Mathematisches Forschungsinstitut Oberwolfach

Report No. 48/2006

Combinatorics, Probability and Computing

Organised by
Noga Alon (Tel Aviv)
Béla Bollobás (Cambridge/Memphis)
Ingo Wegener (Dortmund)

October 29th – November 4th, 2006

Abstract. One of the exciting phenomena in mathematics in recent years has been the widespread and surprisingly effective use of probabilistic methods in diverse areas. The probabilistic point of view has turned out to be very profitable in Discrete Mathematics, Analysis and Theoretical Computer Science. The meeting was dedicated to recent developments in these areas, focusing on the investigation of random graphs and probabilistic methods, on the study of stochastic processes including questions on percolation and random triangulations, on the design and analysis of randomized algorithms, and on the relationship between complexity and randomness.

Mathematics Subject Classification (2000): 05D05, 05D10, 05D40, 68Q25.

Introduction by the Organisers

The conference was organized by Noga Alon (Tel Aviv), Béla Bollobás (Cambridge and Memphis) and Ingo Wegener (Dortmund). The programme consisted of 8 main lectures, supplemented by 21 shorter contributions, and covered many areas in Extremal and Probabilistic Combinatorics as well as in Theoretical Computer Science.

In its simplest form, the Probabilistic Method refers to the technique of proving the existence of structures with unexpected properties by showing that a randomly chosen object from an appropriate probability distribution has the properties with positive probability. The probabilistic method has been strikingly successful in Combinatorics, Graph Theory and Combinatorial Number Theory, and the probabilistic point of view has had an enormous influence on theoretical computer science.

The speakers reported on recent developments in Ramsey theory, including variants of these problems for Random Graphs and for Geometric questions, extending classical work in Functional Analysis. Extremal graph and hypergraph theory has also been discussed extensively, combining combinatorial, probabilistic and analytic tools. Additional active topics discussed included Percolation on quasi-random graphs and the properties of random triangulations of point sets in the plane as well as novel results in Complexity Theory including connections between communication complexity and geometric problems, and the study of refutation proofs for random instances of satisfiability.

The aim of the workshop was to focus on the connections and common themes of Combinatorics, Discrete Probability and Theoretical Computer Science. The diversity of the topics and participants stimulated a lot of fruitful discussions between researchers in different and yet related branches and gave rise to interesting discussions and new collaborations between the participants including the younger generation of researchers.

49 scientists, including 36 from countries other than Germany participated in the meeting. The organizers and participants would like to thank the Mathematisches Forschungsinstitut Oberwolfach for providing an inspiring setting for this conference.

Workshop: Combinatorics, Probability and Computing

Table of Contents

Abstracts

Majority bootstrap percolation on the hypercube

József Balogh
(joint work with Béla Bollobás and Robert Morris)

1. Introduction

In a seminal paper, McCullogh and Pitts [16] attempted to describe the brain as a network of interacting neurons. The neurons are visualized as the vertices of a directed graph with weighted edges. At each moment a vertex can be in one of two states, 0 ("passive") and 1 ("active"). If the weighted sum of the states of the neighbours of a vertex exceeds (resp. falls below) a given threshold value, the vertex state at the next moment is 1 (resp 0). Given some initial data, this switching rule determines, in principle, the time evolution of the configuration of active vertices. McCullogh and Pitts proved that this scheme is broad enough to include a universal Turing machine as a particular case. The neural network model is mathematically close to the cellular automata introduced later by Ulam [18] and von Neumann, see [10]. A popular cellular automaton is Conway's 'Game of Life' (see [14] and [9], Ch. 19). Unlike neural networks, the underlying graph of a cellular automaton is usually taken to be the two-dimensional grid, and thus has a rigid built-in structure. As often with appealing mathematical models, similar dynamic processes have been suggested in other venues, such as the spread of epidemics, systems of interacting particles, percolation, and voter models, to give just a few examples.

In this talk we study a dynamic version of percolation known as *bootstrap* percolation. Let G be a connected graph, and assume that each vertex can be in one of two possible states, *passive* or *active*. At time 0, each vertex is active with probability p, independently of all other vertices. At time t, where t is a positive integer, the state of each vertex is updated: if a vertex was active at time $t-1$ then it stays active; otherwise it will be activated if it had at least k active neighbors at time $t-1$. Here, the excitation threshold $k \geq 2$ is given. Let $P(G,p)$ be the probability that sooner or later each vertex will become active. Note that $P(G,p)$ is a nondecreasing function of p. For $0 \leq \rho < 1$, define $p_\rho = \inf\{p : P(G,p) > \rho\}$. Given a (small) positive constant ε one would like to determine the *ε-window* of the process, i.e., to find the interval $[p_\varepsilon, p_{1-\varepsilon}]$.

Bootstrap percolation was first mentioned and studied — in statistical physics context — by Chalupa, Reich and Leath [12] for the d-regular infinite tree T_d. In fact, they considered a "dual" process: given an integer m, in each round every active vertex with fewer than m active neighbors is deactivated. However, for $k = d - m + 1$ the two processes are equivalent. Very recently, Balogh, Peres and

Pete [6] studied a broader class of infinite trees (in fact all bounded degree trees) and graphs.

Van Enter [13] and Schonmann [17] studied bootstrap percolation on the infinite 2-dimensional grid, and on the d-dimensional grid, respectively. In particular, Schonmann proved that if $k \leq d$ then the critical probability of the process is 0, and if $k > d$ then it is 1.

A substantial effort has been put into analysis of the bootstrap process for the case when G belongs to a sequence $\{G_n\}_{n\geq 1}$ of ever-larger finite graphs G_n, like the $n \times n$ square grid or torus, the n-dimensional hypercube, etc. In this case, p is allowed to depend on n, and the central issue is an asymptotic behavior of $p_\varepsilon(n)$ and $p_{1-\varepsilon}(n)$. Is $p_{1-\varepsilon}(n) - p_\varepsilon(n)$ going to zero as $n \to \infty$, and if so, is $\lim_{n\to\infty} p_{1-\varepsilon}(n)/p_\varepsilon(n)$ equal to 1, or at least bounded? "Yes" means intuitively that there is a sharp phase transition in the limiting behavior of the percolation process. The function $p_{1/2}(n)$ is usually called the *critical probability*, often denoted as $p_c(n)$, and if $p_{1/2}(n)/f(n) \to 1$ as $n \to \infty$ for some function $f(n)$, then f is interpreted as the limiting critical threshold for percolation.

Aizenmann and Lebowitz [1] found bounds for $p_c(n)$ in the case where $k = 2$ and G_n is the d-dimensional grid $[1, n]^d$, and Cerf and Manzo [11] extended these bounds for parameters $2 \leq k \leq d$. The central case $k = d = 2$ was also studied in [7], and Holroyd [15] discovered a sharp asymptotic formula for the critical probability. Balogh and Bollobás [3] determined $p_{1/2}(n)$ up to a constant factor for the case $k = 2$ of the n-dimensional hypercube, and Balogh, Bollobás and Morris [4] improved this result by determining a sharp asymptotic formula in this case as well. However, the question of whether the transition is sharp, i.e., whether $\lim_{n\to\infty} p_{1-\varepsilon}(n)/p_\varepsilon(n) = 1$ for all $\varepsilon > 0$, remains wide open in most of the cases. An exception, Balogh and Bollobás [2] proved that for $[1, n]^d$ and $k = 2$, the transition is indeed sharp. It is widely believed that the transition is sharp for all $k \in [2, d]$. Finally, a recent result of Balogh and Pittel [8] gives a sharp estimate on the critical probability of the k-parameter bootstrap percolation on a d-regular random graph on n vertices (where d is fixed, while $n \to \infty$). Using the method of "approximation by differential equations", possibly sharp bounds were given for the size of the transition window.

Here we study "majority" (i.e., $n/2$-neighbour) bootstrap percolation on the hypercube, Q_n. (Similarly, the work of Cerf and Manzo [11] included the "majority" case $k = d$ on the d-dimensional grid, for fixed d.)

The following theorem is our main result (see also [5].

Theorem 1. *Let $p_c = p_c(Q_n, n/2)$ denote the critical probability for $n/2$-neighbour bootstrap percolation on Q_n, and let n be sufficiently large. Then*

$$\frac{1}{2} - \frac{1}{2}\sqrt{\frac{\log n}{n}} - \frac{2\log\log n}{\sqrt{n\log n}} \leq p_c \leq \frac{1}{2} - \frac{1}{2}\sqrt{\frac{\log n}{n}} + \frac{3\log\log n}{4\sqrt{n\log n}}.$$

2. Some ideas of the proof

Let $n \in \mathbb{N}$ be sufficiently large, and let $p(n) = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{\log n}{n}} + \frac{\lambda \log\log n}{\sqrt{n \log n}}$, for some $\lambda \in \mathbb{R}$. Let the elements of $A^{(0)} \subset Q_n$ be chosen independently at random, each with probability p defined above, and let $A^{(i)}$ be the set of infected vertices after i rounds.

2.1. **Upper bound.** The proof comes in three stages. First we show that if $\lambda \geq 3/4$ then $\Pr(x \in A^{(2)}) \geq 3/4 + o(1)$. Next we show that if $\Pr(x \in A^{(2)}) > 1/2 + \delta$ then $\Pr(x \notin A^{(5)}) < \exp(-cn)$ for some constant $c > 0$. Finally, we show that if $\Pr(x \notin A^{(k)}) < \exp(-cn)$ for some $k \in \mathbb{N}$ and some constant $c > 0$, then $\Pr(x \notin A^{(2k+1)}) < \exp(-dn^2)$ for some constant $d > 0$. It is easy to see that these results imply

$$\Pr(A^{(11)} \neq Q_n) \leq 2^n \Pr(x \notin A^{(11)}) < e^{-dn^2+n} = o(1),$$

so, with probability $1 - o(1)$, the bootstrap process percolates.

2.2. **Lower bound.** Let p be in Theorem 1, and let $\lambda \leq -2$. We cannot prove that the bootstrap process stops quickly; instead, we couple (dominate) the bootstrap process with a process which stops within only three rounds, and (with high probability) does not activate the entire cube. Let $t \in \mathbb{N}$. The dominant process, which we call **Boot3(t)**, is defined as follows.

- The elements of $A^{(0)}$ are chosen as described above.
- $x \in A^{(1)}$ if $x \in A^{(0)}$ or $|\Gamma(x) \cap A^{(0)}| \geq n/2 - 3t$.
- $x \in A^{(2)}$ if $x \in A^{(1)}$ or $|\Gamma(x) \cap A^{(1)}| \geq n/2 - 2t$.
- $x \in A^{(3)}$ if $x \in A^{(2)}$ or $|\Gamma(x) \cap A^{(2)}| \geq n/2 - t$.
- If $i \geq 3$, then $x \in A^{(i+1)}$ if $x \in A^{(i)}$ or $|\Gamma(x) \cap A^{(i)}| \geq n/2$.

We prove that in **Boot3(t)**, with high probability we have $A^{(4)} = A^{(3)} \neq Q_n$. It follows that with high probability $A^{(0)}$ does not percolate in the original bootstrap process.

References

[1] M. Aizenman and J.L. Lebowitz, Metastability effects in bootstrap percolation, *J. Phys. A* **21** (1988), 3801–3813.

[2] J. Balogh and B. Bollobás, Sharp thresholds in bootstrap percolation, *Physics A* **326** (2003), 305–312.

[3] J. Balogh and B. Bollobás, Bootstrap percolation on the hypercube, *Prob. Theory Rel. Field* **134** (2006), 624–648.

[4] J. Balogh, B. Bollobás, and R. Morris, Bootstrap percolation on $[n]^d$, in preparation.

[5] J. Balogh, B. Bollobás, and R. Morris, Majority bootstrap percolation on the hypercube, in preparation.

[6] J. Balogh, Y. Peres, and G. Pete, Bootstrap percolation on infinite trees and non-amenable groups, *Comb. Prob. Comp.* **15** (2006), 715–730.

[7] J. Balogh and G. Pete, Random disease on the square grid, *Random Structures Algorithms* **13** (1998), 409–422.

[8] J. Balogh and B. G. Pittel, Bootstrap percolation on the random regular graph, to appear in *Random Structures Algorithms*.
[9] E. Berlekamp, J.H. Conway, and R.K. Guy, *Winning Ways for Your Mathematical Plays*, vol. 2, Academic Press, London, 1982.
[10] A.W. Burks (ed.), *Essays on Cellular Automata*, University of Illinois Press, 1970.
[11] R. Cerf and F. Manzo, The threshold regime of finite volume bootstrap percolation, *Stochastic Process. Appl.* **101** (2002), 69–82.
[12] J. Chalupa, G.R. Reich, and P.L. Leath, Bootstrap percolation on a Bethe Lattice, *J. Phys. C* **12** (1979), L31–L35.
[13] A.C.D. van Enter, Proof of Straley's argument for bootstrap percolation, *J. Statist. Phys.* **48** (1987), 943–945.
[14] M. Gardner, *Wheels, Life and Other Mathematical Amusements*, W. H. Freeman and Co., 1983.
[15] A.E. Holroyd, Sharp metastability threshold for two-dimensional bootstrap percolation, *Prob. Theory Rel. Fields* **125** (2003), 195–224.
[16] W. S. McCulloch and W. Pitt, A logical calculus of ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics* **5** (1943), 115–133.
[17] R.H. Schonmann, On the behavior of some cellular automata related to bootstrap percolation, *Ann. Probab.* **20** (1992), 174–193.
[18] S. Ulam, *Random processes and transformations*, Proceedings of the International Congress of Mathematicians, Cambridge, Mass., 1950, Amer. Math. Soc., Providence, R.I., vol. 2, 1952, pp. 264–275.

Convergent sequences of dense graphs II.

Christian Borgs
(joint work with J.T. Chayes, L. Lovász, V.T. Sós and K. Vesztergombi)

1. Introduction

Consider a dense sequence of simple graphs (G_n) such that the number of nodes in G_n goes to infinity with n. In two papers, [1] and [2], we considered several natural notions of convergence for such a sequence — some motivated by combinatorics and others by statistical physics. In Part I of our work [1], summarized in the accompanying abstract, we showed the equivalence of several of these notions. This extended abstract summarizes Part II of our work [2], as discussed by one of us (C. Borgs) at Oberwolfach, in which we show equivalence of the remaining notions.

From the point of view of combinatorics, our theory can be viewed as a substantial generalization of the theory of quasirandom graphs, which are sequences of graphs which "look like" random graphs. Obviously, there are many ways in which one could make this precise, but interestingly, many natural ways in which a sequence of graphs could be defined to be quasirandom turn out to be equivalent [5, 4].

Here we prove similar equivalences for the notion of convergent graph sequences. In fact, most of the equivalences for quasirandom graphs are immediate corollaries of the general theory developed here and in our companion paper [1]. A notable exception is the spectral representation of quasirandom graphs: while it turns out

that convergence of the spectrum is implied by our other conditions of convergence, it is not equivalent in our general setting. Indeed, already in the setting of generalized quasirandom graph sequences considered in [6] neither the knowledge of the limiting spectrum of the adjacency matrices nor the knowledge of the limiting spectrum of the Laplacians is enough to characterize the sequences.

From the viewpoint of physics, our results show that convergence of various thermodynamic quantities, notably microcanonical free energies or ground state energies for all so-called "soft-core" models, is equivalent to convergence of apparently more local graph properties, as defined below.

2. Definitions and Results

2.1. **Left Convergence.** The first notion of convergence for a sequence (G_n) we consider is what we call "left convergence". It was introduced in [1] and is a way of characterizing a large graph G in terms of the number of copies of a small graph F that are contained in G.

Given two simple graphs F and G, we denote the number homomorphisms from F to G by $\hom(F,G)$. Let $t(F,G)$ be the probability that a random map $\phi : V(F) \to V(G)$ is a homomorphism,

$$t(F,G) = \frac{1}{|V(G)|^{|V(F)|}} \hom(F,G), \tag{1}$$

where $V(G)$ and $V(F)$ are the set of vertices in G and F, respectively. We then called a sequence (G_n) of simple graphs *left-convergent* if the "homomorphism densities" $t(F,G_n)$ converge for all simple graphs F.

2.2. **Naive Right Convergence.** Instead of testing a graph sequence (G_n) with homomorphisms "from the left," i.e., with homomorphisms from a small graph F into the graphs (G_n), one might want to test (G_n) with homomorphisms "from the right," i.e., one might want to consider the homomorphisms from G_n into some small graph H. For this to be interesting, we have to work with weighted graphs, i.e., graphs H with nodeweights $\alpha_i(H) > 0$ for the nodes $i \in V(H)$ and edgeweights $\beta_{ij}(H) \in \mathbb{R}$ for the edges $ij \in E(H)$. We will usually work with soft-core graphs, i.e., weighted graphs H with $\beta_{ij}(H) > 0$ for all $i,j \in V(H)$. The homomorphism number from a simple graph G into a soft-core graph H is then defined as

$$\hom(G,H) = \sum_{\phi: V(G) \to V(H)} \prod_{u \in V(G)} \alpha_{\phi(u)}(H) \prod_{uv \in E(G)} \beta_{\phi(u),\phi(v)}(H), \tag{2}$$

where $E(G)$ denotes the set of edges in G. These homomorphism numbers "from the right" typically grow or fall exponentially in the number of edges of G. It therefore seems natural to define a sequence (G_n) of graphs to be right-convergent if $\frac{1}{|V(G_n)|^2} \ln \hom(G_n,H)$ converges for every soft-core graph H. For reasons explained below, we will actually call such a sequence *naively right-convergent*.

Example (The Ising Model)**.** *The simplest model that fits into our framework is the so-called Ising model. Indeed, consider the soft-core graph H with $V(H) =$*

$\{-,+\}$, $\alpha_+(H) = e^h$, $\alpha_i(H) = e^h$, $\beta_{--}(H) = \beta_{++}(H) = e^K$, *and* $\beta_{-+}(H) = \beta_{+-}(H) = e^{-K}$. *Maps* $\phi : V(G) \to V(H)$ *can then be identified with spin-configurations, and the homomorphism number,*

$$\hom(G,H) = \sum_{\phi:V(G)\to V(H)} \exp\Big(K \sum_{uv\in E(G)} \phi(u)\phi(v) + h \sum_{u\in V(G)} \phi(u)\Big),$$

is nothing but the partition function of the Ising model with magnetic field h *and coupling constant* K.

2.3. **Ground State Energies.** The next notion we consider is the notion of ground state energies. To motivate this notion, let us observe that for a graph G on n nodes, $\hom(G_n, H)$ is a sum of q^n terms (where q is the number of nodes in H) which all are exponential in the number of edges in G, which is in turn proportional to n^2. As a consequence, $\hom(G,H)$ is dominated by the largest term. More precisely, let H have weights $\beta_{ij}(H) = e^{2J_{ij}}$ and $\alpha_i(H) = e^{h_i}$, and let

$$\mathcal{E}(G,J,h) = \min_{\phi:V(G)\to[q]} \Big(-\frac{1}{|V(G)|} \sum_{u\in V(G)} h_{\phi(u)} - \frac{2}{|V(G)|^2} \sum_{uv\in E(G)} J_{\phi(u)\phi(v)}\Big). \tag{3}$$

Then

$$\frac{1}{n^2} \log \hom(G,H) = -\mathcal{E}(G,J,0) + O(n^{-1}). \tag{4}$$

Motivated by notions from statistical physics, we call the quantity $\mathcal{E}(G,J,h)$ the *ground state energy of the model* (J,h) *on* G. We say that a sequence of graphs (G_n) has *convergent ground state energies*, if $\mathcal{E}(G_n,J,h)$ converges for all q, all symmetric, real valued $q \times q$ matrices J, and all $h \in \mathbb{R}^q$.

Example (Max-Cut)**.** *Let* $q = 2$, $h_1 = h_2 = 0$ *and* $J_{ij} = \frac{1}{2}(1 - \delta_{ij})$, *corresponding to the soft-core graph* H *with nodeweights* $\alpha_1(H) = \alpha_2(H) = 1/2$ *and edgeweights* $\beta_{11}(H) = \beta_{22}(H) = 1$ *and* $\beta_{12}(H) = e$. *Up to a sign, the ground state energy of the model* $(J,0)$ *is then precisely the size of the largest cut in* G, $\mathcal{E}(G,J,0) = -maxcut(G)$, *and the relation* (4) *expresses the fact that the homomorphism number* $\hom(G,H)$ *is dominated by the largest cut in* G. *For general soft-core graphs* H, *the ground state energies* $\mathcal{E}(G,J,0)$ *correspond to weighted versions of certain multi-way cut-problems, and by* (4), *naive right convergence is equivalent to the convergence of the cut-density for all such multi-way cut-problems.*

2.4. **Right Convergence and Microcanonical Ground State Energies.** In contrast to the notion of left convergence, which corresponds to the convergence of local properties like the density of triangles or the density of 4-cycles, naive right convergence corresponds to convergence of global properties like the density of the largest cut and the ground state energies of suitable soft-core models. This raises the question whether the *a priori* quite different notions of left- and right convergence are equivalent. While it turns out that left convergence implies naive right convergence, the converse is not true (hence the term naive right convergence).

But a slight modification involving homomorphisms for which the number of vertices in G_n that map onto a given $i \in V(H)$ is restricted to be a given fraction of $V(G_n)$ gives equivalence.

To formalize this, let us label the nodes of H as $1, \dots, q$, and define Pd_q to be the set of vectors $\mathbf{a} \in \mathbb{R}_q$ for which $\sum_i a_i = 1$ and $a_i \geq 0$ for all $i \in [q]$. Given a probability distribution $\mathbf{a} \in \mathsf{Pd}_q$, we set

$$\Omega_{\mathbf{a}}(G) = \Big\{\phi : V(G) \to [q] \colon \Big||\phi^{-1}(\{i\})| - a_i|V(G)|\Big| \leq 1 \quad \text{for all} \quad i \in [q]\Big\}. \tag{5}$$

and define a constrained version of the homomorphism numbers by

$$\hom_{\mathbf{a}}(G, H) = \sum_{\phi \in \Omega_{\mathbf{a}}(G)} \prod_{uv \in E(G)} \beta_{\phi(u)\phi(v)}(H). \tag{6}$$

We then say that a sequence of simple graphs (G_n) is *right-convergent* if if

$$\frac{1}{|V(G_n)|^2} \ln \hom_{\mathbf{a}}(G_n, H)$$

converges for every soft-core graph H and every probability distribution $\mathbf{a}$ on $V(H)$.

Again motivated by notions from statistical physics, the corresponding constrained version of the ground state energy,

$$\mathcal{E}_{\mathbf{a}}(G, J) = \min_{\phi \in \Omega_{\mathbf{a}}(G)} \Big(-\frac{2}{|V(G)|^2} \sum_{uv \in E(G)} J_{\phi(u)\phi(v)}\Big), \tag{7}$$

is called a *micro-canonical ground state energy*. We say that (G_n) has convergent micro-canonical ground state energies if $\mathcal{E}_{\mathbf{a}}(G_n, J, h)$ converges for all q, all symmetric $q \times q$ matrices J, and all $\mathbf{a} \in \mathsf{Pd}_q$.

2.5. **Graph Quotients.** Another notion of convergence which we will also show to be equivalent is the convergence of "quotients". The quotients of a simple graph G are defined in terms of the partitions $\mathcal{P} = \{V_1, \dots, V_q\}$ of its node set by contracting all nodes in a given group to a new node, leading to a weighted graph $G/\mathcal{P}$ on q nodes. More precisely, we define $G/\mathcal{P}$ as the weighted graph on $[q]$ with weights

$$\alpha_i(G/\mathcal{P}) = \frac{|V_i|}{|V(G)|} \qquad \text{and} \qquad \beta_{ij}(G/\mathcal{P}) = \frac{e_G(V_i, V_j)}{|V_i| \cdot |V_j|}. \tag{8}$$

The quotient graph $G/\mathcal{P}$ thus has nodeweights proportional to the sizes of the classes in $\mathcal{P}$, and edgeweights that are equal to the edgedensities between the different classes of $\mathcal{P}$. We denote the set of quotients obtained by considering all possible partitions of $V(G)$ into q classes by $\widehat{\mathcal{S}}_q(G)$. In order to define convergence, we consider the following version of the ℓ_1 distance between two weighted graphs

H, H' on q nodes

$$(9)\quad d_1(H,H') = \sum_{i,j\in[q]} \Big| \frac{\alpha_i(H)\alpha_j(H)\beta_{ij}(H)}{(\alpha_H)^2} - \frac{\alpha_i(H')\alpha_j(H')\beta_{ij}(H')}{(\alpha_{H'})^2} \Big| + \sum_{i\in[q]} \Big| \frac{\alpha_i(H)}{\alpha_H} - \frac{\alpha_i(H')}{\alpha_{H'}} \Big|.$$

In terms of this distance, we define *Hausdorff metric* d_1^{Hf} between two sets S, S' of weighted graphs by

$$(10)\qquad d^{\mathrm{Hf}}(S,S') = \max\Big\{ \sup_{H\in S} \inf_{H'\in S'} d_1(H,H'),\ \sup_{H'\in S'} \inf_{H\in S} d_1(H,H') \Big\}.$$

We then say that a sequence (G_n) of simple graphs has *convergent quotients* if for all $q \geq 1$, the sequence of sets of quotients $\widehat{\mathcal{S}}_q(G_n)$ is a Cauchy sequence in the Hausdorff distance d_1^{Hf}.

2.6. **Main Results.** Several of the results of Part II of this work [2] are summarized in the following two theorems.

Theorem 2. *Let (G_n) be a sequence of simple graphs such that $|V(G_n)| \to \infty$ as $n \to \infty$. Then the following statements are equivalent:*

(i) The sequence (G_n) is left-convergent.

(ii) The quotients of (G_n) are convergent in the Hausdorff distance d_1^{Hf}.

(iii) The sequence (G_n) is right-convergent.

(iv) The microcanonical ground state energies of (G_n) are convergent.

In contrast to the notions of convergence discussed in Theorem 2, convergence of the energies $\widehat{\mathcal{E}}(G_n, J, h)$ and naive right convergence are not equivalent to left convergence, see [2] for a counterexample. But left convergence does imply convergence of the energies, as well as naive right convergence. It also implies convergence of the spectrum. This is the content of our second theorem.

Theorem 3. *Let (G_n) be a left-convergent sequence of simple graphs such that $|V(G_n)| \to \infty$ as $n \to \infty$. Then the following holds:*

(i) The sequence (G_n) is naively right-convergent.

(ii) The ground state energies of (G_n) are convergent.

(iii) The spectrum of (G_n) is convergent in the sense that if $\lambda_{n,1} \geq \lambda_{n,2} \geq \cdots \geq \lambda_{n,|V(G_n)|}$ are the eigenvalues of the adjacency matrix of G_n, then $|V(G_n)|^{-1}\lambda_{n,i}$ and $|V(G_n)|^{-1}\lambda_{n,|V(G_n)|+1-i}$ converge for all $i > 0$.

2.7. **Other Notions of Convergence.** In addition to the notions of convergence defined above, we consider several other natural notions of convergence in [1, 2]. Among these notions is that of convergence in a suitably defined metric, a concept already considered in [1] and discussed in detail in the accompanying abstract and talk by J. Chayes. Others are the convergence of testable graph parameters, when evaluated on the sequence G_n, or the convergence of micro-canonical free energies of finite spin systems defined on G_n. In [1] and [2], we showed that all these

additional notions of convergence are also equivalent to left-convergence. See [3] for an overview of these and other results.

Finally, in [2], we also generalized Theorems 2 and 3 to weighted graphs.

References

[1] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi: Convergent Graph Sequences I. Subgraph frequencies, metric properties, and testing, preprint (2006) at http://research.microsoft.com/ borgs/Papers/ConvMetric.pdf.

