph on k random vertices in G_n is isomorphic to F tends to a limit as n tends to infinity.

Other equivalent notions involve uniform Szemerédi partitions, the number of graph homomorphisms from G_n into weighted graphs H, properties of small random subgraphs, convergence of the adjacency matrix to a measurable function on $[0,1] \times [0,1]$, and so on. Each of these definitions carries connections to other fields: graph homomorphisms are related to partition functions in statistical physics, small random subgraphs are related to property testing, and other concepts in this new approach lead to a better understanding of some of the results in extremal graph theory, to mention just a few.

This workshop brought together researchers from the fields of graph homomorphisms, extremal graph theory, quasi-random graphs, property testing, Internet modeling, probability theory and statistical physics (as practiced by mathematical physicists as well as theoretical physicists).

Georgia Tech Kickoff for the Special Focus on Discrete Random Systems

Dates: November 20, 2006
Location: Georgia Institute of Technology
Organizer: Dana Randall, Georgia Institute of Technology
Attendance: This was attended by members of the general theory and ACO (algorithms, combinatorics and optimization) communities at large in Atlanta, GA.

This half-day kickoff event identified some of the research directions and open problems to be addressed during the special focus. The talks touched on several major themes of the special focus: properties of large graphs, complex networks and their applications, phase transitions, and Markov chain sampling.

Workshop: Complex Networks and their Applications

Dates: January 22 - 24, 2007
Location: Georgia Institute of Technology
Organizers: Fan Chung Graham, UCSD, Ashish Goel, Stanford University, Milena Mihail, Georgia Institute of Technology, and Chris Wiggins, Columbia University
Attendance: 34

Complex networks and more generally complex systems are pervasive in today's science and technology. They include examples like the Internet, the WWW, peer-to-peer, sensor and ad-hoc networks, as well as biological networks such as gene and protein interactions and others. The study of such networks spans the fields of mathematics, computer science, engineering, biology and the social sciences.

This workshop brought together researchers from mathematics, theoretical computer science, networking, systems and biology. With the common theme of complex networks and systems, the participants shared the state of the art in these fields, the work done so far, the vision and major challenges of the future.

A special issue of the Journal on Internet Mathematics, whose chief editor is Fan Chung Graham, will be devoted to this workshop.

Workshop: Phase Transitions in Random Structures and Algorithms

Dates: March 19 - 23, 2007
Location: Georgia Institute of Technology
Organizers: Gregory Sorkin, IBM Research and Eric Vigoda, Georgia Institute of Technology
Attendance: 60

Many fundamental problems from mathematics and theoretical computer science exhibit “phase transitions” familiar from physics. Phase transitions are observed in physical systems including the Ising, Potts, and lattice-gas models, and in mathematical structures including random graphs, reconstruction problems, formula satisfiability, integer partitions, graph colorings, domino tilings, and geometric graphs. In the last 10 years, huge strides have been made in studying these phenomena --- in particular in uniting tools drawn from physics, theoretical computer science, and discrete mathematics and probability theory --- but the field remains nascent.

The Phase Transitions workshop brought together researchers from theoretical computer science, probabilistic combinatorics and discrete mathematics, and statistical physics. Topics included applications in areas such as those mentioned above; closely related research areas such as Markov Chain Monte Carlo and concentration inequalities; and new techniques, including the porting of established techniques from one field to another. The workshop included seminars on recent research results, but there was also a substantial amount of open time to foster collaborations.

Working Group: Current Topics in Markov Chains and Phase Transitions

Dates: March 26 - 30, 2007
Location: Georgia Institute of Technology
Organizers: Dana Randall and Eric Vigoda, Georgia Institute of Technology
Attendance: 18

Many local Markov chains based on local dynamics are known to undergo a phase transition as a parameter λ of the system is varied. For example, in the context of independent sets, the Gibbs distribution assigns each independent set I a weight $\lambda|I|$, and a natural Markov chain adds or deletes a single vertex at each step. It is believed that there is a critical point λ_c such that for $\lambda < \lambda_c$, local dynamics converge in polynomial time while for $\lambda > \lambda_c$ they require exponential time. A parallel phenomenon arises in statistical physics in the context of determining whether there is a unique limiting distribution (e.g., for independent sets) on the infinite lattice, known as a Gibbs state. While there has been progress in making such observations rigorous, many recent results have been the direct or indirect outcome of interdisciplinary collaboration, building on insights from both disciplines.

