Probabilistic Combinatorics: Recent Progress and New Frontiers

Noga Alon (Tel Aviv University)
Bruce Reed (McGill University, CNRS)
Benny Sudakov (Princeton University)
Van Vu (University of California, San Diego)

October 29 – November 3, 2005

1 Overview

Probabilistic Combinatorics is an interface between Probability and Discrete Mathematics. Initiated by P. Erdős over fifty years ago, it has now become one of the fastest developing areas in all mathematics, with fascinating applications to many other important areas, such as Theoretical Computer Science and Statistical Physics. Roughly speaking, Probabilistic Combinatorics comprises three main topics, for each of which we give a short description. Naturally, there are considerable overlaps between these topics.

The first topic is the application of probability to solve combinatorial problems, and conversely the application of combinatorial methods to prove results in probability theory. Typical examples of the former are the "existence" proofs of Erdős. In general, one wants to show the existence of certain objects by generating an appropriate probabilistic space and proving that the desired object has positive measure in this space. The last twenty years or so have witnessed significant progress in this approach. The development of new and powerful techniques, such as the semi-random method and various sharp concentration inequalities, has enabled researchers to attack many famous open problems, considered intractable not so long ago, with considerable success. Furthermore, many new ideas discovered in this process have turned out to be useful for problems from different areas. For instance, the recent Galvin-Kahn result on Gibb's measures has its roots in an earlier graph colouring result of Kahn. For an example of combinatorics being used in the field of probability, one can look at some recent work of Louigi Addario-Berry and Bruce Reed, which uses combinatorial techniques to bound the point at which a random walk first returns to zero.

The second topic is the study of random combinatorial structures, such as random graphs. The typical question here is to show that at a given density, a random graph has a desired property with very high probability. The study of random graphs has recently received a major boost from industry. It has been discovered that various important real-life graphs (such as the Internet) can be modeled as a random graph of a special type. If one can analyze these graphs, then one can make predictions about the evolution of the real-life networks.

The third topic is the study of randomized algorithms. Here the main question is either to design randomized algorithms for a certain goal or to analyze natural algorithms given special inputs. While this topic can also be considered as a topic in Computer Science, it has turned out quite recently that it also has much to do with Statistical Physics. For instance, there is a natural algorithm (motivated by problems from statistical

physics) for generating a random colouring of a graph. A tantalizing question is to know when this algorithm runs in polynomial time, and a proper bound would have amazing consequences in Physics.

The focus of the workshop lay specifically in the above three main research topics of Probabilistic Combinatorics. One aim of the workshop was simply to foster interaction and collaboration between researchers in these fields, and to discuss recent progress and communicate new results and ideas. To mention an example, the following conjecture of Louigi Addario-Berry (see [1]), communicated during an open problem session, was solved at the workshop by Jacques Verstraete using the technique of combinatorial nullstellensatz:

Theorem 1.1 *Given a graph $G = (V, E)$ and, for every $v \in V$, a list $D_v \subseteq \{0, 1, \ldots, d(v)\}$ satisfying $|D_v| > \lceil d(v)/2 \rceil$, there is a spanning subgraph $H \subseteq G$ such that for all v, $d_H(v) \in D_v$.*

Additionally, this forum was an opportunity to make state-of-the-art probabilistic techniques available to a broader audience, in particular graduate students.

With the rapid development in recent years of probabilistic techniques and their applications to various mathematical disciplines, the workshop was a key opportunity to bring together researchers representing the entire spectrum of Probabilistic Combinatorics, so as to consolidate our knowledge at present and set new horizons for future discoveries.

In the remainder of the report we describe in detail some of the advances presented at the workshop.

2 The Erdős-Rényi Random Graph

Joel Spencer - *Connectedness of $G(n, p)$*

I gave a talk on The Probability of Connectedness, the result being an asymptotic formula for the probability that the random graph Gn, p is connected, for the entire range of p. The key to it is a new analysis of breadth first search over the random graph Gn, p. This is an idea I have been working on for a year or so but it really came together during the workshop. I have given talks on this general topic before, most recently at the CMS Annual Meeting in Waterloo in June, but at this workshop the ideas were clearer than before.

The asymptotic probability of $G(n, p)$ being connected is $A_1 A_2$, with

$$A_1 = A_1(n, p) = (1 - (1 - p)^n)^{n-1}$$

$$A_2 = A_2(n, p) \sim \begin{cases} 1 & \text{for } p \gg n^{-1} \\ 1 - (c+1)e^{-c} & \text{for } p \sim cn^{-1} \\ \frac{1}{2}\epsilon^2 & \text{for } p \sim \epsilon n^{-1} \text{ and } n^{-1/2} \ll \epsilon = o(1) \\ \text{complicated} & \text{for } p \sim cn^{-3/2} \\ n^{-1} & \text{for } 0 < p \ll n^{-3/2} \end{cases}$$

(Note that the probability that there are no isolated vertices if the events of being isolated were independent would be $(1 - (1 - p)^{n-1})^n$ which is quite close.)