[2] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi: Convergent Graph Sequences II. Multiway Cuts and Statistical Physics, preprint (2007),

[3] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi: Counting graph homomorphisms, in: *Topic in Discrete Mathematics* (Klazar, Kratochvil, Loebl, Matousek, Thomas, Valtr, eds.) Springer, Berlin–Heidelberg (2006), 315–371.

[4] F. Chung, R.L. Graham and R.M. Wilson: Quasi-random graphs, *Combinatorica* **9** (1989), 345–362.

[5] A. Thomason: Pseudorandom graphs, in: *Random graphs '85* North-Holland Math. Stud. **144**, North-Holland, Amsterdam, 1987, 307–331.

[6] L. Lovász and V.T. Sós: Generalized quasirandom graphs, preprint (2005) at http://research.microsoft.com/users/lovasz/quasirandom.pdf

The two probe conjecture

Peter Bro Miltersen

Let m be a sufficiently large integer and let $n = \lfloor m^{1/3} \rfloor$. Let s be a positive integer and let $\phi : \binom{[m]}{\leq n} \to \{0,1\}^s$ be any scheme for representing small subsets of $[m]$ as bit strings of length s. We presented and discussed the following *two probe conjecture* first stated by Buhrman *et al* [1]: *If queries "Is x in S"? can be answered by inspecting at most two bits of $\phi(S)$ for all x and for all S, then $s \geq m$ must hold.* The conjecture was known to be true if the two bits are inspected non-adaptively, but not if the location of the second bit probe is allowed to depend upon the result of the first. During the workshop hike, Uri Feige refuted the conjecture in the above strong form by presenting a two-probe scheme for which $s < m$. We believe that it is still a reasonable conjecture that $s \geq m - o(m)$ must hold.

References

[1] Harry Buhrman, Peter Bro Miltersen, Jaikumar Radhakrishnan and Srinivasan Venkatesh, Are bitvectors optimal?, *SIAM Journal on Computing* **31** (2002), 1723–1744

Convergent sequences of dense graphs I.

Jennifer Chayes
(joint work with C. Borgs, L. Lovász, V.T. Sós and K. Vesztergombi)

1. Introduction

We develop a theory of convergence of graph sequences, which works best in two extreme cases: dense graphs (the subject of [6] and [7]) and graphs with bounded degree (the subject of [8]). Convergence of graph sequences was defined by Benjamini and Schramm [3] for graphs with bounded degree, and by the authors of this paper [4] for dense graphs. This extended abstract summarizes the first part of our work on convergent sequences of dense graphs [6], as discussed by one of us (J.T. Chayes) at Oberwolfach.

Our general setup is as following. We have a "large" graph G with node set $V(G)$ and edge set $E(G)$. There are (at least) two ways of studying G using homomorphisms. First, we can count the number of copies of various "small" graphs F in G, more precisely, we count the number of homomorphisms from F to G; this way of looking at G allows us to treat many problems in, e.g., extremal graph theory. Second, we can count homomorphisms from G into various small graphs H; this includes many models in statistical physics and many problems on graph coloring.

These two notions of probing a large graph with a small graph lead to two different notions of convergence of a sequence of graphs (G_n): convergence from the left, corresponding to graphs which look more and more similar when probed with homomorphisms from small graphs into G_n, and convergence from the right, corresponding to graph sequences whose elements look more and more similar when probed with homomorphism from G_n into a small graphs.

In Part I [6], we study convergence from the left, both for sequences of simple graphs and sequences of weighted graphs, and its relations to sampling and testing. Since this abstract focuses on convergence from the left, we'll often omit the phrase "from the left". We also show that convergence from the left is equivalent to convergence in metric for a suitable notion of distance between two weighted graphs. Finally, we show that convergence from the left is equivalent to the property that the graphs in the sequence have asymptotically the same Szemerédi partitions. Convergence from the right is the subject matter Part II of this work [7], and is discussed in the accompanying abstract.

2. Definitions and Results

2.1. **Notation.** We consider both unweighted, simple graphs and weighted graphs, where, as usual, a simple graph G is a graph without loops or multiple edges. We denote the node and edge set of G by $V(G)$ and $E(G)$, respectively.

A *weighted graph* G is a graph with a weight $\alpha_i = \alpha_i(G) > 0$ associated with each node and a weight $\beta_{ij} = \beta_{ij}(G) \in \mathbb{R}$ associated with each edge ij, including

possible loops with $i = j$. For convenience, we set $\beta_{ij} = 0$ if $ij \notin E(G)$. We set

$$\alpha_G = \sum_i \alpha_i(G), \quad \|G\|_\infty = \max_{i,j} |\beta_{ij}(G)|, \quad \text{and} \quad \|G\|_2 = \left(\sum_{i,j} \frac{\alpha_i \alpha_j}{\alpha_G^2} \beta_{ij}^2 \right)^{1/2},$$

and for $S, T \subset V(G)$, we define the weighted number of edges from S to T as

$$e_G(S,T) = \sum_{\substack{i \in S \\ j \in T}} \alpha_i(G)\alpha_j(G)\beta_{ij}(G). \tag{1}$$

A weighted graph G is called *soft-core* if it is a complete graph with loops at each node, and every edgeweight is positive.

As usual, a function from the set of simple graphs into the reals is called a *simple graph parameter* if it is invariant under relabeling of the nodes. Finally, we write $G \cong G'$ if G and G' are isomorphic, i.e., if G' can be obtained from G by a relabeling of its nodes.

2.2. **Homomorphism Numbers and Left Convergence.** Let F and G be two simple graphs. We define $\hom(F,G)$ as the number of homomorphisms from F to G, and the *homomorphism density* of F in G as

$$t(F,G) = \frac{1}{|V(G)|^{|V(F)|}} \hom(F,G).$$

We extend the notion of homomorphism numbers to weighted graphs G by setting

$$\hom(F,G) = \sum_{\phi: V(F) \to V(G)} \prod_{i \in V(F)} \alpha_{\phi(i)}(G) \prod_{ij \in E(F)} \beta_{\phi(i),\phi(j)}(G) \tag{2}$$

where the sum runs over all maps from $V(F)$ to $V(G)$, and define

$$t(F,G) = \frac{\hom(F,G)}{\alpha_G^k}, \tag{3}$$

where k is the number of nodes in F.

Let (G_n) be a sequence of weighted graphs with uniformly bounded edgeweights. We say that (G_n) is *convergent from the left*, or simply *convergent*, if $t(F,G_n)$ converges for any simple graph F.

2.3. **Cut-Distance.** We define a notion of distance between two graphs, which will play a central role throughout this work. This distance is derived from the so-called cut-norm, and expresses similarity of global structure: graphs with small distance in this metric have cuts of similar size. This is easily made precise for two graphs G and G' on the same set V of nodes, with identical nodeweights $\alpha_i(G) = \alpha_i(G')$. We define

$$d_\square(G,G') = \max_{S,T \subset V} \left| \frac{e_G(S,T)}{\alpha_G^2} - \frac{e_{G'}(S,T)}{\alpha_{G'}^2} \right|.$$

But some care is needed when G and G' have different nodesets. Let G, G' be weighted graphs with nodeset V and V', respectively. For $i \in V$ and $u \in V'$, let $\mu_i = \alpha_i(G)/\alpha_G$ and $\mu'_u = \alpha_u(G')/\alpha_{G'}$. We then define the set of fractional

overlays of G and G' as the set $\mathcal{X}(G,G')$ of probability distributions X on $V \times V'$ such that

$$\sum_{u\in V'} X_{iu} = \mu_i \quad \text{for all} \quad i \in V \quad \text{and} \quad \sum_{i\in V} X_{iu} = \mu'_u \quad \text{for all} \quad u \in V',$$

and define our cut-distance as

$$\delta_\square(G,G') = \min_{X\in\mathcal{X}(G,G')} \max_{S,T\subset V\times V'} \Big| \sum_{\substack{(i,u)\in S\\(j,v)\in T}} X_{iu}X_{jv}\big(\beta_{ij}(G) - \beta_{uv}(G')\big)\Big|. \tag{4}$$

2.4. **Main Results.**

2.4.1. *Left Convergence versus Convergence in Metric.* First we state one of the main results of [6], namely, that convergence from the left is equivalent to convergence in the metric $\delta_\square$.

Theorem 1. *Let (G_n) be a sequence of weighted graphs with uniformly bounded edgeweights. Then (G_n) is left convergent if and only if it is a Cauchy sequence in the metric $\delta_\square$.*

In fact, we have the following quantitative version. To simplify our notation, we only give this quantitative version for graphs with edgeweights in $[-1,1]$; the general case follows by simply scaling all edgeweights appropriately.

Theorem 2. *Let G_1, G_2 be weighted graphs with edgeweights in $[-1,1]$.*

(a) *Let F be a simple graph, then*

$$|t(F,G_1) - t(F,G_2)| \le 4|E(F)|\delta_\square(G_1,G_2).$$

(b) *Let $k \ge 1$, and assume that $|t(F,G_1) - t(F,G_2)| \le 3^{-k^2}$ for every simple graph F on k nodes. Then*

$$\delta_\square(G_1,G_2) \le \frac{22}{\sqrt{\log_2 k}}.$$

The first part of this theorem is closely related to the "Counting Lemma" in the theory of Szemerédi partitions.

2.4.2. *Szemerédi Partitions for Graph Sequences.* Convergent graph sequences can also be characterized by the fact that (for any fixed error) they have Szemerédi partitions that become more and more similar.

To make this precise, we need some notation: for a weighted graph G and a partition $\mathcal{P} = \{V_1,\dots,V_k\}$ of $V(G)$ into k classes, we define two weighted graphs $G/\mathcal{P}$ and $G_\mathcal{P}$ as follows: Let $\alpha_{V_i} = \sum_{x\in V_i}\alpha_x(G)$. The *quotient graph* $G/\mathcal{P}$ is a weighted graph on $[k]$, with nodeweights $\alpha_i(G/\mathcal{P}) = \alpha_{V_i}/\alpha_G$ and edgeweights $\beta_{ij}(G/\mathcal{P}) = \frac{e_G(V_i,V_j)}{\alpha_{V_i}\alpha_{V_j}}$, while $G_\mathcal{P}$ is a weighted graph on $V(G)$, with nodeweights $\alpha_x(G_\mathcal{P}) = \alpha_x(G)$ and edgeweights $\beta_{xy}(G_\mathcal{P}) = \beta_{ij}(G/\mathcal{P})$ for $x \in V_i$ and $y \in V_j$. These two graphs have different number of nodes, but they are similar in the sense that $\delta_\square(G/\mathcal{P}, G_\mathcal{P}) = 0$. We call a partition $\mathcal{P}$ *weakly ε-regular* if $d_\square(G,G_\mathcal{P}) \le \varepsilon$.

Theorem 3. *Let (G_n) be a sequence of weighted graphs with nodeweights 1, edgeweights in $[-1,1]$, and $|V(G_n)| \to \infty$ as $n \to \infty$. Then (G_n) is left-convergent if and only if for every $\varepsilon > 0$ we can find an integer $q \le 2^{10/\varepsilon^2}$, and a sequence of partitions, $(\mathcal{P}_n)$ such that the following two conditions hold.*

(i) If $|V(G_n)| \ge q$, then $\mathcal{P}_n$ is a weakly ε-regular partition of G_n into q classes.

(ii) As $n \to \infty$, the quotient graphs $G_n/\mathcal{P}_n$ converge to a weighted graph H_ε on q nodes.

Note that the graphs in (ii) have the same node set $[q]$, so their convergence to H_ε means simply that corresponding nodeweights and edgeweights converge.

2.4.3. *Sampling.* In our language, the Weak Regularity Lemma of Frieze and Kannan [11] states that any graph G can be well approximated by a small graph $H = G/\mathcal{P}$ in the $\delta_\square$ distance. Our next result says that such a small graph can obtained by straightforward sampling. For simplicity, we state the results for graphs with edgeweights in $[-1,1]$. For a graph G and positive integer n, let $\mathbb{G}(n,G)$ denote the (random) induced subgraph $G[S]$, where S is chosen uniformly from all subsets of $V(G)$ of size n.

Theorem 4. *Let G be a weighted graph with nodeweights 1 and edgeweights in $[-1,1]$. Let $k \le |V(G)|$. Then*

$$\delta_\square(G, \mathbb{G}(k,G)) \le \frac{10}{\sqrt{\log_2 k}}. \tag{5}$$

with probability at least $1 - e^{-k^2/(2\log_2 k)}$.

In order to prove this theorem, we will need a theorem which allows us to compare samples from two weighted graphs on the same set of nodes, extending a result by Alon, Fernandez de la Vega, Kannan and Karpinski [1, 2].

Theorem 5. *Let G_1 and G_2 be weighted graphs on a common vertex set V, with nodeweights one and edgeweights in $[-1,1]$. Let $k \le |V|$. If S is chosen uniformly from all subsets of V of size k, then*

$$\Big| d_\square(G_1[S], G_2[S]) - d_\square(G_1, G_2) \Big| \le \frac{20}{k^{1/4}} \tag{6}$$

with probability at least $1 - 2e^{-\sqrt{k}/8}$.

2.4.4. *Testing.* The above theorem allows us to prove several results for testing graph parameters and graph properties in a straightforward way.

In this paper, we only consider parameter testing. We may want to determine some parameter of G; say what is the edge density? How large is the density of the maximum cut? Of course, we'll not be able to determine the exact value of this parameter; the best we can hope for is that if we take a sufficiently large sample, we can find the approximate value of the parameter with large probability.

A graph parameter f is said to be *testable* if for every $\varepsilon > 0$ there is a positive integer k such that if G is a graph with at least k nodes, then

$$\Pr(|f(G) - f(\mathbb{G}(k,G))| > \varepsilon) \le \varepsilon.$$

Testability is related to our framework through the following observation:

Proposition 6. (a) *A simple graph parameter f is testable if and only if $f(G_n)$ converges for every convergent graph sequence (G_n) with $|V(G_n)| \to \infty$.*

(b) *A sequence (G_n) of simple graphs with $|V(G_n)| \to \infty$ is convergent if and only if $f(G_n)$ converges for every testable simple graph parameter f.*

References

[1] N. Alon, W. Fernandez de la Vega, R. Kannan and M. Karpinski: Random sampling and approximation of MAX-CSPs, *J. Comput. System Sci.* **67** (2003), 212–243.

[2] N. Alon, W. Fernandez de la Vega, R. Kannan and M. Karpinski: Random sampling and approximation of MAX-CSP problems; in *Proc. of the 34th ACM Symp. Theor. of Comp. (STOC)* (2002), 232 – 239.

[3] I. Benjamini and O. Schramm: Recurrence of Distributional Limits of Finite Planar Graphs, *Electronic J. Probab.* **6** (2001), paper no. 23, 1–13.

[4] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi, unpublished (2003).

[5] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós and K. Vesztergombi: Counting graph homomorphisms, in: *Topics in Discrete Mathematics* (ed. M. Klazar, J. Kratochvil, M. Loebl, J. Matoušek, R. Thomas, P. Valtr), Springer (2006), 315–371.

[6] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi: Convergent graph sequences I. Subgraph frequencies, metric properties and testing, preprint.

[7] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi: Convergent graph sequences II. Multiway cuts and statistical physics, preprint.

[8] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi: Convergence of sparse graph sequences, in preparation.

[9] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, B. Szegedy and K. Vesztergombi: Graph limits and parameter testing, *Proc. of the 38th ACM Symp. Theory of Comp.* (2006), 261–270.

[10] C. Borgs, J. Chayes, L. Lovász: Unique limits of dense graph sequences, in preparation.

[11] A. Frieze and R. Kannan: Quick approximation to matrices and applications, *Combinatorica* **19** (1999), 175–220.

Local limit theorems for the giant component

Amin Coja-Oghlan

(joint work with Michael Behrisch and Mihyun Kang)

1. Background

We let $\mathcal{N}(H)$ signify the maximum order (i.e., number of vertices) of a component of a d-uniform hypergraph H. Furthermore, for all hypergraphs H we consider the vertex set $V(H)$ will consist of integers. Therefore, the subsets of $V(H)$ can be ordered lexicographically, and we call the lexicographically first component of H that has order $\mathcal{N}(H)$ the *largest component* of H. In addition, we denote by $\mathcal{M}(H)$ the size (i.e., number of edges) of the largest component of H. The random hypergraph $H_d(n,p)$ has the vertex set $V = \{1,\ldots,n\}$, and each of the $\binom{n}{d}$ possible edges is present with probability p independently of all others. Moreover, $H_d(n,m)$

is a uniformly distributed hypergraph with vertex set $V = \{1, \dots, n\}$ and with exactly m edges. Thus, $G(n,p) = H_2(n,p)$ and $G(n,m) = H_2(n,m)$.

In their two pioneering papers [4, 5], Erdős and Rényi studied the component structure of $G(n,m)$. In [4] they obtained the asymptotic number of connected graphs of order n and size $m = \frac{n}{2}(\Theta(1) + \ln n)$. The relevance of this result notwithstanding, possibly the most important contribution of [4] is that Erdős and Rényi solved this *enumerative* problem via *probabilistic* methods. Further, in [5] Erdős and Rényi studied (among other things) the component structure of *sparse* random graphs with $m = O(n)$ edges. The main result is that the order $\mathcal{N}(G(n,m))$ of the largest component undergoes a *phase transition* as $2m/n \sim 1$. Let us state actually state a more general version from [11], which covers d-uniform hypergraphs: let either $H = H_d(n,m)$ and $c = dm/n$, or $H = H_d(n,p)$ and $c = \binom{n-1}{d-1}p$; we refer to c as the *average degree* of H. Then the result is that

- if $c < (d-1)^{-1} - \epsilon$ for an arbitrarily small but fixed $\epsilon > 0$, then we have $\mathcal{N}(G(n,m)) = O(\ln n)$ w.h.p.
- By contrast, if $c > (d-1)^{-1} + \epsilon$, then $G(n,m)$ features a unique component of order $\Omega(n)$ w.h.p., which is called the *giant component*. More precisely, $\mathcal{N}(H) = (1-\rho)n + o(n)$ w.h.p., where ρ is the unique solution to the transcendental equation

$$\rho = \exp(c(\rho^{d-1} - 1)). \tag{1}$$

that lies strictly between 0 and 1. Furthermore, the second largest component has order $O(\ln n)$.

In this work we present a new, purely probabilistic approach for investigating the component structure of sparse random graphs and, more generally, hypergraphs in greater detail. More precisely, we obtain *local limit theorems* for the joint distribution of the order and size of the largest component of $H_d(n,m)$ or $H_d(n,p)$. From these local limit theorems we derive a number of consequences. For instance, we compute the asymptotic probability that H is connected, which yields an asymptotic formula for the number of connected hypergraphs of a given order and size.

In the case of *graphs* (i.e., $d = 2$) these results are either known or can be derived from prior work (in particular, Bender, Canfield, and McKay [2], Pittel and Wormald [8, 9], Stepanov [13]). By contrast, all our results are new for d-uniform hypergraphs with $d > 2$, and complement prior work due to Andriamampianina and Ravelomanana [1], Coja-Oghlan, Moore, and Sanwalani [3], Karoński and Łuczak [6, 7], and Ravelomanana and Rijamamy [10].

2. Results and Techniques

The main result is the following *local limit theorem* for the joint distribution of $\mathcal{N}$ and $\mathcal{M}$.

Theorem 1. *Let $d \geq 2$ be a fixed integer. For any two compact sets $\mathcal{I} \subset \mathbf{R}^2$, $\mathcal{J} \subset ((d-1)^{-1}, \infty)$, and for any $\delta > 0$ there exists $n_0 > 0$ such that the following holds. Let $p = p(n)$ be a sequence such that $c = c(n) = \binom{n-1}{d-1}p \in \mathcal{J}$ for all n and let $0 < \rho = \rho(n) < 1$ be the unique solution to (1). Further, let*

$$\begin{aligned}
\sigma^2_{\mathcal{N}} &= \frac{\rho\left(1-\rho+c(d-1)(\rho-\rho^{d-1})\right)}{(1-c(d-1)\rho^{d-1})^2}n,\\
\sigma^2_{\mathcal{M}} &= c^2\rho^d\frac{2+c(d-1)\rho^{2d-2}-2c(d-1)\rho^{d-1}+c(d-1)\rho^d-\rho^{d-1}-\rho^d}{(1-c(d-1)\rho^{d-1})^2}n\\
&\quad +(1-\rho^d)\frac{cn}{d},\\
\sigma^2_{\mathcal{N}\mathcal{M}} &= c\rho\frac{1-\rho^d-c(d-1)\rho^{d-1}(1-\rho)}{(1-c(d-1)\rho^{d-1})^2}n.
\end{aligned}$$

If $n \geq n_0$ and if ν, μ are integers and $x = \nu - (1-\rho)n$ and $y = \mu - (1-\rho^d)\binom{n}{d}p$ are such that $n^{-\frac{1}{2}}\binom{x}{y} \in \mathcal{I}$, then letting $P(x,y)$ equal

$$\frac{1}{2\pi\sqrt{\sigma^2_{\mathcal{N}}\sigma^2_{\mathcal{M}}-\sigma^2_{\mathcal{N}\mathcal{M}}}}\exp\left[-\frac{\sigma^2_{\mathcal{N}}\sigma^2_{\mathcal{M}}}{2(\sigma^2_{\mathcal{N}}\sigma^2_{\mathcal{M}}-\sigma^2_{\mathcal{N}\mathcal{M}})}\left(\frac{x^2}{\sigma^2_{\mathcal{N}}}-2\sigma^2_{\mathcal{N}\mathcal{M}}\frac{xy}{\sigma^2_{\mathcal{N}}\sigma^2_{\mathcal{M}}}+\frac{y^2}{\sigma^2_{\mathcal{M}}}\right)\right]$$

we have

$$(1-\delta)P(x,y) \leq P\left[\mathcal{N}(H_d(n,m)) = \nu \wedge \mathcal{M}(H_d(n,m)) = \mu\right] \leq (1+\delta)P(x,y).$$

To prove Theorem 1, we first establish a *univariate* local limit theorem for $\mathcal{N}(H_d(n,p))$. To this end, we expose the edges of $H_d(n,p)$ in two rounds, first with probability $p_1 = (1-\epsilon)p$ for some small but fixed $\epsilon > 0$, and then with probability $p_2 \sim \epsilon p_1$. Then, we show via *Stein's method* [12] that $\mathcal{N}(H_d(n,p_1))$ has a central limit theorem. In addition, we prove that *given the specific outcome* $n_1 = \mathcal{N}(H_d(n,p_1))$, the number of vertices that get attached to the largest component of $H_d(n,p_1)$ in the second round has a local limit theorem. Combining these two observations, we obtain the univariate local limit theorem for $\mathcal{N}$.

From this univariate result we then derive the bivariate limit theorem. To this end, we first observe that the univariate limit theorem easily implies a bivariate limit theorem for the joint distribution of $\mathcal{N}(H_d(n,p))$ and the number $\bar{\mathcal{M}}(H_d(n,p))$ of edges *outside* the largest component. Then, we use the relation

$$\begin{aligned}
(2)\qquad & P\left[\mathcal{N}(H_d(n,p)) = \nu \wedge \bar{\mathcal{M}}(H_d(n,p)) = \mu\right]\\
&= \sum_{m=0}^{\binom{n}{d}} P\left[\mathcal{N}(H_d(n,m)) = \nu \wedge \bar{\mathcal{M}}(H_d(n,m)) = \mu\right] \cdot P\left[Bin\left[\binom{n}{d}, p\right] = m\right].
\end{aligned}$$

Since we already know the l.h.s. of (2), and because the equation is true for all values of p, it is possible to solve (2) for $P\left[\mathcal{N}(H_d(n,m)) = \nu \wedge \bar{\mathcal{M}}(H_d(n,m)) = \mu\right]$ via Fourier analysis. Furthermore, as $\bar{\mathcal{M}}(H_d(n,m)) = m - \bar{\mathcal{M}}(H_d(n,m))$, we thus obtain a local limit theorem for the joint distribution of $\mathcal{N}(H_d(n,m))$ and $\mathcal{M}(H_d(n,m))$. Finally, this local limit theorem for $H_d(n,m)$ entails Theorem 1.

Moreover, given that the largest component of $H_d(n,p)$ has order ν and size μ, it is a uniformly distributed connected graph with these parameters. Therefore, Theorem 1 entails the following result on the number of connected hypergraphs of a given order and size.

Theorem 2. *Let $d > 2$ be a fixed integer, and let $m = m(n)$ be a sequence of integers such that $c = c(n) = dm/n > d(d-1)^{-1} + \epsilon$ for an arbitrarily small but fixed $\epsilon > 0$. Further, let $0 < \rho = \rho(n) < 1$ be the unique solution to $\rho = \exp\left[-c(1-\rho)\cdot\frac{1-\rho^{d-1}}{1-\rho^d}\right]$, and set $\phi_d(c) = \rho^{\frac{\rho}{1-\rho}}(1-\rho)\left(\frac{1-\rho^d}{(1-\rho)^d}\right)^{\frac{c}{d}}$. Then the probability $c_d(n,m)$ that $H_d(n,m)$ is connected satisfies*

$$c_d(n,m) \sim \frac{1-\rho^d-(1-\rho)c(d-1)\rho^{d-1}}{\sqrt{(1-\rho^d+c(d-1)(\rho-\rho^{d-1}))(1-\rho^d)-cd\rho(1-\rho^{d-1})^2}} \times \exp\left(\frac{c(d-1)(\rho-2\rho^d+\rho^{d-1})}{2(1-\rho^d)}\right)\cdot\phi_d(\rho,c)^n$$

References

[1] T. Andriamampianina and V. Ravelomanana, Enumeration of connected uniform hypergraphs, Proceedings of FPSAC 2005.

[2] E.A. Bender, E.R. Canfield and B.D. McKay, The asymptotic number of labeled connected graphs with a given number of vertices and edges, *Random Structures and Algorithms* **1** (1990), 127–169.

[3] A. Coja-Oghlan, C. Moore and V. Sanwalani, Counting connected graphs and hypergraphs via the probabilistic method, To appear in *Random Structures and Algorithms*.

[4] P. Erdős and A. Rényi, On random graphs I., *Publicationes Mathematicae Debrecen* **5** (1959), 290–297.

[5] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.* **5** (1960), 17–61.

[6] M. Karoński and T. Łuczak, The number of connected sparsely edged uniform hypergraphs, *Discrete Math.* **171** (1997), 153–168.

[7] M. Karoński and T. Łuczak, The phase transition in a random hypergraph, *J. Comput. Appl. Math.* **142** (2002), 125–135.

[8] B. Pittel and Wormald, N.C., Asymptotic enumeration of sparse graphs with a minimum degree constraint, *J. Combinatorial Theory, Series A* **101** (2003), 249.

[9] B. Pittel and N. C. Wormald, Counting connected graphs inside out, *J. Combin. Theory, Series B* **93** (2005), 127–172.

[10] V. Ravelomanana and A. L. Rijamamy, Creation and growth of components in a random hypergraph process, Preprint (2005).

[11] J. Schmidt-Pruzan and E. Shamir, Component structure in the evolution of random hypergraphs, *Combinatorica* **5** (1985), 81–94.

[12] C. Stein, A bound for the error in the normal approximation to the distribution of a sum of dependent variables, em Proc. 6th Berkeley Symposium on Mathematical Statistics and Probability (1970), 583–602.

[13] V. E. Stepanov, On the probability of connectedness of a random graph $g_m(t)$, *Theory Prob. Appl.* **15** (1970), 55–67.

