

# Computation, Combinatorics and Probability

29 July to 20 December 2002

Report from the Organisers: M Dyer (Leeds), M Jerrum (Edinburgh), P Winkler (Bell Labs)

[Scientific Background](#)

[Structure of the Programme](#)

[Workshops](#)

[Outcome and Achievements](#)

---

## Scientific Background

Over time, the relationships between computer science and mathematics have become more wide ranging and profound. The traffic in ideas has been in both directions. On the one hand, computer science has opened up new areas of study to mathematics; on the other, increasingly sophisticated mathematical methods have been brought to bear on problems in computer science. These developments have been particularly evident in quantitative theoretical computer science, especially in the areas known as computational complexity and the analysis of algorithms. Both of these areas are concerned with the resources (often time or space) required to achieve specified computational tasks; the rough distinction is that the former concerns itself with general phenomena surrounding resource-bounded computation, while the latter probes the inherent difficulty of specific computational tasks.

The aim of this programme was to explore two particularly fruitful interfaces between computer science and mathematics, namely those involving combinatorics and probability theory. Since computers perform manipulations of finite structures (bit patterns or data structures of varying complexity, depending on the level at which we view the computation) it is not at all surprising that combinatorics is pervasive in the design and analysis of algorithms. However, the importance of probability may come as a surprise, given that determinacy and predictability would seem to be desirable features of a real computer! Nevertheless the notion of randomised computation has assumed an increasingly important role, both in theory and practice.

Within the broad area delineated by the title, a number of themes were to receive special emphasis.

- Randomised algorithms. This theme addressed the design and analysis of algorithms that make random choices, and also of deterministic algorithms when run on random instances. The theoretical basis of this study includes, for example, the study of parameters connected with random walks on graphs - coupling time, expansion, multicommodity flow, mixing time, etc. - and the relationships between them.
- Phase transitions in statistical physics and computer science. This heading covers not only finite random systems, but also (infinite sequences of) finite systems that exhibit phenomena

akin to phase transitions. An example of the latter is the satisfiability threshold for random instances of CNF Boolean formulas with  $k$  literals per clause. Particular attention was paid to the apparent link between phase transitions and computational tractability.

- Random graphs and structures. We were concerned here with finite random structures, of which the Erdős-Rényi random graph model is the primal example. Algorithmic aspects received special attention.
- Probabilistic analysis of distributed systems. The central concern was to study the behaviour of distributed systems in the absence of global control. The prime example of such a system is the Internet, though one could also think of traffic on road networks, for example. This is a very new interdisciplinary area bringing together economics, game theory (Nash equilibria of agents in networks) and computer science.

---

## Structure of the Programme

The programme attracted a strong contingent of nearly fifty long-term visitors. It began and ended with very well attended periods of heightened activity built around the workshops. Between these rather hectic periods of concentrated interaction, there was a quieter interlude for collaborative research projects. Momentum was maintained through a programme of regular weekly seminars. We were fortunate in having J Steif (Chalmers) and L Lovász (Microsoft Research) on hand to depict some of our interests to the wider world in the Institute's Monday Seminar series.

---

## Workshops

### Randomised Algorithms

#### **Euroworkshop, 5-16 August 2002**

#### **Organisers: M Dyer, M Jerrum and P Winkler**

One of the appeals of randomised algorithms for the mathematician is that whereas the algorithms themselves are often quite simple, their analysis can be very challenging. Because of this, the study of randomised algorithms has prompted one of the most profound and fruitful interactions between computer science and mathematics. This workshop aimed to stimulate progress in this interdisciplinary area, by bringing together computer scientists with probability theorists, combinatorialists and others working in relevant areas of mathematics.