When $p \ll n^{-3/2}$ it is simpler to write that the probability of $G(n, p)$ being connected is roughly the probability that $G(n, p)$ is precisely a tree, which is $n^{n-2} p^{n-1} (1 - p)^{m-(n-1)}$ with $m = \binom{n}{2}$.

When $p \sim cn^{-3/2}$ let B be the probability $G(n, p)$ is precisely a tree. Then $G(n, p)$ is a tree plus l edges with probability $Bc_l c^{3l/2}$ where the c_l are the "Wright constants". Convergence occurs and the probability that $G(n, p)$ is a tree is $B \sum_{i=0}^{\infty} c_l c^{3l/2}$.

The arrangements were excellent, giving myself and the others plenty of time to "prove and conjecture."

Louigi Addario-Berry - *The Diameter of the Minimum Weight Spanning Tree*

Given a connected graph $G = (V, E)$, $E = \{e_1, \ldots, e_{|E|}\}$, together with edge weights $W = \{w(e) | e \in E\}$, a minimum weight spanning tree of G is a spanning tree $T = (V, E')$ that minimizes

$$\sum_{e \in E'} w(e).$$

If the edge weights are distinct then this tree is unique; in this case we denote it by MWST(G). Minimum spanning trees are at the heart of many combinatorial optimization problems. In particular, they are easy to compute, and may be used to approximate hard problems such as the minimum weight traveling salesman tour. As a consequence, much attention has been given to studying their structure, especially in random settings and under various models of randomness. For instance, Frieze determined the weight of a the MWST of a complete graph whose edges have been weighted by independent and identically distributed (i.i.d.) $[0,1]$-random variables. This result has been reproved and generalized by Frieze and McDiarmid [8] and Aldous [2]. Under the same model, Aldous derived the degree distribution of the MWST. Both these results rely on local properties of minimum spanning trees. We are interested in their global structure.

The *distance* between vertices x and y in a graph H is the length of the shortest path from x to y. The *diameter* diam(H) of a connected graph H is the greatest distance between any two vertices in H. We are interested in the diameters of the minimum weight spanning trees of a clique K_n on n vertices whose edges have been assigned i.i.d. real weights. We use $w(e)$ to denote the weight of e. In Banff we presented our proof of the following theorem, answering a question of Frieze and McDiarmid [9].

Theorem 2.1 *Let $K_n = (V, E)$ be the complete graph on n vertices, and let $\{X_e | e \in E\}$ be independent identically distributed edge-weights. Then conditional upon the event that for all $e \neq f$, $X_e \neq X_f$, it is the case that the expected value of the diameter of MWST(K_n) is $\Theta(n^{1/3})$.*

Benny Sudakov - *Embedding Nearly-Spanning Bounded Degree Trees*

In this talk we describe a sufficient condition for a sparse graph G to contain a copy of every nearly-spanning tree T of bounded maximum degree, in terms of the expansion properties of G. The restriction on the degree of T comes naturally from the fact that we consider graphs of constant degree. Two important examples where our condition applies are random graphs and graphs with a large spectral gap.

The problem of existence of large trees with specified shape in random graphs has a long history starting with conjecture of Erdős that a random graph $G(n, c/n)$ almost surely contains a path of length at least $(1 - \alpha(c))n$, where $\alpha(c)$ is a constant smaller than one for all $c > 1$ and $\lim_{c\to\infty} \alpha(c) = 0$. The question of existence of large trees of bounded degree other than paths in sparse random graphs was studied by de la Vega. He proved that for sufficiently large c one can almost surely embed in $G(n, c/n)$ any tree with maximum degree at most d that occupies a small constant proportion of the random graph. Our first result improves the result of Fernandez de la Vega and generalizes several results on the existence of long paths. It shows that the sparse random graph contains almost surely every nearly-spanning tree of bounded degree, i.e., tree of size $(1 - \epsilon)n$.

For a graph G let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of its adjacency matrix. The quantity $\lambda(G) = \max_{i\geq 2} |\lambda_i|$ is called the *second eigenvalue* of G. A graph $G = (V, E)$ is called an (n, D, λ)-*graph* if it is D-regular, has n vertices and the second eigenvalue of G is at most λ. It is well known that if λ is much smaller than the degree D, then G has strong expansion properties, so the ratio D/λ could serve as some kind of measure of expansion of G. Our second result shows that an (n, D, λ)-graph G with large enough spectral gap D/λ contains a copy of every nearly-spanning tree with bounded degree. This extends a result of Friedman and Pippenger [7].

3 Regular Graphs

Nicholas Wormald - *Large Independent Sets in Regular Graphs of Large Girth*

An *independent set* I of a graph G is a subset of the vertices of G such that no two vertices of I are joined by an edge. The *independence number* of G is the cardinality of a maximum independent set, and is denoted by $\alpha(G)$. The *girth* of G is the length of its shortest cycle.

In 1991, Shearer gave the best known lower bounds on $\alpha(G)$ for G with given maximum degree and large girth. For instance, if G is 3-regular with n vertices, Shearer's results imply that $\alpha(G) \geq \frac{125}{302}n$ provided the girth is sufficiently large, and he gave other results for graphs of maximum degree d in terms of $f(d)$ where the function f is defined iteratively.