Deterministic random walks on the two-dimensional grid

Benjamin Doerr
(joint work with Tobias Friedrich)

The rotor-router model is a simple deterministic process suggested by Jim Propp. It can be viewed as an attempt to derandomize random walks on graphs. So far, the "Propp machine" has mainly been regarded on infinite grids $\mathbb{Z}^d$. There, each vertex $x \in \mathbb{Z}^d$ is associated with a "rotor" and a cyclic permutation (called "rotor sequence") of the $2d$ cardinal directions of $\mathbb{Z}^d$. While in a random walk a particle (we prefer to think of them as "chips") leaves a vertex in a random direction, chips in the Propp model always go into the direction the current rotor is pointing. After a chip is sent, the rotor is rotated according to the fixed rotor sequence. This ensures the chips are distributed highly evenly among the neighbors.

The Propp machine has attracted considerable attention recently. It has been shown that it closely resembles a random walk in several respects. The first result is due to [7, 8] who compared random walk and Propp machine in an aggregating model called *Internal Diffusion-Limited Aggregation (IDLA)* [3]. There, each chip starts at the origin of $\mathbb{Z}^d$ and walks till it reaches an unoccupied site, which it then occupies. In the random walk model it is well known that the shape of the occupied locations converges to a Euclidean ball in $\mathbb{R}^d$ [6]. It has been shown that the fluctuations around the limiting shape are bounded by $O(r^{1/3})$ with high probability [5], where r denotes the radius of the ideal ball. Recently, [7, 8] proved that also for the Propp machine the shape of occupied locations converges to a Euclidean ball. [4] showed experimentally that for circular permutations of the rotors, after three million chips the difference of the radius of the inscribed and circumscribed circle is approximately 1.61. Hence, the occupied locations of the aggregating Propp machine form a perfect circle much better than in the random walk model.

[1] compared the Propp machine and the random walk in terms of the *single vertex discrepancy*. Starting both models with the same initial configurations (that is, chips on vertices and rotor directions) they perform a fixed number of simultaneous steps of all chips. They showed that (apart from a technicality which we defer to the end of this paragraph) for all grids $\mathbb{Z}^d$ the maximal difference of the number of chips of the Propp machine and the expected number of chips of the random walk ("single vertex discrepancy") over all initial configurations on all vertices and at all times is bounded by a constant c_d, which only depends on the dimension. In particular, c_d is independent of the number of chips, their locations and the runtime. It remains to fix the technicality omitted above. Since $\mathbb{Z}^d$ is a bipartite graph, chips on different classes of the bipartition do not interact in the random walk model. This is not true in the Propp model, where they do interact via the rotors. We therefore need to assume that our initial configurations have chips only on one bipartition class.

For the case $d = 1$, that is, the graph being the infinite path, [2] showed among other results that this constant c_1 is approximately 2.29. They further proved

that this discrepancy maximal discrepancy can already show up if each vertex has an odd number of chips only once. Note that if a vertex has an even number of chips at some time, then the Propp machine and the random walk (in expectation) perform identically on these chips.

In this paper, we rigorously analyze the Propp machine on the two-dimensional grid $\mathbb{Z}^2$. A particular difference to the one-dimensional case is that now there are two non-isomorphic orders in which the four neighbors are served. The first are clockwise and counterclockwise orders of the four cardinal directions. These are called *circular* rotor sequences. All other orders turn the rotor by 180° at one time and are called *non-circular* rotor sequences. We prove $7.832 \leq c_2 \leq 7.985$ for circular and by $7.286 \leq c_2 \leq 7.439$ for non-circular rotor sequences. To the best of our knowledge this is the first paper which shows an influence of the rotor sequence. We also characterize the respective worst-case configurations. In particular, we prove that the maximal single vertex discrepancy can only be reached if there are vertices which send a number of chips not divisible by four at at least three different times.

References

[1] Joshua Cooper and Joel Spencer, Simulating a random walk with constant error, *Combinatorics, Probability and Computing*, To appear, preliminary version available from arXiv:math/0402323.

[2] Joshua Cooper, Benjamin Doerr, Joel Spencer, and Gábor Tardos, Deterministic random walks on the integers, *European Journal of Combinatorics*, To appear, preliminary version available from arXiv:math/0602300.

[3] Persi Diaconis and William Fulton, A growth model, a game, an algebra, Lagrange inversion, and characteristic classes, *Rend. Sem. Mat. Univ. Pol. Torino*, 49 (1): 95–119, 1990.

[4] Michael Kleber, Goldbug Variations, *The Mathematical Intelligencer*, 27 (1), 2005.

[5] Gregory F Lawler, Subdiffusive fluctuations for internal diffusion limited aggregation, *Annals of Probability*, 23 (1): 71–86, 1995.

[6] Gregory F. Lawler, Maury Bramson, and David Griffeath, Internal diffusion limited aggregation, *Annals of Probability*, 20 (4): 2117–2140, 1992.

[7] Lionel Levine and Yuval Peres, Spherical asymptotics for the rotor-router model in $\mathbb{Z}^d$, 2005, arXiv:math/0503251.

[8] Lionel Levine and Yuval Peres, The rotor-router shape is spherical, *The Mathematical Intelligencer*, 27 (3): 9–11, 2005.

Witnesses for non-satisfiability of dense random 3CNF formulas

Uriel Feige

(joint work with Jeong Han Kim and Eran Ofek)

We consider random 3CNF formulas with n variables and m clauses. It is well known that when $m > cn$ (for a sufficiently large constant c), most formulas are not satisfiable. However, it is not known whether such formulas are likely to have polynomial size witnesses that certify that they are not satisfiable. A value of $m \simeq n^{3/2}$ was the forefront of our knowledge in this respect. When $m > cn^{3/2}$, such witnesses are known to exist, based on spectral techniques. When $m < n^{3/2-\epsilon}$,

it is known that resolution (which is a common approach for refutation) cannot produce witnesses of size smaller than 2^{n^ϵ}. Likewise, it is known that certain variants of the spectral techniques do not work in this range.

In the current paper we show that when $m > cn^{7/5}$, almost all 3CNF formulas have polynomial size witnesses for non-satisfiability. We also show that such a witness can be found in time $2^{O(n^{0.2}\log n)}$, whenever it exists. Our approach is based on an extension of the known spectral techniques, and involves analyzing a certain fractional packing problem for random 3-uniform hypergraphs.

For more details, see the proceedings of the 47th Annual Symposium on Foundations of Computer Science, 2006, pages 497–506.

Intersecting families are essentially contained in juntas

Ehud Friedgut
(joint work with Irit Dinur)

1. Introduction

The following is based on a paper that is posted on my webpage:
http://www.math.huji.ac.il/∼ehudf/docs/intersecting1.pdf

A family $\mathcal{J}$ of subsets of $\{1,\ldots,n\}$ is called a j-junta if there exists $J \subseteq \{1,\ldots,n\}$, with $|J| = j$, such that the membership of a set S in $\mathcal{J}$ depends only on $S \cap J$.

In this paper we provide a simple description of intersecting families of sets. Let n and k be positive integers with $k < n/2$, and let $\mathcal{A}$ be a family of pairwise intersecting subsets of $\{1,\ldots,n\}$, all of size k. We show that such a family is essentially contained in a j-junta $\mathcal{J}$ where j does not depend on n but only on the ratio k/n and on the interpretation of "essentially".

When $k = o(n)$ we prove that every intersecting family of k-sets is almost contained in a dictatorship, a 1-junta (which by the Erdős-Ko-Rado theorem is a maximal intersecting family): for any such intersecting family $\mathcal{A}$ there exists an element $i \in \{1,\ldots,n\}$ such that the number of sets in $\mathcal{A}$ that do not contain i is of order $\binom{n-2}{k-2}$ (which is approximately $\frac{k}{n-k}$ times the size of a maximal intersecting family).

Our methods combine traditional combinatorics with results stemming from the theory of Boolean functions and discrete Fourier analysis.

Let us begin with some notation. Let $[n] = \{1,2,\ldots n\}$ and $\binom{[n]}{k} = \{T \in [n] : |T| = k\}$. We shall refer to the elements of $\binom{[n]}{k}$ simply as k-sets. If J is a ground set we denote the family of all subsets of J by 2^J. General families of subsets of $[n]$ are denoted by script letters $\mathcal{A}, \mathcal{B}, \mathcal{J}$, etc. and subsets of $[n]$ by capital letters S, T, etc. Two sets will be called intersecting if their intersection is nonempty and a family of sets will be called intersecting if its elements are pairwise intersecting. Two families $\mathcal{A}$ and $\mathcal{B}$ will be called cross intersecting if every member of $\mathcal{A}$

intersects every member of $\mathcal{B}$. The following definition is important enough in this paper to deserve separate indentation.

Definition. *A family of sets $\mathcal{J}$ will be called a j-junta if there exists a subset $J \subseteq [n]$ with $|J| = j$ such that membership of a set in $\mathcal{A}$ is determined only by its intersection with J, in other words there exists a family $\mathcal{J}^* \subseteq 2^J$ such that $\mathcal{J} = \{T \subseteq [n] : (T \cap J) \in \mathcal{J}^*\}$. In this case we will say that $\mathcal{J}$ is the junta generated by $\mathcal{J}^*$.*

2. Results

Questions regarding a family of sets given some simple information regarding their intersection pattern are among the most fundamental in combinatorics. Perhaps the first and most basic of these questions is answered by the Erdős-Ko-Rado theorem [4] which deals with an *intersecting family* of sets. If $\mathcal{A} \subset \binom{[n]}{k}$ is intersecting, with $k \leq n/2$, the EKR theorem states that $\mathcal{A}$ is of size no more than $\binom{n-1}{k-1}$ and that if k is strictly less than $n/2$ this bound is attained only by a family of all k-subsets containing some fixed element. Many other theorems in extremal set theory provide such elegant characterizations of the largest possible family of sets given certain information on their intersections. A notable prototype of such a theorem is the Ahlswede-Khachatrian theorem [1] which deals with a family of k-sets given that the pairwise intersections of its elements are of size at least r for some fixed integer r. There, once again, the extremal families are j-juntas where j can be bounded by some function of r and k/n.

The principal principle that we prove in this paper shows that such results should come as no surprise since (roughly speaking) *every* intersecting family of sets is contained in a family with such a simple description, namely every intersecting family is essentially contained in a junta.

To make this more concrete, we want to state a theorem that says that for every intersecting family $\mathcal{A} \subset \binom{[n]}{k}$ and for every $\varepsilon > 0$ there exists an integer j which depends only on ε and on k/n, a set $J \subset [n]$ with $|J| = j$ and a non-trivial family $\mathcal{J}^* \subset 2^J$ so that if $\mathcal{J}$ is the junta generated by $\mathcal{J}^*$ then $|\mathcal{A} \setminus \mathcal{J}| \leq \varepsilon\binom{n}{k}$. However, when $k = pn$ for constant p, this holds rather trivially by taking a very large junta: if we take $j \gg \log_{k/n}(\varepsilon)$, take an arbitrary $J \subset [n]$ of size j, and define $\mathcal{J}^* = 2^J \setminus \emptyset$ and $\mathcal{J}$ the generated junta this already implies $|\mathcal{A} \setminus \mathcal{J}| < \varepsilon\binom{n}{k}$, for simple statistical reasons that have nothing to do with the structure of $\mathcal{A}$.

Such a trivial hindrance to providing an accurate description of intersecting families does not exist for $k = o(n)$ and at this point we break our treatment of intersecting families contained in $\binom{[n]}{k}$ into two cases: $k = pn$ for some constant $0 < p < 1/2$ and $k = o(n)$.

The case $k = pn$, p constant, $0 < p < \frac{1}{2}$. As illustrated above, finding a junta that contains our intersecting family up to an ε fraction is hardly an accomplishment. It would be far more gratifying if we could find a j-junta $\mathcal{J}$ which is itself intersecting or, in other words, if we could guarantee that $\mathcal{J}^*$ is an intersecting family. This would, in a sense, provide a simple explanation to the fact that $\mathcal{A}$ is intersecting.

We conjecture that this in fact is true, see Conjecture 2 below. But for now we make do with proving that $\mathcal{J}^*$ is almost-intersecting. The definition of being *almost-intersecting* is subtle and can be found in the full paper. Our main theorem in this section is as follows,

Theorem 1. *There exists a function $j(p,\varepsilon)$ such that the following holds. Let $0 < p < 1/2$ and let $\varepsilon > 0$. Let $j = j(\varepsilon, p)$ and let k and n be positive integers with $j \ll k = pn$. Then for every intersecting family $\mathcal{A} \subset \binom{[n]}{k}$ there exists a j-junta $\mathcal{J}$ generated by a family $\mathcal{J}^*$ such that*

(1) $|\mathcal{A} \setminus \mathcal{J}| \leq \varepsilon\binom{n}{k}$.
(2) *$\mathcal{J}^*$ is ε-almost-intersecting in $\mu_p(\{0,1\}^j)$.*

In other words there is a simple (almost) necessary condition for a set to belong to $\mathcal{A}$, where the condition depends only on j elements. Note that finding such a simple sufficient condition is hopeless since one can take arbitrarily complicated subfamilies of $\mathcal{A}$ (e.g. a random subfamily) which, of course, will still be intersecting families. As stated above we hope that Theorem 1 can be improved:

Conjecture 2. *The family $\mathcal{J}^*$ guaranteed by Theorem 1 can be taken to be an intersecting family.*

Since the Erdős-Ko-Rado theorem breaks down when $k > n/2$ it comes as no surprise that the function $j(\varepsilon, p)$ guaranteed by Theorem 1 explodes as p approaches $1/2$. However when p approaches 0 the situation is simpler.
The case $k = o(n)$. As $\frac{k}{n}$ goes to 0, i.e. $k = o(n)$, we can state a theorem that is perhaps more satisfying and surprising. It turns out that when k is small compared to n then *every* intersecting family of k-sets is practically contained in a maximal intersecting family, a family of all sets containing one fixed element. To better appreciate the results that follow let us first calibrate the quantities that appear in the statements. If $k = o(n)$ then, denoting $p = k/n$,

$$\binom{n}{k} \gg \binom{n-1}{k-1} \gg \binom{n-2}{k-2} \gg \ldots \gg \binom{n-r}{k-r},$$

as $\frac{\binom{n-r}{k-r}}{\binom{n}{k}} \approx p^r$. The following theorems hold for any natural numbers $k < n/2$ but are only interesting when $k = o(n)$ otherwise the hidden constants can be set to make the theorems hold trivially. As usual we denote $p = k/n$.

Theorem 3. *There exists a constant $c(1)$ such that if $\mathcal{A} \subset \binom{[n]}{k}$ is an intersecting family then there exists an element $i \in [n]$ such that all but $c(1) \cdot \binom{n-2}{k-2}$ of the sets in $\mathcal{A}$ contain i.*

In other words every intersecting family is practically contained in a maximal intersecting family.

3. Remarks

- Note that up to the value of $c(1)$ the theorem is sharp as shown by the example of all k-sets that contain at least two elements out of a given set of three elements.
- Theorem 3 is a special case of Theorem 4 below, but we state it separately because we find it to be aesthetically pleasing that the junta involved in this case is a dictatorship, a 1-junta.

The following theorem offers a higher degree of precision at the inevitable price of needing a larger junta to capture the intersecting family.

Theorem 4. *There exists functions $j(r), c(r)$ such that for any integers $1 < j(r) < k < n/2$, if $\mathcal{A} \subset \binom{[n]}{k}$ is an intersecting family with $|\mathcal{A}| > c(r) \cdot \binom{n-(r+1)}{k-(r+1)}$ then there exists an intersecting j-junta $\mathcal{J}$ with $j \leq j(r)$ and*

$$|\mathcal{A} \setminus \mathcal{J}| \leq c(r) \cdot \binom{n-(r+1)}{k-(r+1)} = O\left(p^{r+1} \cdot \binom{n}{k}\right).$$

Note that unlike the $k = pn$ case we are able to ensure that the junta $\mathcal{J}$ is in fact intersecting.

The above two theorems also have cross-intersecting-families versions which, besides of being interesting in their own right are also useful for our induction based proofs.

References

[1] R. Ahlswede and L. Khachatrian, The complete intersection theorem for systems of finite sets, *European J. Combin.*, 18:125–136, 1997.

[2] I. Dinur, E. Friedgut, and O. Regev, Independent sets in graph powers are almost contained in juntas, Preprint.

[3] I. Dinur and S. Safra, On the hardness of approximating minimum vertex cover, *Annals of Mathematics*, 162(1):439–485, 2005.

[4] P. Erdős, C. Ko, and R. Rado, Intersection theorems for systems of finite sets, *Quart. J. Math. Oxford, ser. 2*, 12:313–318, 1961.

[5] P. Frankl and N. Tokushige, Some inequalities concerning cross-intersecting families, *Combinatorics Probability and Computing*, 7:247–260, 1998.

[6] E. Friedgut, Boolean functions with low average sensitivity depend on few coordinates, *Combinatorica*, 18(1):27–35, 1998.

[7] Z. Füredi, Cross-intersecting families of finite sets, *J. Combin. Theory Ser. A*, 72:332–339, 1995.

[8] G. O. H. Katona, A theorem of finite sets, In P. Erdo, editor, *Theory of Graphs.* Akade'miai Kiado' and Academic Press, 1968.

[9] J. Kruskal, The number of simplices in a complex, In R. Bellman, editor, *Mathematical Optimization Techniques*, pages 251–278, University of California Press, 1963.

[10] G. Margulis, Probabilistic characteristics of graphs with large connectivity, *Prob. Peredachi Inform.*, 10:101–108, 1974.

[11] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, Noise stability of functions with low influences: invariance and optimality, In *Proc. 46th IEEE Symp. on Foundations of Computer Science*, 2005.

[12] L. Russo, An approximate zero-one law, *Z. Wahrsch. werw. Gebiete*, 61:129–139, 1982.

[13] Z. Tuza, Critical hypergraphs and intersecting set-pair systems, *J. Combin. Theory Ser. B*, 39(2), 1985.

Line-of-sight networks

Alan Frieze
(joint work with Jon Kleinberg, R. Ravi and Warren Debany)

1. Introduction

Most of today's approaches to wireless computing and communications are built on architectures where base stations connect the wireless devices to a supporting infrastructure. However, since the overwhelming trend is to transmit information in packets, over standard protocols, a dominant focus in the wireless research community is on more decentralized approaches where nodes cooperate to relay packets on behalf of other nodes. This focus is at the heart of current work on mobile ad hoc networks (MANETs) [5, 6].

Such networks can be viewed as consisting of a collection of nodes, representing wireless devices, positioned at various points in some physical region. The (wireless) "links" of the network, joining pairs of nodes that can directly communicate with one another, are predominantly short-range and constrained by line-of-sight; this is an inevitable result of the scarcity of radio frequency (RF) spectrum and physical constraints on the propagation of RF and optical signals. The ways in which these physical limits on direct communication affect the overall performance of the network is a fundamental issue that motivates much of the theoretical work in this area.

2. Random Geometric Graphs

Given this framework, random geometric graphs have emerged as a dominant model for theoretical analysis of distributed wireless networks. One places n points uniformly at random in a geometric region (typically a disc or a square), and then, for a *range parameter* r, one connects each pair of nodes that are within distance r of one another. This model is the subject of a recent book by Penrose [8], and we refer the reader there for extensive background; we also note that the enormously influential work of Gupta and Kumar on the capacity of wireless networks is framed this model as well [3, 4].

One of the most basic questions is to determine how the probability of connectivity of a random geometric graph depends on the number of nodes n and the range parameter r. A canonical result here is the following theorem of Penrose [7]. If we place n points uniformly at random in a unit square, and then continuously increase the range parameter r, with high probability the resulting geometric graph becomes k-connected at the smallest value of r for which there are no nodes of degree $< k$. In other words, the graph becomes k-connected at the moment that

all trivial obstacles to k-connectivity (i.e. low-degree nodes) disappear. An analogous type of result is familiar from the theory of classical Erdős-Rényi random graph models [1]. (For further results and discussion concerning thresholds for properties in random geometric graphs, see Goel, Rai, and Krishnamachari [2].)

For modeling distributed wireless networks, the assumption of random node placement has proved to be a reasonable abstraction for the lack of structure in node locations, given that most frameworks for ad hoc networks assume some arbitrary initial "scattering" of nodes, or that nodes reach their positions as a result of arbitrary mobility. More problematic is the fact that the analysis takes place in regions with no obstructions — in other words, that a node can communicate with *all* other nodes within distance r. This is at odds with the underlying constraints in many applications of distributed wireless networks, where there can generally be a large number of obstructions limiting communication between nearby nodes due to a lack of direct line-of-sight contact.

In other words, while random geometric graphs model wireless networks in open spaces, we lack a corresponding model for wireless networks in some of their most common domains: urban settings, large indoor environments, or any other context in which there are obstacles limiting visibility. With such a model would come the ability to address a range of basic theoretical problems. In particular, we are guided by the following genre of question:

> How do connectivity and other structural properties of random geometric graphs change once we introduce line-of-sight constraints?

An understanding of such issues could help provide a framework for reasoning more generally about the performance of distributed wireless networks in obstructed environments.

3. The present work: Connectivity in line-of-sight networks

In this paper, we propose a random-graph model incorporating both range limitations and line-of-sight constraints, and we prove asymptotically tight results for k-connectivity. We also consider related structural questions, including the emergence of a giant component, as well as some of the algorithmic issues raised by the model.

To motivate the model, consider a stylized abstraction of limited-range wireless communication in an urban environment: there are n streets running east-west, n avenues running north-south, and wireless nodes can be placed at intersections of streets and avenues. Each node has *range* ω — it can see up to ω blocks north and south along the avenue it lies on, and up to ω blocks east and west along the street it lies on.

More concretely, we have an underlying set T of lattice points $\{(x, y) : x, y \in \{1, 2, \ldots, n\}\}$. We measure distance using the L_1 metric, though to prevent complications arising from boundary effects in this presentation, we define the distance between points as though they form a torus:

$$d((x, y), (x', y')) = \min(|x - x'|, n - |x - x'|) + \min(|y - y'|, n - |y - y'|).$$

For a specified *range parameter* ω, we say that two points are *mutually visible* if they are in the same row or the same column of the torus, and if they are within distance at most ω from one another. We view the range ω as implicitly being a function of n, and in this paper we will make the assumption that ω is asymptotically bounded below by $\ln n$ and above by some polynomial in n; specifically, we assume $\ln n = o(\omega)$ and that $\omega = O(n^\delta)$ for a value of $\delta < 1$ to be specified below.

We now study the random graph G that results if, for some *placement probability* $p > 0$, we locate a node at each point of T independently with probability p, and then connect those pairs of nodes that are mutually visible. Our main result states, roughly, that the smallest value of p at which G becomes k-connected with high probability is asymptotically the same as the smallest value of p at which the minimum degree in G is k with high probability.

We state this theorem about k-connectivity as follows. First, we write $\omega = n^\delta$ where we assume that $\omega \gg \ln n$ and $\delta < \frac{6}{8k+7}$. Note that we **do not** preclude the case where $\delta = o(1)$.

Theorem 1. *Let $k \geq 1$ be a fixed positive integer and let $p = \frac{(1-\frac{1}{2}\delta)\ln n + \frac{k}{2}\ln\ln n + c_n}{2\omega}$. Then*

$$\lim_{n\to\infty} \Pr(G \text{ is } k\text{-connected}) = \begin{cases} 0 & c_n \to -\infty \\ e^{-\lambda_k} & c_n \to c \\ 1 & c_n \to \infty \end{cases}$$

where

$$\lambda_k = \frac{2^{k-2}(1-\frac{1}{2}\delta)^k e^{-2c}}{(k-1)!}.$$

The proof of this result requires techniques quite different from the analysis of standard geometric random graphs, due to the line-of-sight constraints. One way to appreciate why this appears necessary is to consider that, as we vary ω, the resulting model interpolates between two well-known, but qualitatively different random graph models. When $\omega = 1$, so that a node can only see neighboring points, we have site percolation on a lattice. At the other extreme, when $\omega = n$ and nodes can see all points in their row and column, it is easy to see that the model is equivalent to a purely graph-theoretic one in which we start with the complete bipartite graph $K_{n,n}$ and keep each edge with probability p. Note that our bounds on ω preclude either of these exact extremes, but our analysis for the "middle region" of ω that we consider involves ingredients from both extremes, combining techniques from "classical" random graph analysis with the combinatorics of the underlying grid of points.

We also consider the emergence of a giant component in our model.

Theorem 2. (a) *If $p = \frac{c}{\omega}$ where $c > 1$ and $\omega \to \infty$ then* whp *G contains a component with at least $(1-o(1))(1-x_c^2)n^2/\omega$ vertices, where x_c is the unique solution in $(0,1)$ of $xe^{-x} = ce^{-c}$.*

(b) *If $p = \frac{c}{\omega}$ where $c < 1/(4e)$ and $\omega \to \infty$ then* whp *the largest component in G has size $O(\ln n)$.*

We further consider the problem of how nodes in such a random graph can construct paths between each other, possessing knowledge of their own coordinates but otherwise having only local information.

Theorem 3. *Let $p = C \ln n/\omega$ for a constant C to be specified below. There is a decentralized algorithm that, given s and t, with high probability constructs an s-t path with $O(d(s,t)/\omega + \ln n)$ edges while involving $O(d(s,t)/\omega + \omega \ln n)$ nodes in the computation. (This is nearly optimal, even with global information).*

Finally, we consider a basic algorithmic problem in a non-random version of the line-of-sight model: given an input set of nodes, we would like to add a small set of additional nodes so that the full set becomes connected. More concretely, suppose we are given a set of nodes at points $X \subset T$, such that the graph on X (defined by visibility with respect to the range parameter ω) is not connected. We would like to add further nodes, at a set $Y \subset T$, where Y should be as small as possible subject to the constraint that the graph on $X \cup Y$ should be connected. We think of the additional nodes Y as "relays" that connect the original nodes in X under line-of-sight constraints; as a result, we refer to this as the *Relay Placement* problem.

Theorem 4. *There is a polynomial-time algorithm that produces a solution to the relay replacement problem whose total cost is within a factor of* 6.2 *of optimal.*

References

[1] B. Bollobás. Random Graphs (2nd edition). Cambridge University Press, 2001.

[2] A. Goel, S. Rai, and B. Krishnamachari. Sharp thresholds for monotone properties in random geometric graphs. ACM Symposium on Theory of Computing, 2004.

[3] P. Gupta and P. Kumar, "Critical power for asymptotic connectivity in wireless networks," in Stochastic Analysis, Control,Optimization and Applications: A Volume in Honor of W.H. Fleming, W.M. McEneaney, G. Yin, and Q. Zhang (Eds.), Birkhauser, Boston, 1998.

[4] P. Gupta and P. Kumar, "The capacity of wireless networks," IEEE Transactions on Information Theory, vol. 46, no. 2, March 2000,pp. 388-404, and "A correction to the proof of a lemma in 'The capacity of wireless networks,"' IEEE Transactions on Information Theory, vol. 49, no. 11, November 2000, p. 3117.

[5] M. Mauve, J. Widmer, H. Hartenstein. A Survey on Position-Based Routing in Mobile Ad Hoc Networks. IEEE Network, Nov/Dec 2001.

[6] Mobile Ad-hoc Networks (MANET) Charter, http://www.ietf.org/html.charters/manet-charter.html

[7] M.D. Penrose, "On k-connectivity for a geometric random graph," Random Structures and Algorithms, vol. 15, no. 2, pp. 145-164, 1999.

[8] M.D. Penrose. Random Geometric Graphs, vol. 5 of Oxford Studies in Probability. Oxford University Press, 2003.