The plan was that the first of the two weeks would be concerned with mathematical foundations, for example, random walks on graphs, concentration of measure, etc. The second would deal with the design and analysis of algorithms which make random choices, including Monte Carlo and Markov Chain Monte Carlo (MCMC) algorithms and deterministic algorithms when run on random instances. This plan was roughly followed, though availability of speakers meant in practice that there was some blurring of the theory/application distinction. A typical day was built around four talks of one-hour duration, a timetable that left plenty of time for informal discussions. Amongst the speakers giving talks of a more expository nature were: C Greenhill (Melbourne) who gave a tutorial on the "Differential equations method", an analytical tool whose value was amply illustrated in a

number of other talks during the course of the workshop; M Jerrum (Edinburgh) who surveyed the varied methods, many of them quite recent, for the non-asymptotic analysis of rates of convergence to stationarity of Markov chains; A Sokal (New York) who demonstrated how methods from statistical physics can answer long-standing open questions from combinatorics (and how combinatorics can in the agenda of statistical physics); and P Winkler who untangled for us the web of mixing times (measures of convergence of Markov Chains). Many of the “leading lights” of the field were at the workshop: M Dyer (Leeds) introduced his surprising discovery that dynamic programming is useful for counting as well as optimisation; R Kannan (CMU) described “blocking conductance”, a new and more precise tool for bounding mixing times of Markov chains; and A Sinclair (UC, Berkeley) explained how Clifford algebras might one day provide a practical approach to estimating the permanent of a non-negative matrix.

## **Random Structures**

**Workshop, 26 August-6 September 2002**

**Organisers: M Dyer, M Jerrum and P Winkler**

This two-week workshop was structured into two complementary parts of equal duration: *Combinatorial and Computational Aspects of Statistical Physics*, and *Random Graphs and Structures*. The two parts were to be united by the notion of phase transition, broadly interpreted. Roughly speaking, Part 1 was to concentrate on phase transitions in infinite systems (e.g., the Ising model on  $\mathbf{Z}^2$ ), and Part 2 on “phase transitions” in finite structures (e.g., random graphs or random partial orders). The interdisciplinary aspect was evident, with computer scientists meeting with probabilists, mathematical physicists (particularly in Part 1) and combinatorialists (particularly in Part 2).

In the event there were about sixty officially registered participants, and a significant additional number (particularly from in or around Cambridge) attended on an informal basis. Any selection of talks from the workshop will necessarily be somewhat arbitrary, but here is a brief summary of a few notable contributions of more general interest. G Grimmett (Cambridge) surveyed some open problems in statistical mechanics, reported on recent progress on their solution and expressed optimism that they could be completely resolved in the next few years, in the light of recent work on “stochastic Löwner evolutions”. J Kahn (Rutgers) presented an important new result (with D Galvin) that the hard-core lattice gas model in  $d$  exhibits a phase transition at an activity that decays to 0 with  $\mathbf{Z}^d$ . (This has long been expected, but not proved.)

M Krivelevich (Tel Aviv) gave a personal account of the future of colouring problems on random graphs. S Janson (Uppsala) and P Tetali (Georgia Tech.) described progress on concentration inequalities for combinatorial applications; such inequalities are an essential tool for research in this area. Finally, A Scott (UCL) drew a surprising connection between two seemingly unrelated topics: a theorem of Dobrushin regarding uniqueness of Gibbs measures, and the Lovász Local Lemma. (This was joint work with A Sokal.) The meeting was supported by the MathFIT initiative of EPSRC and the London Mathematical Society.

## **Topics in Computer Communication and Networks**

**Workshop, 16-20 December**

**Organisers: M Dyer, L Goldberg, M Jerrum, F Kelly and P Winkler**

This one-week meeting, partly supported by Cisco Systems Inc., attracted a greater than expected number of participants, suggesting that the topic has captured the imagination of the

community. A number of themes could be identified. One of these was models for the web graph. The World Wide Web can be viewed as an ever-evolving graph, with web pages as nodes and links as directed edges. This graph, though random in some sense, has very different properties to the usual Erdős-Rényi random graph; for example, the vertex degrees obey a power law. A number of participants, among them J Chayes (Microsoft Research) and C Cooper (King's College, London), proposed and analysed graph models that could potentially explain these observed features.