It is known that looking at graphs with maximum degree d for such problems is equivalent to looking at d-regular graphs. In 1995, the speaker analyzed two greedy algorithms which give rise to large independent sets

in random regular graphs, one simple and one more sophisticated. With Joe Lauer, we recently studied the simple greedy algorithm, applied to large girth graphs, and established a result for all regular graphs of large girth, that coincides with the corresponding result for random graphs. We use a "nibble"-type approach but require none of the sophistication of the usual nibble method arguments, using only linearity of expectation. We obtained the following result.

Theorem 3.1 *For all $d \geq 3$, the independence number of a graph with n vertices, maximum degree d and girth g is at least*

$$\big(1-\varepsilon(g)\big)\frac{n}{2}\left(1-(d-1)^{-2/(d-2)}\right),$$

where $\varepsilon(g) \to 0$ as $g \to \infty$.

This improves Shearer's result for all $d \geq 7$.

More recently, with Mohammad Salavatipour, we have analyzed the more sophisticated greedy algorithm mentioned above. The results are stronger but are given in terms of the solutions of differential equations which have only been solved numerically. With Carlos Hoppen we have examined algorithms for finding large induced forests in graphs with bounded degree and large girth. It is believed that, in all cases, the constants obtained for regular graphs of large girth coincide with those already known for random regular graphs.

It was known that, given such a bound for regular graphs of arbitrarily large girth, the same bound carries over to an asymptotic bound for random regular graphs. The current work indicates that for many problems with results on random regular graphs obtained by analyzing greedy algorithms the results can be "explained" in this way, despite the fact that they were first proved directly in the random case. It is not known to what extent this is a general phenomenon. In particular, it is not known if all 4-regular graphs with sufficiently large girth are 3-colourable.

Angelika Steger - *A Probabilistic Counting Lemma for Sparse Regular Graphs*

This is joint work with S. Gerke and M. Marciniszyn.

Over the last decades Szemerédi's regularity lemma [18] has proven to be a very powerful tool in modern graph theory. Unfortunately, in its original setting it only gives nontrivial results for dense graphs, that is graphs with $\Theta(n^2)$ edges. In 1996 Kohayakawa [14] and independently Rödl introduced a variant which holds for sparse graphs, provided they satisfy some additional structural conditions (which essentially mean that the graph does not contain regions that are too dense). However, using this sparse regularity lemma to prove e.g. extremal and Ramsey type results similar to the known results in the dense case requires as an additional step: the existence of appropriate embedding or counting lemmas. For the sparse case this missing step has been formulated as a conjecture by Kohayakawa, Łuczak and Rödl [15]. For a graph H, let $\mathcal{G}(H,n,m)$ be the family of graphs on vertex set $V = \bigcup_{x\in V(H)} V_x$, where the sets V_x are pairwise disjoint sets of vertices of size n, and edge set $E = \bigcup_{\{x,y\}\in E(H)} E_{xy}$, where $E_{xy} \subseteq V_x \times V_y$ and $|E_{xy}| = m$. Let $\mathcal{G}(H,n,m,\varepsilon) \subseteq \mathcal{G}(H,n,m)$ denote the set of graphs in $\mathcal{G}(H,n,m)$ satisfying that each $(V_x \cup V_y, E_{xy})$ is an (ε)-regular graph.

Conjecture 3.2 (KŁR Conjecture [15]) *Let H be a fixed graph and define*

$$\mathcal{F}(H,n,m) = \{G \in \mathcal{G}(H,n,m) : H \text{ is not a subgraph of } G\}.$$

For any $\beta > 0$, there exist constants $\varepsilon_0 > 0$, $C > 0$, $n_0 > 0$ such that for all $m \geq Cn^{2-1/d_2(H)}$, $n \geq n_0$, and $0 < \varepsilon \leq \varepsilon_0$,

$$|\mathcal{F}(H,n,m) \cap \mathcal{G}(H,n,m,\varepsilon)| \leq \beta^m \binom{n^2}{m}^{|E(H)|},$$

where $d_2(H) = \max\left\{\frac{|E(F)|-1}{|V(F)|-2} : F \subseteq H, |V(F)| \geq 3\right\}$.

One of the key difficulties in the proof of the KŁR Conjecture is the fact that for $m = o(n^2)$ the size of a neighbourhood of a vertex is on average $o(n)$. The definition of regularity, however, only deals with linear

sized subsets and thus regularity seems not to be inherited by subgraphs induced on the neighbourhoods of some vertices. In a joint paper [10] with Gerke, Kohayakawa, and Rödl we were recently able to prove that nevertheless in the sparse case a hereditary version holds as well, at least in the probabilistic setting. This result readily implies much shorter and more elegant proofs of the results known so far, namely the case of cycles C_k for all $k \geq 3$ and for $H = K_4$ and K_5. In this talk we show that in fact a much stronger property holds. Namely, small sets not only inherit with high probability the regularity property, but they also satisfy with high probability all properties that regular tuples satisfy with high probability. This allows us to show that the KŁR Conjecture holds for all complete graphs for slightly larger number of edges than the conjectured value. In return, we can show the existence of many copies instead of just one copy. That is, we get a so-called counting lemma.