Connectivity of random addable graphs

Stefanie Gerke
(joint work with Paul Balister and Béla Bollobás)

We call a non-empty class $\mathcal{A}$ of labelled graphs *addable* if for each graph $G \in \mathcal{A}$ and any two vertices u and v in distinct components of G, the graph obtained by adding the edge $\{u, v\}$ to G is in $\mathcal{A}$. Examples of addable graph classes include forests, planar graphs, and triangle-free graphs. The set of graphs in $\mathcal{A}$ with vertex set $\{1, \ldots, n\}$ is denoted by $\mathcal{A}_n$. McDiarmid, Steger, and Welsh [4] conjecture that for an addable class of graphs with the property that $\mathcal{A}_n \neq \emptyset$ for all sufficiently large n, an element R_n drawn uniformly at random from $\mathcal{A}_n$ satisfies

$$\liminf_{n\to\infty} \mathbb{P}[R_n \text{ is connected}] \geq \frac{1}{\sqrt{e}}.$$

Let us remark that one cannot increase $e^{-1/2}$ as the class of forests shows. McDiarmid, Steger, and Welsh [3, 4] proved the conjecture when $e^{-1/2}$ is replaced by e^{-1}. We improve this constant to $e^{-0.7983}$; see [2].

To prove the result we first show that it is sufficient to consider addable classes of graphs that consist of forests only. Let $\mathcal{F}$ be any such addable class of forests, and let $\mathcal{F}_n^i$ be the set of graphs on $\{1, \ldots, n\}$ in $\mathcal{F}$ that have i components. We show that, for $i = 1, \ldots, \lfloor \log n \rfloor$ and for every $\varepsilon > 0$, we have $|\mathcal{F}_n^{i+1}|/|\mathcal{F}_n^i| \leq (0.7983 + \varepsilon)/i$ whenever n is sufficiently large. For $i = \lfloor \log n \rfloor, \ldots, n$, we obtain $|\mathcal{F}_n^{i+1}|/|\mathcal{F}_n^i| \leq 1/i$. The claimed result on the probability that a random element of an addable class is connected then follows by an easy calculation.

To show the upper bounds on $|\mathcal{F}_n^{i+1}|/|\mathcal{F}_n^i|$ we consider the bipartite graph $B = (\mathcal{F}_n^i \cup \mathcal{F}_n^{i+1}, E)$ where two forests $F \in \mathcal{F}_n^i$ and $F' \in \mathcal{F}_n^{i+1}$ are connected by an edge if F' can be obtained by deleting an edge of F. For any weighting $w : E \to R$ of the edges of B, a simple double-counting argument yields

$$|\mathcal{F}_n^i| \max_{u \in \mathcal{F}_n^i} \sum_{x:\{u,x\}\in E} w(\{u, x\}) \geq ||\mathcal{F}_n^{i+1}| \min_{v \in \mathcal{F}_n^{i+1}} \sum_{x:\{v,x\}\in E} |w(\{v, x\}).$$

Thus to give an upper bound on $|\mathcal{F}_n^{i+1}|/|\mathcal{F}_n^i|$ it remains to find a good weighting w of the edges of B, and to determine the above minimum and maximum. For $i = \lfloor \log n \rfloor, \ldots, n$, we give weight 1 to each edge and the claimed result follows easily. For $i = 1, \ldots, \lfloor \log n \rfloor$, we use the following family $(w_\alpha)_{\alpha>0}$ of weightings. For $0 < \alpha < 1$ and $\{F, F'\} \in E$, let $e = \{a, b\}$ be the edge in F the deletion of which yields F'. We assign the weight $(d(a)d(b))^{-\alpha}$ to $\{F, F'\}$ where $d(a)$ and $d(b)$ are the degrees of a and b in F.

It is not hard to see that, for sufficiently large n, $\min_{v \in \mathcal{F}_n^{i+1}} \sum_{x:\{v,x\}\in E} w(\{v, x\})$ is achieved by the forest consisting of i isolated vertices and a path on $n-i$ vertices. Thus

$$\min_{v \in \mathcal{F}_n^{i+1}} \sum_{x:\{v,x\}\in E} w(\{v, x\}) = \binom{i}{2} + i(2^{1-\alpha} + (n - i - 2)3^{-\alpha}) \geq 3^{-\alpha} i \left(n - \frac{i+1}{2} \right).$$

To give a bound on the maximum, it suffices to consider trees T and to find an upper bound on

$$R_{-\alpha}(T) = \sum_{\{a,b\}\in E(T)} (d(a)d(b))^{-\alpha}.$$

This quantity is called the generalized Randić index. Randić [5] introduced this measure, with $\alpha = 1$ and $\alpha = 1/2$, to give a theoretical characterization of molecular branching. For every $\alpha > 0$, we find an effectively computable constant $\beta_c = \beta_c(\alpha)$ such that for all trees T on $n > 2$ vertices, $R_{-\alpha}(T) \le \beta_c(n+1)$ [2]. We also construct infinitely many trees such that $R_{-\alpha}(T) \ge \beta_c(n-1)$. Since $\beta_c < 1$ this means that the upper bound is best possible up to the constant term. Thus

$$\frac{|\mathcal{F}_n^{i+1}|}{|\mathcal{F}_n^i|} \le \frac{\beta_c(\alpha)n + O(i)}{3^{-\alpha}i(n-(i+1)/2)} = \frac{3^\alpha \beta_c(\alpha)}{i} + O(i/n).$$

Setting $\alpha = 0.868$ yields $\beta_c(\alpha) \le 0.30762$ and the claimed result on the connectivity of a random element of an addable class follows.

References

[1] P. Balister, B. Bollobás, S. Gerke, The Randić index of trees, submitted 2006.

[2] P. Balister, B. Bollobás, S. Gerke, On the connectivity of random graphs from addable classes, in preparation.

[3] C. McDiarmid, A. Steger, D.J.A. Welsh, Random planar graphs, *J. Combin. Theory Ser. B* **93** (2005) pp. 187–205.

[4] C. McDiarmid, A. Steger, D.J.A. Welsh, Random graphs from planar and other addable classes in *Topics in Discrete Mathematics, Dedicated to Jarik Nešetril on the Occasion of its 60th birthday*, M. Klazar, J. Kratochvil, M. Loebl, J. Matousek, R. Thomas, P. Valtr, eds., Algorithms and Combinatorics **26**, Springer-Verlag, Berlin, 2006, pp. 231–246.

[5] M. Randić, On characterization of molecular branching, *J. Amer. Chem. Soc.* **97** (1975) pp. 6609–6615.

The probability that a random multigraph is simple

Svante Janson

If $n \ge 1$ and $(d_i)_1^n$ is a sequence of non-negative integers, we let $G(n,(d_i)_1^n)$ be the random (simple) graph with the n vertices $1,\dots,n$, and with vertex degrees $d_1,\dots,d_n$, uniformly chosen among all such graphs (provided that there are any such graphs at all; in particular, $\sum_i d_i$ has to be even). A standard method to study $G(n,(d_i)_1^n)$ is to consider the related random multigraph $G^*(n,(d_i)_1^n)$ defined by taking a set of d_i *half-edges* at each vertex i and then joining the half-edges into edges by taking a random partition of the set of all half-edges into pairs (we tacitly assume that $\sum_i d_i$ is even). This is known as the configuration model, and such a partition of the half-edges is known as a *configuration*; this was introduced by Bollobás [2], see also Section II.4 of [3]. (See Bender and Canfield [1] and Wormald [11, 12] for related arguments.)

We obtain $G(n,(d_i)_1^n)$ by conditioning $G^*(n,(d_i)_1^n)$ on being a simple graph. It is then of crucial importance to be able to estimate the probability that $G^*(n,(d_i)_1^n)$ is simple, and in particular to decide whether

$$\liminf_{n\to\infty} \mathbb{P}\big(G^*(n,(d_i)_1^n) \text{ is simple}\big) > 0 \tag{1}$$

for given sequences $(d_i)_1^n = (d_i^{(n)})_1^n$ (depending on $n \ge 1$). (Note that (1) implies that any statement holding for $G^*(n,(d_i)_1^n)$ with probability tending to 1 does so for $G(n,(d_i)_1^n)$ too.)

A natural condition that has been used by several authors using the configuration method (including myself [6]) as a sufficient condition for (1) is

$$\sum_{i=1}^n d_i = \Theta(n) \qquad \text{and} \qquad \sum_{i=1}^n d_i^2 = O(n) \tag{2}$$

together with some bound on $\max_i d_i$. (Recall that $A = \Theta(B)$ means that both $A = O(B)$ and $B = O(A)$ hold.) Results showing, or implying, that (2) and a condition on $\max_i d_i$ imply (1) have also been given by several authors, for example Bender and Canfield [1] with $\max_i d_i = O(1)$; Bollobás [2], see also Section II.4 in [3], with $\max_i d_i \le \sqrt{2\log n} - 1$; McKay [9] with $\max_i d_i = o(n^{1/4})$; McKay and Wormald [10] with $\max_i d_i = o(n^{1/3})$. (Some of these papers give sharp estimates of the probability that $G^*(n,(d_i)_1^n)$ is simple also when (2) does not hold.) Similar results have also been proved for bipartite graphs [8], digraphs [5], and hypergraphs [4].

Indeed, it is not difficult to see that the method used by Bollobás [2, 3] works, assuming (2), provided only $\max_i d_i = o(n^{1/2})$. This has undoubtedly been noted by several experts, but we have not been able to find a reference to it in print when we have needed one.

One of our main result is that, in fact, (2) is sufficient for (1) without any assumption on $\max_i d_i$. Moreover, (2) is essentially necessary.

Let N be the number of edges in $G^*(n,(d_i)_1^n)$. Thus

$$2N = \sum_{i=1}^n d_i. \tag{3}$$

It turns out that it is more natural to state our results in terms of N than n (the number of vertices).

We consider a sequence of random multigraphs $G_\nu^* = G^*\big(n_\nu,(d_i^{(\nu)})_1^n\big)$ and consider asymptotics as $\nu\to\infty$, but for notational simplicity we will omit the index ν. (In typical application, the graphs are indexed by n, the number of vertices; moreover, typically $N = \Theta(n)$, and then N can be replaced by n in the conditions below.) We can state our first result as follows.

Theorem 1. *Let $N := \frac12\sum_i d_i$ be the number of edges in $G^*(n,(d_i)_1^n)$, and assume that $N\to\infty$. Then*

(i) $\liminf \mathbb{P}(G^*(n,(d_i)_1^n)$ *is simple*$) > 0$ *if and only if* $\sum_i d_i^2 = O(N)$;
(ii) $\lim \mathbb{P}(G^*(n,(d_i)_1^n)$ *is simple*$) = 0$ *if and only if* $\sum_i d_i^2/N \to \infty$.

Our second main result is the following asymptotic formula for the probability that $G^*(n,(d_i)_1^n)$ is simple.

Theorem 2. *Consider $G^*(n,(d_i)_1^n)$ and assume that $N := \frac{1}{2}\sum_i d_i \to \infty$. Let $\lambda_{ij} := \sqrt{d_i(d_i-1)d_j(d_j-1)}/(2N)$; in particular $\lambda_{ii} = d_i(d_i-1)/(2N)$. Then*

$$\mathbb{P}\bigl(G^*(n,(d_i)_1^n) \text{ is simple}\bigr) = \exp\Bigl(-\tfrac{1}{2}\sum_i \lambda_{ii} - \sum_{i<j}\bigl(\lambda_{ij} - \log(1+\lambda_{ij})\bigr)\Bigr) + o(1);$$

equivalently,

$$\begin{aligned}&\mathbb{P}\bigl(G^*(n,(d_i)_1^n) \text{ is simple}\bigr)\\ &= \exp\biggl(-\frac{1}{4}\Bigl(\frac{\sum_i d_i^2}{2N}\Bigr)^2 + \frac{1}{4} + \frac{\sum_i d_i^2(d_i-1)^2}{16N^2} + \sum_{i<j}\bigl(\log(1+\lambda_{ij}) - \lambda_{ij} + \tfrac{1}{2}\lambda_{ij}^2\bigr)\biggr)\\ &\qquad + o(1).\end{aligned}$$

In the case $\max_i d_i = o(N^{1/2})$, Theorem 2 simplifies as follows.

Corollary 3. *Assume that $N \to \infty$ and $\max_i d_i = o(N^{1/2})$. Let*

$$\Lambda := \frac{1}{2N}\sum_{i=1}^n \binom{d_i}{2} = \frac{\sum_i d_i^2}{4N} - \frac{1}{2}.$$

Then

$$\begin{aligned}\mathbb{P}\bigl(G^*(n,(d_i)_1^n) \text{ is simple}\bigr) &= \exp\bigl(-\Lambda - \Lambda^2\bigr) + o(1)\\ &= \exp\Bigl(-\frac{1}{4}\Bigl(\frac{\sum_i d_i^2}{2N}\Bigr)^2 + \frac{1}{4}\Bigr) + o(1).\end{aligned}$$

This formula is well known, at least under stronger conditions on $\max_i d_i$, see, for example, Bender and Canfield [1], Bollobás [3, Theorem II.16], McKay [9] and McKay and Wormald [10, Lemma 5.1]. In this case, one can use the method by Bollobás [2, 3] and show by the method of moments that the number of loops plus the number of pairs of parallel edges is asymptotically Poisson distributed with mean $\Lambda + \Lambda^2$, which yields a direct proof of Corollary 3.

To prove the theorems under the weaker condition $\max_i d_i = O(N^{1/2})$, this method fails, because this number no longer has to be approximatively Poisson distributed; for example, two vertices with degree $N^{1/2}$ will have roughly a Poisson distributed number of edges between them, and thus the number of pairs of such edges will be more like the square of a Poisson variable. Instead we count vertices with at least one loop and pairs of vertices with at least two edges between them, disregarding the number of parallel loops or edges. We show that the indicators that a vertex or a pair of vertices is bad in this sense are asymptotically independent. This uses a more complicated conditioning argument; we would like to condition on the event that a certain pair of vertices is bad, but instead we condition on the event that it has k edges given by specific pairs of half-edges, which is easier to study; by an inclusion-exclusion type argument, we then get the necessary estimates for the conditioning we want.

Details are given in [7].

References

[1] E. A. Bender & E. R. Canfield, The asymptotic number of labeled graphs with given degree sequences. *J. Combin. Theory Ser. A*, **24** (1978), no. 3, 296–307.

[2] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European J. Comb.* **1** (1980), 311–316.

[3] B. Bollobás, *Random Graphs*, 2nd ed., Cambridge Univ. Press, Cambridge, 2001.

[4] C. Cooper, The cores of random hypergraphs with a given degree sequence. *Random Structures Algorithms* **25** (2004), no. 4, 353–375.

[5] C. Cooper & A. Frieze, The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combin. Probab. Comput.* **13** (2004), no. 3, 319–337.

[6] S. Janson & M. Luczak, A simple solution to the k-core problem. *Random Structures Algorithms*, to appear. `http://arxiv.org/math.CO/0508453`

[7] S. Janson, The probability that a random multigraph is simple. `http://arxiv.org/math.CO/0609802`

[8] B. D. McKay, Asymptotics for 0-1 matrices with prescribed line sums. *Enumeration and design (Waterloo, Ont., 1982)*, pp. 225–238, Academic Press, Toronto, ON, 1984.

[9] B. D. McKay, Asymptotics for symmetric 0-1 matrices with prescribed row sums. *Ars Combin.* **19A** (1985), 15–25.

[10] B. D. McKay & N. C. Wormald, Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica* **11** (1991), no. 4, 369–382.

[11] N. C. Wormald, Some problems in the enumeration of labelled graphs. Ph. D. thesis, University of Newcastle, 1978.

[12] N. C. Wormald, The asymptotic distribution of short cycles in random regular graphs. *J. Combin. Theory Ser. B* **31** (1981), no. 2, 168–182.

Rainbow Turán problems

Peter Keevash

(joint work with Dhruv Mubayi, Benny Sudakov and Jacques Verstraëte)

We define the rainbow Turán number $\mathrm{ex}^*(n, H)$ to be the maximum number of edges in a graph G on n vertices that has a proper edge-colouring with no rainbow H. Here H is some fixed graph, and by a rainbow H we mean a copy of H in G all of whose edges have different colours.

There are two main motivations for our study of rainbow Turán numbers. One is the possibility of applying purely combinatorial methods to certain extremal problems in additive number theory. Call a subset A of an abelian group G a B_k^*-set if does not contain disjoint k-subsets B, C with the same sum. Given a set A consider the following edge-coloured bipartite Cayley graph. The two parts X, Y are both copies of G, we join $x \in X$ to $y \in Y$ if $x - y \in A$, and then the edge xy is assigned the colour $x - y$. This is a properly coloured graph, and if A is a B_k^*-set then it does not contain a rainbow C_{2k}, the cycle of length $2k$.

Another motivation is that it seems to be a natural meeting point of two areas of extremal graph theory. Firstly, there is the classical Turán problem, which has a rich history in combinatorics. This asks for the maximum number of edges in a graph on n vertices that contains no copy of some fixed graph H. The maximum here is denoted $\mathrm{ex}(n, H)$, and is known as the Turán number for H. Next there is

the literature on extremal problems for edge-colourings (not necessarily proper). An example is the Canonical Ramsey Theorem, proved by Erdős and Rado [6], a special case of which shows that any proper colouring of K_n produces a rainbow K_m, provided n is large relative to m. Motivated by this and work in [5] and [11], Alon, Jiang, Miller and Pritikin [1] introduced the problem of finding a rainbow copy of a graph H in a colouring of K_n in which each colour appears at most m times at each vertex. The rainbow Turán problem is a natural Turán-type extension of this problem.

There are many results on rainbow Turán problems that can be deduced from results on classical Turán problems. For the reader's convenience we first give some background information on this (ordinary) problem. Its systematic study originated with Turán, who considered forbidding K_r, the complete graph on r vertices. The Turán graph $T_{r-1}(n)$ is the complete $(r-1)$-partite graph with part sizes as equal as possible; we write $t_{r-1}(n)$ for the number of edges in $T_{r-1}(n)$. Then Turán's Theorem [15] states that $\mathrm{ex}(n, K_r) = t_{r-1}(n)$, and $T_{r-1}(n)$ is the unique extremal K_r-free graph. Erdős and Stone [7] showed that the behaviour of the Turán number of a general graph H is determined by its chromatic number. They proved that if $\chi(H) = r$ then $\mathrm{ex}(n, H) = t_{r-1}(n) + o(n^2)$, which gives asymptotics except when H is bipartite.

From this last result it is not hard to show when H is non-bipartite we have $\mathrm{ex}^*(n, H) \sim \mathrm{ex}(n, H)$. This is a consequence of the the previous result and the following lemma: given a proper edge-colouring of the complete r-partite graph with parts of size r^3t^3 it is possible to choose t points from each part that span a rainbow subgraph. The proof of the lemma is by a simple greedy embedding procedure. Full details of this argument and proofs of the other results mentioned here are given in a paper by the authors that will appear in the journal Combinatorics, Probability and Computing.

We can describe a general class of graphs in which $\mathrm{ex}^*(n, H) = \mathrm{ex}(n, H)$ as follows. We say that H is colour-critical if it contains an edge e so that $\chi(H\backslash e) = \chi(H) - 1$, where $\chi(H)$ denotes the chromatic number of H. (Note that our definition is non-standard; often the term means that deleting *any* edge reduces the chromatic number.) If H is colour-critical and $\chi(H) = r$ then a result of Simonovits [14] shows that $\mathrm{ex}(n, H) = t_{r-1}(n)$ for sufficiently large n. Again it is not hard to show that $\mathrm{ex}^*(n, H) = \mathrm{ex}(n, H)$: it follows from the result of Simonovits and a variant of the lemma mentioned in the previous paragraph in which one extra edge is introduced.

For bipartite graphs that can be a considerable gap between $\mathrm{ex}(n, H)$ and $\mathrm{ex}^*(n, H)$. A simple example is when H is a path, when they are both essentially linear functions of n with different constants of proportionality. In general, even the order of magnitude of Turán numbers for bipartite graphs is not well understood. In the case of complete bipartite graphs, Kővári, Sós and Turán [12] showed $\mathrm{ex}(n, K_{s,t}) = O(n^{2-1/s})$, where the implied constant depends only on s and t. We obtain the same bound for $\mathrm{ex}^*(n, K_{s,t})$ (with a different

hidden constant) via another reduction lemma, namely that a properly edge-coloured $K_{s,t'}$ contains a rainbow $K_{s,t}$ for an appropriate $t' = t'(s,t)$. The case of $K_{2,2}$ (a.k.a. the 4-cycle C_4) is particularly interesting. We have the bounds $\frac{1}{2}n^{3/2} \leq \mathrm{ex}(n, K_{2,2}) \leq \mathrm{ex}^*(n, K_{2,2}) \leq \mathrm{ex}(n, K_{2,4}) \leq \frac{\sqrt{3}}{2}n^{3/2}$, but we cannot close the gap between the constants.

This leads us to the rainbow Turán problem for even cycles, which, as mentioned above, has potential applications to bounding B_k^*-sets in Abelian groups. A lower bound $\mathrm{ex}^*(n, C_{2k}) = \Omega(n^{1+1/k})$ can be obtained from the bipartite Cayley graph using the B_k-set in $\mathbb{Z}/n\mathbb{Z}$ of size $\sim n^{1/k}$ constructed by Bose and Chowla [4]. An upper bound of the same order of magnitude was obtained by Ruzsa [13], who showed that a B_k^*-set in the integers $\{1, \cdots, n\}$ has at most $(1 + o(1))k^{2-1/k}n^{1/k}$ elements. One of the outstanding problems in combinatorial number theory is to close the gap between the upper and lower bounds for such sets. We conjecture a corresponding upper bound of $\mathrm{ex}^*(n, C_{2k}) = O(n^{1+1/k})$, which would provide a combinatorial proof of Ruzsa's bound in the setting of general Abelian groups. There is an interesting constrast with the ordinary Turán problem for even cycles, where an upper bound $\mathrm{ex}(n, C_{2k}) = O(n^{1+1/k})$ was obtained by Bondy and Simonovits [3] but a conjectured lower bound of $\mathrm{ex}(n, C_{2k}) = \Omega(n^{1+1/k})$ is known only for $k = 2, 3, 5$.

We have made some progress on our conjecture. As already mentioned, it is true when $k = 2$. We can prove it for $k = 3$, which is already quite difficult and occupies a substantial part of our paper. In fact, we can show the existence of absolute constants $c_2 \geq c_1 > 1$ such that $c_1\mathrm{ex}(n, C_6) \leq \mathrm{ex}^*(n, C_6) \leq c_2\mathrm{ex}(n, C_6)$, i.e. the rainbow Turán number has the same magnitude as the ordinary Turán number, but with a different constant. For the general case we were able to prove the bound $O(n^{1+1/k})$ under the additional assumption that the graph does not contain any cycles of length less than $2k$.

A graph on n vertices without any cycle at all has at most $n - 1$ edges, but how many edges can there be in a properly coloured graph without a rainbow cycle? By contrast with the ordinary Turán problem, we can give a construction with $\Omega(n \log n)$ edges. In fact we can construct graphs with $\Omega(n \log n)$ edges with no cycle that uses more than half as many colours as edges, and we can prove a corresponding $O(n \log n)$ bound for such graphs. However, for the original problem we cannot improve the upper bound of $O(n^{4/3})$ given by our result for C_6. Could there be an $O(n^{1+\epsilon})$ bound? Or even $O(n \log n)$?

References

[1] N. Alon, T. Jiang, Z. Miller and D. Pritikin, Properly coloured subgraphs and rainbow subgraphs in edge-colourings with local constraints, *Random Structures Algorithms* **23** (2003), 409–433.

[2] C. T. Benson, Minimal regular graphs of girths eight and twelve, *Canad. J. Math.* **18** (1966), 1091–1094.

[3] J. A. Bondy, and M. Simonovits, Cycles of even length in graphs, *J. Combin. Theory Ser. B* **16** (1974), 97–105.

[4] R. C. Bose and S. Chowla, Theorems in the additive theory of numbers, *Comment. Math. Helv.* **37** (1962/1963), 141–147.
[5] G. Chen, R. Schelp, and B. Wei, Monochromatic-rainbow Ramsey numbers, 14th Cumberland Conference Abstracts, May 2001, http://www.msci.memphis.edu/balistep/Abstracts.html.
[6] P. Erdős and R. Rado, A combinatorial theorem, *J. London Math. Soc.* **25** (1950), 249–255.
[7] P. Erdős and A. H. Stone, On the structure of linear graphs, *Bull. Amer. Math. Soc.* **52** (1946), 1087–1091.
[8] Z. Füredi, New asymptotics for bipartite Turán numbers, *J. Combin. Theory Ser. A* **75** (1996), 141–144.
[9] Z. Füredi, A. Naor and J. Verstraëte, On the Turán Number for the Hexagon, *Advances in Mathematics*, to appear.
[10] S. W. Graham, B_h sequences, Analytic number theory, Vol. 1, (Allerton Park, IL, 1995), 431–449, *Progr. Math.* **138**, Birkhäuser Boston, Boston, MA, 1996.
[11] R. E. Jamison, T. Jiang, and A. C. H. Ling, Constrained Ramsey numbers of graphs, *J. Graph Theory* **42** (2003), 1–16.
[12] T. Kővári, V. Sós and P. Turán, On a problem of K. Zarankiewicz, *Colloquium Math.* **3** (1954), 50–57.
[13] I. Z. Ruzsa, Solving a linear equation in a set of integers I., *Acta Arith.* **65** (1993), 259–282.
[14] M. Simonovits, A method for solving extremal problems in graph theory, stability problems, in: *Theory of Graphs (Proc. Colloq. Tihany, 1966)*, Academic Press, New York, and Akad. Kiadó, Budapest, 1968, 279–319.
[15] P. Turán, On the theory of graphs, *Colloquium Math.* **3** (1954), 19–30.

The size-Ramsey number of short subdivisions

Yoshiharu Kohayakawa
(joint work with Vojtěch Rödl and Eduardo Tengan)

1. Introduction

Given an integer $q > 0$ and graphs Γ and H we write $\Gamma \to (H)_q$ if Γ contains a *monochromatic* copy of H in any *q-colouring* of the edges of Γ. That is, for any $\varphi\colon E(\Gamma) \to \{1, 2, \dots, q\}$, there is a copy H' of H in Γ (that is, a subgraph of Γ isomorphic to H) such that φ is constant on $E(H')$. For simplicity, we shall always take $q = 2$ in what follows.

The *Ramsey number* $r(H)$ of a graph H is the smallest number of vertices in a graph Γ such that $\Gamma \to (H)_2$. In contrast, the *size-Ramsey number* $r_{\mathrm{e}}(H)$ of a graph H is the smallest number of edges in a graph Γ such that $\Gamma \to (H)_2$, that is,

$$r_{\mathrm{e}}(H) = \min\{|E(\Gamma)|\colon \Gamma \to (H)_2\}\,. \tag{1}$$

The study of size-Ramsey numbers was proposed by Erdős, Faudree, Rousseau and Schelp [8] in 1978 when, for example, they investigated the size-Ramsey number of star forests and raised some questions concerning $r_{\mathrm{e}}(P^n)$, where P^n is the path on n vertices (see also [7]):

Problem 1 (Erdős, Faudree, Rousseau, and Schelp 1978)**.** *Is it true that*

$$r_{\mathrm{e}}(P^n)/n \to \infty \quad \textit{and} \quad r_{\mathrm{e}}(P^n)/n^2 \to 0? \tag{2}$$

Beck [3], using probabilistic methods, proved the surprising fact that $r_{\mathrm{e}}(P^n) \leq cn$, where c is an absolute constant, that is, the size-Ramsey number of paths is 'linear'. Explicit examples of linear sized graphs that are Ramsey for P^n were given by Alon and Chung [1], that is, they showed how to construct explicitly graphs Γ with $O(n)$ edges such that $\Gamma \to (P^n)_q$.

The linearity of the size-Ramsey number of paths was generalized to bounded degree trees by Friedman and Pippenger [10] (see also [11, 16]). (See [13, 14, 15] for more on tree embeddings.) It was proved in [12] that cycles also have linear size-Ramsey numbers.

