

Invited talks

JOHN D. DIXON
Carleton University

Groups, algorithms and probability

In the past 15 years a lot of progress has been made in developing algorithms for computation in groups and other algebraic systems, and many of the algorithms have been implemented in the algebraic computing systems GAP and Magma. Often these algorithms are probabilistic in nature, or require an analysis in terms of average behaviour where this is appropriate. Typically, solutions to the problems which arise in developing such algorithms involve group theory together with applications of techniques such as Markov chains. I shall survey some of the work done and discuss some open problems.

JIM FILL
The Johns Hopkins University

*Precise logarithmic tail asymptotics for some limit random variables for
random trees*

We obtain precise logarithmic asymptotics for the tails of certain random variables that arise as limits of functionals of random finite trees. Our results are based on the facts (i) that the random variables we study can be represented as functionals of a Brownian excursion and (ii) that a large deviation principle with good rate function is known explicitly for Brownian excursion. Examples include limit distributions of the total path length and of the Wiener index in conditioned Galton-Watson trees (also known as simply generated trees), where we recover results proved by Philippe Chassaing and Svante Janson by a different method; in the case of Wiener index (and some other examples), a key constant is expressed as the solution to a certain optimization problem, but the constant's precise value is unknown. (We invite audience participation!) Co-authors: Svante Janson, Department of Mathematics, Uppsala University (Sweden)

FAN CHUNG GRAHAM
University of California at San Diego

Random graphs and Internet graphs

We will discuss some recent developments on random graphs with given expected degree distributions. Such random graphs can be used to model various very large graphs arising in Internet and telecommunications. In turn, these "massive graphs" shed insights and lead to new directions for random graph theory. For example, it can be shown that the sizes of connected components depend primarily on the average degree and the second-order average degree under certain mild conditions. Furthermore, the spectra of the adjacency matrices of some random power law graphs obey the power law while the spectra of the Laplacian follow the semi-circle law. We will mention a number of related results and problems that are suggested by various applications of massive graphs.

QI-MING HE,
Dalhousie University

*Algorithms for Computing Bi-diagonal Representations of
Matrix-Exponential Distributions*

The class of matrix-exponential distributions includes phase-type distributions, Coxian distributions, generalized Erlang distributions, and exponential distributions as its special cases. Applications of those classes of distributions can be found in control theory, biostatistics, telecommunications, risk analysis, queueing theory, and inventory theory.

In the past few years, we studied the representations of matrix-exponential distributions. Our focus was the bi-diagonal representations, including the Coxian representations, of matrix-exponential distributions. A spectral polynomial algorithm (SPA) was introduced for computing a bi-diagonal representation for any matrix-exponential distribution. Based on the SPA, methods for computing the minimal Coxian representations, Coxian representations of Coxian distributions as approximations to matrix-exponential distributions were introduced as well. In our talk, we shall report those results and discuss some problems currently under investigation.

BORIS PITTEL
Ohio State University

Critical percolation on a random regular graph

Let $d \geq 3$ be fixed. Consider G_n , a uniformly random d -regular graph on n vertices. Given $p \in (0, 1)$, introduce $G_{n,p}$, a random subgraph of G_n obtained by opening (closing) each edge of G_n with probability p ($q = 1 - p$ resp.), independently of all other edges. Thus the edge set of $G_{n,p}$ is a random set of open edges of G_n . For $d = n - 1$ this is a well-studied

random subgraph of the complete graph K_n . We show that for $p < p_c := (d-1)^{-1}$ with high probability all the components of $G_{n,p}$ are either trees or unicyclic components, of size $O(\log_b n)$, $b = p_c q_c^{d-2} / p q^{d-2}$. (p_c is the maximum point of $p q^{d-2}$, so $b > 1$.) For $p > p_c$ whp the random graph $G_{n,p}$ contains a unique giant component of size asymptotic to $\alpha(p)n$, where

$$\alpha = 1 - (zp + q)^d, \quad z = (zp + q)^{d-1};$$

explicitly

$$\gamma(p) = 1 - \sum_{j=1}^{\infty} \frac{dq^d}{j(d-1)+1} \cdot \binom{j(d-1)+1}{j-1} (pq^{d-2})^{j-1}.$$

