Summary

Welcome

Welcome to Maresias!

This booklet contains the titles and abstracts of the plenary talks, as well as the schedule of the short talks of the *2008 International Conference on the Analysis of Algorithms* (AofA), Maresias, 12–18 April 2008.

This conference is taking place in Brazil in an attempt to attract more people in Latin America to work in this central area of Theoretical Computer Science.

This edition of AofA will have 8 plenary speakers, 3 minicourses, and 42 short presentations, distributed in 9 sessions. The minicourses were designed to give a general overview, to students and researchers from related areas, of the main ingredients and tools used in the analysis of algorithms. A quick look at the list of invited speakers, their talk titles and their abstracts, and the short talks suffices to show that this will be a great meeting. We are most grateful to our invited speakers for agreeing to come, and to all speakers and participants.

Despite the busy programme, we do hope that you will have some time to see the surroundings and enjoy the beautiful local beaches.

We wish you all a pleasant and very productive conference!

April 2008

Cristina Gomes Fernandes

Yoshiharu Kohayakawa

Carlos Gustavo T. de A. Moreira

Roberto Imbuzeiro de Oliveira

Daniel Panario (Chair)

Sponsors and Support

We are very grateful for the financial support received from the following Brazilian agencies and institutions: *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES), *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP), *Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro* (FAPERJ), *Instituto de Matemática e Estatística da Universidade de São Paulo* (IME-USP), and *Instituto Nacional de Matemática Pura e Aplicada* (IMPA). We have also received the support of IM-AGIMB, the Mathematics Millennium Institute *Avanço Global e Integrado da Matemática Brasileira* (`http://milenio.impa.br/`), and of NUMEC-USP, the *Núcleo de Apoio à Pesquisa em Modelagem Estocástica e Complexidade* (`http://www.numec.prp.usp.br/`).

Overview of Program

Time	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday
8:30 - 9:30		Bruno Salvy					
9:30 - 10:30 Plenaries			Jean-François Marckert	Robert Sedgewick	Gaston Gonnet	Conrado Martínez	Short Talks (ST9)
10:30 - 11:00		Coffee break	Coffee break	Coffee break	Coffee break	Coffee break	
11:00 - 13:00 Sessions		Michael Drmota	Short Talks (ST2)	Short Talks (ST4)	Short Talks (ST5)	Short Talks (ST7)	Farewell lunch
13:00 - 15:00		Lunch	Lunch	Lunch	Lunch	Lunch	
15:00 - 16:00 Plenaries		Brendan McKay	Andrea Montanari	Free	John Dixon	Brigitte Vallée	
16:00 - 16:30		Coffee break	Coffee break		Coffee break	Coffee break	
16:30 - 18:30 Sessions	Registration	Short Talks (ST1)	Short Talks (ST3)		Short Talks (ST6)	Short Talks (ST8)	
18:30 - 20:00	Zhicheng Gao	Reception			Business Meeting		

Reception

There will be a reception cocktail on Sunday, after ST1.

Dinner

Dinner will be at 20:00h. The conference dinner will be a barbecue on Tuesday night.

Business Meeting

The business meeting will take place on Wednesday night, after ST6.

Bus back to São Paulo

The bus going back to São Paulo will leave on Friday, at 13:00h.

Minicourses

Saturday, 18:00 - 20:00

Symbolic methods in Combinatorics
Zhicheng Gao
Carleton University

Chair: Daniel Panario

"Analytic Combinatorics aims at predicting precisely the properties of large structured combinatorial configurations, through an approach based extensively on analytic methods. Generating functions are the central objects of study of the theory" (P. Flajolet and R. Sedgewick. Analytic Combinatorics, web edition 2007).

In this lecture we present an introduction to the combinatorial and enumerative points of view to the area based on generating functions. The main tools that have been developed under this approach are a collection of theorems oriented to provide a systematic translation mechanism between combinatorial constructions and operations on generating functions. We present the connection between unlabeled structures and ordinary generating functions as well as the one relating labelled structures and exponential generating functions. Moreover, we also present the relation between multivariate generating functions and the analysis of parameters of combinatorial structures. Several applications to the analysis of algorithms involving words, partitions, tree structures, and hashing will also be presented.

Sunday, 8:30 - 10:30

Asymptotic techniques for the analysis of algorithms
Bruno Salvy
INRIA

Chair: Julien Clement

In this minicourse, we show how characteristic parameters of a large family of combinatorial structures can be encoded into generating functions, from which asymptotic information can then be extracted by singularity analysis. Examples and exercises will help understand the basic techniques.

Sunday, 11:00 - 13:00

Limiting distributions in Combinatorics
Michael Drmota
TU Wien

Chair: Zhicheng Gao

The purpose of this talk is to present analytic and probabilistic methods in order to characterize the "typical behaviour" of certain parameters in "random discrete structures". A typical example is the number of cycles in the cycle decomposition of random permutations of n elements. A very powerful tool in this context are multivariate generating functions, the analysis of singularities and saddle point asymptotics.

Plenary talks

Sunday, 15:00 - 16:00

Correlation immune Boolean functions: asymptotic enumeration

Brendan D. McKay
Australian National University

Chair: Wojciech Szpankowski

A boolean function of n boolean variables is *correlation-immune* of order k if the function value is uncorrelated with the values of any k of the arguments. Such functions are of considerable interest due to their cryptographic properties, and are also related to the orthogonal arrays of statistics and the balanced hypercube colourings of combinatorics. The *weight* of a boolean function is the number of argument values that produce a function value of 1. If this is exactly half the function values, that is, 2^{n-1} values, a correlation-immune function is called *resilient*.

An asymptotic estimate of the number $N(n, k)$ of n-variable boolean functions of order k was obtained in 1992 by Denisov for the constant k. Denisov repudiated that estimate in 2000, but we will show that the repudiation was a mistake.

The main contribution of this paper is an asymptotic estimate of $N(n, k)$ which holds if k increases with n within generous limits and specialises to functions with a given weight, including the resilient functions. In the case of $k = 1$, our estimates are valid for all weights.

(This is joint work with *E. Rodney Canfield*, *Catherine Greenhill*, *Jason Gao*, and *Robert W. Robertson*.)