This working group promoted collaboration by bringing together researchers from computer science, combinatorics, probability, and mathematical physics to study topics central to the interplay between work on Markov chains and on phase transitions in random structures. Topics emphasized include

- Connections between equilibrium properties of Gibbs distributions in statistical physics models and the efficiency of Markov chain algorithms in computer science.
- Relationships between phylogeny algorithms in biology and properties of the random cluster model from statistical physics.
- Applications of advanced coupling techniques.
- Cutting-edge geometric techniques for bounding mixing rates, such as evolving sets.

Workshop: Markov Chain Monte Carlo: Synthesizing Theory and Practice

Dates: June 4 - 7, 2007
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Jim Fill, Johns Hopkins University, Jim Hobert, University of Florida, Antonietta Mira, University of Insubria (Italy), Luke Tierney, University of Iowa, Peter Winkler, Dartmouth
Attendance: 71

In the past two decades researchers in discrete mathematics and the theory of computing have made enormous theoretical strides in sampling via Markov chains, sometimes known as MCMC (Markov Chain Monte Carlo). A huge variety of computations is now proved to be doable in randomized polynomial time to arbitrary accuracy; included are the volumes of convex bodies in high dimension, the permanent of a matrix, the partition function for some models in statistical physics, and many more. At the same time, statisticians have made huge strides in implementing MCMC in complex applications; the applications of MCMC to Bayesian statistical inference probably outnumber all the computer science and physics applications combined.

Yet there has been little interaction between the theoretical mathematics community and the practical statistics one. In many cases, the theoreticians have gone off in directions of little interest in practical statistics; in others, they may have overlooked directions that could have changed the way things are done. On the other side, statisticians may not be aware of advances that really could impact their methods. Communication between theoreticians and researchers who actually handle data is extremely beneficial.

Here are some issues dividing theory and practice, where intriguing questions remain for both sides:

1. Distance to stationarity: In MCMC a Markov chain is run until its state distribution is as close as needed to stationarity. Theoreticians typically require closeness in total variation, meaning that *any* event has nearly the same probability as it would in stationarity. But, in practice (and even in theory --- for example in computing volumes) only a very limited class of events is of interest. Bayer and Diaconis recognized this fact in recommending seven riffle shuffles of a deck of cards; this still leaves a huge total variation of~30 but in practice is good enough.

2. Diagnosis of mixing: Closely related is the issue of deciding when, in running an actual Markov chain, to stop and sample. If coupling is easy to diagnose, as in a monotone chain, there may be a convenient criterion. But in many cases, there is no good bound on the mixing time and we don't really know when to stop. For example, László Lovász has asked the following simple question: Suppose that proper colorings of a graph are being sampled by means of local recoloring (heat bath), and suppose we have reached the point where the color of a particular vertex is close to uniformly random. Does this mean the coloring of the whole graph is now uniformly random? What criteria are used in practice to diagnose mixing?

3. Exact sampling: Propp, Wilson, Fill and others have developed some startling methods of sampling from the stationary distribution of a Markov chain. Are these methods useful in practice? If so, are they being used?

4. Random updates vs. sweeps: In many cases, Markov chains proceeding by local dynamics, with sites selected uniformly at random, are known to be rapidly mixing. However, in practice, it is easier to do updates by sweeping through the sites repeatedly in some fixed order. It appears empirically (and intuitively) that sweeping is just as good as random updates, but proof is lacking except in special cases.