Beck [4] asked whether $r_{\mathrm{e}}(H)$ is always linear in the size of H for graphs H of bounded degree, and this was settled in the negative by Rödl and Szemerédi [22], who proved that there are graphs of order n, maximum degree 3, and size-Ramsey number $\Omega(n(\log n)^{1/60})$. It is conjectured in [22] that, for some $\varepsilon = \varepsilon(\Delta) > 0$, we have

$$n^{1+\varepsilon} \leq r_{\mathrm{e}}(n, \Delta) \leq n^{2-\varepsilon}, \tag{3}$$

where $r_{\mathrm{e}}(n, \Delta)$ is the maximum of $r_{\mathrm{e}}(H)$ over all graphs H on n vertices and of maximum degree at most Δ. The upper bound in (3) has been proved by Rödl, Schacht, Szemerédi and the speaker [17]. For further results on size-Ramsey numbers, see [9, 19, 20, 21].

2. Results on subdivision of graphs

Let I be a graph and h a positive integer. We denote by $I^{(h)}$ the *h-subdivision* of I, namely the graph $I^{(h)}$ is obtained by replacing each edge of I by a path with $h+2$ vertices (so, for instance, $I^{(0)} = I$). The following result, which confirms a conjecture of Burr and Erdős [5], was proved by Alon [2].

Theorem 2 (Alon 1994)**.** *If an n-vertex graph H has no two vertices of degree at least 3 adjacent, then its Ramsey number is at most $12n$.*

Therefore, if we subdivide every edge of a graph I at least once, then we obtain a graph with linear Ramsey number. Thus, clearly, the size-Ramsey number of such a subdivision is at most quadratic. Pak [18] put forward the following conjecture.

Conjecture 3 (Pak 2002)**.** *There is an absolute constant c for which the following holds. For every integer D, there is a constant C_D such that if H is a graph with $\Delta(H) = D$ and h is an integer with $h > c \log N$, where*

$$N = |V(H^{(h)})| = |V(H)| + h|E(H)|, \tag{4}$$

then $r_{\mathrm{e}}(H^{(h)}) \leq C_D N$.

Making use of results on mixing times of random walks on expanders, Pak [18] proved Conjecture 3 in a weaker form (he obtained the desired upper bound for $r_{\mathrm{e}}(H^{(h)})$ up to a polylogarithmic factor in N). Donadelli, Haxell, and the

speaker [6] observed that Conjecture 3 holds in the case in which H is a fixed graph and $h \to \infty$.

We are now able to state our main result. Below, an n-vertex graph with maximum degree at most D will be called an (n, D)-*graph*.

Theorem 4. *For any $h \geq 1$ and $D \geq 1$, there exist C and n_0 such that, for all $n \geq n_0$, any (n, D)-graph I is such that*

$$r_{\mathrm{e}}(I^{(h)}) \leq n^{1+1/(h+1)}(\log n)^C. \tag{5}$$

Our proof of Theorem 4 is based on showing that a suitable binomial random graph $G(N, p)$ will do for any (n, D)-graph I. We in fact obtain a stronger result.

Definition 5. *Let $h \geq 1$, $D \geq 1$, and $q \geq 2$ be given. For $n \geq 1$, we define the (h, D, q)-universal size Ramsey number of n, denoted* $\mathrm{USR}(h, D, q, n)$*, to be the smallest number of edges in a graph Γ with the following 'universal' Ramsey property: $\Gamma \to (I^{(h)})_q$ for any (n, D)-graph I.*

Theorem 6. *Let $h \geq 1$, $D \geq 1$, and $q \geq 2$ be given. Then, for some absolute constant C, we have*

$$\mathrm{USR}(h, D, q, n) \leq n^{1+1/(h+1)}(\log n)^C. \tag{6}$$

We are not able to give interesting lower bounds for $r_{\mathrm{e}}(I^{(h)})$ for a given (n, D)-graph I, but we remark that a simple counting argument shows the following.

Theorem 7 (Lower bound for universal graphs)**.** *Let $D \geq 3$ be an integer and let Γ be a graph that contains at least one copy of $I^{(h)}$ for each (n, D)-graph I. Then, for n large enough,*

$$|E(\Gamma)| \geq \frac{1}{2} n^{1+1/(h+1)-2/D(h+1)}. \tag{7}$$

References

[1] N. Alon and F. R. K. Chung, *Explicit construction of linear sized tolerant networks*, Discrete Mathematics **72** (1988), no. 1-3, 15–19. MR 90f:05080

[2] Noga Alon, *Subdivided graphs have linear Ramsey numbers*, J. Graph Theory **18** (1994), no. 4, 343–347. MR 95b:05144

[3] József Beck, *On size Ramsey number of paths, trees, and circuits. I*, J. Graph Theory **7** (1983), no. 1, 115–129. MR 84g:05102

[4] ———, *On size Ramsey number of paths, trees and circuits. II*, Mathematics of Ramsey theory, Algorithms Combin., vol. 5, Springer, Berlin, 1990, pp. 34–45. MR 1 083 592

[5] S. A. Burr and P. Erdős, *On the magnitude of generalized Ramsey numbers for graphs*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. 1, North-Holland, Amsterdam, 1975, pp. 215–240. Colloq. Math. Soc. János Bolyai, Vol. 10. MR 51 #7918

[6] Jair Donadelli, Penny E. Haxell, and Yoshiharu Kohayakawa, *A note on the size-Ramsey number of long subdivisions of graphs*, Theor. Inform. Appl. **39** (2005), no. 1, 191–206. MR 2005k:05160

[7] P. Erdős, *On the combinatorial problems which I would most like to see solved*, Combinatorica **1** (1981), no. 1, 25–42. MR 82k:05001

[8] P. Erdős, R. J. Faudree, C. C. Rousseau, and R. H. Schelp, *The size Ramsey number*, Periodica Mathematica Hungarica **9** (1978), no. 2-2, 145–161. MR 80h:05037

[9] R. J. Faudree and R. H. Schelp, *A survey of results on the size Ramsey number*, Paul Erdős and his mathematics, II (Budapest, 1999), Bolyai Soc. Math. Stud., vol. 11, János Bolyai Math. Soc., Budapest, 2002, pp. 291–309. MR 2003k:05083
[10] J. Friedman and N. Pippenger, *Expanding graphs contain all small trees*, Combinatorica **7** (1987), no. 1, 71–76. MR 88k:05063
[11] P. E. Haxell and Y. Kohayakawa, *The size-Ramsey number of trees*, Israel J. Math. **89** (1995), no. 1-3, 261–274. MR 96c:05128
[12] P. E. Haxell, Y. Kohayakawa, and T. Łuczak, *The induced size-Ramsey number of cycles*, Combin. Probab. Comput. **4** (1995), no. 3, 217–239. MR 96h:05140
[13] P. E. Haxell and T. Łuczak, *Embedding trees into graphs of large girth*, Discrete Math. **216** (2000), no. 1-3, 273–278. MR 2000m:05073
[14] P. E. Haxell, T. Łuczak, and P. W. Tingley, *Ramsey numbers for trees of small maximum degree*, Combinatorica **22** (2002), no. 2, 287–320, Special issue: Paul Erdős and his mathematics. MR 2003d:05141
[15] Tao Jiang, *On a conjecture about trees in graphs with large girth*, J. Combin. Theory Ser. B **83** (2001), no. 2, 221–232. MR 2002h:05045
[16] Xin Ke, *The size Ramsey number of trees with bounded degree*, Random Structures Algorithms **4** (1993), no. 1, 85–97. MR 94a:05149
[17] Y. Kohayakawa, V. Rödl, M. Schacht, and E. Szemerédi, *The size-Ramsey number of graphs of bounded degree*, in preparation.
[18] Igor Pak, *Mixing time and long paths in graphs*, Proceedings of the 13th annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2002), 2002, pp. 321–328.
[19] Oleg Pikhurko, *Asymptotic size Ramsey results for bipartite graphs*, SIAM J. Discrete Math. **16** (2002), no. 1, 99–113 (electronic). MR 2004b:05140
[20] ———, *Size Ramsey numbers of stars versus 4-chromatic graphs*, J. Graph Theory **42** (2003), no. 3, 220–233. MR 2003m:05130
[21] David Reimer, *The Ramsey size number of dipaths*, Discrete Math. **257** (2002), no. 1, 173–175. MR 2003h:05143
[22] Vojtěch Rödl and Endre Szemerédi, *On size Ramsey numbers of graphs with bounded degree*, Combinatorica **20** (2000), no. 2, 257–262. MR 2001d:05122

Hamilton cycles in expanding and highly connected graphs

Michael Krivelevich
(joint work with Dan Hefetz and Tibor Szabó)

We prove the following sufficient condition for a graph to be hamiltonian. Let $G = (V, E)$ be a graph on n vertices, and let $12 \leq d = d(n) \leq e^{\sqrt[3]{\log n}}$ be a parameter. Suppose that:

- Every subset $S \subset V$ of size $|S| \leq \frac{n \log\log n \log d}{d \log n \log\log\log n}$ has at least $d|S|$ outside neighbors in G;
- There is an edge of G between every pair of disjoint subsets $A, B \subset V$ of size $|A|, |B| \geq \frac{n \log\log n \log d}{4130 \log n \log\log\log n}$.

Then G is Hamiltonian, for large enough n.

The obtained criterion improves on previous results of similar nature, as in particular it is much more natural and straighforward to apply. It is applicable to a wide variety of graphs, including relatively sparse graphs, like those with degrees polylogarithmic in the number of vertices.

The proof relies on the celebrated rotation-extension technique of Pósa [5].

We also discuss several applications of our result. The first group of applications belongs to the field of Positional Games (see, e.g., [1] for a background). In [3] the criterion was used by the authors to prove that Enforcer can win the $(1, q)$ Avoider-Enforcer Hamilton cycle game, played on the edges of K_n, for every $q \leq \frac{cn \log\log\log\log n}{\log n \log\log\log n}$ where c is an appropriate constant; this is presently the best known bound. A similar result can be obtained for Maker in the corresponding Maker-Breaker game.

The second group of applications is in Extremal Graph Theory. We prove that a graph G satisfying the above two conditions is Hamilton-connected (i.e., every pair of vertices $u, v \in V(G)$ can be connected by a Hamilton path). We also prove that, for certain choices of the parameter $d = d(n)$, the second condition alone is enough to guarantee the existence of cycles of many lengths in G, including almost Hamilton cycles. We also come close to solving a recent conjecture of Brandt, Broersma, Diestel and Kriesell [2], connecting connectivity/expansion and Hamiltonicity.

Finally, our criterion (with minor changes in the proof) can be used to reprove a classical result in the theory of Random Graphs (see, e.g., [4]), postulating that if the edge probability $p = p(n)$ satisfies $p = (\log n + \log\log n + \omega(1))/n$, then the random graph $G(n, p)$ is almost surely Hamiltonian.

References

[1] J. Beck, *Positional games*, Combinatorics, Probability and Computing **14** (2005), 649–696.

[2] S. Brandt, H. Broersma, R. Diestel and M. Kriesell, *Global connectivity and expansion: long cycles and factors in f-connected graphs*, Combinatorica **26** (2006), 17–36.

[3] D. Hefetz, M. Krivelevich and T. Szabó, *Avoider-Enforcer games*, Journal of Combinatorial Theory Series A, to appear.

[4] J. Komlós and E. Szemerédi, *Limit distributions for the existence of Hamilton circuits in a random graph*, Discrete Mathematics **43** (1983), 55–63.

[5] L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. **14** (1976), 359–364.

Complexity measures of sign matrices

Nati Linial

(joint work with Adi Shraibman)

What is *complexity*, and how should it be studied mathematically? In the interpretation that we adopt, there are several underlying common themes to complexity theories. The basic ground rules are these: There is a family $\mathcal{F}$ of some mathematical objects under consideration. The elements of some subset $\mathcal{S} \subseteq \mathcal{F}$ are deemed *simple*. Also, there are certain composition rules that allow one to put together objects in order to generate other objects in $\mathcal{F}$. The complexity of an object is determined by the length of the shortest chain of steps to generate it from simple objects. In full generality one would want to get good estimates for all or many

objects in the family $\mathcal{F}$. Specifically, a major challenge is to be able to point out specific concrete objects that have high complexity. That is, elements that cannot be generated from simple objects using only a small number of composition steps.

Arguably the currently most developed mathematical theory of complexity is to be found in the field of computational complexity. Typically (but not exclusively), $\mathcal{F}$ consists of all boolean functions $f : \{0,1\}^m \to \{0,1\}$. The class $\mathcal{S}$ of simple objects contains the constant functions, and the functions $x \to x_i$ (the i-th coordinate). Functions can be composed using the basic logical operations (**or, and, not**).

Another interesting and important approach to complexity starts as above, but adds another ingredient into the game, namely, a metric. In this approach the complexity of a member in $\mathcal{F}$ is defined as its distance to the set $\mathcal{S}$ of simple objects. This approach leads to the notion of *matrix rigidity* as well as to the notion of approximating a given matrix by a low rank matrix, a problem that can be solved using Singular Value Decomposition.

In view of the fundamental importance and the apparent great difficulty of the problems of computational complexity we suggest to address issues of complexity in other mathematical fields. Aside from the inherent interest in understanding complexity in general, insights gained from such investigations are likely to help in speeding up progress in computational complexity. We seek to develop a complexity theory for sign matrices (matrices all of whose entries are ± 1). There are several good reasons why this should be a good place to start. First, a number of hard and concrete problems in computational complexity proper can be stated in this language. Two notable examples are (i) The log-rank conjecture and (ii) The matrix rigidity problem. Also, matrices come with a complexity measure that we all know, namely, the *rank*. To see why, let us declare the class $\mathcal{S}$ of simple matrices to be those matrices (not necessarily with ± 1 entries) that have rank one. Suppose, furthermore, that the composition rule is matrix sum. We recall a theorem from linear algebra that the rank of a matrix A equals the least number of rank-one matrices whose sum is A. This shows that rank indeed fits the definition of a complexity measure for matrices.

One important lesson from the experience gathered in computational complexity, is that it is beneficial to study a variety of complexity measures in order to understand the behavior of the main quantities of interest. Thus, aside from circuit complexity (the "real" thing), people are investigating communication complexity, proof complexity, decision tree models etc. This is the direction we take too, as we conduct a comparative study of several measures of complexity for sign matrices.

Our main discovery in these papers is that factorization norms of matrices (as studied in Banach Space theory) are very powerful complexity measures. In particular we are able to show the connection of these parameters to two well-studied areas of computational complexity, namely, machine learning and communication complexity. Specifically, we obtain new lower bounds in various models of (randomized and quantum) commincation complexity.

Many fascinating problems remain open.

Uniformly cross intersecting families

Eyal Lubetzky
(joint work with Noga Alon)

Let $\mathcal{A}$ and $\mathcal{B}$ denote two families of subsets of an n-element set. We say that the pair $(\mathcal{A},\mathcal{B})$ is ℓ-cross-intersecting iff $|A\cap B|=\ell$ for all $A\in\mathcal{A}$ and $B\in\mathcal{B}$. Let $P_\ell(n)$ denote the maximum possible value of $|\mathcal{A}||\mathcal{B}|$ over all ℓ-cross-intersecting pairs $(\mathcal{A},\mathcal{B})$. We are interested in finding the precise value of $P_\ell(n)$, and in characterizing all the extremal pairs $\mathcal{A},\mathcal{B}$ which achieve this maximum.

A well known conjecture of Erdős [2] stated that if $\mathcal{F}\subset 2^{[n]}$ is a family satisfying $|F\cap F'|\neq\lfloor\frac{n}{4}\rfloor$ for all $F,F'\in\mathcal{F}$, then $|\mathcal{F}|<(2-\epsilon)^n$ for some $\epsilon>0$. This was proved by Frankl and Rödl [3], by considering the corresponding variant on two families: it is shown in [3], that if $\mathcal{A},\mathcal{B}\subset 2^{[n]}$ and $|A\cap B|\neq l$, where $\eta n\le l\le(\frac{1}{2}-\eta)n$ for some $\eta<\frac{1}{4}$, then $|\mathcal{A}||\mathcal{B}|\le(4-\epsilon(\eta))^n$. The authors of [3] studied several additional problems related to cross-intersections of two families of sets, and among their results, they provided the following upper bound on $P_\ell(n)$, which was later reproved in [1]:

$$\begin{cases} P_0(n)\le 2^n \\ P_\ell(n)\le 2^{n-1} \quad \text{for } \ell\ge 1 \end{cases} . \tag{1}$$

The argument which gives the upper bound of 2^n is simple: consider the characteristic vectors of the sets in $\mathcal{A},\mathcal{B}$ as vectors in $\mathbb{Z}_2^n$. Notice that the intersection of two sets is equal to the inner product of the two corresponding vectors modulo 2. Therefore, if ℓ is even, then the families $\mathcal{A},\mathcal{B}$ belong to two orthogonal linear spaces, giving $|\mathcal{A}||\mathcal{B}|\le 2^n$. Otherwise, we may add an additional coordinate of 1 to all vectors, and repeat (carefully) the above argument, gaining a slight improvement: $|\mathcal{A}||\mathcal{B}|\le 2^{n-1}$. Similar ideas are used to show that the upper bound 2^{n-1} holds for even values of $\ell>0$ as well, by performing the analysis over $GF(p)$ for some prime $p>2$ instead of over $\mathbb{Z}_2$.

As part of their study of questions in Coding Theory, Ahlswede, Cai and Zhang [1] gave the following simple construction of an ℓ-cross-intersecting pair: for $n\ge 2\ell$, let $\mathcal{A}$ contain a single 2ℓ-element set, A, and let $\mathcal{B}$ contain all the sets which contain precisely ℓ elements of A. This gives:

$$|\mathcal{A}||\mathcal{B}|=\binom{2\ell}{\ell}2^{n-2\ell}=(1+o(1))\frac{2^n}{\sqrt{\pi\ell}} , \tag{2}$$

where the $o(1)$-term tends to 0 as $\ell\to\infty$. The upper bound (1) implies that this construction achieves the maximum of $P_\ell(n)$ for $\ell\in\{0,1\}$, and the authors of [1] conjectured that this in fact holds for all ℓ.

As the upper bound (1) is independent of ℓ, compared to the above lower bound of $\Theta(2^n/\sqrt{\ell})$, Sgall [5] asked whether or not $P_\ell(n)$ is bounded from above by some decreasing function of ℓ. One of the motivations of [5] was a relation between problems of restricted cross-intersections of two families of sets and problems in Communication Complexity; see [5] for more details.

In [4], the authors verified the above conjecture of [1] for the case $\ell = 2$, by showing that $P_2(n) \leq 3 \cdot 2^{n-3}$. However, for any $\ell > 2$ the best known upper bound on $P_\ell(n)$ remained 2^{n-1}.

The following theorem confirms the above conjecture of [1] for all sufficiently large values of ℓ, and thus provides also a positive answer to the above question of Sgall.

Theorem 1. *There exists some $\ell_0 > 0$ such that, for all $\ell \geq \ell_0$, every ℓ-cross-intersecting pair $\mathcal{A}, \mathcal{B} \subset 2^{[n]}$ satisfies:*

$$|\mathcal{A}||\mathcal{B}| \leq \binom{2\ell}{\ell} 2^{n-2\ell} \ . \tag{3}$$

Furthermore, if $|\mathcal{A}||\mathcal{B}| = \binom{2\ell}{\ell}2^{n-\ell}$, then there exists some choice of parameters κ, τ, n':

$$\begin{array}{l} \kappa \in \{2\ell - 1, 2\ell\} \ , \ \tau \in \{0, \ldots, \kappa\} \ , \\ \kappa + \tau \leq n' \leq n, \end{array} \tag{4}$$

such that, up to a relabeling of the elements of $[n]$ and swapping $\mathcal{A}, \mathcal{B}$, the following holds:

(5)

$$\begin{aligned} \mathcal{A} &= \left\{ \bigcup_{T \in J} T \ : \ J \subset \left\{ \begin{array}{c} \{1, \kappa+1\}, \ldots, \{\tau, \kappa+\tau\}, \\ \{\tau+1\}, \ldots, \{\kappa\} \end{array} \right\} , \ |J| = \ell \right\} \times 2^X \ , \\ \mathcal{B} &= \left\{ L \cup \{\tau+1, \ldots, \kappa\} : \begin{array}{l} L \subset \{1, \ldots, \tau, \kappa+1, \ldots, \kappa+\tau\} \\ |L \cap \{i, \kappa+i\}| = 1 \text{ for all } i \in [\tau] \end{array} \right\} \times 2^Y \ . \end{aligned}$$

where $X = \{\kappa + \tau + 1, \ldots, n'\}$ and $Y = \{n' + 1, \ldots, n\}$.

Indeed, this family satisfies:

$$|\mathcal{A}||\mathcal{B}| = \binom{\kappa}{\ell} \cdot 2^{|X|} \cdot 2^{\tau + |Y|} = \binom{\kappa}{\ell} 2^{n-\kappa} = \binom{2\ell}{\ell} 2^{n-2\ell} \ ,$$

where the last inequality is by the choice of $\kappa \in \{2\ell - 1, 2\ell\}$. The construction of [1] fits the special case $\tau = 0$, $\kappa = 2\ell$.

The proof of Theorem 1 combines tools from linear algebra with techniques from extremal combinatorics, including the Littlewood-Offord Lemma, extensions of Sperner's Theorem and some large deviation estimates.

Remarks:

- Our results show that if two families of subsets of an n-element set, $\mathcal{A}, \mathcal{B}$, are ℓ-cross-intersecting, and ℓ is sufficiently large, then $|\mathcal{A}||\mathcal{B}| \leq \binom{2\ell}{\ell}2^{n-2\ell}$, and in addition, we have a complete characterization of all the extremal pairs $\mathcal{A}, \mathcal{B}$ for which equality is achieved.
- It would be interesting to prove that the above result holds for all values of ℓ (instead of all $\ell \geq \ell_0$ for some ℓ_0). Perhaps knowing the precise structure of the extremal pairs $\mathcal{A}, \mathcal{B}$, as described in Theorem 1 (assuming that this holds for all ℓ), will assist in proving this result.

- Finally, one may consider the corresponding problem where the pair $\mathcal{A}, \mathcal{B}$ does not have one possible cross-intersection, but rather a set L of legal cross-intersections. Such notions have been studied in [1], [5], [4], with different restrictions on L, and it would be interesting to derive tight bounds on $|\mathcal{A}||\mathcal{B}|$, and possibly describe the structure of all the extremal pairs, when in addition, each member of L is larger than some predefined integer ℓ.

References

[1] R. Ahlswede, N. Cai, and Z. Zhang, A general 4-words inequality with consequences for 2-way communication complexity, Adv. in Appl. Math. 10 (1989), 75-94.

[2] P. Erdős, Problems and results in graph theory and combinatorial analysis, Proc. of the Fifth British Comb. Conf. 1975 Aberdeen, 169-192. Congressus Numerantium, No. XV, Utilitas Math., Winnipeg, Man., 1976.

[3] P. Frankl and V. Rödl, Forbidden intersections, Trans. Amer. Math. Soc. 300 (1987), 259-286.

[4] P. Keevash and B. Sudakov, On a restricted cross-intersection problem, J. Combinatorial Theory Ser. A, 113 (2006), 1536-1542.

[5] J. Sgall, Bounds on pairs of families with restricted intersections, Combinatorica 19 (1999), 555-566.

Erdős-Ko-Rado for three sets

Dhruv Mubayi

Katona formulated the following definition to generalize the Erdős-Ko-Rado theorem to more than two sets.

Definition. Let $k \leq s \leq 3k$. Then $f(n,k,s)$ denotes the maximum size of a family $\mathcal{G}$ of k-element subsets of $[n]$ so that whenever $A, B, C \in \mathcal{G}$ satisfy $|A \cup B \cup C| \leq s$, we have $A \cap B \cap C \neq \emptyset$.

Katona asked for the determination of $f(n,k,s)$. Frankl and Füredi [3] proved the surprising fact that for *every* $2k \leq s \leq 3k$, $f(n,k,s) = \binom{n-1}{k-1}$ as long as $n \geq k^2 + 3k$. They observed that $f(n,k,2k-1) = \Omega(n^k)$ for fixed k (the lower bound $f(n,k,s) \geq \binom{n-1}{k-1}$ is trivial). Moreover, by definition $f(n,k,s+1) \leq f(n,k,s)$, hence Frankl and Füredi's first result follows by proving the upper bound just for $s = 2k$. They conjectured that $f(n,k,2k) = \binom{n-1}{k-1}$ for all $k \geq 3$ and $n \geq 3k/2$.

Frankl and Füredi [3] proved their conjecture for $k = 3$, and commented (without proof) that their approach also works for $k = 4, 5$ and more generally for $n > k^2/\log k$. We prove Frankl and Füredi's conjecture.

Theorem 1. ([4]) *Fix integers $k \geq 3$ and $n \geq 3k/2$. Let $\mathcal{G}$ be a family of k-element subsets of an n-element set so that whenever $A, B, C \in \mathcal{G}$ satisfy $|A \cup B \cup C| \leq 2k$, we have $A \cap B \cap C \neq \emptyset$. Then $|\mathcal{G}| \leq \binom{n-1}{k-1}$ with equality only when $\bigcap_{S \in \mathcal{G}} S \neq \emptyset$.*

One can consider a common generalization of both the Erdős-Ko-Rado theorem and Katona's problem.

Definition. Fix $k \geq 3$. A family $\mathcal{G}$ of k-element sets is a $K(d)$-family if, whenever distinct sets $A_1, \ldots, A_d \in \mathcal{G}$ satisfy $|\bigcup_i A_i| \leq 2k$, we have $\bigcap_i A_i \neq \emptyset$.

Note that $K(2)$-families are just intersecting families, and $f(n, k, 2k)$ is the maximum size of a $K(3)$-family of k-element sets of $[n]$. Also note that there exist families of size $\lfloor n/k \rfloor^k = \Omega(n^k)$ such that every $d \geq 3$ sets have empty intersection provided their union is at most $2k - 1$ (simply partition $[n]$ into k almost equal parts, and take all k-sets with exactly one point in each part). This is the reason for the threshold $2k$ in our definition. The following result determines the asymptotic maximum size of $K(d)$-families, since it is obvious that this maximum is at least $\binom{n-1}{k-1}$.

Theorem 2. *Fix $k \geq d \geq 3$. For every $\delta > 0$, there exists n_0 such that if $n > n_0$ and $\mathcal{G}$ is a $K(d)$-family of k-element subsets of $[n]$, then $|\mathcal{G}| \leq (1 + \delta)\binom{n-1}{k-1}$.*

It is possible that Theorem 2 holds for $K(d)$-families even when $d > k$. In fact, it is an interesting open problem to determine the largest $d = d(k)$ for which Theorem 2 holds.

The main open problem here is to extend Theorem 2 from an asymptotic to an exact result.

Conjecture 3. *Let $k \geq d \geq 3$ and $n \geq dk/(d-1)$. Suppose that $\mathcal{G}$ is a $K(d)$-family of k-element subsets of $[n]$. Then $|\mathcal{G}| \leq \binom{n-1}{k-1}$, with equality only when $\bigcap_{S \in \mathcal{G}} S \neq \emptyset$.*

The $d = 2$ case of Conjecture 3 is the Erdős-Ko-Rado theorem, and the $d = 3$ case is Theorem 1. The next case, $d = 4$, has been recently settled [5], but only for n sufficiently large. Finally, a very recent paper of Chen and Liu [1], reduces the $d = k$ case of Conjecture 3 to an old conjecture of Chvátal on $(k-1)$-simplices. Since this case was already settled by Chvátal [2], the reduction results in a proof of Conjecture 3 when $d = k$.

References

[1] W. Y. C. Chen, J. Liu, Families of Sets with Intersecting Clusters, preprint (available at http://front.math.ucdavis.edu/math.CO/0605171).

[2] V. Chvátal, An extremal set-intersection theorem, J. London Math. Soc. (2) 9 (1974/75), 355–359.