Monday, 9:30 - 10:30

Weak convergence of discrete objects to continuous objects in Combinatorics

Jean-François Marckert
University of Bordeaux 1

Chair: Guy Louchard

A prominent result in probability theory is the weak convergence of rescaled random walk to the Brownian motion. In the last years, some results of the same type have been proved about combinatorial structures: properly rescaled, some combinatorial objects as trees, maps, special families of paths, discrete measures... converge in distribution to some continuous counterparts. In this talk we will explain what are the benefits and the limitation of such results, and also, around examples, we will try to see what is needed to obtain such limit theorems.

Monday, 15:00 - 16:00

Hard constraint satisfaction problems, message passing algorithms, and statistical physics

Andrea Montanari
Stanford University

Chair: Roberto Imbuzeiro de Oliveira

Over the last few years there has been considerable success in applying message passing algorithms to treat hard combinatorial problems on sparse random graphs. A part of this progress has been triggered by the use of mathematically sophisticated but non-rigorous ideas from statistical physics. I will overview the on-going effort in understanding and rigorizing these ideas and their role in algorithm performances.

Tuesday, 9:30 - 10:30

Left-leaning red-black trees

Robert Sedgewick
Princeton University

Chair: Philippe Flajolet

The red-black tree model for implementing balanced search trees, introduced by Guibas and Sedgewick 30 years ago, is now found throughout our computational infrastructure. Red-black trees are the underlying data structure in the symbol-table implementation in C++, Java, Python, BSD Unix, and many other modern systems. However, current implementations have sacrificed some of the original design goals (primarily in order to develop an effective delete implementation, which was incompletely specified in the original paper), so a new look is worthwhile. In this talk, we describe a new variant of red-black trees that meets many of the original design goals and leads to substantially simpler code for insert/delete, less than one-fourth as much code as in implementations in common use. Can we analyze average-case performance for this new, simpler version?

Wednesday, 9:30 - 10:30

The impact of analysis of algorithms on bioinformatics

Gaston Gonnet
ETH Zurich

Chair: Nivio Ziviani

In principle, analysis of algorithms and bioinformatics share few tools and methods. This is only true when we look at the surface, deeper inspection shows many points of convergence, in particular in asymptotic analysis and model development. We would also like to stress the importance and usefulness of Maximum Likelihood for modelling in bioinformatics and the relation to problems in analysis of algorithms.

Wednesday, 15:00 - 16:00

Groups and algorithms

John Dixon
Carleton University

Chair: Daniel Panario

Study of algorithms in group theory and other algebraic structures dates back at least to the early 1960s. Computational group theory is now an active branch of mathematics and its practical success is illustrated by the computations possible with the two large suites of programs GAP and MAGMA which are available. Because groups have more structure than many other combinatorial objects it is possible to deal with larger objects; for example, sophisticated operations can be routinely carried out on groups of size around 10^{10} and sometimes much larger. As in any study of effective computation, the development of good algorithms is based on using as much of the underlying theory as possible, including some of the deepest theorems in group theory. On the other hand, some questions in computational group theory are essentially combinatorial in nature and can be described in quite an elementary manner. For example, how can we generate random elements in a group when all we know about the group is a small set of generating permutations or matrices? For this question, there remains a large gap between what is known and proved, and the kind of algorithms which programmers would like to have. I shall discuss this, and a number of related open questions.

Thursday, 9:30 - 10:30

The role of experiments in Analysis of Algorithms

Conrado Martinez
University Politecnica de Catalunya

Chair: Robert Sedgewick

There are two fundamental roles played by experimentation in connection to analysis of algorithms. Experiments are of great help in exploratory stages to build up hypothesis and develop intuitions. After the mathematical analysis has been completed they provide valuable support and validation for the theoretical model, when does the asymptotic regime start, etc. I'll discuss and give illustrative examples for these applications of experiments; however, my talk will be mostly focuses in one of the often unspoken goals of Analysis of Algorithms: to explain observed behavior.

Like in physics and the other natural sciences, the mathematical analysis of algorithms builds a theory to explain observations (experiments). Further experimentation is then necessary to validate or correct our models.

Thursday, 15:00 - 16:00

Analysis of Algorithms and Dynamical Systems: the example of the Euclid Algorithm

Brigitte Vallée
GREYC, CNRS and University of Caen

Chair: Michael Drmota

The Euclid algorithm is, according to Knuth himself, "the grandfather of all the algorithms". It is of great use, and the basic "parameters" of the algorithm, namely the number of iterations, the evolution of remainders, or the total number of elementary operations are of great interest, since they describe the precise complexity of the algorithm. Moreover, such a basic algorithm admits numerous variations, as the binary algorithm, the version which deals with least significant bits, the fast versions due to Knuth and Schönhage, etc. There exist very natural questions about the probabilistic behaviour of all these algorithms which had not received a precise answer until recently.

All the recent results are obtained by using the same general method, where an algorithm is viewed as a dynamical system. As usual in analysis of algorithms, generating functions of the main parameters are considered, but, now, they are generated themselves by the transfer operator of the underlying dynamical system, which plays the rôle of a generating operator. Even if the (various) Euclid algorithms give rise to very different dynamical systems, such a general methodology provides a powerful tool for answering (almost) all the main questions about the analysis of all these Euclid algorithms.

Overview of Short Talks

Sunday

ST1: Sunday, 16:30 - 18:30	*Chair: Gabor Lugosi*
Manuel Lladser, Analysis of patterns and minimal embeddings of non-Markovian sequences	
Yongwook Choi, Constrained pattern matching	
Antonio Galves, Stochastic chains with memory of variable length	
Hosam M. Mahmoud, On the average similarity and self-similarity of digital strings	

Monday

ST2: Monday, 11:00 - 13:00	*Chair: Hosam M. Mahmoud*
Vytas Zacharovas, A new approach to Poisson approximation and de-Poissonization	
Roberto Imbuzeiro de Oliveira, Random projections and nonlinear structure	
Vladimir Pestov, Analysis of hierarchical metric-tree based indexing schemes for similarity search in high-dimensional datasets	
Michele Soria, Boltzman sampling and properties of trees	

