Mathematisches Forschungsinstitut Oberwolfach

Report No. 18/2013

DOI: 10.4171/OWR/2013/18

Combinatorics and Probability

Organised by
Béla Bollobás, Cambridge and Memphis
Michael Krivelevich, Tel Aviv
Emo Welzl, Zürich

14 April – 20 April 2013

Abstract. The main theme of this workshop was the use of probabilistic methods in combinatorics and theoretical computer science. Although these methods have been around for decades, they are being refined all the time: they are getting more and more sophisticated and powerful. Another theme was the study of random combinatorial structures, either for their own sake, or to tackle extremal questions. Both themes were richly represented at the workshop, with many recent exciting results presented by the lecturers.

Mathematics Subject Classification (2010): 05-06, 60-06.

Introduction by the Organisers

The workshop was organized by Béla Bollobás (Cambridge and Memphis), Michael Krivelevich (Tel Aviv) and Emo Welzl (Zürich). There were 52 participants from 14 countries. The scientific programme consisted of 8 main lectures, 19 shorter talks, and a problem session on Tuesday evening.

The lectures were about some of the most important recent developments in combinatorial mathematics, probabilistic combinatorics, and the related parts of theoretical computer science, and attested to the exciting and important activities in these fields. In fact, there is an embarrassment of riches: it would have been good to invite twice as many people as the Institute can accommodate.

The workshop started with a beautiful talk by Noga Alon on an extremal problem suggested by questions arising in Social Choice. This talk set the tone for the lectures that followed. Let us mention three of the particularly exciting talks.

Yufei Zhao reported on his joint work with Eyal Lubetzky on large deviations of random graphs. Given $0 < p < r < 1$, what can we say about a random

graph $G_{n,p}$ conditioned on the rare event that there are at least $\binom{n}{3}r^3$ triangles? Lubetzky and Zhao have greatly improved on the recent results of Chatterjee and Varadhan that for given r, if p is small enough, symmetry breaking takes place. Among other results, they determined the exact replica symmetric phase.

Robert Morris gave an excellent lecture on the major results he obtained recently with Gonzalo Fiz Pontiveros and Simon Griffiths on the triangle process and its consequences concerning the Ramsey number $R(3, k)$. The paper containing these results is well over 100 pages long and is full of intricate arguments, but Morris gave a masterful lecture giving us glimpses of the main ideas.

The third lecture we shall mention was given by Jeff Kahn on Turán's theorem for random graphs. Given $r \geq 3$, for what values of $p = p(n)$ is it true that whp the maximal size of a K_r-free subgraph of $G_{n,p}$ is equal to the maximal size of an $(r-1)$-partite subgraph? (Turán's theorem tells us that $p = 1$ is such a value.) This question has been studied for over two decades, with better and better bounds on the threshold. In his lecture, Jeff Kahn sketched a proof of his recent result with Robert DeMarco that pins down this threshold precisely.

All in all, the workshop was a great success, perhaps an even greater success than any of its predecessors. This is a testament to the vigorous activities in combinatorics and the related areas of probability theory and theoretical computer science.

Workshop: Combinatorics and Probability

Table of Contents

Abstracts

Minimal Majority Sequences

Noga Alon
(joint work with Robert Bredereck, Jiehua Chen, Stefan Kratsch, Rolf Niedermeier, Gerhard J. Woeginger)

We study an extremal combinatorial problem motivated by questions arising in the theory of Social Choice. A brief description of a possible motivation follows. In [1] there is a more detailed description of variants of this motivation.

Consider elections in which n parliament members are chosen. A prospective candidate for prime minister is trying to obtain the support of a majority of the members. To do so he has to select an agenda consisting of a subset of a given set M of m issues. The opinions of the parliament members on the issues are represented by an n by m matrix $A = (a_{ij})$ with $\{-1, 1\}$-entries, where $a_{ij} = 1$ iff member i supports issue j and $a_{ij} = -1$ otherwise. For $Q \subset [m] = \{1, 2, \ldots, m\}$, member i likes Q iff he supports more than half the issues indexed by the elements of Q, that is, iff $\sum_{j \in Q} a_{ij} > 0$. Call Q a winning agenda if more than $n/2$ members like it.

In [1] it is proved that the computational problem of deciding whether a given input n by m matrix A with $\{-1, 1\}$-entries admits a winning agenda is NP-complete, but fixed-parameter tractable with respect to the parameter n (or the parameter m).

Let $g(n, m)$ denote the minimum number g so that for any n by m matrix that admits a winning agenda, there exists such an agenda Q satisfying $|Q| \leq g$. Define, also $G(n) = \mathrm{Sup}\ g(n, m)$, where the supremum is taken over all values of m.