Theorem 3.3 ([11]) *For all $\ell \geq 3$, $\delta > 0$, and $\beta > 0$, there exist constants $n_0 \in \mathbb{N}$, $C > 0$, and $\varepsilon > 0$ such that*

$$|\mathcal{F}(K_\ell, n, m, \delta) \cap \mathcal{G}(K_\ell, n, m, \varepsilon)| \leq \beta^m \cdot \binom{n^2}{m}^{\binom{\ell}{2}}$$

provided that $m \geq Cn^{2-1/(\ell-1)}$, $n \geq n_0$, and $0 < \varepsilon \leq \varepsilon_0$ and where $\mathcal{F}(K_\ell, n, m, \delta)$ denotes the family of graphs in $\mathcal{G}(K_\ell, n, m)$ that contain less than $(1-\delta)n^{|V(H)|}(\frac{m}{n^2})^{|E(H)|}$ copies of H.

4 Graph Colouring

Andrew King - *Advances Towards Reed's Conjecture*

My current research includes several problems: partial results towards Reed's Conjecture, probabilistic colouring work to similar ends, and the reconciliation of probabilistic models via rapidly-mixing Markov chains.

Reed's Conjecture states that for any graph G, $\chi(G) \leq \lceil (1/2)(\Delta(G) + 1 + \omega(G)) \rceil$ [19]. Generally speaking, there are two ways to work towards this result. The first involves proving it outright for certain classes of graphs, and the second involves proving that it is not far from the truth. That is, $\chi(G) \leq \lceil (1/2 + o(1))(\Delta(G) + 1 + \omega(G)) \rceil$, meaning that $\chi(G) \leq \lceil (1/2 + f(\Delta(G)))(\Delta(G) + 1 + \omega(G)) \rceil$ where f tends to 0 as Δ tends to infinity. There are partial results of this flavour, and I am working towards broadening this body of work as well as finding ways to colour graphs with few colours in polynomial time.

Since the workshop, Bruce Reed and I have proved that Reed's Conjecture holds for quasi-line graphs, improving upon a result of Chudnovsky and Ovetsky [3]. Furthermore, for these graphs a colouring using at most $\lceil (1/2)(\Delta(G) + 1 + \omega(G)) \rceil$ colours can be found in polynomial time.

5 Pseudorandom Graphs

Yoshiharu Kohayakawa - *Turán's Theorem for Pseudorandom Graphs*

This is joint work with V. Rödl (Emory University), M. Schacht (Humboldt-Universität zu Berlin), P. Sissokho (Illinois State University), and J. Skokan (Universidade de São Paulo).

The *generalized Turán number* $ex(G, H)$ of two graphs G and H is the maximal number of edges in a subgraph of G not containing H. If G is the complete graph K_n on n vertices, then, by the Erdős–Stone–Simonovits theorem, we have $\mathrm{ex}(K_n, H) = \Big(1 - 1/(\chi(H) - 1) + o(1)\Big)\binom{n}{2}$, where $o(1) \to 0$ as $n \to \infty$.

We give an analogous result for triangle-free graphs H and pseudorandom graphs G. Our concept of pseudorandomness is inspired by the *jumbled* graphs introduced by A. Thomason. We say that a graph G is (q, α)-*bijumbled* if

$$\Big| e_G(X, Y) - q|X||Y| \Big| \leq \alpha\sqrt{|X||Y|}$$

for every pair of sets $X, Y \subset V(G)$, where $e_G(X, Y)$ denotes the number of pairs $(x, y) \in X \times Y$ with $xy \in E(G)$.

For simplicity, here we only state a consequence of our main result: for any triangle-free graph H with maximum degree Δ and for any $\delta > 0$, there exists $\gamma > 0$ such that any large enough n-vertex, $(q, \gamma q^{\Delta+1/2}n)$-bijumbled graph G satisfies

$$\mathrm{ex}(G,H) \le \left(1 - \frac{1}{\chi(H)-1} + \delta\right)|E(G)|.$$

Jan Vondrák - *2-Colourability of Randomly Perturbed Hypergraphs*

This is joint work with Benny Sudakov.

In the classical Erdős-Rényi model, a random graph is generated by starting from an empty graph and then adding a certain number of random edges. More recently, Bohman, Frieze and Martin considered a generalized model where one starts with a fixed graph $G = (V, E)$ and then inserts a collection R of additional random edges. We denote the resulting random graph by $G + R$. The initial graph G can be regarded as given by an adversary, while the random perturbation R represents noise or uncertainty, independent of the initial choice. This scenario is analogous to the *smoothed analysis* of algorithms proposed by Spielman and Teng, where an algorithm is assumed to run on the worst-case input, modified by a small random perturbation.