ST3: Monday, 16:30 - 18:45	*Chair: Alois Panholzer*
Guy Louchard, The odds-algorithm based on sequential updating and its performance	
Thomas Bruss, On a class of optimal stopping problems with mixed constraints	
Hsien-Kuei Hwang, Analysis of an exhaustive search algorithm in random graphs	
Amalia Duch, Pirates games (15min)	
Fabiano Botelho, Minimal perfect hash functions based on random graphs containing cycles (15min)	
Vlady Ravelomanana, Max-2-CSP in expected polynomial time (15min)	

Tuesday

ST4: Tuesday, 11:00 - 13:00	*Chair: Bruno Salvy*
Eugenijus Manstavicius, On analytic methods in combinatorics	
Zhicheng Gao, The smallest component size in decomposable structures of alg-log type	
Javiera Barrera, A law of large numbers for random partitions of the interval and the limiting search-cost of the move-to-front strategy	
Yuliy Baryshnikov, Snakes and Young tableaux	

Wednesday

ST5: Wednesday, 11:00 - 13:00	*Chair: Luc Devroye*
Michael Fuchs, On the number of subtrees on the fringe of random trees	
Gabor Lugosi, The effective resistance of random trees	
Bernhard Gittenberger, The shape of unlabeled trees	
Philippe Flajolet, Analytic combinatorics of two non-standard urn models, the Mabinogion and the OK Corral	

ST6: Wednesday, 16:30 - 18:30	*Chair: Hsien-Kuei Hwang*
Danièle Gardy, Asymptotic density of Boolean functions over implication	
Christian Mauduit, On the digital representation of squares	
Clemens Heuberger, Hamming weight of the non-adjacent-form under various input statistics and a two-dimensional version of Hwang's Quasi-Power-Theorem	
Daniel Panario, Analysis of some revocation schemes for stateless receivers (15min)	
Joachim von zur Gathen, Counting reducible and singular polynomials (15min)	

Thursday

ST7: Thursday, 11:00 - 13:00	*Chair: Conrado Martinez*
Uwe Roesler, Find revisited	
Jim Fill, The numbers of bit comparisons required by Quicksort and by Quickselect for Bernoulli-string input: average-case analysis	
Alois Panholzer, On a discrete parking problem	
Nivio Ziviani, Near space-optimal perfect hashing algorithms	

ST8: Thursday, 16:30 - 18:45	*Chair: Frédérique Bassino*
Brigitte Chauvin, ADN sequences and digital trees	
Robin Pemantle, An upper bound on the time to find a square in Pomerance's model for quadratic sieving	
Guillermo Matera, Fast computation of rational solutions to polynomial systems over a finite field	
Helmut Prodinger, Identities involving Harmonic numbers revisited	
Bruce Richmond, Maximum Stirling numbers of the second kind (15min)	

Friday

ST9: Friday, 9:30 - 11:30	*Chair: Jim Fill*
Tsung-Hsi Tsai, Fast forward property and decompositions of graphs of functions	
Louigi Addario-Berry, Killed branching random walks	
Markus Kuba, Area of lattice paths associated with urn models	
Philippe Chassaing, On statistics of some factorisations of random words	

Short Talks

ST1

Time: Sunday, 16:30 - 18:30
Chair: Gabor Lugosi

16:30 - 17:00:
Analysis of patterns and minimal embeddings of non-Markovian sequences
Manuel Lladser
The University of Colorado

Consider an infinite sequence $X = (X_n)_{n\geq 1}$ of symbols in a finite alphabet. This sequence may have a non-Markovian structure, i.e., for each $n > 1$ the distribution of X_n may depend in a non-trivial manner on each of the values of $X_1, \ldots, X_{n-1}$. The concatenation of the first n characters of X, i.e., the random word $X_1 \cdots X_n$ may be used as a model of a random text of length n. For example, in applied areas such as computer security, random text are used to model audit files. Since hackers leave traces of their intrusions in audit files, the identification of unusual patterns may be used to warn of potential intrusions in secured databases. Of course how unusual a pattern is observed in a text depends both on the pattern and the correlation structure between the characters in that text. In this talk we will show that regardless of the probabilistic model of X (Markovian or not) and of the pattern considered (regular or not), there exists an optimal first-order homogeneous Markov chain where one may embed X to keep track of the number of occurrences of that pattern in the random text. Since in many situations of interest this embedding will not be analytically tractable, we introduce an alternative embedding that—while not as optimal—it is amenable to keeping track of regular patterns in non-Markovian sequences. A case study of a non-Markovian sequence of 0's and 1's is analyzed and reveals new unexpected behavior: both discrete and Gaussian asymptotic distributions are encountered for the number of matches with two primitive regular patterns in $X_1 \cdots X_n$, as $n \to \infty$.

17:00 - 17:30:
Constrained pattern matching
Yongwook Choi
Purdue University

Constrained sequences find applications in communication, magnetic recording, and biology. In this paper, we restrict our attention to the so-called (d, k) constrained binary sequences in which any run of zeros must be of length at least d and at most k, where $0 \leq d < k$. In some applications one needs to know the number of occurrences of a given pattern w in such sequences, for which we coin the term *constrained pattern matching*. For a given word w, we first estimate the mean and the variance of the number of occurrences of w in a (d, k) sequence generated by a memoryless source. Then we present the (conditional) probability that w occurs r times in such a (d, k) sequence for

various ranges of r including central limit and large deviations regimes. Our experiment shows how these results can be applied to analyze neural spike data. As a by-product, we enumerate asymptotically the number of (d, k) sequences with exactly r occurrences of w, and compute Shannon entropy of (d, k) sequences with a given number of occurrences of w. Throughout this paper we use techniques of analytic information theory such as combinatorial calculus, generating functions, and complex asymptotics.

17:30 - 18:00:
Stochastic chains with memory of variable length
Antonio Galves
University of São Paulo

Stochastic chains with memory of variable length constitute an interesting family of stochastic chains of infinite order on a finite alphabet. The idea is that for each past, only a finite suffix of the past, called context, is enough to predict the next symbol. These models were first introduced in the information theory literature by Rissanen (1983) as a universal tool to perform data compression. Recently, they have been used to model up scientific data in areas as different as biology, linguistics and music. This paper presents a personal introductory guide to this class of models focusing on the algorithm Context and its rate of convergence.

18:00 - 18:30:
On the average similarity and self-similarity of digital strings
Hosam M. Mahmoud
The George Washington University

We investigate the average similarity of random strings as captured by the average number of "cousins" in the underlying tree structures. Two types of random data are considered: independent strings, and the suffixes of one random string, the latter is important in assessment of the degree of autocorrelation or self-similarity of a random string.