It is not difficult to check that $G(1) = G(2) = G(3) = 1, G(4) = G(5) = 3$ and $G(6) = G(7) = 5$, but it is not even clear whether or not $G(n)$ is finite for all n. The following result determines the asymptotic behavior of the function $G(n)$.

Theorem. $G(n) = n^{(1+o(1))n/4}$.

The proof combines combinatorial and geometric ideas with tools from linear algebra and discrepancy theory, and is related to results by Huckeman, Jurkat and Shapley on indecomposable hypergraphs (c.f. [3]), of Graham and Sloane on anti-Hadamard matrices ([4]), of Hastad on threshold gates requiring large weights ([5]), and of Alon and Vu on a certain coin weighing problem ([2]). The details appear in [1].

One can extend the discussion above to the weighted case, where there is an n by m matrix $A = (a_{ij})$ with $\{-1, 1\}$-entries, and there are also weights $w_1, \ldots, w_n$ and $z_1, \ldots, z_m$. Here $Q \subset [m]$ is a winning agenda iff the sum of weights w_i over all i so that $\sum_{j \in Q} a_{ij} z_j > 0$ exceeds $0.5 \sum_i w_i$. The weighted case corresponds to the motivating example in which there are n parties instead of parliament members,

each party is weighted by the number of members in it, and the issues are weighted according to their significance.

It can be shown that if the weights z_j are all equal, then the behavior is similar to the unweighted case. On the other hand when the weights are general positive integers the situation is very simple for $n \leq 3$ in the sense that if there is a winning agenda there is one consisting of a single issue. In sharp contrast, if $n = 4$ (or $n > 4$) then the problem of deciding if there is a winning agenda is NP-complete, and for every integer $p > 3$ there is an n by p matrix with $\{-1, 1\}$-entries and weights w_i, z_j so that there is a winning agenda and yet the minimum size of such an agenda is p.

References

[1] N. Alon, R. Bredereck, J. Chen, S. Kratsch, R. Niedermeier and G. J. Woeginger, How to (ab)use a multiple referendum for your hidden agenda, manusript, 2013.

[2] N. Alon and V. H. Vu, Anti-Hadamard matrices, coin weighing, threshold gates and indecomposable hypergraphs, *J. Combin. Theory Ser. A* **79** (1997), 133–160.

[3] J. E. Graver, A survey of the maximum depth problem for indecomposable exact covers, in: *Infinite and finite sets, Colloq. Math. Soc. János Bolyai* **10**, North Holland, Amsterdam (1973), pp. 731–743.

[4] R. L. Graham and N. J. A. Sloane, Anti-Hadamard matrices, *Linear Algebra Appl.* **62** (1984), 113–137.

[5] J. Håstad, On the size of weights for threshold gates, *SIAM J. Discrete Math.* **7** (1994), 484–492.

The largest triangular submatrix of a random matrix

Svante Janson

(joint work with Zur Izhakian and John Rhodes)

Consider a random random $n \times n$ matrix $X_n = (x_{ij})$, where the entries are i.i.d. and, for example, $\{0, 1\}$-valued with a fixed probability $\mathbb{P}(x_{ij} = 0) = p_0 > 0$.

Problem 1. What is the size, T_n, of the largest (lower) triangular submatrix of X_n?

A submatrix of order m is defined by selecting subsets of m rows and m columns. We may also allow reordering of the rows and columns, and consider the largest permuted triangular submatrix. This problem originally came from a question in supertropical algebra; in that setting the matrix elements can take three values, and we are really interested in triangular submatrices with 1's on the diagonal, see [3] and the references there for details. Asymptotically, to the first order treated here, the different versions have the same answer, and we consider, for simplicity, only the version stated above.

Theorem 2. Let $Q = 1/p_0 > 1$. Then, as $n \to \infty$,

$$T_n / \log_Q n \xrightarrow{\mathrm{p}} 2 + \sqrt{2},$$

where $\xrightarrow{\mathrm{p}}$ denotes convergence in probability.

Remark. The corresponding problem of the largest square submatrix with only 0's (or, equivalently, after interchange of 0 and 1, with only 1's) has been studied by several authors, see [6] and the references therein. It is shown in [6] that if S_n is the size of the largest such matrix, then $S_n/\log_Q n \xrightarrow{\mathrm{p}} 2$. This problem can be seen as finding the largest balanced complete subgraph of a random bipartite graph. The analogous problem of finding the largest complete set in a random graph $G(n,p)$ (or, equivalently, the largest independent set in $G(n,1-p)$) was solved by [2] and [5], see also [1] and [4]; again the size, C_n say, is asymptotically $2\log_Q n$, where $Q = 1/p$.