(Here $z = z(p)$ is the non-extinction probability for a branching process with immediate family size distributed binomially, with parameters $d-1$ and p . Intuitively this should be expected since within a relatively small distance from a vertex whp the edges of G_n form a d -regular subtree.) Whp all other components are again of size $O(\log_b n)$. In fact the j -th term in the sum is the limiting fraction of vertices in the tree components of size j . We also show that a unique component, dwarfing in size the other components, whp exists when $p - p_c \gg n^{-1/4}$.

NICK WORMALD
University of Waterloo

Exploring the cores of random graphs

A pseudograph model of random graphs introduced by Bollobás and Frieze, and also Chvátal, is especially useful for obtaining properties of sparse random graphs, when the average degree is roughly constant. Some recent results (with many coauthors) contain analysis which involves the k -core (largest subgraph of minimum degree at least k).

SUSAN H. XU
The Pennsylvania State University

Service performance analysis and improvement for a ticket queue with balking customers

This paper considers a queue managed by numbered tickets: each customer is at arrival issued a numbered ticket while the number being served is broadcast on display. In the ticket queue customers estimate their expected waiting time based on the difference between the ticket number issued and the number being served and would desert the queue if the difference exceeds a tolerance level. We construct a Markov chain model of the ticket queue and develop a two-step procedure to compute its steady state distribution. As the exact computation is exponentially time and memory space consuming, we propose an approximation to the ticket queue by eliminating the states with negligible steady

state probabilities. A polynomial time algorithm is designed to compute the steady state distribution for the approximate system and shown to be highly effective. We then benchmark the balking probability of the ticket queue against that of its counterpart physical queue in which customers have full information of the queueing position. We show that the two systems would have significantly different balking probabilities when customers' patience level is low and the traffic intensity is high. We also propose an improvement to the ticket queue by providing each customer with his expected waiting time conditioned on his observed number difference; the improvement is shown to raise the performance of the ticket queue on par with that of the physical queue.

Contributed talks

LI DONG

Carleton University

Combinatorial Decomposable Structures with Restricted Pattern

We study generalizations of the cycle index of the symmetric group (also called permutation pattern) to other decomposable structures like polynomials over finite fields, 2-regular graphs, random mappings and so on. The first generalization is to restricted patterns, that is, when part of the pattern is fixed. We then consider restricted patterns with the additional condition that the r th smallest (or largest) size of a component be larger (or smaller) than m , $1 \leq m \leq n$. We provide generating functions in both the labelled and the unlabelled case for these generalizations. We also present some asymptotic results.

MOHAMED HAMDOUNI

Ecole Polytechnique de Montreal

Parking buses in a depot with stochastic arrival times

Given buses of different types arriving at a depot during the evening, the bus parking problem consists of assigning these buses to parking slots in such a way that they can be dispatched adequately to the next morning routes without moving them between their arrivals and departures. In practice, the bus arrival times deviate stochastically from the planned schedule. In this paper, we introduce for this problem two solution approaches that produce solutions which are robust to variations in the arrival times. The first approach considers that each arrival can deviate from its planned arrival order (sooner or later) by at most k positions, where k is a predefined parameter. In the second approach, the objective aims at minimizing the expectation of a function positively correlated with the number of buses that make the planned solution infeasible because they arrive too late or too early. In both approaches, the problem is modeled as an integer linear program that can be solved by a commercial mip solver. Computational results obtained on instances derived from a real-world dataset are reported.