Analytical techniques including Poissonization and the Mellin transform are used for the accurate calculation of the mean. The string alphabets we consider are m-ary, and the corresponding trees are m-ary trees. Certain analytic issues arise in the m-ary case, and do not have an analog in the binary case.

(This is joint work with *Mark D. Ward*.)

ST2

Time: Monday, 11:00 - 13:00
Chair: Hosam M. Mahmoud

11:00 - 11:30:
A new approach to Poisson approximation and de-Poissonization
Vytas Zacharovas
Academia Sinica

A new approach to Poisson approximation is presented. It is based on properties of Charlier polynomials and Parseval's identity, and applicable to de-Poissonization problems.

(This is joint work with *Hsien-Kuei Hwang*.)

11:30 - 12:00:
Random projections and nonlinear structure
Roberto Imbuzeiro de Oliveira
IMPA

Dimensionality reduction methods are an important topic in algorithmic research and practice, with connections to machine learning, statistics, geometric computing and signal processing. Random projections stand out among such methods as simple, efficient and generic procedures to map sets $M \subset \mathbb{R}^D$ down to $d \ll D$ dimensions while approximately preserving much of the structure of M. For instance, if M is a finite set, setting $d = C \ln |M|/\varepsilon^2$ preserves all pairwise distances in M up to $1 \pm \varepsilon$ factors, with high probability. Notice that there is no dependence on D, the ambient dimension.

However, in some applications one needs to preserve the structure of a infinite set $M \subset \mathbb{R}^D$, such as a manifold on which the dataset approximately lies. We will show in this talk that if M is k-dimensional, then one can project it down $d = C(\kappa)\, k/\varepsilon^2$ in a almost-distance preserving manner, where $C(\kappa)$ only depends on the (extrinsic) curvature of M. The best previous result (due to Baraniuk and Wakin) had an extra $\ln D$ factor depending on the dimension of the ambient space $\mathbb{R}^D$. Our main technique is the so-called chaining method for proving concentration, which has a long history in high-dimensional probability.

12:00 - 12:30:
Analysis of hierarchical metric-tree based indexing schemes for similarity search in high-dimensional datasets
Vladimir Pestov
University of Ottawa

The curse of dimensionality is a well-known phenomenon across the entire computer science, negatively affecting, in particular, the performance of indexing schemes into large datasets for the purpose of similarity-based information retrieval. Paradoxically, there is still no mathematical proof that the above phenomenon really exists, that is, is in the nature of high-dimensional datasets. Of course, the very concept of intrinsic dimension of a dataset is still open to a discussion, but even in the cases that would be commonly accepted as "high-dimensional" (e.g., the uniformly distributed data in the Hamming cube $\{0,1\}^n$ as $n \to \infty$), the presence of the curse of dimensionality in the sense of proximity search remains unproved at a rigorous level (the "curse of dimensionality conjecture"). Various existing results in this direction are still very preliminary, in part because getting lower complexity bounds is always a hard task.

In this talk we will discuss a general setting for similarity search in the context of metric spaces equipped with probability distribution, where datapoints are modelled by i.i.d. random variables, and explain how various concrete indexing schemes fall into a general pattern of a hierarchical metric tree equipped with 1-Lipschitz decision functions. We assume, as is common, that the number n of datapoints grows faster than the dimension, d, but is subexponential in d. Using techniques from statistical learning, we obtain asymptotic lower bounds on the average-case performance of hierarchical metric tree indexing schemes that are superpolynomial in $\log n$.

12:30 - 13:00:
Boltzman sampling and properties of trees
Michele Soria
Université Paris 6, LIP6 and INRIA

We show that, for any simple family, trees generated with a singular Boltzmann sampler have the same distribution as subtrees in a large random tree. As a consequence, we easily get some statistical properties concerning global parameters, such as various weights, of subtrees in a random tree.

ST3

Time: Monday, 16:30 - 18:45
Chair: Alois Panholzer

16:30 - 17:00:

The odds-algorithm based on sequential updating and its performance

Guy Louchard
University Libre de Bruxelles

Let $I_1, \ldots, I_n$ be independent indicator functions on some probability space $(\Omega, \mathcal{A}, \Pr)$. Suppose that these indicators can be observed sequentially. Let T be the set of stopping times on $(I_k), k = 1, \ldots, n$ adapted to the increasing filtration $(\mathcal{F}_k)$, where $\mathcal{F}_k = \sigma(I_1, \ldots, I_k)$. The odds-algorithm solves the problem of finding a stopping time $\tau \in T$ which maximizes the probability of stopping on the last $I_k = 1$, if any. To apply the algorithm one needs only the odds for the events $\{I_k = 1\}$, that is $r_k = p_k/(1 - p_k)$, where $p_k = \mathrm{E}(I_k)$, $k = 1, \ldots, n$, or at least a certain number of them. The goal here is to offer tractable solutions for the case where the p_k are unknown and must be sequentially estimated. The motivation is that this case is important for many real word applications of optimal stopping. Our main result is a new version of the odds-algorithm based on online observation and sequential updating. Questions of speed and performance of this algorithm are studied in detail.

(This is joint work with *F.T. Bruss*.)

17:00 - 17:30:

On a class of optimal stopping problems with mixed constraints

Thomas Bruss
Universite Libre de Bruxelles

Let $X_1, X_2, \ldots, X_n$ be i.i.d. positive random variables having absolute continuous distribution function F. We can observe the X_k's sequentially and must decide at each time (instant) k whether to select (accept) X_k, or else to refuse and continue. The number of selections is subject to two different types of constraints. On the one hand, there is a prescribed minimal number r of selections, on the other hand, the selection strategy must accept, in expectation, a minimum number of μ of the X_k's. We interpret X_i as the cost of item i. Our objective is to minimize the expected total cost of selected items under the two constraints over the set of all non-anticipative online selection strategies. To solve the problem—which was motivated by a question about spanning trees—we prove that the two constraints can be put into a natural hierarchy. This allows to obtain optimality equations which we can solve precisely for any n via a double recursion. This determines the (unique) optimal strategy. The asymptotic behavior of the optimal value as a function of μ and r (normalized with respect to n) is proved to be the uniquely relevant solution of a corresponding Riccati differential equation.