5. Heuristically fast chains: Similarly, theoreticians have shown that certain Markov chains (like the “ball walk” inside a convex body in euclidean space) mix rapidly. But other chains --- e.g., the “hit-and-run” chain where a random line through the current point is chosen, then a point taken uniformly from the intersection of the line with the body --- may well be faster, yet harder to analyze. More generally, there is often some flexibility in designing a Markov chain to achieve a given stationary distribution. For analysis purposes, it is desirable to make the chain simple; for practical purposes, one wishes to design the chain to admit easy updates and move quickly around the state space. Can these be reconciled?

6. Competing methods: MCMC is popular in the “theory community” of computer science, but in the statistical world it is only one of many methods for random sampling. Even when a Markov chain is already in the picture, heuristic approaches like sequential importance sampling and variations of simulated annealing or simulated tempering may be superior in practice to straight, analyzable running of the chain. Is there any way to know when this will be the case?

In the workshop, we addressed these issues both generally and in special cases; for example, in the generation of random contingency tables (matrices with given row and column sums); in genomics; and in graphical models. Our invitees included many of the major contributors to MCMC, but leaning toward those who have demonstrated interest in practical considerations; and a sampling, hopefully somewhat representative, of potential consumers of MCMC theory from the statistics community and elsewhere.

We began with a one-day tutorial on MCMC and on sampling issues in general, understandable to graduate students and a general math/statistics/computer science audience. The tutorial was about half devoted to mathematical results on MCMC and half to MCMC in statistics.

Workshop on Puzzling Mathematics and Mathematical Puzzles: a Gathering in Honor of Peter Winkler's 60th Birthday

Dates: June 8 - 9, 2007

Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Graham Brightwell, London School of Economics; Dana Randall and Tom Trotter, Georgia Institute of Technology
Attendance: 78

Examples abound of deep mathematical results whose study was originally inspired by simple and appealing puzzles: questions that can be understood and appreciated by the layman, but whose solution proves challenging even to experts. For instance: problems of counting seating arrangements motivate problems in combinatorics, the bus waiting-time paradox is a cautionary example and a compelling introduction to size-biased sampling, while a problem as easy to explain as Fermat's Last Theorem inspires some of the deepest mathematics.

Puzzles can be engaging to professional mathematicians as well as amateur enthusiasts. They can be conduits to deep results or, on occasion, extremely challenging and time-consuming mathematical dead ends. Likewise, some of the deepest and most useful results in mathematics and computer science contain essential components that have the elegance and appeal of simple brain teasers. Through these parallels, the interplay of mathematics and recreational puzzles is often a central component of research, especially for discrete mathematicians and computer scientists.

The subject of discrete random systems is a particularly rich source of important mathematical questions that can be presented as intriguing puzzles. One of many examples is Peter Winkler's "Clairvoyant Demon" problem: Can two random walks on the same graph be scheduled so that they never collide? Problems arising from card shuffling, from the study of epidemics, or from the behavior of traffic in networks, also often prove easy to describe, yet hard to analyze.

This workshop was timed to celebrate Peter Winkler's 60th birthday, with emphasis on puzzles, and puzzling phenomena, especially as related to the special focus themes. Peter has been a pioneer in many areas of computer science and combinatorics. His work ranges over pure mathematics, combinatorial curiosities, applications of discrete mathematics, connections between discrete mathematics and computer science, and---the aspect singled out for this workshop---games and puzzles.

Findings

Anomalous heat-kernel decay for random walk among bounded random conductances

Gady Kozma (Weizmann Institute of Science) completed work with Marek Biskup (UCLA) on a problem involving random walks as a result of their participation in the workshop on properties of large graphs. Kozma and Biskup, along with co-authors Noam Berger and Christopher E. Hoffman, considered the nearest-neighbor simple random walk on Z^d, d greater than or equal to 2, driven by a field of bounded random conductances between 0 and 1. The conductance law is i.i.d., subject to the condition that the probability that the random conductance is greater than 0 exceeds the threshold for bond percolation on Z^d. For environments in which the origin is connected to infinity by bonds with positive conductances, Kozma, Biskup, Berger, and Hoffman studied the decay of the 2n-step return probability. They proved that the 2n-step return probability is bounded by a random constant times $n^{-d/2}$ in d=2,3, while it is $o(n^{-2})$ in d greater than or equal to 5 and $O(n^{-2})$ \log n) in d = 4. By producing examples with anomalous heat-kernel decay approaching $1/ n^2$ they proved that the $o(n^{-2})$ bound in d greater than or equal to 5 is the best possible. They also constructed natural n-dependent environments that exhibit the extra log n factor in d=4.