[3] P. Frankl, Z. Füredi, A new generalization of the Erdős-Ko-Rado theorem, Combinatorica 3 (1983), no. 3-4, 341–349

[4] D. Mubayi, Erdős-Ko-Rado for three sets, J. Combin. Theory Ser. A, 113 (2006), no. 3, 547–550.

[5] D. Mubayi, A stability method in extremal set theory, submitted (available at http://www.math.uic.edu/∼mubayi/papers.html).

The complexity of monotone linear programming

Pavel Pudlák

This lecture presents a work in progress done together with Michal Koucký, Pierre McKenzie and Ronald de Wolf. Our goal is to prove an exponential lower bound on linear programs that represent monotone boolean functions. Such a lower bound would have consequences concerning the complexity of integer linear programing. In particular it would imply a lower bound on the Lovasz-Schrijver proof system.

Here we present a lower bound for a rather restricted system that does not seem to have such applications.

A general *monotone LP* (applied to compute boolean functions) is determined by a finite set of inequalities

$$\sum_j a_{ij} z_j \leq \sum_k b_{ik} x_k + c_i$$

where $a_{ij}, c_i \in \mathbf{R}$, $b_{ik} \in \mathbf{R}^+$ are constants and $z_j \in \mathbf{R}^+$, $x_k \in \{0,1\}$ are variables.

Such a program P computes boolean function $f(\bar{x})$, if for every assignment to $\bar{x}$

$$P \text{ has a solution, iff } f(\bar{x}) = 1$$

We present a *restricted model* using zero-sum games instead of LPs. It is determined by a matrix A and an ϵ satisfying:

- $A = \{a_{ij}\}$, $|a_{ij}| \leq 1$;
- $\epsilon > 0$ – a constant, *the gap*;
- columns of A are marked by x_k's;
- given an assignment to $\bar{x}$, Column player can only use columns with $x_k = 1$;
- if $f(\bar{x}) = 1$ then Column player wins $\geq \epsilon$;
- if $f(\bar{x}) = 0$ then Row player wins $\leq -\epsilon$.

Let $MATCHING_n$ denote the monotone function which is 1 iff a graph on $3n$ vertices has a matching with n edges.

Theorem 1. *Let $\epsilon > 0$ be a constant. Suppose monotone LP given by matrix A computes the function $MATCHING_n$ with gap ϵ. Then A must have dimensions $2^{\Omega(n^\delta)}$, $\delta > 0$ constant.*

The proof is based on reducing the probabilistic communication complexity of disjointness to such LPs, in a similar way as in the paper of Raz and Wigderson [1].

References

[1] R. Raz and A. Wigderson, Monotone circuits for matching require linear depth, Journal of the ACM 39(3), 1992, 736-744.

Learning boolean functions

Rüdiger Reischuk

For many application areas, greedy strategies are natural, important, and efficient heuristics. In some cases, such as simple scheduling problems, it has been shown that greedy strategies actually find a global optimum. For the vast majority of optimization problems, however, greedy heuristics do not always achieve optimal solutions. In such cases, the behaviour of greedy algorithms is hardly understood. One notable exception is the characterization of transportation problems using the Monge property [14]. Sometimes one can at least show that a specific greedy algorithm achieves a certain nontrivial approximation ratio. This, for example, holds for the Set Cover problem with a logarithmic approximation factor [10, 7, 15], which has been proven to be best possible in [8].

Confronted with an unknown Boolean function $f : \{0,1\}^n \to \{0,1\}$ on n variables, the problem of detecting which variables x_i are relevant to f is known as *relevant feature selection*. This problem lies at the heart of many data mining applications. This is particularly the case if f depends only on a small number d of all n attributes— such concepts are called *d-juntas* (for a survey see [6]).

To infer relevant attributes from a randomly drawn sample $S = (x^k, y^k)_{k=1,\ldots,m}$ with $x^k = (x_1^k, \ldots, x_n^k) \in \{0,1\}^n$ and $y^k = f(x^k) \in \{0,1\}$, the key task is to find a minimal set of attributes $R \subseteq \{x_1, \ldots, x_n\}$ such that S admits a consistent hypothesis h (i.e., $h(x^k) = y^k$ for all k) that depends only on the variables in R. By standard arguments, once the sample size m exceeds $\mathrm{poly}(2^d, \log n)$, with high probability there remains only one such hypothesis—the target concept f itself. Finding such a set R is equivalent to solving the following Set Cover instance. The ground set is the set of all pairs $\{k, \ell\}$ such that $y^k \neq y^\ell$. A pair $\{k, \ell\}$ may be covered by any attribute x_i such that $x_i^k \neq x_i^\ell$. The goal is to cover the ground set by as few attributes as possible. This reduction opens the door to apply well-known greedy heuristics: the most generic one, which we call Greedy, successively selects the largest remaining set and deletes all covered elements.

For relevant feature selection, this approach has been proposed among others in [2]. In. [1] it is proved that with high probability, Greedy successfully infers the relevant variables for the concept class of conjunctions of attributes or their negations (i.e., monomials) and that a small sample of size $\mathrm{poly}(2^d, \log n)$already suffices. In [9] this has been extended to functions f that are unbalanced with respect to all of their relevant variables (i.e., for x uniformly chosen at random, $\Pr[f(x) = 1 | x_i = 0] \neq \Pr[f(x) = 1 | x_i = 1]$ for each relevant x_i).

We give a concise characterization of the class of target concepts for which Greedy is able to infer the relevant variables. The new characterization is based on a property of the Fourier spectrum of the target concept, which we call *Fourier-accessibility*. A function $f : \{0,1\}^n \to \{0,1\}$ is *Fourier-accessible* if for each relevant variable, one can find a sequence $\emptyset \subsetneq I_1 \subsetneq \ldots \subsetneq I_s \subseteq [n]$ such that $i \in I_s$ and for all $j \in \{1, \ldots, s\}$, $|I_j \setminus I_{j-1}| = 1$ and $\hat{f}(I_j) \neq 0$. Equivalently, $f \not\equiv 0$ is Fourier-accessible iff for every x_i that is relevant to f, there exists $I \subseteq [n]$ with

$i \in I$ and a path from $\emptyset$ to I in the *Fourier support graph*. This graph is the subgraph of the n-dimensional Hamming cube induced by the *Fourier support* of f, i.e., the set of subsets I of $[n]$ such that $\hat{f}(I) \neq 0$.

We then prove that GREEDY correctly infers all relevant variables of Fourier-accessible d-juntas $f : \{0,1\}^n \to \{0,1\}$ under the uniform distribution from $m = \text{poly}(2^d, \log n, \log(1/\delta))$ examples with probability at least $1 - \delta$. On the other hand, we show that if a function f is *not* Fourier-accessible, then the error probability of GREEDY is at least $1 - d^2/(n-d)$. In particular, this probability tends to 1 as d is fixed and $n \to \infty$ or as $d \to \infty$ and $n \in \omega(d^2)$.

For $t \in \{1, \ldots, n\}$, a function $f : \{0,1\}^n \to \{0,1\}$ is said to be *t-low* [4] if for each relevant variable x_i of f, there exists $I \subseteq [n]$ with $i \in I$ such that $|I| \leq t$ and $\hat{f}(I) \neq 0$. Similarly to the analysis of GREEDY, we obtain that a variant called GREEDY RANKING learns *exactly* the class of 1-low functions. Since all 1-low functions are Fourier-accessible (but not vice versa), these results also provide an analytic argument that the dynamic variant GREEDY (that adjusts the covering sets after each round) is in general preferrable to the static version GREEDY RANKING.

The relation between algorithmic learning of Boolean fucntions and their spectral properties has been investigated before [11, 12, 5]. Specifically, in [13] spectral and algebraic methods are combined to reduce the worst-case running time for learning the class of all n-ary d-juntas to roughly $n^{0.7\,d}$ (a trivial way is to test all $\Theta(n^d)$ sets of potentially relevant variables). We show that GREEDY as well as the Fourier method are quite fault-tolerant. They learn correctly and efficiently even when the sample is corrupted by random noise [3, 4]. Some results can even be generalized to nonuniform distributions.

REFERENCES

[1] T. Akutsu,, S. Miyano, S. Kuhara. A Simple Greedy Algorithm for Finding Functional Relations: Efficient Implementation and Average Case Analysis. *Theoret. Comput. Science 292,* 481–495, 2003.

[2] H. Almuallim, T Dietterich. Learning Boolean Concepts in the Presence of Many Irrelevant Features, *Artificial Intelligence 69,* 279–305, 1994.

[3] J. Arpe, R. Reischuk. Robust Inference of Relevant Attributes. *Proc. 14. Conf. on Algorithmic Learning Theory, Springer Lecture Notes in Artificial Intelligence 2942,* 99–113, 2003.

[4] J. Arpe, R. Reischuk. Learning Juntas in the Presence of Noise. *Proc. 3. Conf. on Theory and Applications of Models of Computation and Logic, Springer Lecture Notes in Computer Science 3959*, 387–398, 2006.

[5] A. Blum, M. Furst, J. Jackson, M. Kearns, Y. Mansour, S. Rudich. Weakly Learning DNF and Characterizing Statistical Query Learning Using Fourier Analysis. *Proc. 26. ACM Symposium on Theory of Computing,* 253–262, 1994.

[6] A, Blum, P. Langley. Selection of Relevant Features and Examples in Machine Learning. *Artificial Intelligence 97,* 245–271, 1997.

[7] V. Chvatal. A Greedy Heuristic for the Set Covering Problem. *Math. Oper. Res. 4,* 233–235, 1979.

[8] U. Feige. A Threshold of $\ln n$ for Approximating Set Cover. *J. ACM 45,* 634–652, 1998.

[9] D. Fukagawa, T. Akutsu. Performance Analysis of a Greedy Algorithm for Inferring Boolean Functions. *Inform. Process. Lett. 93,* 7–12, 2005.
[10] D. Johnson. Approximation Algorithms for Combinatorial Problems. *J. Comput. System Sciences 9,* 256–278, 1974.
[11] N. Linial, Y. Mansour, N. Nisan. Constant Depth Circuits, Fourier Transform, and Learnability. *J. ACM 40,* 607–620, 1993.
[12] Y. Mansour. Learning Boolean Functions via the Fourier Transform. In V. Roychodhury, K. Siu, A. Orlitsky (editors), *Theoretical Advances in Neural Computation and Learning*, 391–424. Kluwer Academics, 1994.
[13] E. Mossel, R. O'Donnell, R. Servedio. Learning Functions of k Relevant Variables. *J. Comput. System Sciences 69,* 421–434, 2004.
[14] R. Shamir, B. Dietrich. Characterization and Algorithms for Greedily Solvable Transportation Problems. *Proc. 1. ACM-SIAM Symposium on Discrete Algorithms,* 358–366, 1990.
[15] P. Slavík. A Tight Analysis of the Greedy Algorithm for Set Cover. *Proc. 28. ACM Symposium on Theory of Computing,* 435–441, 1996.

Percolation on non-uniform quasi-random graphs

Oliver Riordan

(joint work with Béla Bollobás, Christian Borgs and Jennifer Chayes)

Given two (simple, undirected) graphs F and G, let $\hom(F,G)$ denote the number of homomorphisms from F to G, i.e., the number of maps ϕ from the vertex set of F to that of G with the property that $\phi(u)\phi(v)$ is an edge of G whenever uv is an edge of F. Let

$$t(F,G) = |V(F)|^{-|V(G)|}\hom(F,G)$$

denote the normalized form of $\hom(F,G)$. When F is 'small' and G is 'large', then up to a $o(1)$ term, $t(F,G)$ just counts the number of copies of F in G, normalized by dividing by the number of copies in the complete graph on $V(G)$.

A sequence G_n of quasi-random graphs with density p is simply a sequence of graphs whose subgraph densities approach that of the random graph $G(n,p)$; in other words, $t(F,G_n) \to p^{e(F)}$ for every fixed graph F. Such sequences were introduced by Thomason [13] under a different name, and have been studied in great detail by Chung, Graham and Wilson [8], Chung and Graham [6, 7], Simonovits and Sós [11], and many others; it turns out that they behave in many respects like random graphs.

The question of other possible limits of the numbers $t(F,G_n)$ was studied by Erdős, Lovász and Spencer [9]. Following Borgs, Chayes, Lovász, Sós and Vesztergombi [2], see also [3], we call a sequence (G_n) of graphs *left convergent* if $t(F,G_n)$ converges (to some non-negative real number) for every graph F. Building on this notion, Lovász and Szegedy [10] showed that if (G_n) is left convergent, then there is a symmetric measurable function W on $[0,1]^2$ such that

$$t(F,G_n) \to t(F,W) = \int_{[0,1]^{V(F)}} \prod_{ij\in E(F)} W(x_i,x_j) \prod_{i\in V(F)} \mathrm{d}x_i$$

holds for every fixed graph F.

The quantity $t(F, W)$ above is exactly the normalized expected number of copies of F in a certain dense *inhomogeneous* random graph $G(n, W)$, in which edges are present independently, and the edge probabilities are given by the values of W at suitably chosen points. Thus, a left convergent sequence G_n may be thought of as a sequence of inhomogeneous quasi-random graphs. Indeed, Borgs, Chayes, Lovász, Sós and Vesztergombi [4, 5] have shown that many natural notions of graph convergence are equivalent, and that such convergent sequences share many properties of the inhomogeneous random graphs $G(n, W)$.

Sparse inhomogeneous random graphs have been studied in hundreds of papers, since they arise very naturally as models of real world networks. Bollobás, Janson and Riordan [1] introduced a very general model of such graphs, closely related to $G(n, W)$, but with greater generality and with edge probabilities scaling as $1/n$. One of the most important properties of such sparse graphs is their *phase transition*, i.e., the point at which a giant component emerges as a density parameter is varied. Bollobás, Janson and Riordan showed that for their model, the critical point is determined by a certain multi-type branching process associated to W.

Of course, one may pass from a dense random graph to a sparse random graph by selecting edges with probability $1/n$, where n is the number of vertices. Here we perform this operation on the graphs G_n in a convergent sequence. Thus, given an arbitrary left-convergent sequence G_n, we study 'percolation' on G_n, in particular, asking what fraction of the edges must be retained to ensure a giant component.

It turns out that G_n behaves like $G(n, W)$ even in this complicated respect: the critical point is again given by the branching process associated to W. Since we have no local control on the (deterministic, but quasi-random) graphs G_n, the arguments are rather different from those in [1], for example; the main tools are local counting and Szemerédi's Lemma [12]. It turns out that strong results can be obtained, including logarithmic bounds on the small components away from the threshold; this does not hold for the more general model studied in [1].

References

[1] B. Bollobás, S. Janson and O. Riordan, The phase transition in inhomogeneous random graphs, to appear in *Random Structures and Algorithms*. Preprint available from `http://www.arxiv.org/abs/math.PR/0504589`.

[2] C. Borgs, J. Chayes, L. Lovász, V.T. Sós and K. Vesztergombi, unpublished, 2004.

[3] C. Borgs, J. Chayes, L. Lovász, V.T. Sós and K. Vesztergombi, Counting graph homomorphisms, in: *Topics in Discrete Mathematics* (Klazar, Kratochvil, Loebl, Matousek, Thomas, Valtr, eds.) Springer, Berlin–Heidelberg (2006), 315–371.

[4] C. Borgs, J. Chayes, L. Lovász, V.T. Sós and K. Vesztergombi, Convergent Sequences of Dense Graphs I: Subgraph Frequencies, Metric Properties and Testing, preprint (2006), `http://research.microsoft.com/~borgs/Papers/ConvMetric.pdf`.

[5] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi: Convergent Graph Sequences II. H-Colorings, Statistical Physics and Quotients, manuscript (2006).

[6] F.R.K. Chung and R.L. Graham, Quasi-random hypergraphs, *Random Struct. Algorithms* **1** (1990), 105–124.

[7] F.R.K. Chung and R.L. Graham, Sparse quasi-random graphs, in 'Special issue: Paul Erdős and his Mathematics', *Combinatorica* **22** (2002), 217–244.

[8] F.R.K. Chung, R.L. Graham and R.M. Wilson, Quasi-random graphs, *Combinatorica* **9** (1989), 345–362.
[9] P. Erdős, L. Lovász and J. Spencer, Strong independence of graphcopy functions, in *Graph Theory and Related Topics* (Proc. Conf., Univ. Waterloo, Waterloo, Ont., 1977), pp. 165–172, Academic Press, New York-London, 1979.
[10] L. Lovász and B. Szegedy, Limits of dense graph sequences, MSR Tech Report # MSR-TR-2004-79,
`ftp://ftp.research.microsoft.com/pub/tr/TR-2004-79.pdf`.
[11] M. Simonovits and V.T. Sós, Ramsey-Turán theory, in 'Combinatorics, Graph theory, Algorithms and Applications', *Discrete Math.* **229** (2001), 293–340.
[12] E. Szemerédi, Regular partitions of graphs, in *Problèmes Combinatoires et Théorie des Graphes* (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), pp. 399–401, Colloq. Internat. CNRS **260**, CNRS, Paris, 1978.
[13] A. Thomason, Pseudo-random graphs, in *Proceedings of Random Graphs* (M. Karonski, ed.), Poznań, 1985, *Annals of Discrete Mathematics*, **33** (1987) 307–331.

Hamilton cycles and perfect matchings in hypergraphs

Andrzej Ruciński

1. Hamiltonian cycles

A classic theorem of Dirac states that a sufficient condition for an n-vertex graph to be hamiltonian is that the minimum degree is at least $n/2$. Given an integer $k \geq 2$, a k-uniform hypergraph (shortly: *k-graph*) is a hypergraph where every edge is of size k.

By *a cycle* we mean a k-graph with vertices $v_0, \ldots, v_{l-1}$ and edges $\{v_i, v_{i+1}, \ldots, v_{i+k-1}\}$, $i = 0, \ldots, l-1$, where the indices are computed modulo l. *A hamiltonian cycle* in a k-graph H is a sub-hypergraph of H which is a cycle and contains all vertices of H. Given a k-graph H, let $\delta_{k-1}(H)$ be the largest integer d such that every $(k-1)$-element set of vertices of H belongs to at least d edges of H.

An extension of Dirac's theorem to k-graphs, $k \geq 2$, has been conjectured in [1].

Conjecture 1. *If $\delta_{k-1}(H) \geq \lfloor (n-k+3)/2 \rfloor$ then H is hamiltonian.*

Due to a construction in [1], if true, the above conjecture is best possible. In [4] we proved an approximate version of the conjecture for all $k \geq 3$.

Theorem 2 ([4])**.** *Let $k \geq 3$ and $\epsilon > 0$. Then, for sufficiently large n, every k-graph on n-vertices such that $\delta_{k-1}(H) \geq (1/2 + \epsilon)n$ is hamiltonian.*

Here is the idea of the proof. As a preliminary step, we find in H a powerful path A, called *absorbing*, which has the property that *every* not too large subset of vertices can be "absorbed" by that path. In the sub-hypergraph $H' = H - V(A)$ we find a collection of long, disjoint paths which cover almost all vertices of H'. (The main tool allowing us to do that is a weak hypergraph generalization of the regularity lemma from [6].) Then, we "glue" these paths and the absorbing path

A together to form a long cycle in H. In the final step, the vertices which are not yet on the cycle are absorbed by A to form a hamiltonian cycle in H.

2. Perfect matchings

A matching in a k-graph is a set of disjoint edges. Let $\nu(H)$ be the size of a largest matching in H. Clearly, if H is hamiltonian then $\nu(H) = \lfloor n/k \rfloor$. For all integer $n \geq k \geq 2$, denote by $t_k(n)$ the smallest integer t such that for every k-graph H on n vertices and with $\delta_{k-1}(H) \geq t$ we have $\nu(H) = \lfloor n/k \rfloor$.

In the case of graphs we have $t_2(n) = \lfloor n/2 \rfloor$. For $k \geq 3$, as a by-product of Theorem 2 above, it follows that $t_k(n) = n/2 + o(n)$. Kühn and Osthus proved in [2] that $t_k(n) \leq n/2 + O(\sqrt{n \log n})$, which we further improved in [3] to $t_k(n) \leq n/2 + O(\log n)$. Recently, we have determined precisely the values of $t_k(n)$ in the case when n is large and divisible by k.

Theorem 3 ([5]). *For all $k \geq 3$ and sufficiently large n which is divisible by k, we have $t_k(n) = n/2 - k + c(k,n)$, where $c(k,n) \in \{\frac{3}{2}, 2, \frac{5}{2}, 3\}$.*

The constructions yielding the lower bounds are similar to those in [2]. For instance, when k and $n/2$ are both odd, split the vertex set evenly into A and B, and take as edges all k-tuples of vertices whose intersection with A has even size.

Interestingly, if n is not divisible by k, the values of $t_k(n)$ are substantially smaller than in the divisible case. Indeed, if $n \not\equiv 0 \pmod k$ then $\lfloor n/k \rfloor \leq t_k(n) \leq n/k + O(\log n)$.

Conjecture 4. *If $n \not\equiv 0 \pmod k$ then $t_k(n) = \lfloor n/k \rfloor$.*

Our proofs of upper bounds on $t_k(n)$ also involve an "absorbing" technique, similar to that used for hamiltonian cycles. Roughly speaking, in order to construct a perfect matching, we begin with a powerful but relatively small matching M'. This matching has the property that for any set S of k unmatched vertices, one can slightly alter M' so that, for the resulting matching M'_S, $V(M'_S) = V(M') \cup S$. Next, we find an *almost* perfect matching M'' in $H - V(M')$ which leaves some set S of k vertices unmatched. The matching $M'_S \cup M''$ is then perfect. In the non-divisible case this technique is suitably modified.

Algorithmic aspects. For integers $k \geq 2$ and $0 \leq r \leq k-1$ and for $0 < c < 1$, let $PM(k,r,c)$ be the problem of deciding if for an input k-graph H with $|V(H)| = n \equiv r \pmod k$ and $\delta_{k-1}(H) \geq cn$ we have $\nu(H) = \lfloor n/k \rfloor$. It is proved in [7] that $PM(k,r,c)$ is NP-complete for all k, r and all $c < 1/k$. On the other hand, in view of our results, for $c > 1/k$ and $r \neq 0$ as well as for $c \geq 1/2$ and $r = 0$, the problem becomes trivial. Moreover, it is shown in [7] that both our proofs of the upper bounds on $t_k(n)$ can be turned into polynomial time algorithms finding a matching of size $\lfloor n/k \rfloor$.

Problem 5. *Determine the complexity of $PM(k,0,c)$ for $1/k \leq c < 1/2$.*

References

[1] G. Y. Katona and H. A. Kierstead, Hamiltonian chains in hypergraphs, *J. Graph Theory* 30 (1999) 205-212.

[2] D. Kühn and D. Osthus, Matchings in hypergraphs of large minimum degree, *J. Graph Th.*, 51 (2006), 269-280

[3] V. Rödl, A. Ruciński, and E. Szemerédi, Perfect matchings in uniform hypergraphs with large minimum degree *Europ. J. Combin.*, 27 (2006) 1333-1349.

[4] V. Rödl, A. Ruciński and E. Szemerédi, An approximative Dirac-type theorem for k-uniform hypergraphs, submitted.

[5] V. Rödl, A. Ruciński and E. Szemerédi, Perfect matchings in large uniform hypergraphs with large minimum collective degree, submitted.

[6] E. Szemerédi, Regular partitions of graphs. Problemes combinatoires et theorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), pp. 399–401, Colloq. Internat. CNRS, 260, CNRS, Paris, 1978.

[7] E. Szymańska, The complexity of perfect matchings in hypergraphs, in preparation.

Ramsey type problems in random graphs

Angelika Steger

(joint work with M. Marciniszyn, J. Skokan, R. Spöhel)

It follows from Ramsey's celebrated result that *every* k-coloring of the edges of the complete graph on n vertices contains a monochromatic copy of F if n is sufficiently large. While this seems to rely on the fact that K_n is a very dense graph, Folkman [1] and, in a more general setting, Nešetřil and Rödl [8] showed that there also exist locally sparse graphs $G = G(F)$ with the property that every k-coloring of the edges of G contains a monochromatic copy of F. By transferring the problem into a random setting, Rödl and Ruciński [9, 10] showed that in fact such graphs G are quite frequent. More precisely, they showed that $p = n^{-1/m_2(F)}$ is a threshold for the property $G_{n,p} \to (F)^e_k$, where $m_2(F) := \max\left\{\frac{|E(H)|-1}{|V(H)|-2} : H \subseteq F \wedge |V(H)| \geq 3\right\}$.

This setup can be generalized by allowing k different forbidden graphs, one per color. Within classical Ramsey theory the study of these so-called asymmetric Ramsey properties led to many interesting questions and results. Within the random setting only very little is known about asymmetric Ramsey properties. Kohayakawa and Kreuter formulated the following conjecture and proved it for the case of cycles:

Conjecture 1. [4] *Let F_1, F_2 be graphs with $1 < m_2(F_1) \leq m_2(F_2)$. Then there exists a constant $b > 0$ such that for all $\varepsilon > 0$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left[G_{n,p} \to (F_1, F_2)^e\right] = \begin{cases} 0 & \text{if } p \leq (1-\varepsilon)bn^{-1/m_2(F_1,F_2)} \\ 1 & \text{if } p \geq (1+\varepsilon)bn^{-1/m_2(F_1,F_2)} \end{cases},$$

where

$$m_2(F_1, F_2) := \max\left\{\frac{|E(H)|}{|V(H)| - 2 + 1/m_2(F_1)} : H \subseteq F_2 \wedge |V(H)| \geq 2\right\}.$$

The threshold function from Conjecture 1 is supported by the following observation. The expected number of copies of F_2 in $G_{n,p}$ with $p = \Theta\big(n^{-1/m_2(F_1,F_2)}\big)$ is

$$\Theta\left(n^{|V(F_2)|}p^{|E(F_2)|}\right) = \Omega\left(n^{2-1/m_2(F_1)}\right) \quad .$$

Since every edge-coloring of $G_{n,p}$ must avoid monochromatic copies of F_2 in color 2, there is at least one edge of color 1 in every subgraph of $G_{n,p}$ isomorphic to F_2. Select one such edge from each copy of F_2 arbitrarily. It is plausible that these edges span a graph G' with edge density $\Omega\big(n^{-1/m_2(F_1)}\big)$ that satisfies certain pseudo-random properties. As it turns out, that seems just about the right density in order to embed a copy of F_1 into G', no matter which edges were selected from the original graph.

In [5] we consider the threshold function p_0 for cliques $K_{\ell_1}, \ldots, K_{\ell_k}$, where $\ell_1 \geq \cdots \geq \ell_k \geq 3$. For the 0-statement we suppose that $p \leq bn^{-1/m_2(K_{\ell_2},K_{\ell_1})}$ for some sufficiently small constant $b > 0$ and provide an algorithm that a.a.s. finds a k-edge-coloring of a random graph $G_{n,p}$ that avoids every forbidden clique K_{ℓ_i}, $1 \leq i \leq k$, in the corresponding color class i. A standard way of attacking the 1-statement, which was also pursued in [4], is via the sparse version of Szemerédi's regularity lemma, which was independently developed by Kohayakawa [2] and Rödl (unpublished). Using properties of regularity, one can find a monochromatic copy of a forbidden subgraph in the colored graph $G_{n,p}$. Unfortunately, generalizing this argument from cycles to cliques requires a proof of Conjecture 23 in [3] of Kohayakawa, Łuczak, and Rödl. This so-called KŁR-Conjecture is a probabilistic version of the classical embedding lemma for dense graphs. Assuming that this conjecture is true, a proof of the 1-statement is routinely obtained.