In subsequent work, Krivelevich, Sudakov and Tetali [16] considered random formulas obtained by adding random k-clauses (disjunctions of k literals) to a fixed k-SAT formula. They proved that for any formula with at least $n^{k-\epsilon}$ k-clauses, adding $\omega(n^{k\epsilon})$ random clauses of size k makes the formula almost surely unsatisfiable. This is tight, since there is a k-SAT formula with $n^{k-\epsilon}$ clauses which almost surely remains satisfiable after adding $o(n^{k\epsilon})$ random clauses. A related question, which was raised in this paper, is to find a threshold for non-2-colourability of a random hypergraph obtained by adding random edges to a large hypergraph of a given density.

While 2-colourability of graphs is well understood, being equivalent to non-existence of odd cycles, for k-uniform hypergraphs with $k \ge 3$ it is already NP-complete to decide whether a 2-colouring exists. Consequently, there is no efficient characterization of 2-colourable hypergraphs. The problem of 2-colourability of random k-uniform hypergraphs for $k \ge 3$ was first studied by Alon and Spencer. Recently, the threshold for 2-colourability has been determined very precisely. Achlioptas and Moore proved that the number of edges for which a random k-uniform hypergraph becomes almost surely non-2-colourable is $(2^{k-1}\ln 2 - O(1))n$. Interestingly, the threshold for non-2-colourability is roughly one half of the threshold for k-SAT. Achlioptas and Peres proved that a formula with m random k-clauses becomes almost surely unsatisfiable for $m = (2^k \ln 2 - O(k))n$. The two problems seem to be intimately related and it is natural to ask what is their relationship in the case of a random perturbation of a fixed instance.

The proof of Krivelevich et al. (for randomly perturbed k-SAT) also yields that for any k-uniform hypergraph H with $n^{k-\epsilon}$ edges, adding $\omega(n^{k\epsilon})$ random edges destroys 2-colourability almost surely. Nonetheless, it turns out that this is not the right answer. It is enough to use substantially fewer random edges to destroy 2-colourability: roughly a square root of the number of random clauses necessary to destroy satisfiability. Our main result is that for any k-uniform hypergraph with $\Omega(n^{k-\epsilon})$ edges, adding $\omega(n^{k\epsilon/2})$ random edges makes it almost surely non-2-colourable. This is almost tight in the sense that adding $o(n^{k\epsilon/2})$ random edges is not sufficient in general.

6 First Order Graph Properties

Oleg Pikhurko - *First Order Graph Properties*

Graph properties expressible in first order logic were studied. The vocabulary consists of variables, connectives ($\vee$, $\wedge$ and $\neg$), quantifiers ($\exists$ and $\forall$), and two binary relations: the equality and the graph adjacency ($=$ and $\sim$ respectively). The variables denote vertices only so we are not allowed to quantify over sets or relations. The notation $G \models A$ means that a graph G is a model for a *sentence* A (a first order formula without free variables); in other words, A is true for the graph G.

A first order sentence A *defines* G if G is the unique (up to an isomorphism) finite model for A. The *quantifier depth* (or simply *depth*) $D(A)$ is the largest number of nested quantifiers in A. This parameter is closely related to the complexity of checking whether $G \models A$. Let $D(G)$ be the smallest quantifier depth of a first order formula defining G.

In a sense, a defining formula A can be viewed as the canonical form for G (except that A is not unique): in order to check whether $G \cong H$ it suffices to check whether $H \models A$. Unfortunately this approach does

not seem to lead to better isomorphism algorithms, but this notion, being on the borderline of combinatorics, logic and computer science, is interesting on its own and might yield unforeseen applications.

Recently, various results on the values of $D(G)$ for order-n graphs appeared. The paper of Pikhurko, Veith and Verbitsky studied the maximum of $D(G)$ (the 'worst' case). The 'best' case is considered by Pikhurko, Spencer, and Verbitsky, while Kim, Pikhurko, Spencer and Verbitsky obtained various results for the random graph $G(n,p)$.

Pikhurko presented new results for random sparse structures obtained jointly with Bohman, Frieze, Łuczak, Smyth, Spencer, and Verbitsky. Specifically, it was proved that almost surely

- $D(G) = \Theta(\frac{\ln n}{\ln \ln n})$, where G is the giant component of a random graph $G(n, \frac{c}{n})$ with constant $c > 1$;
- $D(T) = (1 + o(1))\frac{\ln n}{\ln \ln n}$ where T is a random tree of order n.

These results rely on computing the maximum of $D(T)$ for a tree T of order n and maximum degree l, so this problem was studied as well.

7 Combinatorial Games

Thomas Bohman - *Making and Breaking the Giant Component*

I presented the following results at the workshop. We consider a game that can be viewed as a random graph process. The game has two players and begins with the empty graph on a set of n vertices. During each turn a pair of random edges is generated and one of the players chooses one of these edges to be an edge in the graph. Thus the players guide the evolution of the graph as the game is played. One player controls the even rounds with the goal of creating a so-called giant component as quickly as possible. The other player controls the odd rounds and has the goal of keeping the giant from forming for as long as possible. We show that the product rule is an asymptotically optimal strategy for both players. (The product rule chooses between two edges by comparing the products of the sizes of the components joined. For example, the player who is trying to create a giant component would choose the edge that maximized the product of the sizes of the components joined.)

