Randomised Algorithms (RAND2) Workshop
ICMS, Edinburgh, 27th–29th March 1998
List of Abstracts

Invited talks

Disagreement percolation and mixing properties of Markov fields

Jacob van den Berg (CWI, Amsterdam, The Netherlands)

ABSTRACT. Markov random fields (or finite-range Gibbs measures) have, since a long time, been studied in Statistical Physics. More recently, much interest has been shown from researchers in Image Analysis and Theoretical Computer Science. Typical examples are the Ising model, the hard-core model (independent sets) and the monomer-dimer model. We show how percolation ideas (combined with coupling arguments), can be used to obtain bounds for spatial dependencies. This is relevant for the question how efficiently (approximately) random samples can be generated.

Randomized embeddings of metrics

Michel Goemans (Université Catholique de Louvain, Belgium, and MIT, USA)

ABSTRACT. In this talk, I will describe several results concerning the embedding of certain types of metric spaces into others, either isometrically or with distortion. I will also discuss applications of such results to the design of approximation algorithms.

Fast approximate singular value decomposition of a matrix

Ravi Kannan (Yale University, New Haven, USA)

ABSTRACT. Recently, there has been a flurry of "constant time" algorithms for several Graph problems, like finding the approximate maximum cut in a dense graph. The running time of these algorithms depends exponentially on the allowed relative error epsilon. Here, we present another constant time algorithm which has a running time only polynomially dependent on epsilon.

The problem, we consider arises in many practical situations. We are given an $m \times n$ matrix A and have to find a matrix D of rank at most k (k is given and $k \ll m, n$) so that $\mathrm{norm}(A - D)$ (the sum of squares of all the entries) is as small as possible.

We describe a randomized algorithm which given k, ε, finds a D of rank at most k so that $\mathrm{norm}(A - D)$ is at most the least possible plus $\varepsilon\, \mathrm{norm}\, A$. If we can sample the entries of A with probabilities proportional to the square of the entry, then the algorithm works in time polynomial in k and $1/\varepsilon$ only (independent of m, n).

Joint Work with Alan Frieze and Santosh Vempala.

Contributed talks

Learning generalized Cavender-Farris trees

Mary Cryan (University of Warwick, UK)

Abstract. A Generalized Cavender-Farris tree (GCFT) on k states is a rooted tree T in which the root has some probability of being in each state in $\{0, \ldots, k-1\}$, and in which each edge is associated with a transition matrix on the states of $\{0, ..., k-1\}$. The GCFT can be associated with a stochastic process as follows: the root produces a state from its probability distribution. This state proceeds along the edges of the trees towards all the leaves of T, acted on by the transition matrices of each edge that it traverses. Therefore a GCFT with n leaves defines a distribution on strings of length n over the alphabet $\{0, ..., k-1\}$.

We consider the problem of PAC-learning GCFTs on 2 states in the variational distance measure when we already know the topology of T. We show that for any topology, and any $c, d > 0$, we can take f samples from the distribution generated by T, where f is polynomial in n, $1/d$ and $1/c$, and find a root probability for T and transition matrices for the edges of T so that the distribution generated by the hypothesis is within c of the distribution produced by T, with probability at least $1-d$.

This is joint work with Leslie Goldberg and Paul Goldberg.

Delayed path coupling and its applications

Artur Czumaj (Heinz Nixdorf Institute, Paderborn, Germany)

Abstract. Bubley and Dyer (FOCS'97) have recently proposed a path coupling method, a new approach to the classical coupling technique for bounding mixing rates of Markov chains. In this talk we will discuss applications of a simple refinement of path coupling that we call delayed path coupling. We illustrate delayed path coupling on few examples related to mixing items in distributed and parallel systems, and to estimate the recovery time in the edge orientation problem.

Part of this work has been done jointly with P. Kanarek, M. Kutylowski, and K. Lorys.