17:30 - 18:00:

Analysis of an exhaustive search algorithm in random graphs

Hsien-Kuei Hwang
Academia Sinica

What is the complexity of a naive exhaustive search for finding a maximum independent set in random graphs $G_{n,p}$? We show that the expected complexity is of order $n^{c \log n}$, surprisingly comparable with the use of the powerful Metropolis process. In addition to the mean value, we also present asymptotic normality result for the distribution of the cost. The approach we develop can be quickly extended to several other graph algorithms.

(This is joint work with *Cyril Banderier*, *Vlady Ravelomanana*, and *Vytas Zacharovas*.)

18:00 - 18:15:
Pirates games
Amalia Duch
Universitat Politecnica de Catalunya

In this ongoing work we consider n pirates roaming a weighted graph where m treasures are hidden in m selected vertices. All pirates start their roaming from given starting vertices and for traversing a given edge in the graph all pirates pay the same cost which is the weight of that edge. We analyze this game over two different models:

1. a pirate may get a benefit when reaching a treasure which will be equally shared with all those pirates reaching the treasure and
2. a pirate may get a benefit when reaching a treasure which will be the whole treasure when he arrives first or a proportional part of the treasure if he gets there first but simultaneously with other pirates.

We show the relation of this game with congestion games coming from algorithmic game theory and we propose and analyze algorithms that find Nash equilibrium when it is possible.

18:15 - 18:30:
Minimal perfect hash functions based on random graphs containing cycles
Fabiano Botelho
Federal University of Minas Gerais

Minimal perfect hash functions are widely used for memory efficient storage and fast retrieval of items from static sets. A perfect hash function (PHF) h from a key universe U to a range $[0, m-1]$ for a subset S of U, where $|S| = n$, is a function that maps the keys of S to unique values. A minimal perfect hash function (MPHF) is a PHF with the smallest range, i.e., $m = n$. An efficient approach to generate PHFs and MPHFs is based on random graphs with no cycles. By using random graphs with no cycles Botelho, Pagh and Ziviani proposed in WADs 2007 an efficient algorithm to generate near-space optimal PHFs and MPHFs that are stored in $O(n)$ bits. A similar approach was used by Majewski, Wormald, Havas and Czech in TCJ 1996 to efficiently obtain MPHFs that are stored in $O(n \log n)$ bits. The objective of this talk is to present two algorithms that generates PHFs and MPHFs based on random graphs with cycles instead. The first one, previously published by Botelho, Kohayakawa and Ziviani in WEA 2005, improves the space requirement to 55% of one instance of the algorithms designed by Majewski, Wormald, Havas and Czech. However, it still generates functions that are stored in $O(n \log n)$ bits and it is not completely analyzed. Here we present the main open problem of the result by Botelho, Kohayakawa and Ziviani. The second algorithm is also based on random graphs with no cycles and considerably improves the efficiency of one instance of the algorithms designed by Botelho, Pagh and Ziviani to build functions in the same format. The amount of space needed to represent a PHF is approximately $2n$ bits and a MPHF is approximately $3.3n$ bits, which is around a factor of 2 from the information theoretical minimum of approximately $0.89n$ and $1.44n$ bits, respectively.

18:30 - 18:45:
Max-2-CSP in expected polynomial time
Vlady Ravelomanana
University of Paris 13

Using tools from enumerative/analytic combinatorics and the Scott/Sorkin's reduction, we show that many optimization and constraints satisfaction problems including Max-Cut, Max-Dicut, Max-2-Lin, Max-2-Sat, Max-Ones-2-Sat, Max-Independent-Set, Minimum-Vertex-Cover are solved in expected polynomial time whenever their underlying graphs have n vertices and $n/2+O(\log n^{1/3})n^{2/3}$ edges.

ST4

Time: Tuesday, 11:00 - 13:00
Chair: Bruno Salvy

11:00 - 11:30:

On analytic methods in combinatorics

Eugenijus Manstavicius
Vilnius University

We intend to point out some obstacles in applying the classical and more recent analytic methods in combinatorics. First, dealing with generating functions depending on some extra parameters, we face the lack of uniformity in the approximations of the nth Taylor coefficient. The situation becomes even worse if, having an array of these coefficients, we examine the diagonal ones. The most typical problem of this kind is enumeration of decomposable combinatorial structures of size n with some component constraints also depending on n. Next, many value distribution problems for the mappings defined on random structures lead to analysis of generating series which do not have any analytic extension outside the convergence disk. Using some ideas going back to number theory, we have developed an approach capable to overcome these obstacles (see, for instance, E. Manstavicius, Mappings on decomposable combinatorial structures: analytic approach, *Combinatorics, Probab. Computing*, 11 (2002), 61-78, and E. Manstavicius, An analytic method in probabilistic combinatorics, *Osaka J. Math.* (2006), submitted). Under mild conditions, the approach can yield even sharp remainder term estimates. That has been demonstrated in the case of permutations (see E. Manstavicius, The BerryEsseen bound in the theory of random permutations, *The Ramanujan J.*, 2 (1998), 185-199, and more recent papers by V. Zacharovas). The further perspectives of the method will be discussed in the talk.

11:30 - 12:00:

The smallest component size in decomposable structures of alg-log type

Zhicheng Gao
Carleton University

A *decomposable combinatorial structure* consists of simpler objects called *components* which by themselves can not be further decomposed. In this talk, we assume that the component generating function $C(z)$ is of alg-log type, that is, $C(z)$ behaves like

$$(1-z/\rho)^{-\alpha}\left(\ln\frac{1}{1-z/\rho}\right)^{\beta}$$

near its dominant singularity ρ.

We present some results about the size of the rth smallest component in a random structure. For example, if $X_n^{[r]}$ denotes the rth smallest component size in a random labeled structure of size n, and if $-1<\alpha<0$, then

$$P\left(X_n^{[r]}>k\right)\sim\exp(-C_k(\rho))\sum_{j=0}^{r-1}C_k(\rho)^j/j!$$

uniformly for all $k=o(n/\log n)$, where r is any fixed positive integer and $C_k(z)$ is the partial sum of the first $k+1$ terms of $C(z)$. Moreover the expected smallest component size is asymptotically equivalent to $n\exp(-C(\rho))$.