8 Geometric Problems

Imre Bárány - *On the Randomized Integer Convex Hull*

This is joint work with J. Matoušek.

Assume $K \subset R^d$ is a convex body. Its integer convex hull is, by definition, the convex hull of $K \cap Z^d$ where Z^d is the usual integer lattice. Notation: $I(K) = \text{conv}(K \cap Z^d)$. The integer convex hull is of central interest in integer programming. Define the lattice $L_{\rho,t} = \rho(Z^d + t)$ where $t \in [0,1)^D$ and $\rho \in SO(d)$, which is an isometric copy of Z^d. The set of lattices $\mathcal{L} = \{L_{\rho,t}\}$ is a probability space with probability measure equal to the product of the Lebesgue measure on $[0,1)^d$ and the Haar measure on $SO(d)$. The randomized integer convex hull is $I_L(K) = \text{conv}(K \cap L)$, where L is a random element of $\mathcal{L}$. $I_L(K)$ is a polytope.

Motivated by integer programming, we estimate the expected number of vertices of $I_L(K)$, and also the expected missed volume, that is, the expectation of $\text{vol}(K \setminus I_L(K))$. One of our results says that the expected number of vertices of $I_L(K)$ is of order $(\text{vol}(K))^{(d-1)/(d+1)}$ when K is smooth, and is of order $(\log \text{vol}(K))^{d-1}$ when K is a polytope. The expected missed volume problem leads to the following question which is a distant relative of Buffon's needle problem. Given a convex body $K \subset R^d$, what is the probability that a randomly chosen congruent copy of K is lattice point free? We show that this probability (1) is always smaller than $c_1/\text{vol}(K)$ for c_1 constant, and (2) is larger than $c_2/\text{vol}(K)$ for c_2 constant if the width of K is small enough. The constants depend only on dimension.

Ross M. Richardson - *Random Inscribing Polytopes*

This is joint work with Van Vu and Lei Wu.

Let K be a compact convex body in $\mathbb{R}^d$. Choose n points uniformly in K. The convex hull of these n points is referred to as a *random polytope*. The study of random polytopes is the study of certain key

functionals of these polytopes; the volume of the random polytope and the number of $i-$dimensional faces are the most studied. There has been much recent progress in their characterization, and a broad range of techniques have arisen out of the intersection of geometry, probability, and combinatorics. A comprehensive survey by I. Bárány will soon appear in the volume *Stochastic Geometry*.

Now restrict K to have smooth boundary and everywhere positive Gaussian curvature. We define a new model of random polytopes where we now choose points on the boundary ∂K according to some positive continuous distribution. The convex hull of n points chosen in this manner is referred to as the *random inscribing polytope*.

Our work focuses on determining the distribution of the volume functional, which we denote by Z. We prove a concentration result of the following form:

$$P\left(|Z - EZ| \geq \sqrt{\lambda V}\right) \leq 2\exp(-\lambda/4) + \exp(-c\epsilon n),$$

where here $\epsilon \geq \alpha \ln n/n$, $V = \Theta(\epsilon^{(d+3)/(d-1)})$ and c, α are constants. We can use this result to show that the k^{th} moment M_k satisfies

$$M_k = O(V^{k/2}).$$

We can also prove better bounds, though with more complicated error terms.

In contrast to the integral geometric methods typically employed to study random polytopes, we rely on the notion of $\epsilon-$nets and VC-dimension to control the relevant geometry. Our concentration result employs a special instance of a more general martingale concentration theorem due to Kim and Vu. In particular we provide a quantitative notion of the volume added with the addition of a new point to the random polytope and show how this implies sharp concentration via the aforementioned tools.

We also provide a lower bound on the variance of the volume functional as well as showing the volume satisfies a central limit theorem.

9 Random Matrices

Van H. Vu - *Singularity of Random Matrices*

The study of random matrices is an important area of mathematics, with strong connections to various other fields. One of the main objects in this area is matrices whose entries are i.i.d. random variables. We focus on the basic model in which M_n is an n by n matrix whose entries are i.i.d. variables with Bernoulli distribution (taking values -1 and 1 with probability $1/2$).

A famous problem is to estimate the probability that M_n is singular. Let us denote by p_n this probability. Since M_n is singular if it has two identical rows, it is trivial that $p_n \geq (1/2 + o(1))^n$. A notorious conjecture in the field is that this bound is sharp:

Conjecture 9.1 $p_n = (1/2 + o(1))^n$.

The first result concerning singularity was obtained by Komlós in 1967, who proved $p_n = o(1)$. Later, he improved the bound to $O(n^{-1/2})$. A significant progress was made in 1995, when Kahn, Komlós and Szemerédi proved that $p_n \leq .999^n$ (see [13] and the references therein).

Recently, T. Tao and I made progress by further improving the upper bound to $(3/4 + o(1))^n$ [20]. We discovered a surprising connection between problems on random matrices and additive combinatorics. In particular, the proof of the new bound uses various ingredients from additive combinatorics (in particular, Freiman's theorem).