On approximation hardness of dense TSP and other path problems

Wenceslas Fernandez de la Vega (Université Paris-Sud, France)

Abstract. TSP(1,2) is the problem of finding a tour with minimum length in a complete weighted graph where each edge has length 1 or 2. Let d_o satisfy $0 < d_o < \frac{1}{2}$. We show that TSP(1,2) has no PTAS on the set of instances where the density of the subgraph spanned by the edges with length 1 is bounded below by d_o. We also show that LONGEST PATH has no PTAS on the set of instances with density bounded below by d_o.

Joint work with Marek Karpinski.

Contention resolution in multiple-access channels

Leslie Ann Goldberg (University of Warwick)

Abstract. A multiple-access channel is a broadcast channel (such as an Ethernet) that allows multiple users to communicate with each other by sending messages onto the channel. If two or more users simultaneously send messages, then the messages interfere with each other (collide), and the messages are not transmitted successfully. The channel is not centrally controlled. Instead, the users use a contention-resolution protocol to resolve collisions. Thus, after a collision, each user involved in the collision waits a random amount of time (which is determined by the protocol) before resending. In this talk, we focus on the expected (average) waiting time of messages. We show that some simple contention-resolution protocols cause messages to wait a long time (on average), and we describe a protocol due to the speaker, MacKenzie, Paterson and Srinivasan in which the expected average waiting time is at most a constant, independent of the number of users, as long as the mean arrival rate is at most $1/e$.

A genuinely polynomial-time algorithm for sampling two-rowed contingency tables

Catherine Greenhill (University of Leeds, UK)

Abstract. We define a new Markov chain for contingency tables with two rows. The chain is shown to be rapidly mixing using the path coupling method. The mixing time of the chain is quadratic in the number of columns and linear in the logarithm of the table sum. Two extensions of the new chain are discussed: one for three-rowed contingency tables and one for m-rowed contingency tables. Unfortunately, it is not possible to prove rapid mixing for these chains by simply extending the path coupling approach used in the two-rowed case.

Joint work with Martin Dyer.

Some upper and lower bounds on randomized branching programs

Marek Karpinski (Universität Bonn, Germany)

Abstract. We survey both upper and lower bounds established recently on the sizes of randomized read-once branching programs computing explicit boolean functions, like the middle bit of integer multiplication or the graph of integer multiplication. The underlying proof techniques are based on the randomized communication complexity of certain functions, and display for the first time exponential gaps in the randomized, deterministic, and nondeterministic size complexities of computing explicit boolean functions by ordered read-once branching programs (OBDDs).

On Learning Read-k-Satisfy-j DNF

Roni Khardon (University of Edinburgh, UK)

Abstract. We study the learnability of Read-k-Satisfy-j (RkSj) DNF formulas using the well-studied protocol of equivalence and membership queries. These are boolean formulas in disjunctive normal form (DNF), in which the maximum number of occurrences of a variable is bounded by k, and the number of terms satisfied by any assignment is at most j. We present an algorithm that with high probability finds a logically equivalent DNF formula, for any unknown RkSj DNF formula to be learned. The algorithm runs in polynomial time for $kj = O(\log n/\log\log n)$, where n is the number of input variables.

Joint work with H. Aizenstein, A. Blum, E. Kushilevitz L. Pitt and D. Roth.

On probabilistic model checking

Marta Kwiatkowska (University of Birmingham, UK)

Abstract. Model checking tools and techniques have seen increased industrial use in recent years, particularly in hardware design: given a formula of temporal or modal logic and a model of the system, one can algorithmically check whether the formula is satisfied in the model or not, and if not a counter-example can also be produced. Probabilistic model checking allows to verify systems containing probability information, such as Markov chains and randomized distributed algorithms, where instead of satisfiability checking the model checker outputs and estimate of the probability of the formula being satisfied. In this talk we will will propose a logic and a method applicable in the context of randomized distributed algorithms, and also discuss ways for making probabilistic calculations more efficient.