A second theme was the “cost of dispensing with global control”. In other words, how far can a Nash equilibrium (obtained when agents optimise their own utility) be from a globally optimal solution imposed by centralised control? This question was taken up by T Roughgarden (Cornell) in the context of routing in networks.

The final theme in this non-exhaustive list was the complexity of computing Nash equilibria. Problems involving these equilibria are interesting and challenging from a computational viewpoint, because they do not seem to fall naturally into the usual complexity-theoretic classes such as “NP-complete”. V Vazirani (Georgia Tech.) presented one of the first moves in the direction of positive results.

---

## Outcome and Achievements

The programme was broad in scope, covering work in computer science, probability, statistical physics, graph theory and combinatorics. Consequently the research achievements of the programme were varied in topic. The work was highly collaborative, and often bridged two or more areas. We summarise the outputs below, alphabetically by first collaborator. Most have resulted in one or more publications, and several formed the topic of a seminar during the programme. The overall success of the research programme can be judged by the volume and quality of the output. Further details of many of the items mentioned below may be found in the growing list of Newton Institute preprints, and in due course in a special issue of *Random Structures and Algorithms*.

Bollobás collaborated with Brightwell on the paper *How Many Graphs are Unions of  $k$ -cliques?*, and with Simonovits on extending their theorem, with Balogh, concerning the number of graphs that exclude a certain set  $L$  of subgraphs. (When  $L$  is the singleton set containing the complete graph on  $k$  vertices, this is the Erdős-Kleitman-Rothschild theorem.) Brightwell worked with Winkler on problems at the interface between graph theory and statistical physics. Asymptotics were obtained for the following interesting problem: When does the unique simple invariant Gibbs measure for the hard-core gas model on the Bethe lattice (i.e., infinite regular tree) cease to be extremal?

Cryan worked with Dyer and Randall on the counting problem for a cell-bounded version of the contingency table problem, and with Dyer, Jerrum and Karpinski on polynomial time generation of independent sets in hypergraphs. Contingency tables are matrices of non-negative integers with given row- and column-sums, which arise in statistics; in determining whether an apparent correlation is statistically significant it is necessary to be able to count and/or sample such tables.

Dyer completed work on a new approach using dynamic programming to polynomial time approximate counting for the knapsack and related problems. This gives simpler, and more efficient, algorithms than were known previously. He worked with Frieze and Vigoda on

sampling from the Ising and hard-core models on “large girth” graphs. Related work, undertaken with Goldberg and Greenhill, attempts generalisations to (other)  $H$ -colouring (also known as graph homomorphism) problems. Also, with Goldberg and Jerrum, he investigated alternative dynamics for sampling particle systems on “simple” graphs, in particular  $H$ -colourings. With Sinclair and Vigoda, he completed some work on the relationship between spatial and temporal mixing in physical systems.

Fill continued collaboration with Janson on a series of papers giving refined analysis of the Quicksort algorithm. Quicksort is a classical randomised algorithm for sorting items using pairwise comparisons. Describing the distribution of the number of comparisons used in sorting  $n$  items is a challenging problem.

Goldberg, Martin and Paterson completed work on Markov chains on three-colourings of the rectangular grid in  $\mathbf{Z}^2$ , and showed rapid mixing of the Glauber dynamics. Somewhat surprisingly, the question of rapid mixing for four to six colours remains open, even though movements within the state space appear then to be less constrained.

Greenhill worked with Rucinski (and Wormald) on applying the “differential equation approach” to random hypergraph processes. The idea, made rigorous by Kurtz, and introduced to the algorithms community by Karp and Sipser, is to approximate the evolution of a Markov chain on  $\mathbf{Z}^d$  by the solution to a differential equation. As the size of the state space increases, the approximation becomes more exact.