The details are somewhat technical, but my feeling is that the optimal bound $(1/2 + o(1))^n$ might be within sight. In fact, I believe that any improvement upon the constant $3/4$ could perhaps lead to the solution of the conjecture. Furthermore, our techniques can be used for other discrete distributions as well and in certain cases we can obtain sharp results.

A closely related question is to estimate the probability that a random symmetric matrix is singular. Let Q_n be the random symmetric n by n matrix whose upper diagonal entries are i.i.d. Bernoulli random variables. Weiss conjectured in the 1980s that Q_n is almost surely non-singular. Recently, Costello, Tao and I confirmed this conjecture. Our proof again makes a detour to additive combinatorics, with the main lemma being a quadratic version of the classical Littlewood-Offord-Erdős problem [5].

There have been several further developments in the research of random matrices reported at BIRS:

(1) The singularity problem: Costello, Tao and I generalized the singularity result for random matrices with arbitrary distribution. It seems that for any (discrete) random matrix with independent entries with distributions not concentrated on one value, the probability that the matrix is singular is exponentially small.

(2) Rank of random graphs: Costello reported a result showing that the threshold for singularity of (the adjacency matrix of) a random graph is $(\log n)/n$. (It is clear that below $(\log n)/n$, the graph has isolated vertices which correspond to all zero row; the main part is to handle the other side of the threshold.) We have extended this result to the following: For any $p > (\log n)/2n$, the corank of $G(n,p)$ equals the number of isolated vertices. As a corollary, it follows that the giant component has full rank.

(3) Richardson and Wu reported a result showing central limit theorems for random inscribing polytopes. Bárány and I extended these results for random polytopes spanned by points sampled from the Gaussian distribution.

10 Sequential Growth Models

Graham Brightwell - *Classical Sequential Growth Models*

Graham Brightwell gave a talk entitled "Classical Sequential Growth Models", including a discussion of joint work with Nicholas Georgiou.

Classical sequential growth models were introduced by Rideout and Sorkin in 2000; they are of particular interest as they are the only models satisfying some natural-looking conditions for discrete random models of space-time.

A particular classical sequential growth model is defined by a sequence $\mathbf{t} = (t_0, t_1, \ldots)$ of non-negative constants. The process starts with the partial order P_0 with one element labeled 0. At stage $n = 1, 2, \ldots$, the element n is added to P_{n-1} and placed above all elements in D_n, where D_n is a random subset of $\{0, 1, \ldots, n-1\}$, the probability that D_n is equal to a set D being proportional to $t_{|D|}$. The transitive closure is taken to form the partial order P_n.

One can either stop after stage n and study the finite partial order, or continue to get a partial order on the set of non-negative integers.

Special cases include random forests ($t_0 = t_1 = 1$, $t_i = 0$ for $i \geq 2$), and random binary orders (t_2 is the highest non-zero entry). Although random binary orders are very sparse, it is nevertheless the case that, a.s., in the infinite partial order, every element is incomparable with finitely many others. In a recent paper, Georgiou proves that, for any $\varepsilon > 0$, most elements r are incomparable with at most $r^{2+\varepsilon}$ other elements.

A random graph order, also known as a transitive percolation process, is defined by taking a random graph $G(n,p)$ on the vertex set $\{0, \ldots, n-1\}$, and putting i below j if there is a path $i = i_1, \ldots, i_k = j$ in the graph with $i_1 < \ldots < i_k$. This is equivalent to a classical sequential growth model with $t_n = t^n$, $t = p/(1-p)$.

In a later paper, Rideout and Sorkin provide computational evidence that suitably normalized sequences of random graph orders have a "continuum limit". Brightwell and Georgiou use results about the structure of random graph orders to confirm that this is indeed the case, and showed that the continuum limit is always a *semiorder*, i.e., a partial order representable by unit intervals in the line, one below another if it lies entirely to the left. Alternatively, a semiorder is a partial order containing no induced copy of either of the two specific partial orders $\mathbf{1} + \mathbf{3}$ and $\mathbf{2} + \mathbf{2}$.

It might be hoped that sequences of classical sequential growth models can have more interesting continuum limits, in particular ones that bear a closer resemblance to 4-dimensional Minkowski space-time. However, Brightwell and Georgiou show that classical sequential growth models are all "almost" semiorders, so that any continuum limit must also be very close to being a semiorder.

To be more precise, Brightwell and Georgiou show that, for any sequence $\{P_n\}_{n=0}^{\infty}$, where P_n is a classical sequential growth model stopped at stage n, the proportion of 4-element subsets isomorphic to either $\mathbf{1} + \mathbf{3}$ or $\mathbf{2} + \mathbf{2}$ tends to 0 as n tends to infinity.

11 Markov Chain Mixing Times

Prasad Tetali - *Analysis of Markov Chain Mixing Times*

Prasad Tetali gave a brief update on some recent progress in the analysis of Markov chain mixing times. The update included the status of several long-standing open problems, as well as recent theoretical developments in the topic.