(This is joint work with *Li Dong*, *Daniel Panario*, and *Bruce Richmond*.)

12:00 - 12:30:
A law of large numbers for random partitions of the interval and the limiting search-cost of the move-to-front strategy
Javiera Barrera
Universidad Técnica Federico Santa María

We prove a law of large numbers for certain finite random partitions of $[0,1]$, when the number of fragments goes to ∞. Then, we apply it to compute the limiting distribution of the transient search-cost of the move-to-front rule for general classes of random and deterministic request probabilities, when the list size goes to ∞.

(This is joint work with *Joaquin Fontbona*.)

12:30 - 13:00:
Snakes and Young tableaux
Yuliy Baryshnikov
Bell Labs

Snake calculus is a generalization of the well known problem—to count up-down permutations. Up-down permutations can be interpreted as the standard Young tableaux with the stair-like shape with steps north-east-north-east-$\cdots = (NE)^k$. In this work we will count tableaux for the shapes encoded by arbitrary periodic words like $(\mathrm{NNENEEENN})^k$ (and their thickened generalizations) using an explicit diagonalization of the transfer operator: the problem turns out to be exactly solvable in Baxter's sense.

ST5

Time: Wednesday, 11:00 - 13:00
Chair: Luc Devroye

11:00 - 11:30:
On the number of subtrees on the fringe of random trees
Michael Fuchs
National Chiao Tung University

In a recent paper, Feng, Mahmoud, and Panholzer showed that the limit distribution of the number of subtrees on the fringe of random recursive trees and random binary search trees undergoes a phase change from normal to Poisson as the subtree size is growing. In this talk, we introduce a different approach which can be used to shed further light on the above phase change phenomena. Moreover, we will discuss applications of our approach to other classes of random trees as well as applications to parameters arising from genetics and molecular biology.

11:30 - 12:00:
The effective resistance of random trees
Gabor Lugosi
Pompeu Fabra University

An electrical network is a graph whose edges are equipped with positive real numbers (the resistances). The effective resistance between two vertices is determined by Ohm's law. Electrical networks have been known to be intimately related with random walks on the corresponding graph. In this talk we investigate the effective resistance of electrical networks represented by trees when resistances are assigned randomly to each edge. We give precise estimates of the expected value and variance of the effective resistance between the root and leaves of a complete binary tree with random edge-resistances. Some results are generalized to networks given by supercritical branching processes.

(This is joint work with *Nicolas Broutin* and *Louigi Addario-Berry*.)

12:00 - 12:30:
The shape of unlabeled trees
Bernhard Gittenberger
TU Wien

We consider the number of nodes in the levels of unlabeled rooted random trees and show that the stochastic process given by the properly scaled level sizes weakly converges to the local time of a standard Brownian excursion. Furthermore we compute the average height of such trees. These results extend existing results for simply generated trees and forests to the case of unlabeled rooted trees and show that they behave essentially like a Galton-Watson process.

(This is joint work with *Michael Drmota*.)

12:30 - 13:00:
Analytic combinatorics of two non-standard urn models, the Mabinogion and the OK Corral

Philippe Flajolet
INRIA

This presentation is relative to two urn models of the Polya type that are non-standard in the sense that balls of certain types may disappear and the global population does not increase. The OK Corral model is one of the simplest models in the theory of conflicts, where annihilation between opposite groups take place. The Mabinogion urn models a less brutal model of the spread of influences. Both models admit explicit analytic solutions in the form of combinatorial sums and integrals and are related to classical combinatorics (e.g., the Eulerian numbers and the Ehrenfest model). The asymptotic treatment, including the extraction of limit laws, can then be based on methods of analytic combinatorics like the saddle-point method, but it also leads to some non-classical asymptotic problems. The final outcome is a precise characterization of the probabilistic laws involved, including local limit and speed of convergence estimates.

(This is based on joint works with *Vincent Puyhaubert* and *Thierry Huillet*.)

ST6

Time: Wednesday, 16:30 - 18:30
Chair: Hsien-Kuei Hwang

16:30 - 17:00:
Asymptotic density of Boolean functions over implication

Danièle Gardy
Université de Versailles Saint-Quentin

Consider Boolean expressions built on implication and n (fixed) positive literals: There is a many-to-one correspondence between such expressions and the Boolean functions of the Post class S_0. Now assume a uniform distribution on the expressions of given (large) size m: This induces a probability distribution on S_0. What is this probability distribution? How does the probability of a function relate to its tree complexity (minimal size of an expression associated with the function)? What can we say about the average complexity of a random function under this distribution?

(This is joint work with *Hervé Fournier*, *Bernhard Gittenberger* and *Antoine Genitrini*.)

17:00 - 17:30:
On the digital representation of squares

Christian Mauduit
Université de la Méditerranée

The main purpose of this talk is to present a joint work with Joel Rivat concerning the study of the q-adic representation of square numbers. In particular we can prove that (under some obvious necessary conditions), the sum of digits of squares is well distributed in arithmetic progression, which answer to a question asked by Alexander Gelfond 40 years ago. The techniques we use to estimate the associated exponential sums can also be applied to give precise estimates for the number of squares with an average sum of digits.

(This is work in progress with *Michael Drmota* and *Joel Rivat*.)

17:30 - 18:00:
Hamming weight of the non-adjacent-form under various input statistics and a two-dimensional version of Hwang's Quasi-Power-Theorem

Clemens Heuberger
TU Graz

Public key cryptography relies on the efficient calculation of one-way functions, for instance, scalar multiplication in the point group of an elliptic curve. One well-known method to perform this operation is the use of a signed binary expansion for the scalar and Horner's scheme. The number of operations is minimized if the Hamming weight (number of non-zero digits) is minimized. This is achieved by the non-adjacent form, a signed binary expansion without adjacent non-zero digits.

The Hamming weight of the non-adjacent form has been studied from an analysis-of-algorithms point of view before; however, a detailed study in relation to the Hamming weight of the standard binary expansion was still unavailable. In particular, we investigate the expected Hamming weight of the NAF of a n-digit binary expansion with k ones where k is either fixed or proportional to n. The expected Hamming weight of NAFs of binary expansions with large ($\geq n/2$) Hamming weight is studied. Finally, the covariance of the Hamming weights of the binary expansion and the NAF is computed. Asymptotically, these Hamming weights become independent and normally distributed.