CARLOS HOPPEN
University of Waterloo

Bounding the size of an induced forest in a graph with large girth

In the present talk, I shall discuss the problem of finding a large induced forest, that is, a large acyclic induced subgraph of a given graph. A simple randomized algorithm is devised for this purpose, and its output provides a lower bound on the maximum number of vertices in an induced forest of a regular graph. When the girth is large and the degree is at least 4, this bound coincides with the best bound known to hold asymptotically almost surely for random regular graphs.

GRAEME KEMKES
University of Waterloo

Colouring random regular graphs

The chromatic number of a graph is the smallest number of colours needed to colour the vertices so that adjacent vertices receive different colours. In the uniform probability space of d -regular graphs on n vertices, what can we say about the chromatic number, with high probability? This presentation will describe recent work in this area, including probabilistic tools such as the small subgraph conditioning method. We will also see a new result about the chromatic number of the random 5-regular graph.

GANG LI
Carleton University

Variance reduction and long-range dependence preservation by leaky bucket

Consider a leaky bucket $L(\mu, b)$ with stationary input process A and stationary output process D . We show that D is less variable than A in both time domain and count domain and the property of long range dependence of the input process, if any, is preserved by a leaky bucket.

JUN LI
Carleton University

Mean Value Analysis of Resequencing Delay in a Discrete-time Queueing Model

Resequencing, which ensures received objects resequenced in an appropriate order before they are fed to the output of a system, is one kind of synchronization mechanisms in modeling telecommunication systems and parallel and distributed systems. Motivated by the facts of retransmission and in-sequence delivery of data packets in some telecommunication network protocols, one discrete-time disordering queueing model is considered, and we are interested in the mean value of resequencing delay of a packet in equilibrium regime. The stability condition and a conditional probability used to evaluate a packet's resequencing delay are discussed.

CONRADO MARTINEZ
University Politecnica de Catalunya

On the Variance of Quickselect

Quickselect with median-of-three is routinely used as the method of choice for selection of the m th element out of n in general-purpose libraries such as the C++ Standard Template Library. Its average behavior is fairly well understood and has been shown to outperform that of the standard variant, which chooses a random pivot on each stage. However, no results were previously known about the variance of the median-of-three variant, other than for the number of comparisons made when the rank m of the sought element is given by a uniform random variable. Here, we consider the variance of the number of comparisons made by quickselect with median-of-three and other quickselect variants when selecting the m th element for $m/n \rightarrow \alpha$ as $n \rightarrow \infty$. We also investigate the behavior of proportion-from- s sampling as $s \rightarrow \infty$.

DANIEL PANARIO
Carleton University

Degree Distribution of the Greatest Common Divisor of Polynomials over \mathbb{F}_q

We study the degree distribution of the greatest common divisor of two or more random polynomials over a finite field \mathbb{F}_q . We provide estimates for several parameters like number of distinct common irreducible factors, number of irreducible factors counting repetitions and total degree of the gcd of two or more polynomials. We show that the limiting distribution of a random variable counting the total degree of the gcd is geometric, and that the distributions of random variables counting the number of common factors (with and without repetitions) are very close to Poisson distributions when q is large.

ALOIS PANHOLZER

Vienna University of Technology, Institute of Discrete Mathematics and Geometry

Recent results on the analysis of increasing trees

Several important tree models, e.g., recursive trees, plane-oriented recursive trees and binary increasing trees are members of the family of increasing trees. These tree models turned out to be appropriate in order to describe the behaviour of a lot of quantities in various applications, e.g., they are used to describe the spread of epidemics, for pyramid schemes, and they are used as a simplified growth model of the world wide web, since plane-oriented recursive trees are a special instance of the so called Albert-Barabási model for scale-free networks.

Increasing trees can be considered as labeled trees, where the nodes of a tree of size n are labeled by distinct integers of the set $\{1, \dots, n\}$ in such a way that each sequence of labels along any path starting at the root is increasing. It is well known that certain members of increasing tree families can be described via a tree evolution process, i.e., with a probabilistic growth rule. This property of a tree family gives access to several probabilistic techniques, like using Pólya-Eggenberger urn models or translating results from continuous time branching processes, and is thus a very useful property of a tree structure. A lot of previous analyzes of increasing tree families are using this property. However only for very few members of increasing tree families a description via a probabilistic growth rule was known, and furthermore a characterization of tree families with this property was missing. Also by means of methods mentioned above there are obtained relatively few results that give insight into the behavior of the node j (= the j -th individual) during the growth process in a tree of size n , in particular if the label $j = j(n)$ is growing with n .