The update on the theoretical development focused on isoperimetric and functional approaches to bounding mixing times. It is well known that the spectral gap of a Markov chain can be estimated in terms of conductance, facilitating isoperimetric bounds on mixing time. Observing that small sets often have large conductance, Lovász and Kannan refined this result by bounding the total variation mixing time for reversible chains in terms of the "average conductance" taken over sets of various sizes. Morris and Peres introduced the idea of evolving sets and strengthened the Lovász-Kannan result by extending the results to bound the L^∞ mixing time. Side-stepping conductance (and using a more direct functional approach, along the lines of the works on manifolds by Coulhon, Grigor'yan, and Pittet), Goel, Montenegro, and Tetali recently introduced the notion of "spectral profile" to bound L^∞ mixing time. Standard Cheeger-type inequalities show that the spectral profile bounds imply the conductance bounds. Furthermore, the known estimates on mixing times using Logarithmic Sobolev inequalities and Nash inequalities can also be derived easily with the spectral profile approach.

The strength of the above isoperimetric and spectral profile techniques has further been demonstrated in card-shuffling: A recent breakthrough result of Ben Morris provides an upper bound of d^{44} on the mixing time of the so-called Thorp shuffle on a card-deck of size 2^d, resolving a long-standing conjecture. The result of Morris has already been improved to d^{29} using the new technique of spectral profile. Morris used coupling and evolving sets techniques to prove his result, while a recent survey-style article by Montenegro and Tetali illustrates the derivation of the d^{29} mixing time for the Thorp shuffle using each technique – spectral profile as well as the evolving sets.

Tetali's report also mentioned that progress has been slow on other problems, most notably (random) sampling of contingency tables, which are of interest in statistics. The same is true for acyclic orientations, matroid bases, and Euler tours, all of which are of interest to combinatorialists. The need for new techniques in facilitating a tighter analysis of additional Markov chains such as triangulations of regular polygons and card-shuffling on general graphs has also been made clear.

References

[1] L. Addario-Berry, K. Dalal, C. McDiarmid, B. Reed, and A. Thomason, Vertex-colouring edge-weightings, *Combinatorica*, to appear.

[2] D. Aldous, The random walk construction of uniform spanning trees and uniform labelled trees, *SIAM J. Disc. Math.* **3** (1990), 450–465.

[3] M. Chudnovsky and A. Ovetsky, Coloring quasi-line graphs, *Discrete Math.*, to appear.

[4] K. Costello, T. Tao and V. Vu, Random symmetric matrices are almost surely singular, *submitted*.

[5] P. Erdős, On a lemma of Littlewood and Offord, *Bull. Amer. Math. Soc.* **51** (1945), 898–902.

[6] G. Freiman, Foundations of a structural theory of set addition. Translated from the Russian. *Translations of Mathematical Monographs* **37**, American Mathematical Society, Providence, 1973.

[7] J. Friedman and N. Pippenger, Expanding graphs contain all small trees, *Combinatorica* **7** (1987), 71–76.

[8] A. Frieze and C. McDiarmid, On random minimum length spanning trees, *Combinatorica* **9** (1989), 363–374.

[9] A. Frieze and C. McDiarmid, Algorithmic theory of random graphs, *Rand. Struct. Alg.* **10** (1997), 5–42.

[10] S. Gerke, Y. Kohayakawa, V. Rödl, and A. Steger, Small subsets inherit sparse ϵ-regularity, Submitted (2004).

[11] S. Gerke, M. Marciniszyn, and A. Steger, Probabilistic Counting Lemma for Complete Graphs, Submitted (2005).

[12] S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs*, John Wiley & Sons, New York, 2000.

[13] J. Kahn, J. Komlós, E. Szemerédi, On the probability that a random ± 1 matrix is singular, *J. Amer. Math. Soc.* **8** (1995), 223–240.

[14] Y. Kohayakawa, Szemerédi's regularity lemma for sparse graphs. In *Foundations of Computational Mathematics, (Berlin, Heidelberg) (F. Cucker and M. Shub, eds.)*, Springer-Verlag, Heidelberg, 216–230, 1997.

[15] Y. Kohayakawa, T. Łuczak, and V. Rödl, On K^4-free subgraphs of random graphs, *Combinatorica* **17** (1997), 173–213.

[16] M. Krivelevich, B. Sudakov and P. Tetali, On smoothed analysis in dense graphs and formulas, *Rand. Struct. Alg.* **29** (2006), 180–193.

[17] M. Mehta, Random matrices. Third edition. Pure and Applied Mathematics (Amsterdam), 142. Elsevier/Academic Press, Amsterdam, 2004.

[18] E. Szeméredi, Regular partitions of graphs. In *Problémes Combinatoires et Théorie des Graphes*, Colloques Internationaux CNRS **260**, 399–452, 1978.

[19] B. Reed, ω, Δ, and χ, *J. Graph Th.* **31** (1998), 177–212.

[20] T. Tao and V. Vu, On the singularity probability of random Bernoulli matrices, *submitted*.