NP for combinatorialists

Gabor Kun (Eötvös University, Budapest) and Jaroslav Nesetril (Charles University, Prague) have shown that to every problem in NP there is a computationally equivalent language of digraphs determined by forbidden injective or full homomorphic images, respectively. In the case of usual homomorphism one gets the class MMSNP (Monotone Monadic SNP) introduced by Feder and Vardi. A typical language in this class is the language of undirected graphs that are two colorable without a monochromatic triangle. Kun and Nesetril give an easy characterization of the MMSNP languages, which are actually CSP (Constraint Satisfaction Problem) languages. They show that every MMSNP language restricted to a bounded expansion class (generalization of minor-closed and bounded degree classes) of digraphs is actually equal to the restriction of a CSP language to the same class.

The Elusive Behavior of Lattice Gases

Dana Randall (Georgia Institute of Technology), Sam Greenberg (Georgia Institute of Technology), and Peter Winkler (Dartmouth) collaborated on a discrete Markov chain problem that seems to capture the elusive behavior of lattice gases on the reals. Over the reals a lattice gas is modeled by a set of particles that occupy space, and particles are required to be nonoverlapping. The discrete version is typically modeled as an independent set on a lattice, where the lattice edges enforce the distance constraint. Local Markov chains on each of these models seem to undergo a phase transition whereby their convergence is fast at high temperature (the sparse case) and slow at low temperature (the dense cases), but this is only understood for the discrete model. In the discrete case the slow mixing comes from the parity of the lattice that forces very dense independent sets to lie almost entirely on one of the two sublattices. In the real case the slow mixing seems to come from dense packings choosing a preferred angle of alignment. Randall, Greenberg, and Winkler are trying to capture this behavior on a new discrete model that packs 2 by 2 squares on lattice points so that the squares do not overlap. It seems that the squares will predominantly align horizontally or vertically for reasons similar to the expected behavior on the reals, and with a discrete model there is more hope of rigorously analyzing the Markov chain. This is an ongoing collaboration on which Randall, Greenberg, and Winkler focused during the working group Current Topics in Markov Chains and Phase Transitions in March 2007.

Markov random fields (MRF) models for chromosomal breakpoint prediction

George Popescu (Yale) has developed a new statistical algorithm for estimation of chromosomal breakpoint's genomic coordinates and for detection of copy number variation from comparative genomic hybridization (CGH) data. The algorithm is robust to several sources of noise typically found in DNA micro-arrays. The breakpoint detection algorithm uses a Markov Random Field that models spatial coherence of neighbor oligo-nucleotide probes and contains a regularization term derived from structural coherence of genomic segments with higher degree of similarity (low copy repeats - LCR regions). Popescu showed that the algorithm can localize chromosomal breakpoints with high accuracy for several types of chromosomal aberrations. Popescu demonstrated a reduction in breakpoint estimator variance when using optimized isoTM arrays. He analyzed the variation of statistical properties of breakpoint estimator with micro-array tilling density. Finally, he optimized the detection of chromosomal copy number changes with CGH MRF by computing an efficient sampling of target genomic regions with structural and tilling density constraints.

Training and Development

The workshops in this program offered many students the opportunity to meet senior researchers and develop new collaborations. Here are several examples.