In the second part of our talk we study online-colorings of random graphs. We consider the following two graph processes starting with the empty graph G_0. In the vertex case in each step $i > 0$, a new vertex v_i and is randomly connected to the vertices of G_{i-1} by independently including each of the edges $\{v_i, v_j\}$, $1 \leq j < i$ with probability p into a new graph G_i. After n steps this process results in a random graph $G = G_{n,p}$, in which each edge is present with probability p independently of all other edges. An online coloring strategy must assign a color to v_i in each step i immediately, thereby extending the coloring of G_{i-1} to G_i in a proper way. Similarly, in the edge case in each step, a new edge e_i, randomly chosen form all edges not yet present in the graph, arrives. After m steps this process results in a random graph $G = G_{n,m}$. An online coloring strategy must assign a color to each edge immediately when it arrives.

We call a k-coloring valid w.r.t. a fixed nonempty graph F if no color class induces a monochromatic copy of F. The simple greedy F-avoidance strategy of selecting the largest color index that allows for a valid extension of the coloring turns out to be not optimal in general. For example, consider the graph F consisting of a triangle with one edge attached to it. Here our results yield that greedily avoiding triangles and forgetting about the additional edge is better than greedily avoiding the entire graph. We therefore consider the following refinement. Let $H_1, \ldots, H_k$ be a sequence of fixed nonempty subgraphs $H_r \subseteq F$, $1 \leq r \leq k$.

Then the greedy $\langle H_1, \ldots, H_k \rangle$-avoidance strategy chooses, in each step i, the *maximal* color index r such that v_i is not contained in a copy of H_r induced by the color class r in G_i. In other words, it avoids monochromatic copies of H_r in color k, $1 \leq r \leq k$. Clearly, if we succeed in assigning a color to each vertex of G in this way, we obtain a valid k-coloring w.r.t. the graph F.

Let

$$\overline{m}_1^0(F) := 0 \ , \qquad \overline{m}_1^r(F) := \max_{H \subseteq F} \frac{e_H + \overline{m}_1^{r-1}(F)}{v_H} \quad \forall r \geq 1 \ .$$

For any given graph F, we define the following infinite sequence of graphs accordingly:

$$H_r = H_r(F) \ := \ \arg\max_{H \subseteq F} \frac{e_H + \overline{m}_1^{r-1}(F)}{v_H} \quad \forall r \geq 1 \ . \tag{1}$$

Marciniszyn and Spöhel showed:

Theorem 2. [7] *Let F be a nonempty graph and let $k \geq 1$ be an integer. Let $H_1, \ldots, H_k \subseteq F$ be chosen according to* (1)*. Then the greedy $\langle H_1, \ldots, H_k \rangle$-avoidance strategy a.a.s. produces an k-coloring of $G_{n,p}$ without any monochromatic copy of F online provided we have*

$$p = p(n) \ll n^{-1/\overline{m}_1^k(F)} \ .$$

Moreover, whenever there exists an induced subgraph $F^\circ \subseteq F$ on $v_F - 1$ vertices that satisfies

$$\max_{H \subseteq F^\circ} \frac{e_H}{v_H - 1} \leq \overline{m}_1^2(F) \ ,$$

this strategy is optimal.

For the edge case we have a similar lower bound for all $k \geq 2$ and a matching upper bound for $k = 2$ colors, cf. [6].

References

[1] J. Folkman. Graphs with monochromatic complete subgraphs in every edge coloring. *SIAM J. Appl. Math.* **18**, 1970, 19–24.

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

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

[4] Y. Kohayakawa, B. Kreuter. Threshold functions for asymmetric Ramsey properties involving cycles. *Random Structures Algorithms* **11**, 1997, 245–276.

[5] M. Marcinszyn, J. Skokan, R. Spöhel, A. Steger, The online clique avoidance game on random graphs. In: Proceedings of the 10th. International Workshop on Randomized Techniques in Computation (RANDOM '06), 2006, 462-474. Journal version in preparation.

[6] M. Marcinszyn, R. Spöhel, A. Steger, The online clique avoidance game on random graphs. In: Proceedings of the 9th. International Workshop on Randomized Techniques in Computation (RANDOM '05), 2005, 390 - 401. Journal version in preparation.

[7] M. Marcinszyn, R. Spöhel. The online clique avoidance game on random graphs. In: Proceedings of the 18th annual ACM-SIAM Symposium on Discrete Algorithms (SODA '07), 2007, accepted for publication.

[8] J. Nešetřil, V. Rödl. The Ramsey property for graphs with forbidden complete subgraphs. *J. Combinatorial Theory Ser. B* **20**, 1976, 243–249.
[9] V. Rödl, A. Ruciński. Lower bounds on probability thresholds for Ramsey properties. In: Combinatorics, Paul Erdős is eighty, Vol. 1. Bolyai Soc. Math. Stud., János Bolyai Math. Soc., Budapest, 1993, 317–346.
[10] V. Rödl, A. Ruciński. Threshold functions for Ramsey properties. *J. Amer. Math. Soc.* **8**, 1995, 917–942.

Expansion of H-free graphs and its applications

Benny Sudakov

Let G be a graph which contains no subgraphs isomorphic to fixed bipartite graph H. Using well known results from Extremal Graph theory, one can show that such G has certain expansion properties, i.e., all small subsets of G have large boundary. This simple observation appears to be a powerful tool in attacking several extremal problems. In what follows next we discuss recent results in which this observation was used to resolve conjectures about cycle lengths and large clique-minors in H-free graphs.

1. Cycles in sparse graphs

Let $C(G)$ denote the set of lengths of cycles in a graph G. In [12], together with Jacques Verstraëte, we study the minimum possible value of $|C(G)|$ over all graphs G of average degree d and girth g. Erdős [4] conjectured that $|C(G)| = \Omega\big(d^{\lfloor (g-1)/2 \rfloor}\big)$ for all such graphs, and we prove this conjecture. We also show that this is a lower bound for the number of odd cycle lengths in a graph of chromatic number d and girth g. Further results are obtained for the number of cycle lengths in H-free graphs of average degree d.

Also in [12], motivated by the conjecture of Erdős and Gyárfás [5] (see also Erdős [6]) that every graph of minimum degree at least three contains a cycle of length a power of two, we prove a general theorem which gives an upper bound on the average degree of an n-vertex graph with no cycle of even length in a prescribed infinite sequence of integers. For many sequences, including the powers of two, our theorem gives the upper bound $e^{O(\log^* n)}$ on the average degree of graph of order n with no cycle of length in the sequence, where $\log^* n$ is the number of times the binary logarithm must be applied to n to get a number which is at most one.

2. Minors in expanding graphs

In another recent paper with Michael Krivelevich [8], we address several extremal problems related to graph minors. In all of our results we assume essentially that a given graph G is expanding, where expansion is either postulated directly, or G can be shown to contain a large expanding subgraph, or G is locally expanding due to the fact that G does not contain a copy of a fixed bipartite graph H. We need the following definitions to state our results. A graph $\Gamma = (U, F)$ with vertex set $U = \{u_1, \ldots, u_k\}$ is a *minor* of a graph $G = (V, E)$ if the vertex set V of G

contains a sequence of disjoint subsets $A_1, \ldots, A_k$ such that the induced subgraphs $G[A_i]$ are connected, and there is an edge of G between A_i and A_j whenever the corresponding vertices u_i, u_j of Γ are connected by an edge. A graph $G = (V, E)$ is (t, α)-*expanding* if every subset $X \subset V$ of size $|X| \leq \alpha|V|/t$ has at least $t|X|$ external neighbors in G. A graph $G = (V, E)$ is called (p, β)-*jumbled* if

$$\left| e(X) - p\frac{|X|^2}{2} \right| \leq \beta|X|$$

for every subset $X \subseteq V$, where $e(X)$ stands for the number of edges spanned by X in G. Informally, this definition indicates that the edge distribution of G is similar to that of the random graph $G_{|V|,p}$, where the degree of similarity is controlled by parameter β.

Here are the main results which we obtain together with Krivelevich [8].

Theorem 1. *Let $0 < \alpha < 1$ be a constant. Let G be a (t, α)-expanding graph of order n, and let $t \geq 10$. Then G contains a minor with average degree at least*

$$c\frac{\sqrt{nt \log t}}{\sqrt{\log n}},$$

where $c = c(\alpha) > 0$ is a constant.

This is an extension of results of Alon, Seymour and Thomas [1], Plotkin, Rao and Smith [11], and of Kleinberg and Rubinfeld [7], who cover basically the case of expansion by a constant factor $t = \Theta(1)$.

Theorem 2. *Let G be a (p, β)-jumbled graph of order n such that $\beta = o(np)$. Then G contains a minor with average degree $cn\sqrt{p}$, for an absolute constant $c > 0$.*

This statement is an extension of results of A. Thomason [14, 15], who studied the case of constant p. It can be also used to derive some of the results of Drier and Linial [3].

Theorem 3. *Let $2 \leq s \leq s'$ be integers. Let G be a $K_{s,s'}$-free graph with average degree r. Then G contains a minor with average degree $cr^{1+\frac{1}{2(s-1)}}$, where $c = c(s, s') > 0$ is a constant.*

This confirms a conjecture of Kühn and Osthus from [9].

Theorem 4. *Let $k \geq 2$ and let G be a C_{2k}-free graph with average degree r. Then G contains a minor with average degree $cr^{\frac{k+1}{2}}$, where $c = c(k) > 0$ is a constant.*

This theorem generalizes results of Thomassen [13], Diestel and Rompel [2], and Kühn and Osthus [10], who proved similar statements under the (much more restrictive) assumption that G has girth at least $2k + 1$.

All of the above results are, up to a constant factor, asymptotically tight (Theorems 1, 2), or are allegedly tight (Theorems 3, 4), where in the latter case the tightness hinges upon widely accepted conjectures from Extremal Graph Theory about the asymptotic behavior of the Turán numbers of $K_{s,s'}$ and of C_{2k}.

References

[1] N. Alon, P. Seymour and R. Thomas, A separator theorem for nonplanar graphs, *J. Amer. Math. Soc.* 3 (1990), 801–808.
[2] R. Diestel and C. Rempel, Dense minors in graphs of large girth, *Combinatorica* 25 (2005), 111–116.
[3] Y. Drier and N. Linial, Minor in lifts of graphs, *Random Structures and Algorithms* 29 (2006), 208–225.
[4] P. Erdős, Some of my favourite problems in various branches of combinatorics, *Matematiche (Catania)* 47 (1992), 231–240.
[5] P. Erdős, Some of my favorite solved and unsolved problems in graph theory, *Quaestiones Math.* 16 (1993), 333–350.
[6] P. Erdős, Some old and new problems in various branches of combinatorics. Graphs and combinatorics (Marseille, 1995). Discrete Math. 165/166 (1997), 227–231.
[7] J. Kleinberg and R. Rubinfeld, Short paths in expander graphs, *Proc.* 37^th *Symposium on Foundations of Computer Science (FOCS)*, IEEE Comput. Soc. Press, 1996, 86–95.
[8] M. Krivelevich and B. Sudakov, Minors in expanding graphs, submitted.
[9] D. Kühn and D. Osthus, Complete minors in $K_{s,s}$-free graphs, *Combinatorica* 25 (2005), 49–64.
[10] D. Kühn and D. Osthus, Minors in graphs of large girth, *Random Structures Algorithms* 22 (2003), 213–225.
[11] S. Plotkin, S. Rao and W. Smith, Shallow excluded minors and improved graph decompositions, *Proc.* 5^th *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, ACM, New York, 1994, 462–470.
[12] B. Sudakov and J. Verstraëte, Cycle lengths in sparse graphs, submitted.
[13] C. Thomassen, Girth in graphs, *J. Combin. Theory Ser. B* 35 (1983), 129–141.
[14] A. Thomason, Pseudorandom graphs, in: *Random graphs '85 (Poznań)*, North-Holland Math. Stud., 144, Amsterdam, 1987, 307–331.
[15] A. Thomason, Random graphs, strongly regular graphs and pseudorandom graphs, in: *Surveys in combinatorics*, London Math. Soc. Lecture Note Ser., 123, Cambridge Univ. Press, 1987, 173–195.

Relaxed colorings of graphs with bounded maximum degree — extremal graph theory meeting complexity theory

Tibor Szabó

(joint work with Robert Berke)

A coloring of the vertices of a graph is proper if each color class induces connected components of order one (where the *order* of a graph is its number of vertices). Here we study relaxations of proper two-colorings, such that the order of the induced monochromatic components in one (or both) of the color classes is bounded by a constant. In a (C_1, C_2)-*relaxed coloring* of a graph G every monochromatic component induced by vertices of the first (second) color is of order at most C_1 (C_2, resp.). We concentrate on the *symmetric* and *asymmetric case*, i.e., when $C_1 = C_2$ and $C_1 = 1$, respectively.

Symmetric case. The largest maximum degree which still guarantees the existence of a proper coloring is obviously 1: odd cycles are 2-regular and not properly two-colorable. In contrast, every graph of maximum degree 3 is two-colorable if we allow monochromatic components of order two. That is, in our terminology

every graph of maximum degree 3 can be $(2,2)$-relaxed colored. (The two sides of a max-cut will provide a $(2,2)$-relaxed coloring.) Alon, Ding, Oporowski and Vertigan [1] investigated what is the largest Δ for which there exists a constant C_Δ such that every graph of maximum degree Δ is (C_Δ, C_Δ)-relaxed colorable. On the one hand they showed that any graph of maximum degre 4 is $(57,57)$-relaxed colorable, on the other hand for any constant C they gave an example of a 6-regular graph which is not (C,C)-relaxed colorable. Haxell, Szabó, and Tardos [3] answered the open case of degree 5 by showing that every graph of degree maximum 5 is $(17617, 17617)$-relaxed colorable. In [3] also the bound on the component order for graphs of maximum degree 4 was improved from 57 to 6.

Asymmetric case. The problem of $(1,C)$-relaxed colorings is equivalent to determining when it is possible to break up a graph into small (i.e., of order at most C) components with the removal of an independent set. Again, we are concerned about giving a condition based on the maximum degree.

It is easy to see that every graph of maximum degree 2 is $(1,2)$-relaxed colorable. An adaptation of the construction of [1] shows that there is no constant $f(4)$ such that every 4-regular graph is $(1, f(4))$-relaxed colorable. In an earlier work [2] we showed that there exists a constant $f(3)$ such that every graph of maximum degree 3 is $(1, f(3))$-relaxed colorable.

Triviality and hardness. In the present work we study relaxed colorings from two points of view, extremal graph theory and complexity theory, and find that these points eventually meet for asymmetric relaxed colorings. We also make the first steps for a similar connection in the symmetric case.

On the one hand there is the purely graph theoretic question:

> For a given maximum degree Δ what is the smallest component order $f(\Delta) \in \mathbb{N} \cup \{\infty\}$ such that every graph of maximum degree Δ is $(1, f(\Delta))$-relaxed colorable?

On the other hand, for fixed Δ and C one can study the computational complexity question:

> What's the complexity of the decision problem: Given a graph of maximum degree Δ, is there a $(1,C)$-relaxed coloring?

Obviously, for the critical component order $f(\Delta)$ which answers the extremal graph theory question, the answer is *trivial* for the complexity question: every instance is a YES-instance. Note also, that for $C = 1$ the complexity question is polynomial-time solvable, as it is equivalent to deciding whether a graph is bipartite.

We investigate the complexity question in the range between 2 and the critical component order $f(\Delta)$. It is maybe worthwhile to note that at the moment we do not see any *a priori* reason why the hardness of the decision problem should even be monotone in the component order C, i.e. why the hardness of the problem for component order $C+1$ should imply the hardness for component order C. In fact the problem is obviously polynomial time decidable for $C = 1$, while for $C = 2$ we show NP-completeness. Here we establish the monotonicity of the hardness of the problem in the interval $C \geq 2$ and prove a very sharp "hardness

jump". By this we mean that the problem is NP-hard for every component order $2 \leq C < f(\Delta)$, while, of course, the problem becomes trivial (i.e. all instances are "YES"-instances) for component order $f(\Delta)$.

The other main contribution of our paper concerns the extremal graph theory question and obtains significant improvements over previously known bounds and algorithms. This result becomes particularly important in light of our NP-hardness results, as the exact determination of the place of the jump from NP-hard to trivial gets within reach.

To formalize our theorems we need further definitions. Let us denote by (Δ, C)-AsymRelCol the decision problem whether a given graph G of maximum degree at most Δ allows a $(1, C)$-relaxed coloring. Analogously, let us denote by (Δ, C)-SymRelCol the decision problem whether a given graph G of maximum degree at most Δ allows a (C, C)-relaxed coloring. Note here that both $(\Delta, 1)$-AsymRelCol and $(\Delta, 1)$-SymRelCol is simply testing whether a graph of maximum degree Δ is bipartite.

We prove the following.

Theorem 1. *(k, C)-AsymRelCol is* NP*-complete for every $2 \leq C < f(k)$.*

We also reduce the bounds involving the point of the hardness-jump and give a quazi-linear time algorithm for finding a relaxed coloring of cubic graphs.

Theorem 2. *We have*

$$\begin{aligned} 6 \leq f(3) &\leq 22 \\ f(4) &= \infty. \end{aligned}$$

Moreover there is an $O(n \log^4 n)$ algorithm which finds a $(1, 22)$-relaxed coloring of any n-vertex graph of maximum degree 3.

For the symmetric case we know much less.

Theorem 3. *The problems $(4, 2)$-SymRelCol, $(4, 3)$-SymRelCol and, for $C \geq 2$, $(6, C)$-SymRelCol are* NP*-complete.*

Open problems. In light of our hardness-jump result it would be very interesting to determine exactly the critical monochromatic component order $f(3)$ from where the problem $(3, C)$-AsymRelCol becomes trivial.

We conjecture that the problem $(4, C)$-SymRelCol exhibits a similar sudden hardness-jump. Such a result would particularly be interesting, since here the determination of the critical component order is even more within reach (between 4 and 6.)

Let $g(\Delta, n)$ be the smallest integer g such that every n-vertex graph of maximum degree Δ is (g, g)-relaxed colorable. Motivated by the fact that $g(n, 5) = O(1)$ [3] and their result [5] showing that $g(n, 7) = \Omega(n)$, in the open problem session Nati Linial raised the problem of determining the order of $g(n, 6)$. By a theorem of Hochberg, McDiarmid and Saks [4], for any two-coloring the triangular lattice

(which has maximum degree 6) does contain a monochromatic component of order $\Omega(\sqrt{n})$. Linial conjectures $g(n,6) = \Theta(\sqrt{n})$.

For colorings with more than two colors we know much less. Even the graph theoretic questions about interesting maximum degrees are open. We list here the most natural question together with its simplest open special case.

- Determine asymptotically the largest Δ_k for which there exists a constant C_k such that every graph of maximum degree Δ_k can be k-colored such that every monochromatic component is of order at most C_k. The current bounds are $3 < \Delta_k/k \leq 4$ (see [3]).
- Is there a constant C such that every graph with maximum degree 9 can be three-colored such that every monochromatic component is of order at most C? The answer is "yes" for graphs with maximum degree 8 and "no" for graphs of maximum degree 10 (see [3]).

References

[1] N. Alon, G. Ding, B. Oporowski, D. Vertigan, Partitioning into graphs with only small components, *Journal of Combinatorial Theory, (Series B)*, **87**(2003) 231-243.

[2] R. Berke, T. Szabó, Relaxed Coloring of Cubic Graphs, *submitted*; extended abstract in *Proceedings EuroComb* (2005), 341-344.

[3] P. Haxell, T. Szabó, G. Tardos, Bounded size components — partitions and transversals, *Journal of Combinatorial Theory (Series B)*, **88** no.2. (2003) 281-297.

[4] R. Hochberg, C. McDiarmid, M. Saks, On the bandwidth of triangulated triangles. *Discrete Mathematics* **138** (1995), no. 1-3, 261–265.

[5] N. Linial, personal communication.

Spanning 3-colourable subgraphs of small bandwidth in dense graphs

Anusch Taraz

(joint work with Julia Böttcher and Mathias Schacht)

The study of sufficient degree conditions which imply that a given graph G satisfies a certain property is one of the central themes in extremal graph theory. Here we will be concerned with conditions on the minimum degree of G which guarantee that G contains a copy of a particular spanning subgraph H.

A well known example of such a result is Dirac's theorem [5]. It asserts that any graph G on n vertices with minimum degree $\delta(G) \geq n/2$ contains a spanning, so called Hamiltonian, cycle. Another classical result of that type by Corrádi and Hajnal [4] states that every graph G with n vertices and $\delta(G) \geq 2n/3$ contains $\lfloor n/3 \rfloor$ vertex disjoint triangles. This was generalised by Hajnal and Szemerédi [10], who proved that every graph G with $\delta(G) \geq (r-1)n/r$ must contain a family of $\lfloor n/r \rfloor$ vertex disjoint cliques, each of size r.

Pósa (see, e.g., [6]) and Seymour [19] indicated how these theorems could actually fit into a common framework. They conjectured that, at the same threshold $\delta(G) \geq (r-1)n/r$, one can in fact ask for 'well-connected' cliques, more precisely that such a graph G contains a copy of the $(r-1)$-st power of a Hamiltonian cycle

(where the $(r-1)$-st power of an arbitrary graph is obtained by inserting an edge between every two vertices of distance at most $r-1$ in the original graph). Following earlier partial results [8, 9, 15], this conjecture was proven by Komlós, Sárközy, and Szemerédi [13, 16] for sufficiently large graphs G.

Recently, several other results of a similar flavour have been obtained which deal with a variety of spanning subgraphs H, such as, e.g., trees, F-factors, and planar graphs (see the surveys [17, 18]). Facing this wealth of results, there seems to be a need for a unifying generalisation. Which parameter(s) of H determine the minimum degree threshold for G to guarantee a spanning copy of H as a subgraph? The results above indicate that the chromatic number of H plays a crucial rôle. Obviously, by the classical results of Turán [21] and of Erdős and Stone [7], any graph H of *constant size* with $\chi(H) = r$, is forced to appear as a subgraph in any sufficiently large graph G if $\delta(G) \geq (\frac{r-2}{r-1} + \gamma)n$. However, if H has *as many vertices as* G and if in every r-colouring of H the colour classes are of the same size, then it is clear that we do indeed need $\delta(G) \geq \frac{r-1}{r}n$. For example, let G be the complete r-partite graph with partition classes almost, but not exactly, of the same size and let H be the union of vertex disjoint r-cliques.

Thus, in an attempt to move away from results that concern only graphs H with a special, rigid structure, a naïve conjecture could be that $\delta(G) \geq (\frac{r-1}{r} + \gamma)n$ suffices to guarantee that G contains a spanning copy of any r-chromatic graph H of bounded maximum degree. While the results mentioned above are in accordance with this idea, it is known that it fails in general as the following simple example shows. Let H be a random bipartite graph with bounded maximum degree and partition classes of size $n/2$ each, and let G be the graph formed by two cliques of size $(1/2 + \gamma)n$ each, which share exactly $2\gamma n$ vertices. It is then easy to see that G cannot contain a copy of H, since in H every set of vertices of size $(1/2 - \gamma)n$ has more than $2\gamma n$ neighbours.

One way to rule out such expansion properties for H, is to restrict the *bandwidth* of H. A graph is said to have bandwidth at most b, if there exists a labelling of the vertices by numbers $1, \ldots, n$, such that for every edge $\{i, j\}$ of the graph we have $|i - j| \leq b$. Bollobás and Komlós [12, Conjecture 16] conjectured that every r-chromatic graph on n vertices of bounded degree and bandwidth limited by $o(n)$, can be embedded into any graph G on n vertices with $\delta(G) \geq (\frac{r-1}{r} + \gamma)n$. Recently [3], we managed to prove this conjecture for the case $r = 3$.

Theorem 1. *For all $\Delta \in \mathbb{N}$ and $\gamma > 0$, there exist constants $\beta > 0$ and $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$ the following holds.*

If H is a 3-chromatic graph on n vertices with $\Delta(H) \leq \Delta$, and bandwidth at most βn and if G is a graph on n vertices with minimum degree $\delta(G) \geq (2/3+\gamma)n$, then G contains a copy of H.

We note that our proof can be turned into an algorithm. More precisely, an embedding of H can be found in $O(n^{3.376})$ if H is given along with a valid 3-colouring and a labelling of the vertices respecting the bandwidth bound βn.

Theorem 1 embraces a fairly large class of 3-chromatic graphs H. In fact, most graphs H considered so far were of constant bandwidth, whereas Theorem 1 includes for example (higher-dimensional) grid graphs as possible graphs H.

The analogue of Theorem 1 for bipartite H was announced by Abbasi [1] in 1998, and can now easily be obtained by our methods (see [11]), too. In [2] it is shown that in this case no sharp version of Theorem 1 (with $\gamma = 0$) is possible. More precisely, it is shown that if $\gamma \to 0$ and $\Delta \to \infty$ then β must tend to 0 in Theorem 1. However, the bound on β coming from our proof is rather poor, having a tower-type dependence on $1/\gamma$.

The proof of Theorem 1 is based on the *regularity method* and uses, in particular, the regularity lemma [20] and the blow-up lemma [14]. There is a well established strategy for proofs of this kind, which, as described by Komlós in his survey [12], proceeds in several steps: First, prepare the graph H by dividing it into a constant number of smaller pieces, which is usually possible and not too difficult by calling upon the structural properties guaranteed for H. Secondly, prepare the graph G by applying the regularity lemma and thus obtaining a sufficiently regular vertex partition. Thirdly, find an assignment that maps vertices of H to the partition classes of G. Fourthly, ensure that the edges between the different parts of H are mapped to edges in G. Finally, complete the embedding by applying the blow-up lemma to the individual pieces of H and their counterparts in G.

Steps 2, 3, and 5 have been standardised by the use of the powerful tools mentioned above, but the proofs are still technically rather involved: although H and G have been 'prepared' roughly for each other, there is still a great deal of details that have to be carefully adjusted and fitted, especially in step 4. Since, in our case, we have very little control about the structure of H, this difficulty becomes particularly pressing. In order to avoid having to deal with many separate cases, we have pushed the agenda described above a bit further by splitting the proof into two main lemmas.

While they deal exclusively with the graph G and the graph H respectively, they are linked to each other in the following way: the *lemma for G* will suggest a partition of G and communicate the structure of this partition (but not the graph G) to the *lemma for H*. The lemma for H will then try to find a partition of H with a very similar structure, and return the sizes of its partition classes to the lemma for G. The latter will then adjust its partition classes by shifting a few vertices of G, until they fit exactly the class sizes of H. The embedding of H into G can then be found using (a slight variant of) the embedding lemma first used by Chvátal et al. for step 4 and the blow-up lemma for step 5.

This approach provides a very modular proof strategy that can easily be checked and may be of further use for other similar problems. For example, our current work-in-progress indicates that a proof of the Bollobás-Komlós conjecture for general r-chromatic graphs H is now within reach.

References

[1] S. Abbasi, *Embedding low bandwidth bipartite graphs*, unpublished.