Hubenko worked (with Gyárfás) on certain factors of the  $n$ -cube. The particular feature of these factors is that they arise as induced subgraphs of the  $n$ -cube. The prototype is the “Snake in a box” (largest induced cycle in the  $n$ -cube) problem that arises in coding theory. Janson collaborated with Rucinski (and Oleskiewicz) on large deviation inequalities for counts of (not necessarily induced) subgraphs in a random graph. He also worked with Chassaing on the Brownian snake. Both of these investigations have led to subsequent publications.

Járai worked with van den Berg on a “forest fire” model, and (with Athreya) on the abelian sandpile model in  $\mathbf{Z}^d$ ; both of these are models of self-organised criticality.

Jerrum (with Son) continued investigations of the spectral gap and log-Sobolev constants of Markov chains. One outcome, with Tetali and Vigoda, is a remarkable decomposition theorem for the log-Sobolev constant which closely mirrors that known for spectral gap. He worked with Sinclair and Vigoda on (non-trivially) extending their celebrated solution of the permanent problem to the case of non-bipartite matchings. This work remains in progress. He also completed a monograph, *Counting, Sampling and Integration*, which has since been published by Birkhäuser.

Kahn, during a short visit, worked with Randall and Sorkin on phase transitions for three-colourings of the lattice  $\mathbf{Z}^d$  in high dimension  $d$ . The expectation, which is consistent with what has been discovered so far, is that there should be six Gibbs states; a typical such state has the even sublattice coloured largely “red” and the odd sublattice largely “blue” and “green”. He also proved a conjecture of van den Berg.

Karonski worked with T Łuczak and Thomason on local irregularity strength of graphs, with Pittel on perfect matchings in a class of sparse random graphs, and with Krivelevich and Vazirani on information hiding in graphs and its potential applications to software watermarking.

Kendall studied problems arising in image analysis concerning phase transitions in the Ising model defined on “augmented quad trees”, and the form of the resulting interfaces.

Lovász wrote a paper (with Vempala) on the geometry of log-concave distributions: these occur widely in practice, the most familiar being the Gaussian distribution. A concrete outcome of their study is a particularly efficient algorithm for sampling a point from a log-concave distribution. Lovász also collaborated on other work in discrete probability and combinatorics.

M Luczak worked with T Luczak on phase transitions in the cluster-scaled model of a random graph, and with Winkler on showing that a random walk on the non-negative integers produces a uniform random subtree of the Cayley tree.

T Luczak and Simonovits collaborated on generalising results of Andrásfai-Erdős-Sós concerning the minimum degree of a graph of chromatic number  $k$  that contains no clique on  $k$  vertices. Pikhurko and Simonovits (with Füredi) investigated extremal hypergraph problems, and proved a conjecture of Mubayi and Rödl as to the maximum size of a 3-graph on  $n$  vertices that excludes a specific 3-graph on five vertices.

Randall continued her work on simulated tempering, a general method that has been proposed for speeding up simulations of “low temperature” systems. The idea is to introduce a dynamics that is able to move between low temperature states and high temperature states, where mixing occurs more freely. Preliminary results, both positive and negative, have been obtained. Scott worked with Wilmer on a wide range of problems (extremal, algorithmic and complexity-theoretic) concerning classes of permutations which occur in large-scale genome changes.

Stougie proved a polynomial upper bound on the edge-diameter of transportation polyhedra. His initial results were improved during the prog-ramme in collaboration with van den Heuvel, and subsequently further improved in collaboration with van den Heuvel and Brightwell. Van den Berg (with Brouwer) studied “self-destructive” percolation in which all sites in an infinite cluster become vacant through some catastrophe. With what probability do vacant sites need to be re-populated before an infinite cluster re-appears? They conclude the answer may be larger than one might suspect. Van den Heuvel extended some of his own results on cyclic and distant 2-colourings of planar graphs. Vigoda (with Hayes) developed a new tool for analysing non-Markovian couplings, giving an application to sampling graph colourings. The idea should have many other applications.

---