"My graduate student Yun Long participated in two meetings in the special focus and has benefited greatly. A student from MIT that I am mentoring (K. Meszaros) benefited greatly from a talk by J. Propp in the meeting in honor of P. Winkler which affected her choice of thesis topic." Yuval Peres, Berkeley

"I'm Yun Long, a graduate student from U.C.Berkeley, that have attended the DIMACS workshops on 'Phase Transitions in Random Structures and Algorithms' as well as the 'Markov Chain Monte Carlo: Synthesizing Theory and Practice' one. In these workshops, I like the tutorial lectures most which provided me with leading progress in the areas. Moreover, I discussed with professor A. Frieze, from whose paper my research topic on the mixing time of the Swendsen-Wang dynamics was originated. I also got some ideas of possible further improvements from him. From the talk of D. Randall, I learned more about sampling from the Ising model, and noticed that studying the complexity of various sampling algorithms is of great importance in practice, from which I was motivated a lot. For me, these are really great workshops." Yun Long, graduate student, U.C.Berkeley

Outreach Activities

This project, which has a strong biomathematics component, is a relevant source of information for DIMACS efforts to link mathematics and computer science with biology in the high schools. Fred Roberts, the PI for this project, is also PI for the DIMACS Bio-Math Connection (BMC) project, which is aimed at introducing high school math/CS and Bio teachers to topics at the interface. Indeed, bioinformatics has been a key area for educational module development at DIMACS during the past year.

Papers

D. Achlioptas F. Ricci-Tersenghi, "Constraint Satisfaction Problems: From Physics to Algorithms," *Abstract Book of the XXIII IUPAP International Conference on Statistical Physics* (2007)

N. Berger, M. Biskup, C. E. Hoffman, and G. Kozma, "Anomalous heat-kernel decay for random walk among bounded random conductances," *Annales De L'institut Henri Poincaré (B)Probabilitése et Statistiques*, to appear.

D. Bokal, E. Czabarka, L. A. Szekely, and I. Vrvto, "Graph minors and the crossing number of graphs," *Electronic Notes on Discrete Math*, **28** (2007) 169-175.

G. Kun and J. Nesetril, "Lifts and shadows," *European Journal of Combinatorics*, to appear.

G. Kun and J. Nesetril, "NP for combinatorialists," *Proceedings of EUROCOMB07*, to appear.

G. Kun and J. Nesetril, "Forbidden lifts (NP and CSP for combinatorialists)," *KAM-DIMATIA Series* 2006-775, (2006).

G. Kun and J. Nesetril, "NP and CSP by means of lifts and shadows," *Proceedings of the Mathematical Foundations of Computer Science 2007*, accepted.

G. Popescu, "Markov random fields (MRF) models for chromosomal breakpoint prediction," in preparation.

Talks

Dimitris Achlioptas, "Constraint Satisfaction Problems: From Physics to Algorithms," Statphys23, Genova, Italy, July 9-13, 2007.

Gabor Kun and Jaroslav Nesetril, "NP for combinatorialists," European Conference on Combinatorics, Graph Theory and Applications, Seville, September 11 - 15, 2007.

Gabor Kun and Jaroslav Nesetril, "NP and CSP by means of lifts and shadows," Mathematical Foundations of Computer Science, Český Krumlov Czech Republic, August 26 – August 31, 2007.

Dana Randall, "Markov chain convergence and the efficiency of some self-assembly models," keynote address, Foundations of Nanotechnology: Self-Assembled Architectures and Devices, Snowbird, Utah, April 2007.

Other Specific Products

A special issue of the *Journal on Internet Mathematics*, whose chief editor is Fan Chung Graham, will be devoted to the workshop "Complex Networks and their Applications" that was held in January, 2007 at Georgia Institute of Technology.