This last result is obtained by a two-dimensional version of H.K. Hwang's Quasi Power Theorem, which shall also be discussed in this talk.

(This is joint work with *Helmut Prodinger*.)

18:00 - 18:15:
Analysis of some revocation schemes for stateless receivers
Daniel Panario
Carleton University

We consider the problem of a center broadcasting an encrypted message to a group of users such that some subset is considered revoked and should not be able to obtain the content of the broadcasted message even if all revoked users collaborate. Various encryption schemes have been proposed to solve this problem which arises, for example, with pay-TV and satellite communications.

In one class of proposed schemes the center distributes a unique combination of keys to each user who decrypts the message individually. If keys cannot be updated once distributed the receivers are called *stateless*.

Several key distribution schemes use a balanced binary tree structure. Some examples that we consider are the subset-difference scheme (SD), introduced by Naor, Naor and Lotspiech (2003), the complete subtree scheme (CST), introduced independently by Wallner, Harder and Agee (1998), and Wong, Gouda and Lam (1998), and the layered subset-difference scheme (LSD) by Halevy and Shamir (2002). Park and Blake (2006) give generating functions that entail the exact mean number of encryptions for the above key distribution schemes. We extend their results by showing that the distribution of the number of encryptions is asymptotically normal with mean and variance both proportional to 2^n.

(This is joint work with *Christopher Eagle*, *Zhicheng Gao*, *Mohammed Omar*, and *Bruce Richmond*.)

18:15 - 18:30:
Counting reducible and singular polynomials
Joachim von zur Gathen
Universität Paderborn

Among the bivariate polynomials over a finite field, most are irreducible. We count some classes of special polynomials, namely the reducible ones, those with a square factor, the 'relatively irreducible' ones which are irreducible but factor over an extension field, and the singular ones, which have a root at which both partial derivatives vanish.

ST7

Time: Thursday, 11:00 - 13:00
Chair: Conrado Martinez

11:00 - 11:30:
Find revisited
Uwe Roesler
Universitat Kiel

The running time analysis of the algorithms Find or Quickselect leads to a stochastic fixed point equation for distributions on some function space. It is the first one of this kind for processes. We will review the specific results for Find, view the technical progress made by studying this example and discuss various Find versions, including the asymptotically best.

11:30 - 12:00:
The numbers of bit comparisons required by Quicksort and by Quickselect for Bernoulli-string input: average-case analysis
Jim Fill
The Johns Hopkins University

The analysis of many algorithms and data structures (such as digital search trees) for searching and sorting are based on the representation of the keys involved as bit strings and so count the number of bit comparisons. On the other hand, the standard analysis of many other algorithms (such as Quicksort and Quickselect) are performed in terms of the number of key comparisons. Fill and Janson (SODA04) introduced the prospect of a fair comparison between algorithms of the two types by providing an average-case analysis of the number of bit comparisons required by Quicksort, and Fill and Nakama (ANALCO08) did similarly for Quickselect.

Both pairs of authors treated the case that the keys to be sorted are independent random variables uniformly distributed over $(0, 1)$. For Quicksort Fill and Janson considered rather general continuous distributions over $(0, 1)$ in place of the uniform; this is a natural input-model generalization to consider from a key-comparisons point of view. In this talk I will show how to extend the Quicksort and Quickselect results in a direction more natural from a bit-comparisons orientation by assuming that the successive bits in each key are independent Bernoulli random variables with success probability p different from $1/2$.

12:00 - 12:30:
On a discrete parking problem
Alois Panholzer
TU Wien

Konheim and Weiss introduced in the Sixtees the notion of parking functions during their studies of a linear probing hashing algorithm. Using their vivid description we consider a one-way street with m (originally empty) parking lots with addresses $1, 2, \ldots, m$ on the side. Now n cars are arriving sequentially and each car wants to park on a specific address $x_i \in \{1, 2, \ldots, m\}$. If this place is occupied he drives on and parks on the first empty lot. If he does not find an empty lot he drives on without parking. A parking function is then a sequence $x_1, \ldots, x_n$ of addresses such that all n cars can be parked. Parking functions and various generalizations have been studied extensively and many connections to various other combinatorial objects, e.g., to trees, hyperplane arrangements and non-crossing partitions, have been revealed.

Here we go another direction and consider, for m parking lots and n cars, the problem: "How many cars are unsuccessful, i.e., cannot be parked?". We first give an exact formula for the number of sequences $x_1, \ldots, x_n \in \{1, \ldots, m\}^n$ of addresses, where exactly k cars are unsuccessful. This allows us to study the random variable $X_{m,n}$, which denotes the number of unsuccessful cars for a random sequence of addresses for n cars and m parking lots. A first study of $X_{m,n}$ leading to exact and asymptotic results for the expectation has been carried out by Gonnet and Munro [1984] during their analysis of a certain sorting algorithm. Here we focus on limiting distribution results and locate, depending on the growth of m and n, a total of nine different phases in the limiting behaviour of $X_{m,n}$, and for each phase we can give a characterization of the limit law appearing.

12:30 - 13:00:

Near space-optimal perfect hashing algorithms

Nivio Ziviani

Federal University of Minas Gerais

Minimal perfect hash functions are widely used for memory efficient storage and fast retrieval of items from static sets. A perfect hash function (PHF) $h : U \rightarrow [0, m-1]$ for a key set S is a function that maps the keys of S to unique values. A minimal perfect hash function (MPHF) is a PHF with the smallest range, i.e., $m = n$. The objective of this paper is to present two time efficient and near space-optimal perfect hashing algorithms. We present: (1) an internal memory based algorithm that assume uniform hashing to build a family of PHFs (for $m = 1.23n$) or MPHFs (for $m = n$) based on random graphs, and (2) an external memory based algorithm that has experimentally proven practicality for sets in the order of billions of keys. Both internal and external algorithms have the following properties: (i) evaluation of a PHF or a MPHF requires constant time, (ii) the algorithms are simple to describe and implement, and generate the functions in linear time, (iii) for the internal algorithm, the amount of space needed to represent a PHF is approximately $1.95n$ bits and a MPHF is approximately $2.62n$ bits, which is around a factor of 2 from the information theoretical minimum of approximately $0.89n$ and $1.44n$ bits, respectively; for the external algorithm, the amount of space needed to represent a PHF is approximately $2.7n$ bits and a MPHF is approximately $3.8n$ bits. To our knowledge, no previously known algorithm has these properties and any algorithm in the literature with the property (iii) either requires exponential time for construction and evaluation, or uses near-optimal space only asymptotically, for extremely large n. The internal algorithm is used in one phase of the external algorithm. We demonstrate the scalability of the external algorithm by constructing MPHFs for a set of 1.024 billion URLs from the World Wide Web of average length 64 characters in approximately 62 minutes, using a commodity PC.

