Mathematisches Forschungsinstitut Oberwolfach

Report No. 23/2011

DOI: 10.4171/OWR/2011/23

Mini-Workshop: Random Trees, Information and Algorithms

Organised by
Ralph Neininger (Frankfurt)
Wojciech Szpanowski (West Lafayette)

April 24th – April 30th, 2011

Abstract. The subject of this Mini-Workshop is the probabilistic analysis of random tree models that originate from applications in Computer Science. Emphasis is put on their connections to algorithms and information theory. Trees with a stochastic growth dynamic appear in Computer Science as data structures, in the context of coding schemes as well as connected to fundamental algorithms such as sorting, searching and selecting. The focus of this Mini-Workshop is on probabilistic and analytic techniques that have been developed recently in the asymptotic analysis of random trees such as martingale methods, connections to branching random walks, the contraction method, the method of moments as well as various techniques based on generating functions.

Mathematics Subject Classification (2000): 60F05, 68W40.

Introduction by the Organisers

In the asymptotic analysis of random tree models that originate from Computer Science diverse probabilistic and analytic techniques have been developed in the last decades with a strongly increasing interest during the last years. The techniques being developed include methods based on martingales, connections to branching random walks, the contraction method, techniques using generating functions, and the method of moments. Classically, in Computer Science random trees appear in the performance analysis of data structures, in the context of coding schemes as well as connected to fundamental algorithms such as sorting, searching and selecting. However, in the last years also fascinating connections of these random tree models to coalescent processes, fragmentation theory and

other combinatorial stochastic processes have been found. The aims of this Mini-Workshop are a deeper understanding and advances in the probabilistic analysis of random tree models as well as to discover their connections beyond the realm of models motivated from Computer Science.

The topics discussed at the workshop are briefly summarized as follows: For the model of simply generated trees Svante Janson discussed convergence of the size n trees to limit objects in a topology corresponding to convergence of all out-degrees. In particular, he discussed the cases where there is a representation of the simply generated tree as conditioned subcritical Galton-Watson tree or there is no representation as a conditioned Galton-Watson tree, the generating function associated having zero radius of convergence. Louigi Addario-Berry considered the problem of cutting down such simply generated trees (for finite variance offspring distributions) and its connection to the distance between two independent, uniformly chosen nodes in the tree by a coupling method. Here, the Rayleigh distribution appears in a $\sqrt{n}$ scaling and connections to Brownian excursions are found. Christina Goldschmidt also discussed the cutting of random trees to isolate the root. However, in contrast to the $\sqrt{n}$-height simply generated trees she considered a $\log n$-height tree, the random recursive tree. The stochastic process describing the cutting procedure on the set of partitions of $\{1, \dots, n\}$ turns out to be the Bolthausen-Sznitman coalescent. Also asymptotic frequencies of blocks were discussed and the open problem was raised to find other coalescents that can be represented by cutting down a combinatorial tree.

An important tree structure to store bit-strings are tries. In the talk of Wojciech Szpankowski tries were considered under the symmetric and asymmetric Bernoulli model (which describe memoryless sources). The quantities under consideration are the internal and external profile of tries. Results on asymptotic expectations with their phase changes were presented together with asymptotic variances and limit laws. In the talk of Mark Ward the internal profile of tries was considered under a different probabilistic model that makes the trie a suffix tree, another important data structure in applications. A related quantity is the subword complexity of the given string; the approach is based on generating functions.

A couple of talks discussed the analysis of search trees, mainly by the use of the contraction method. Ludger Rüschendorf presented results on depths and various distance measures of the weighted b-ary tree together with applications to special kinds of random trees that are covered by the class of weighted b-ary trees. Uwe Rösler discussed a functional limit law for a process version associated to the Quicksort algorithm: the algorithm on size n input always first recursively sorts the left sublist generated and stops once the smallest $\ell \in \{1, \dots, n\}$ items are sorted. The number of key comparisons as a process in ℓ is considered asymptotically in n. Ralph Neininger developed the contraction method on the spaces $C[0,1]$ and $D[0,1]$ of continuous resp. cádlág functions with the supremum norm by use of the Zolotarev metric. This leads to a framework that covers as an application Donsker's invariance principle. Another application was presented by

Henning Sulzbach. He showed a functional limit law for the complexity of partial match queries of the form $(s, *)$, $s \in [0, 1]$ in random two-dimensional (point) quadtrees, the process being in s. As corollaries, open questions regarding the variance and limit law for uniform queries and the order of worst case queries could be solved. The approach also covers the k-d trees for $k = 2$. Related to the contraction method Gerold Alsmeyer characterized the set of solutions of general smoothing transforms. By asking for fixed points of a functional equation under various constraints on the solutions he covers recursive distributional equations of sum and max type. The relation between solutions of homogeneous and inhomogeneous equations was also discussed as well as the phenomenon of endogeneity that plays a role in cases also important in applications. Rudolf Grübel viewed the stochastic evolution of random (search) trees as a transient Markov chain and used connections to discrete potential theory. He discussed almost sure convergence of the normalized trees to limiting random measures. Also a metric on trees based on subtree sizes was introduced and the resulting limiting metric space was explored.

Nicolas Broutin considered algorithms to resolve collisions in communication of multiple users on one broadcast channel. He studied protocols for which the execution of the algorithms can be represented by a tree. The main focus was on the stability of the protocols, analyzed via the long-term behavior of an associated conditioned Markov chain.

Talks with geometrically motivated topics started with Hsien-Kuei Hwang discussing different notions of dominance in random point sets in space. He studied threshold phenomena and uniform estimates for the expected number of such points among n iid. points in d-dimensional cubes and simplices in various asymptotic settings. Yuliy Baryshnikov discussed bounding the unimodal category of functions, in particular addressed the case of a simple random function in dimension 1, the uniformly chosen random Dyck path of length $2n$, for which a limit law was derived. Gábor Lugosi presented a random geometric graph model on high-dimensional spheres that is motivated from a statistical hypothesis testing problem. In particular the clique number and its dependence on the dimension of the spheres were addressed.

There was a special session on Wednesday morning dedicated to the memory of Philippe Flajolet who passed away about a month before the workshop. Hsien-Kuei Hwang gave a survey on the subjects of Philippe Flajolet's research continued by the talk of Wojciech Szpankowski.