Web pages

http://dimacs.rutgers.edu/SpecialYears/2005_DRS/
Main web page for the program

http://dimacs.rutgers.edu/Workshops/Spencer/
Web page for Conference on Probabilistic Combinatorics & Algorithms: a Conference in Honor of Joel Spencer's 60th Birthday

http://dimacs.rutgers.edu/Events/2006/abstracts/sorkin.html
Web page for Discrete Random Systems Special Focus Kickoff Lecture

http://dimacs.rutgers.edu/Workshops/LargeGraphs/
Web page for Workshop: Properties of Large Graphs: From Combinatorics to Statistical Physics and Back

http://dimacs.rutgers.edu/Workshops/Art
Web page for Georgia Tech Kickoff for the Special Focus on Discrete Random Systems

http://dimacs.rutgers.edu/Workshops/ComplexNetworks/
Web page for Workshop: Complex Networks and their Applications

http://dimacs.rutgers.edu/Workshops/RandomStructures/
Web page for Workshop: Phase Transitions in Random Structures and Algorithms

http://dimacs.rutgers.edu/Workshops/MarkovChains/
Web page for Working Group: Current Topics in Markov Chains and Phase Transitions

http://dimacs.rutgers.edu/Workshops/MonteCarlo/
Web page for Workshop: Markov Chain Monte Carlo: Synthesizing Theory and Practice

http://dimacs.rutgers.edu/Workshops/Puzzles/
Web page for Workshop on Puzzling Mathematics and Mathematical Puzzles: a Gathering in Honor of Peter Winkler's 60th Birthday

Contributions

Contributions within Discipline

The “discipline” is by definition a combination of disciplines. Many of the results described in the Activities and Findings section of this report illustrate this combination of disciplines, in particular the application of methods of discrete mathematics, theoretical computer science, and statistical physics to problems of biomathematics. An example is George Popescu’s work on Markov random fields models for chromosomal breakpoint prediction. This is described in more detail earlier in this report.

Introducing people to this combination of disciplines has been a key goal and a key accomplishment of this project. Below is a selection of comments from project participants indicating their assessment of the impact of the project:

Greg Sorkin, IBM Research, reports that at the kickoff meeting he began collaboration with Eduardo Sontag, Rutgers University. The subject arises from a gene-network question, but can be translated into a question about random-interaction Ising models on sparse random graphs. The hypothesis to be confirmed is that real gene networks are much closer to being monotone (in the sense of dynamical systems) than almost any random graph would be. To verify this hypothesis, one needs to estimate the “distance to monotone” for random graphs.

László A. Székely, University of South Carolina, reports that he and Drago Bokal, Simon Fraser University, began a collaboration at the Spencer workshop that they continued with Éva Czabarka and Imrich Vrťo on graph minors and the crossing number of graphs. There are three general lower bound techniques for the crossing numbers of graphs, all of which can be traced back to Leighton's work on applications of crossing number in VLSI: the Crossing Lemma, the Bisection Method, and the Embedding Method. Bokal, Székely, Czabarka, and Vrťo developed an adaptation to the minor crossing number.

Prior to attending a program in the Special Focus, Silvio Franz, Abdus Salam International Center for Theoretical Physics, worked with Michele Leone, INFM and SISSA, to generalize to the case of diluted spin models and random combinatorial optimization problems a technique recently introduced by Guerra to prove that the replica method generates variational bounds for disordered systems. They analyzed a family of models that includes the Viana-Bray model, the diluted p-spin model or random XOR-SAT problem, and the random K-SAT problem, showing that the replica method provides an improvable scheme to obtain lower bounds of the free-energy at all temperatures and of the ground state energy. One very important consequence is that in the case of K-SAT the replica method gives upper bounds for the satisfiability threshold. Franz's participation in a Special Focus program led to a communication via another participant that Franz and Leone's proofs for even values of p and K appear to apply to odd values as well. Franz is working to check this seemingly simple but very significant observation, one consequence of which would be an upper bound for the 3-SAT threshold.

George V. Popescu, Yale University, reports that he developed collaborations with Joel Lebowitz, Rutgers University, James Propp, University of Massachusetts Lowell, and Gregory Sorkin, IBM Research, in which new research ideas were developed. One area of future research is to explore MCMC statistical algorithms for cellular network inference centered on two ideas:

- A hierachical Bayes model for protein interactions in cellular networks.
- An MCMC algorithm to converge to the most probable state of the network given the evidence and prior beliefs.