ST8

Time: Thursday, 16:30 - 18:30
Chair: Frédérique Bassino

16:30 - 17:00:
ADN sequences and digital trees

Brigitte Chauvin
University of Versailles Saint-Quentin

Digital trees (digital search trees or suffix tries) can be used to represent an ADN sequence, so that usual characteristics of random trees (height, saturation level, path length, ...) have a new interest. The study is delicate (and challenging) since the random data giving the tree are strongly dependent. The "good" frame is that of dynamical sources.

17:00 - 17:30:
An upper bound on the time to find a square in Pomerance's model for quadratic sieving

Robin Pemantle
University of Pennsylvania

Pomerance formulated the following problem which arises in quadratic sieving (the run time of quadratic sieving reduces to this as long as quadratic residues have the same asymptotic factor distribution as nonresidues).

> Generate random numbers in the interval $[1, N]$ until some subset has a product which is a square (and there is a computable witness to this): how long does this take?

Naive probabilistic models for the distribution of prime factors suggest that this stopping time has a very sharp threshold. Based on more sophisticated probabilistic models, we prove an upper bound that is within a factor of $4/\pi$ of a proven lower bound, and conjecture that our upper bound is in fact asymptotically sharp.

(This is joint work with *Andrew Granville*, *Ernie Croot*, and *Prasad Tetali*.)

17:30 - 18:00:
Fast computation of rational solutions to polynomial systems over a finite field

Guillermo Matera
Universidad Nacional de General Sarmiento

This talk will be concerned with the computation of rational solutions to a polynomial system defined over a finite field. We shall comment on estimates on the number of rational points on an algebraic variety defined over a finite field. Then we shall present algorithms for computing a rational point of an absolutely irreducible variety defined over a finite field, and extensions to more general situations.

18:00 - 18:30:
Identities involving Harmonic numbers revisited
Helmut Prodinger
Stellenbosch University

People who analyse algorithms encounter harmonic numbers frequently. But I was also involved a few years ago in a project about Pade approximations to the logarithm where some nontrivial identities had to be proved. I will give a survey of identities that played a role in my life, also some q-versions, and spiced up with some recent results which I obtained using a clever but very simple method due to Wenchang Chu.

18:30 - 18:45:
Maximum Stirling numbers of the second kind
Bruce Richmond
University of Waterloo

Say an integer n is *exceptional* if the maximum Stirling number of the second kind $S(n,k)$ occurs for two (of necessity consecutive) values of k. We prove that the number of exceptional integers less than or equal to x is $O(x^{1/2+\epsilon})$, for any $\epsilon > 0$. We derive a similar result for partitions of n into exactly k integers.

ST9

Time: Friday, 9:30 - 11:30
Chair: Jim Fill

9:30 - 10:00:

Fast forward property and decompositions of graphs of functions

Tsung-Hsi Tsai
Academia Sinica

A function $f : \{1, \ldots, N\} \to \{1, \ldots, N\}$ is fast forward if for each natural number m and each $x = 1, \ldots, N$, the computational complexity of evaluating $f^m(x)$, the m-th iterate of f at x, is polynomial in $\log N$. This is useful in simulations and cryptographic applications. Tsaban (2007) proposed an approach: orbit decompositions of graphs of functions with lookup tables of size N implemented. Experimental results indicate that most functions satisfy the fast forward property with lookup tables. We derive that all functions satisfy the fast forward property. We also propose optimal decompositions of graphs of functions for the computational complexity of evaluating the iteration.

10:00 - 10:30:

Killed branching random walks

Louigi Addario-Berry
University of Oxford

The problem is related to searching in trees. Suppose we are given a complete binary tree with i.i.d. random edge weights (say copies of some random variable X). The depth $d(v)$ of a vertex v is the number of edges on the path to the root. We give each vertex v the label S_v which is the sum of the edge weights on the path from v to the root. For positive integers n, we let M_n be the maximum label of any vertex at depth n, and let $M^* = \max\{M_n : n = 0, 1, \ldots\}$. It is of course possible that M^* is infinity.

Under suitable moment assumptions on X, it is known that there is a constant A such that $M_n/n \longrightarrow A$ almost surely and in expectation. We call the cases $A > 0$, $A = 0$, and $A < 0$ supercritical, critical, and subcritical, respectively. When $A \leq 0$ it makes sense to try to find the vertex of maximum weight M^* in the whole tree. One possible strategy is to only explore the subtree T_0 containing the root consisting only of vertices of non-negative weight. With probability bounded away from zero this strategy finds the vertex of maximum weight. We derive precise information about the expected running time strategy. Equivalently, we derive precise information about the random variable $|T_0|$. In the process, we also derive rather precise information about M^*. This answers a question of David Aldous.

(This is joint work with *Nicolas Broutin*.)

10:30 - 11:00:
Area of lattice paths associated with urn models
Markus Kuba
TU Wien

The study of the area under lattice paths, measured either continuous, or discrete as the number of lattice points below the sample path, has a long history. We are interested in the distribution of the area under lattice paths associated with certain diminishing urn models, in particular the urns specified by the 2×2 ball replacement matrix $M = \left(\begin{smallmatrix} -a & 0 \\ c & -d \end{smallmatrix}\right)$, where $a, d \in N$ and $c = p \cdot a$. We give a full characterization of the arising limiting distributions.

(This is joint work with *Alois Panholzer*.)

11:00 - 11:30:
On statistics of some factorisations of random words
Philippe Chassaing
Université Henri Poincaré

We give some results on the asymptotic behavior of the Lyndon factorisation of random words, extending results by Clement, Bassino and Nicaud, or Diaconis, Pitman and McGrath or again Marchand & Zohoorian.

(This is joint work with *Elahe Zohoorian-Azad*.)