[2] S. Abbasi, *How tight is the Bollobás-Komlós conjecture?*, Graphs Combin. **16** (2000), no. 2, 129–137.
[3] J. Böttcher, M. Schacht, and A. Taraz, *The bandwidth conjecture for 3-colourable graphs*, Extended abstract, to appear in Proc. of the ACM-SIAM Symposium on Discrete Algorithms (SODA), 2007.
[4] K. Corradi and A. Hajnal, *On the maximal number of independent circuits in a graph*, Acta Math. Acad. Sci. Hungar. **14** (1963), 423–439.
[5] G. A. Dirac, *Some theorems on abstract graphs*, Proc. London Math. Soc. (3) **2** (1952), 69–81.
[6] P. Erdős, *Problem 9*, Theory of Graphs and its Applications (Proc. Sympos. Smolenice, 1963) (M. Fiedler, ed.), Publ. House Czechoslovak Acad. Sci., Prague, 1964, p. 159.
[7] P. Erdös and A. H. Stone, *On the structure of linear graphs*, Bull. Amer. Math. Soc. **52** (1946), 1087–1091.
[8] G. Fan and H. A. Kierstead, *The square of paths and cycles*, J. Combin. Theory Ser. B **63** (1995), no. 1, 55–64.
[9] G. Fan and H. A. Kierstead, *Hamiltonian square-paths*, J. Combin. Theory Ser. B **67** (1996), no. 2, 167–182.
[10] A. Hajnal and E. Szemerédi, *Proof of a conjecture of P. Erdős*, Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969), North-Holland, Amsterdam, 1970, pp. 601–623.
[11] H. Hàn, *Einbettungen bipartiter Graphen mit kleiner Bandbreite*, Master's thesis, Humboldt-Universität zu Berlin, Institut für Informatik, Fall 2006, expected.
[12] J. Komlós, *The blow-up lemma*, Combin. Probab. Comput. **8** (1999), no. 1-2, 161–176, Recent trends in combinatorics (Mátraháza, 1995).
[13] J. Komlós, G. N. Sárközy, and E. Szemerédi, *On the square of a Hamiltonian cycle in dense graphs*, Random Structures Algorithms **9** (1996), no. 1-2, 193–211.
[14] J. Komlós, G. N. Sárközy, and E. Szemerédi, *Blow-up lemma*, Combinatorica **17** (1997), no. 1, 109–123.
[15] J. Komlós, G. N. Sárközy, and E. Szemerédi, *On the Pósa-Seymour conjecture*, J. Graph Theory **29** (1998), no. 3, 167–176.
[16] J. Komlós, G. N. Sárközy, and E. Szemerédi, *Proof of the Seymour conjecture for large graphs*, Ann. Comb. **2** (1998), no. 1, 43–60.
[17] J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi, *The regularity lemma and its applications in graph theory*, Lecture Notes in Comput. Sci. **2292** (2002), 84–112.
[18] J. Komlós and M. Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*, Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993) **2** (1996), 295–352.
[19] P. Seymour, *Problem section*, Combinatorics (Proc. British Combinatorial Conf. 1973) (T. P. McDonough and V. C. Mavron, eds.), Cambridge Univ. Press, London, 1974, pp. 201–204. London Math. Soc. Lecture Note Ser., No. 13.
[20] E. Szemerédi, *Regular partitions of graphs*, Problèmes combinatoires et théorie des graphes (Orsay, 1976), Colloques Internationaux CNRS, vol. 260, CNRS, 1978, pp. 399–401.
[21] P. Turán, *Eine Extremalaufgabe aus der Graphentheorie*, Mat. Fiz. Lapok **48** (1941), 436–452.

On the number of lattice triangulations

Emo Welzl
(joint work with Jiří Matoušek and Pavel Valtr)

For n a positive integer, we consider triangulations of the $n \times n$ lattice set $\{0, 1, 2, \ldots, n\}^2$, i.e. crossing-free straight line embedded geometric graphs on this point set—thus with $(n+1)^2$ vertices, $3n^2 + 2n$ edges and $2n^2$ triangular faces.

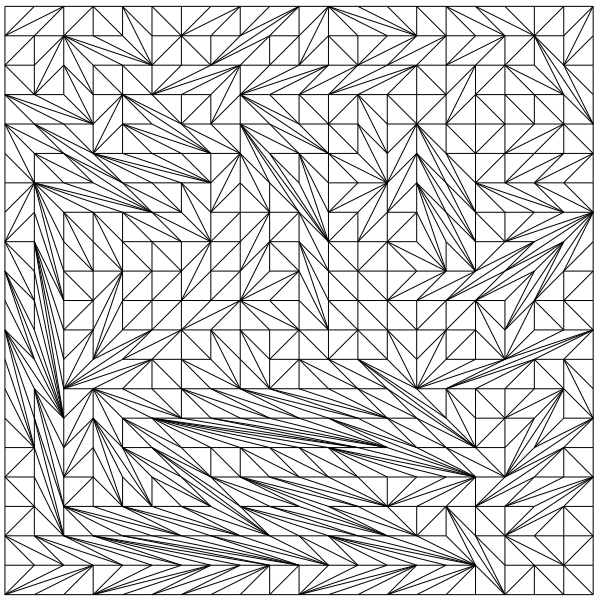

Figure 1. A triangulation of the 20×20 lattice.

Extending a previous argument by Emile Anclin [1], we show that the number of triangulations of the $n \times n$ lattice is at most $O(6.86^{n^2})$, improving on the previous bounds of $O(64^{n^2})$ and $O(8^{n^2})$ in [4] and [1], respectively. It compares to a lower bound of $\Omega(4.15^n)$ given in [2].

The flip-graph has the triangulations as vertices, and it has two triangulations adjacent if one can be obtained from the other by replacing one single edge (an edge whose incident triangles form a convex quadrilateral, called a flippable edge). We demonstrate that the flip-graph of the triangulations of the $n \times n$ lattice is an induced subgraph of the $(3n^2 - 2n)$-dimensional hypercube (no such embedding in a hypercube of smaller dimension is possible). We also show that the diameter of the flip-graph is $\Theta(n^3)$, and in a random triangulation (uniformly from all triangulations), the expected number of flippable edges is $\Theta(n^2)$ (while there exist triangulations with as few as $O(n)$ flippable edges).

The main proofs are based on particular binary encodings of lattice triangulations which readily yield the respective results.

References

[1] E. E. Anclin, An upper bound for the number of planar lattice triangulations, *J. Combinat. Theory, Ser. A* **103** (2003), 383–386.
[2] V. Kaibel, G. Ziegler, Counting lattice triangulations, *British Combinatorial Surveys* (C.D. Wensley, ed.), Cambridge University Press, 2003.
[3] J. Matoušek, P. Valtr, E. Welzl, On two encodings of lattice triangulations, manuscript (2006).
[4] S. Yu. Orevkov, Asymptotic number of triangulations in $\mathbf{Z}^2$, *J. Combinat. Theory, Ser. A* **86** (1999), 200–203.

Runtime analysis of a modern randomized search heuristic

Carsten Witt

The rigorous mathematical analysis of randomized search heuristics (RSHs) with respect to their expected runtime is a growing research area where many results have been obtained in recent years. This class of heuristics contains well-known approaches such as Randomized Local Search (RLS), the Metropolis Algorithm (MA), Simulated Annealing (SA), and Evolutionary Algorithms (EAs). Such heuristics are often applied to problems whose structure is not known or if there are not enough resources such as time, money, or knowledge to obtain good specific algorithms. It is widely acknowledged that a solid mathematical foundation for such heuristics is needed.

Most designers of RSHs, however, rather focused on mimicking processes in nature (such as evolution) rather than making the heuristics amenable to a mathematical analysis. This is different to the classical design of (randomized) algorithms which are developed with their theoretical analysis of runtime (and proof of correctness) in mind. Despite these obstacles, research from the last about 15 years has shown how to apply the methods for the probabilistic analysis of randomized algorithms to RSHs. Mostly, the expected runtime of RSHs on selected problems is analzyed. Thereby, we understand why and when RSHs are efficient optimizers and, conversely, when they cannot be efficient.

Some general results on the expected runtime of RLS can be found in [11]. The graph bisection problem has been subject to analysis of SA [9] and MA [8], where MA can be seen as SA with a fixed temperature. For a long time, it was an open question whether there is a natural example where SA outperforms MA for all fixed temperatures. This question has recently been answered positively by [12] for instances of the minimum spanning tree problem.

In this study (joint work with Frank Neumann, Max-Planck-Institut für Informatik Saarbrücken), we focus on a recently developed kind of RSHs, namely Ant Colony Optimization (ACO). Like EAs, these heuristics imitate optimization processes from nature, in this case the search of an ant colony for a common source of food. Solving problems by ACO techniques has become quite popular in recent years. Developed by [2], they have shown to be a powerful heuristic approach to solve combinatorial optimization problems (for an overview on the problems that

these heuristics have been applied to, see [3]). From a theoretical point of view, there are no results that provide estimates of the expected runtime of ACO algorithms. Despite interesting theoretical investigations of models and dynamics of ACO algorithms [1], convergence results are so far the only results related to their runtimes. [1] explicitly formulate the open problem to determine the expected runtime of ACO algorithms on simple problems in a similar fashion to what has been done for EAs.

We solve this problem, starting the analysis of ACO algorithms with respect to their expected runtimes and success probability after a specific number of steps. RLS, SA, MA, and simple EAs search more or less locally, and runtime bounds are often obtained by considering the neighborhood structure of the considered problem. Considering ACO algorithms, this is different as search points are obtained by random walks of ants on a so-called construction graph. The traversal of an ant on this graph is determined by values on the edges which are called pheromone values. Larger pheromone values correspond to a higher probability of traversing a certain edge, where the choice of an edge usually fixes a parameter in the current search space. The pheromone values are updated if a good solution has been constructed in this random walk. This update depends on the traversal of the ant and a so-called evaporation factor ρ.

The choice of ρ seems to be a crucial parameter in an ACO algorithm. Using a large value of ρ, the last accepted solution changes the pheromone values by a large amount such that there is a large probability of producing this solution in the next step. In contrast to this, the use of a small evaporation factor leads to a small effect of the last accepted solution such that an improvement may be hard to find in the next step. We show that a simple ACO algorithm behaves for very large values of ρ (namely $\rho \geq 1/3$) as the simplest EA called (1+1) EA. This algorithm has been studied extensively with respect to its runtime on pseudo-boolean functions $f\colon \{0,1\}^n \to \mathbb{R}$ (see e.g. [4]) as well as on combinatorial optimization problems. The list of problems where runtime bounds have been obtained include some of the best-known polynomially solvable problems such as maximum matchings [5] and minimum spanning trees [10]. It should be clear that we cannot expect such general heuristics to outperform the best-known algorithms for these mentioned problems. The main aim of such analyses is to get an understanding how these heuristics work. In the case of NP-hard problems, one is usually interested in good approximations of optimal solutions. [13] has presented a worst-case and average-case analysis of the (1+1) EA for the partition problem, which is one of the first results on NP-hard problems. All these results immediately transfer to our ACO algorithm with very large ρ.

After these general results, we consider the effect of the evaporation factor ρ on the runtime of our ACO algorithm in detail. As proposed in the open problem stated by [1], we examine the simplest non-trivial pseudo-boolean function called OneMax and analyze for the first time for which choices of ρ the runtime with high probability is upper bounded by a polynomial and for which choices it is exponential. We observe a phase transition from exponential to small polynomial

runtime when ρ crosses the threshold value $1/n$. Larger values of ρ imply that the expected function value of a new solution is determined by the function value of the best seen solution. Then an improvement will be achieved after an expected polynomial number of steps. In the case of smaller ρ, an improvement does not increase the expected function value sufficiently. Here exponential lower bounds are obtained by showing that there is a large gap between the expected value and the best-so-far function value.

In the analysis, we are confronted with challenging mathematical questions. The stochastic process induced by our ACO algorithm is governed by a sum of independent Poisson trials with different success probabilities. Usually, the tails of such random variables are upper bounded by Chernoff-Hoeffding bounds. In our analysis, however, we need lower bounds on these tails and develop new analytical tools to arrive at such bounds. In particular, we extend a less-known result by [6] that characterizes the extreme values of the considered type of random variables and can be used as a counterpart to the upper bounds provided by Hoeffding bounds [7]. Our new tools may be of independent interest in other probabilistic analyses.

References

[1] Dorigo, M. and Blum, C. (2005), Ant colony optimization theory: A survey, *Theor. Comput. Sci.*, **344**, 243–278.

[2] Dorigo, M., Maniezzo, V., and Colorni, A. (1991), The ant system: An autocatalytic optimizing process, Tech. Rep. 91-016 Revised, Politecnico di Milano.

[3] Dorigo, M. and Stützle, T. (2004), *Ant Colony Optimization*, MIT Press.

[4] Droste, S., Jansen, T., and Wegener, I. (2002), On the analysis of the (1+1) evolutionary algorithm, *Theor. Comput. Sci.*, **276**, 51–81.

[5] Giel, O. and Wegener, I. (2003), Evolutionary algorithms and the maximum matching problem, In *Proc. of STACS '03*, vol. 2607 of *LNCS*, 415–426.

[6] Hoeffding, W. (1956), On the distribution of the number of successes in independent trials, *Ann. Math. Stat.*, **27**, 713–721.

[7] Hoeffding, W. (1963), Probability inequalities for sums of bounded random variables, *American Statistical Association Journal*, **58**(301), 13–30.

[8] Jerrum, M. and Sorkin, G. B. (1998), The metropolis algorithm for graph bisection, *Discrete Appl. Math.*, **82**(1–3), 155–175.

[9] Jerrum, M. R. and Sorkin, G. B. (1993), Simulating annealing for graph bisection, In *Proc. of FOCS '93*, 94–103.

[10] Neumann, F. and Wegener, I. (2004), Randomized local search, evolutionary algorithms, and the minimum spanning tree problem, In *Proc. of GECCO 04*, vol. 3102 of *LNCS*, 713–724.

[11] Papadimitriou, C. H., Schäffer, A. A., and Yannakakis, M. (1990), On the complexity of local search, In *Proc. of STOC '90*, 438–445. ACM Press.

[12] Wegener, I. (2005), Simulated annealing beats metropolis in combinatorial optimization, In *Proc. of ICALP '05*, vol. 3580 of *LNCS*, 589–601.

[13] Witt, C. (2005), Worst-case and average-case approximations by simple randomized search heuristics, In *Proc. of STACS '05*, vol. 3404 of *LNCS*, 44–56.

Maintaining external memory efficient hash tables

Philipp Woelfel

We present randomized algorithms which maintain hash tables efficiently under circumstances typical for applications dealing with massive data. Consider a set S of n keys from a finite universe U and assume that each key $x \in S$ has some data D_x associated with it. A static dictionary for S supports a query operation which returns the data D_x for a given key x. A dynamic dictionary also supports update operations which allow to insert new data into the dictionary or to remove data from it. A typical solution is to maintain a hash function h mapping each key $x \in S$ to an entry of a hash table T. Such a hash function h is called *perfect for* S if it is injective on S. If in addition h has range $[n]$, where $n = |S|$, then h is called *minimal perfect*.

The following assumptions are typical for many dictionary and hashing applications: Firstly, efficiency is much more a concern for lookups than for update operations. Secondly, the data set is so massive that even the description of the hash function does not fit into the internal memory – then usually the number of non-consecutive accesses to external memory dominate the evaluation time of the hash function. Finally, the data D_x associated with a key x requires much more space than its key x. Therefore it is especially important that a hash table implementation achieves a high utilization, since we have to reserve a fixed amount of space for each table entry (if we want to avoid another level of indirection). Assuming that the hash table is implemented by an extendible array $T[0], T[1], \ldots$, its utilization is given as $|S|/(\max\{h(S)\}+1)$. In particular, even a small constant utilization seems infeasible, and a utilization as close to 100% as possible should be achieved. A minimal perfect hash function for the set S achieves 100% utilization.

For the following discussion we assume that the size of the universe is $|U| = n^{O(1)}$ and that the size of the internal memory is bounded by n^ϵ, $\epsilon < 1$. Fredman, Komlós, and Szemerédi [6] were the first to devise an algorithm which constructs a perfect hash function (with $O(n \log n)$ bits) in expected linear time such that the hash function can be evaluated in constant time. The utilization is less than 0.2 in the case that only consecutive probes into external memory are allowed for hash function evaluation. Dietzfelbinger, Karlin, Mehlhorn, Meyer auf der Heide, Rohnert, and Tarjan [3] have devised a dynamic version of that scheme with essentially the same parameters, but which also supports updates in expected amortized constant time. Later improvements have either focused on reducing the space requirements or on obtaining a constant update time even with high probability. Demaine, Meyer auf der Heide, Pagh, and Patrăscu [1] show an upper space bound of essentially $O(n \log\log(u/n) + n \log n - n \log t)$ for maintaining a perfect hash function with range $[n+t]$ and $O(n \log\log(u/n) + n \cdot r)$ for a dynamic dictionary where the data associated with each key comprises r bits. Update operations are supported with high probability in constant time and the algorithm is stable (stable means that $h(x)$ remains fixed for the duration that x is in S). For the static case, Hagerup and Tholey [7] hold the space record: They show how to

construct a minimal perfect hash function in expected $O(n + \log\log|U|)$ time such that its encoding requires only almost optimal $(1+o(1))(n\cdot\log e+\log\log|U|)$ space. Multiple non-oblivious probes into external memory are required for lookups in these space efficient dynamic or static schemes.

Dictionary algorithms such as Cuckoo-Hashing [9] and its extensions [5, 4] also allow the retrieval of data with few non-consecutive probes into external memory. Especially space and external memory efficient is the Cuckoo-Hashing variant of Dietzfelbinger and Weidling [4], where two hash functions h_1 and h_2 and two hash tables T_1 and T_2 are used. A table position consists of d consecutive memory cells, and the data D_x is stored in one of the $2 \cdot d$ memory cells from $T_1[h_1(x)]$ and $T_2[h_2(x)]$. For $d \geq 90 \cdot \ln(1/\epsilon)$ a utilization of $1-\epsilon$ can be achieved and the data can be retrieved with only two non-consecutive probes into external memory. Due to the large constant for d, this scheme may be disadvantageous if the data associated with the keys is very large (now the time for finding D_x depends on the size of the data). Moreover, such dictionary solutions do not provide perfect hash functions without an additional level of indirection.

Pagh [8] showed how to construct in linear time a minimal perfect hash function which can be evaluated very efficiently with simple arithmetics (essentially one or two multiplications) and by probing only one word from external memory. The hash function itself can be encoded in $(2+\epsilon)\cdot n\cdot\log n$ bits. Dietzfelbinger and Hagerup [2] improved Pagh's scheme so that the resulting hash function can be encoded with $(1+\epsilon)\cdot n\cdot\log n$ bits. Both schemes yield a static dictionary with 100% utilization.

In this talk we present a dynamic variant of Pagh's scheme. Maintaining 100% utilization and using exactly the same hash functions, we show how to perform updates in expected amortized constant time. I.e., the hash functions can be evaluated very efficiently in constant time and with only one probe into external memory. In addition to the $(2+\epsilon)\cdot n\cdot\log n$ bits for encoding of the hash function we also need an auxiliary data structure comprising $(3+\epsilon)\cdot n\cdot\log n$ bit. However, this auxiliary data structure is only needed for update operations and not for lookups. For many applications updates occur infrequently, e.g., at night time, so that the auxiliary data structure may be swapped out (or it can be removed and later be rebuild from scratch in expected linear time if needed). We believe that this scheme is quite practical if the main focus is on lookup performance, although the algorithm for updates is not very simple.

We also investigate how much the space for the hash function description can be reduced under the constraint that evaluation requires only consecutive probes into external memory. We show that it is possible to reduce the encoding size of the hash functions and the space for the auxiliary data structure to $O(n\log\log n)$ bits. In the dynamic case we obtain a utilization of $1-\epsilon$, for arbitrary small $\epsilon > 0$. In the static case we still achieve 100% utilization, hence we even have a minimal perfect hash function. For both implicit versions the corresponding hash functions can be evaluated in constant time and by probing $O(1)$ consecutive words from external memory.

References

[1] E. D. Demaine, F. Meyer auf der Heide, R. Pagh, and M. Pătrascu, De dictionariis dynamicis pauco spatio utentibus (lat. on dynamic dictionaries using little space), In *Proc. of Proc. of 7th LATIN*, volume 3887 of *LNCS*, pp. 349–361. 2006.

[2] M. Dietzfelbinger and T. Hagerup, Simple minimal perfect hashing in less space, In *Proc. of 9th ESA*, volume 2161 of *LNCS*, pp. 109–120. 2001.

[3] M. Dietzfelbinger, A. Karlin, K. Mehlhorn, F. Meyer auf der Heide, H. Rohnert, and R. E. Tarjan, Dynamic perfect hashing: Upper and lower bounds, *SIAM J. on Comp.*, 23:738–761, 1994.

[4] M. Dietzfelbinger and C. Weidling, Balanced allocation and dictionaries with tightly packed constant size bins, In *Proc. of 32nd ICALP*, volume 3580 of *LNCS*, pp. 166–178. 2005.

[5] D. Fotakis, R. Pagh, P. Sanders, and P. G. Spirakis, Space efficient hash tables with worst case constant access time, *Theory of Comp. Syst.*, 38:229–248, 2005.

[6] M. L. Fredman, J. Komlós, and E. Szemerédi, Storing a sparse table with $O(1)$ worst case access time, *J. of the ACM*, 31:538–544, 1984.

[7] T. Hagerup and T. Tholey, Efficient minimal perfect hashing in nearly minimal space, In *Proc. of 18th STACS*, volume 2010 of *LNCS*, pp. 317–326. 2001.

[8] R. Pagh, Hash and displace: Efficient evaluation of minimal perfect hash functions, In *Proc. of 6th WADS*, volume 1663 of *LNCS*, pp. 49–54, 1999.

[9] R. Pagh and F. F. Rodler, Cuckoo hashing, *J. of Alg.*, 51:122–144, 2004.

Reporter: Alexander D. Scott

Participants

Prof. Dr. Noga Alon
Department of Mathematics
Sackler Faculty of Exact Sciences
Tel Aviv University
Tel Aviv 69978
ISRAEL

Prof. Dr. Jozsef Balogh
Department of Mathematics
University of Illinois at
Urbana-Champaign
1409 West Green Street
Urbana IL 61801
USA

Prof. Dr. Thomas A. Bohman
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15213-3890
USA

Prof. Dr. Bela Bollobas
Dept. of Pure Mathematics and
Mathematical Statistics
University of Cambridge
Wilberforce Road
GB-Cambridge CB3 OWB

Prof. Dr. Christian Borgs
Microsoft Research
One Microsoft Way
Redmond, WA 98052-6399
USA

Prof. Graham R. Brightwell
Dept. of Mathematics
London School of Economics
Houghton Street
GB-London WC2A 2AE

Prof. Dr. Jennifer Chayes
Microsoft Research
One Microsoft Way
Redmond, WA 98052-6399
USA

Dr. Amin Coja-Oghlan
Institut für Informatik
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin

Prof. Dr. Artur Czumaj
Department of Computer Science
University of Warwick
GB-Coventry CV4 7AL

Prof. Dr. Martin Dietzfelbinger
Fakultät für Informatik und
Automatisierung
Technische Universität Ilmenau
Postfach 100565
98684 Ilmenau

Priv-Doz.Dr. Benjamin Doerr
Max-Planck-Institut für Informatik
Geb. 46
Stuhlsatzenhausweg 85
66123 Saarbrücken

Prof. Dr. Uriel Feige
Microsoft Research
One Microsoft Way
Redmond, WA 98052
USA

Prof. Dr. Ehud Friedgut
Institute of Mathematics
The Hebrew University
Givat-Ram
91904 Jerusalem
ISRAEL

Prof. Dr. Alan M. Frieze
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15213-3890
USA

Prof. Dr. Anna Gal
Department of Computer Science
University of Texas at Austin
Austin, TX 78712
USA

Dr. Stefanie Gerke
Institut für theoretische Informatik
ETH-Zentrum
Universitätstr.6
CH-8092 Zürich

Prof. Dr. Andreas Goerdt
Fakultät für Informatik
TU Chemnitz
Str. der Nationen 62
09107 Chemnitz

Prof. Dr. Svante Janson
Matematiska institutionen
Uppsala Universitet
Box 480
S-751 06 Uppsala

Prof. Dr. Mark R. Jerrum
School of Mathematical Sciences
Queen Mary
University of London
Mile End Road
GB-London E1 4NS

Prof. Dr. Jeff Kahn
Dept. of Mathematics
Rutgers University
Busch Campus, Hill Center
New Brunswick, NJ 08903
USA

Prof. Dr. Ravindran Kannan
Department of Computer Science
Yale University
P.O.Box 208285
New Haven CT 06520-8285
USA

Prof. Dr. Michal Karonski
Faculty of Mathematics and Computer Science
A. Mickiewicz University
ul. Umultowska 87
61-614 Poznan
POLAND

Dr. Peter Keevash
Dept. of Mathematics
California Institute of Technology
Pasadena, CA 91125
USA

Dr. Yoshiharu Kohayakawa
Instituto de Matematica e Estatistica
Universidade de Sao Paulo (IME-USP)
rua do Matao 1010
Sao Paulo 05508-090 - SP
BRAZIL

Prof. Dr. Matthias Krause
Fakultät für Mathematik und Informatik
Universität Mannheim
68131 Mannheim

Prof. Dr. Michael Krivelevich
Department of Mathematics
Sackler Faculty of Exact Sciences
Tel Aviv University
Tel Aviv 69978
ISRAEL

Prof. Dr. Nathan Linial
School of Computer Science and Engineering
The Hebrew University
Givat-Ram
91904 Jerusalem
ISRAEL

Dr. Eyal Lubetzky
Department of Mathematics
School of Mathematical Sciences
Tel Aviv University
Ramat Aviv, P.O. Box 39040
Tel Aviv 69978
ISRAEL

Prof. Dr. Peter Bro Miltersen
Dept. of Computer Science
University of Aarhus
IT-Parken, Aabogade 34
DK-8200 Aarhus N

Dr. Dhruv Mubayi
Dept. of Mathematics, Statistics and Computer Science, M/C 249
University of Illinois at Chicago
851 S. Morgan Street
Chicago, IL 60607-7045
USA

Prof. Dr. Assaf Naor
Microsoft Research
One Microsoft Way
Redmond, WA 98052-6399
USA

Prof. Dr. Hans Jürgen Prömel
Institut für Informatik
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin

Dr. Pavel Pudlak
Institute of Mathematics of the AV CR
Zitna 25
115 67 Praha 1
CZECH REPUBLIC

Prof. Dr. Rüdiger Reischuk
Institut für Theoretische Informatik
Universität Lübeck, Geb. 64
Ratzeburger Allee 160
23538 Lübeck

Dr. Oliver M. Riordan
Dept. of Pure Mathematics and Mathematical Statistics
University of Cambridge
Wilberforce Road
GB-Cambridge CB3 OWB

Prof. Dr. Andrzej Rucinski
Faculty of Mathematics and Computer Science
A. Mickiewicz University
ul. Umultowska 87
61-614 Poznan
POLAND

Prof. Dr. Alex Scott
Mathematical Institute,
University of Oxford,
24-29 St Giles',
Oxford, OX1 3LB, UK

Prof. Dr. Alistair J. Sinclair
Computer Science Division
University of California
at Berkeley
615 Soda Hall
Berkeley, CA 94720-1776
USA

Dr. Gregory B. Sorkin
Dept. of Mathematical Sciences
IBM Thomas J. Watson Research Center
P.O.Box 218
Yorktown Heights, NY 10598
USA

Prof. Dr. Angelika Steger
Institut für theoretische Informatik
ETH-Zentrum
Universitätstr.6
CH-8092 Zürich

Dr. Benjamin Sudakov
Department of Mathematics
Princeton University
Fine Hall
Washington Road
Princeton, NJ 08544
USA

Prof. Dr. Tibor Szabo
Institute of Theoretical Computer Science
ETH Zentrum
CH-8092 Zürich

Dr. Anusch Taraz
Zentrum Mathematik
Kombinatorische Geometrie (M9)
Technische Universität München
Boltzmannstr. 3
85747 Garching bei München

Prof. Dr. Ingo Wegener
Lehrstuhl für Informatik II
Universität Dortmund
44221 Dortmund

Prof. Dr. Emo Welzl
Theoretische Informatik
ETH Zürich
CH-8092 Zürich

Dr. Carsten Witt
Lehrstuhl für Informatik II
Universität Dortmund
44221 Dortmund

Prof. Dr. Philip Woelfel
Dept. of Computer Science
University of Toronto
10 King's College Road
Toronto Ontario M5S 3G4
CANADA