Another research direction Popescu plans to pursue is the use of graph spectrum for MCMC convergence analysis.

Popescu was also involved in other discussions on computing bounds on MCMC performance, statistical tests from Poisson distributions, and using expansions on the basis of orthogonal polynomials for image reconstruction. He described his experience overall as follows:

"The tutorials were useful to set a common ground for discussions. The presentations were diverse, ranging from very applied work to recent advances in the theory of discrete random systems. From a personal viewpoint, the impact was significant. I was able to get an up to date view of research directions and start planning my potential contributions. I had made several contacts and had exchanged some research ideas which may result in future collaborations."
George V. Popescu, Yale University

Other participants reported:

"During the program I had the opportunity to initiate a collaboration and work extensively with Amin Coja Oghlan. This collaboration has led to significant advances in rigorizing a number of hypotheses put forward by physicists with respect to the solution-space geometry of random Constraint Satisfaction Problems. We expect at least two or three articles to result from this work. Also, during July 2007, I had the opportunity to announce some of these results at Statphys 23, the premier statistical physics conference." Dimitris Achlioptas, UC Santa Cruz

"The program led to a new collaboration with P. Tetali, J.H. Kim and R. Montenegro on the Pollard random walk used to solve the discrete log problem. It also led to substantial progress in an ongoing project involving Glauber dynamics for the Ising model with M. Luczak and D. Levin." Yuval Peres, Berkeley

"A collaboration started in April 2006 between Drago Bokal and Laszlo Szekely at the Spencer birthday conference and conference that immediately followed. The work is about finding analogues of the classical lower bound techniques for the crossing number in the case of the even more elusive minor crossing number. One paper is already published and an extended journal version is under preparation."
Laszlo Szekely, University of South Carolina

Contributions to Other Disciplines

This "discipline" is inherently multidisciplinary. Examples of the type of synergy across disciplines that has been created include some of the presentations in the DIMACS - Georgia Tech Workshop on Complex Networks and their Applications. Joel Bader, Johns Hopkins, discussed his group's work on understanding how genes and proteins assemble in a network. They have developed data-mining algorithms for inferring gene modules from high-throughput genetic screens and a computational scheme to compute gene regulatory network wiring diagrams directly from DNA and protein sequence using all-atom simulations. Badar described an extension of capture-recapture statistics that he developed to estimate false-positive and false-negative rates for network edges (biological interactions) revealed by noisy high-throughput experiments. Meredith Betterton, University of Colorado, Boulder talked about how the balance between activating and inhibiting connections by, for example, deleting a protein from a signaling network, is important in determining whether a network reaches steady state or oscillates. She used a simplified mathematical model that captures the essential behavior of a network of interacting genes or proteins to study networks subjected to selection for specific properties and topologies of real biological networks. Guillermo Cecci, IBM, described how complex biological systems and man-made complex systems can be studied by abstracting them to directed networks and that the statistics of directional correlations (i.e. links coming in and out of each node) can be captured by a simple physical model, a Pott's spin system.

Interactions with other DIMACS activities such as the multi-year special foci on Computational and Mathematical Epidemiology, Information Processing in Biology, and Computation and the Socio-economic Sciences are providing additional directions for collaborations involving connections between discrete random systems and several other disciplines.

Contributions beyond Science and Engineering

The research in discrete random systems is of increasing commercial, economic, and societal relevance as large-scale real-world systems are successfully modeled stochastically. Examples of these diverse applications include discrete models of biological epidemics, structural analysis of the world-wide web, and stochastic optimization methods for operations research problems in transportation and scheduling. It is early in the program to have specific contributions beyond science and engineering but industry researchers have been active participants of this program, which leads us to expect that such contributions will develop in later years. For example, the workshop “Complex Networks and their Applications” in January, 2007 was attended by four members of Yahoo Research, three of whom gave presentations on applications of interest to Yahoo.

Contributions to Human Resources Development

Many graduate students, undergraduates, and several postdocs participated in the program. Local graduate students and many non-local students were also involved as workshop/working group attendees.