Bounds on constructing phylogenetic trees using balanced randomized splitting

Andrzej Lingas (Lund University, Sweden)

Abstract. We present new upper and lower bounds on the problem of constructing a phylogenetic tree from experiments, and a new lower bound on the problem of constructing a tree from an ultrametric distance matrix. The upper bound relies on a new technique called *balanced randomized tree splitting* that can be used to construct trees recursively by querying.

Joint work with Ming-Yang Kao and Anna Östlin.

Optimal data arrangement in a tree directory

Malwina Luczak (University of Oxford, UK)

Abstract. The problem of designing efficient data structures to store user information is crucial in mobile communications. For quick and reliable access, present databases are hierarchically distributed, usually in the form of a logical rooted tree.

Due to high demand for communication, the arrangement of data at tree nodes is crucial in reducing traffic volumes in the signalling network. The assignment of data records to physical locations should reflect geographical considerations, the pattern of call arrivals and user mobility characteristics.

In this paper, we define the problem OPTIMAL DATA ARRANGEMENT. The input is a graph G on d^h vertices, for some integers d and h, and symmetric traffic intensities λ_{ij} for each pair i, j of vertices of G. The question is how to assign the vertices of G to the leaves of a perfect d-ary tree of height h, so as to minimise the total sum, weighted by the traffic rates λ_{ij}, of their distances measured with respect to the tree topology.

We show that the decision version of OPTIMAL DATA ARRANGEMENT is strongly NP-complete for any fixed $d \geq 3$. The complexity of the binary tree case is an open problem. However, we derive two partial results which together suggest that the binary tree problem is strongly NP-complete also.

We also consider OPTIMAL GROUPING of data in databases, and show it is, in general, NP-hard.

Minimal spanning trees on random points

Mathew Penrose (University of Durham)

ABSTRACT. Given a set of points in Euclidean space, the minimum spanning tree is the graph connecting them with minimum total edge-length. We discuss asymptotic large-sample properties of the MST on *random* points, assumed to be a sample from a continuous distribution. In particular we look at the proportionate number of points of given degree, and the maximum edge-length, in the MST. Early impetus for studying this type of problem came from Karp's polynomial-time near-solution for the travelling salesman problem on random points. Further motivation comes from Statistics.

Testing approximate linearity with relative errors over finite rational domains

Miklos Santha (CNRS, Université Paris-Sud, France)

ABSTRACT. We present a self-tester to decide if a given program approximately computes with high probability a linear function on some finite rational domain. The novelty of our tester is that we consider relative errors in the computation. In the talk I will point out analogies and differences with existing testers for group homomorphisms and for linear functions over finite rational domains when the errors are absolute.

Joint work with Marcos Kiwi and Frédéric Magniez.

Random 2SAT and unsatisfiability

Yann Verhoeven (Université Paris-Sud, France)

ABSTRACT. It is a known result that almost all 2CNF formulas with $n(1+\varepsilon)$ clauses are not satisfiable, $\varepsilon > 0$ a constant and n the number of variables. We strenghten this result, showing that it still holds with $\varepsilon = \varepsilon(n) \geq w(n)n^{1/4}$ where $w(n)$ is any function tending to 0 as n tends to infinity.

On approximation intractability of the bandwidth problem

Jürgen Wirtgen (Universität Bonn)

ABSTRACT. The bandwidth problem is the problem of enumerating the vertices of a given graph G such that the maximum difference between the numbers of adjacent vertices is minimal. The problem has a long history and a number of applications. There was not much known about approximation hardness of this problem till recently. In this talk we show, that there is no PTAS for the bandwidth problem unless P = NP, even for the trees. More precisely we prove that there are no polynomial time approximation algorithms for general graphs with an approximation ratio better than 1·5, and for the trees with an approximation ratio better than 4/3.

Joint work with Gunter Blache and Marek Karpinski.