Here we want to present recent progress on the analysis of increasing tree families. First we can completely characterize the class of increasing tree families that are growing, i.e., those increasing tree families which can be described via a tree evolution process. It turns out that for all such tree families the growth rule is quite simple. However the main topic is to present a method suitable for a distributional analysis of label-based parameters, which relies on a combinatorial description of increasing tree families. This combinatorial/analytic approach gives access to a quite detailed description of the behavior of such parameters of interest, as the degree of node j , the depth of node j , the size of the subtree of node j , the distance between two nodes j_1 and j_2 , etc. Dependent on the growth of $j = j(n)$ we can localize the phase changes (if appearing) of the quantities mentioned before and characterize the limiting distributions in the different phases.

Furthermore the combinatorial approach can be extended to obtain joint distributions for several quantities depending on node j and to derive joint distributions for several nodes j_1, j_2, \dots, j_r .

WEI XU

University of Ottawa

Comparative genomics and the distribution of the number of cycles and paths in fragmented random 3-regular graphs.

We represent a “circular” genome, such as those in bacteria or in mitochondria and chloroplasts, as a 2-regular Hamiltonian graph on $2n$ vertices. One of the perfect matchings (the black matching) in this graph connects the two ends of each gene (or chromosomal segment) in the genome. The other matching (the red matching) determines the adjacencies, i.e., the ordering, of the different genes (or segments) in the genome. The Hamiltonian condition ensures that the genome is made up of only one circular genome; in the more general, non-Hamiltonian case, there may be several such circles, sometimes called “plasmids” in the biological literature. We “fragment” the genome by deleting any m_r of the red edges from the graph, we obtain a representation of a genome consisting of m_r linear chromosomes. Without the Hamiltonian condition, this genome could also have some circular plasmids in addition to the m linear chromosomes.

In comparative genomics, we add a third perfect matching (blue edges), representing a different ordering of the same genes or segments, from which m_b edges are omitted if the genome is fragmented into $m_b > 0$ linear chromosomes.

For circular genomes, the number of cycles κ in the red-blue subgraph is well-known to be a key indicator of the similarity of the two genomes being compared, and we are interested in calculating the distribution of the number of cycles for random genomes. This would enable statistical testing of the biological relationship between two genomes.

It is known that in the case of unrestricted circular ($m_b = m_r = 0$) chromosomes, $\lim_{n \rightarrow \infty} E(\kappa) = \log(2) + \frac{1}{2}(\gamma + \log n)$, where γ is Euler’s constant. Moreover, the Kim-Wormald theorem on Hamiltonians in 2-regular subgraphs of random 3-regular graphs of may be applied to show that under mild conditions, this extends to the particularly interesting, Hamiltonian, case.

We are currently interested in genomes having linear chromosomes, where ($m_b > 0, m_r > 0$). Here, the red-blue subgraph contains both alternating colour cycles and alternating colour paths. The study of the paths is straightforward. For the cycles, in the general case (not necessarily Hamiltonian), we find that the limiting value of κ is $\frac{1}{2} \log \frac{n+m}{2m}$, where ($m = m_b = m_r$), based on precise evaluation of the recurrence

$$(2n + 2m)(2n + 2m - 1)\kappa(n, m) = 2m(2m - 1)\kappa(n, m - 1)(2) + ((2n + 2m)(2n + 2m - 1) - 2m(2m - 1))\kappa(n - 1, m) + 2n,$$

up to very large n , as well as simulations. For the moment, we have proved this limit analytically only up to an additive constant.