Note that $T_n \geq S_n \geq \lfloor T_n/2 \rfloor$, which shows that T_n and S_n are equal within a factor of $2 + o(1)$, and in particular of the same order of magnitude. However, it does not seem possible to get the right constant in front of $\log_Q n$ for one of these problems from the other.

A simple calculation shows that the expected number of triangular submatrices of size $c\log_Q n$ tends to infinity if $c < 4$; hence the first moment method is not useful here. The reason is that triangular submatrices of size $c\log_Q n$ with $2 + \log_Q 2 < c < 4$ are unlikely, but if they occur, they tend to occur in large groups.

The proof is therefore based on studying a truncated version of triangular submatrices, and then extend these to triangular matrices. See [3] for details.

Open problems

For the largest square zero submatrix and the largest cliques in $G(n,p)$, much more precise estimates are known, see [6] and [1, 4]; for example, it follows that if

$$s(n) = 2\log_Q n - 2\log_Q \log_Q n + 2\log_Q(e/2),$$

then for any $\epsilon > 0$, $\lfloor s(n) - \epsilon \rfloor \leq S_n \leq \lfloor s(n) + \epsilon \rfloor$ and $\lfloor s(n) + 1 - \epsilon \rfloor \leq C_n \leq \lfloor s(n) + 1 + \epsilon \rfloor$ w.h.p. (and, in fact, almost surely); in particular the sizes are concentrated on one or at most two values. It would be interesting to find similar sharper versions of the result above, which leads to the following open problems.

Problem 3. Find second order terms for T_n.

Problem 4. Is T_n concentrated on at most two values?

Problem 5. Prove a version of Theorem 2 (or a stronger result) with convergence almost surely instead of just in probability.

Problem 6. Find corresponding results when p_0 and p_1 depend on n.

Problem 7. Find corresponding results for rectangular matrices.

Acknowledgement. This work started during a chance meeting of researchers from two different groups at a supper table during a previous visit to Mathematisches Forschungsinstitut Oberwolfach, Germany, and the work was essentially completed there.

References

[1] B. Bollobás, *Random Graphs*, 2nd ed. Cambridge Univ. Press, Cambridge, 2001.
[2] B. Bollobás and P. Erdős, Cliques in random graphs. *Math. Proc. Camb. Phil. Soc.* **80** (1976), 419–427.
[3] Z. Izhakian, S. Janson and J. Rhodes, Superboolean rank and the size of the largest triangular submatrix of a random matrix. *Proc. Amer. Math. Soc.*, to appear.
[4] S. Janson, T. Łuczak & A. Ruciński, *Random Graphs*. Wiley, New York, 2000.
[5] D. Matula, The largest clique size in a random graph. Tech. Rep., Dept. Comp. Sci., Southern Methodist University, Dallas, Texas, 1976.
[6] X. Sun and A. B. Nobel, On the size and recovery of submatrices of ones in a random binary matrix. *J. Mach. Learn. Res.* **9** (2008), 2431–2453.

Recent developments in phase transitions and critical phenomena

Mihyun Kang

The phase transition is a phenomenon observed in mathematics and natural sciences in many different contexts. It deals with a sudden change in the properties of a large structure caused by altering a critical parameter. The phase transition in random discrete structures (e.g. random graphs, random satisfiability problems, Ising/Potts model, percolation) has captured the attention of many scientists in recent years.

The phase transition in random graphs was first discussed in 1959 by Erdős and Rényi in a series of papers. As Bollobás wrote in one of his books, *the most drastic example of a phase transition discovered by Erdős and Rényi concerns the order of the largest component of a random graph*. Since the seminal work of Erdős and Rényi [7], the phase transitions in random graphs have been extensively studied [1, 2, 3, 8, 9, 12, 13]. One of the reason why the phase transition has been among the main subjects of the random graph theory is due to its close relation to statistical physics and percolation theory [5] as well as computational problems in computer science such as random k-SAT [4].

Over the past few years, the so-called Achlioptas processes have gained increasing attention. The basic idea of Achlioptas processes is to use the power of two choices in order to create random graph processes with different behavior. An Achlioptas process starts with an empty graph on n vertices, and in each step two potential edges are chosen uniformly at random, and one of them is chosen according to a given rule and added to the graph. The key topics of recent results on Achlioptas processes include the acceleration or delaying of the phase transition; the avoidance of small subgraphs; and the acceleration of the appearance of Hamiltonian cycles.

In this talk we discuss new approaches to the study of the size and structure of components near the critical point of the phase transition in simple Achlioptas processes (e.g. bounded-size rules) [10]: key techniques are the classical ordinary differential equations method, a quasi-linear partial differential equation that tracks key statistics of the process, and singularity analysis. We also discuss a new proof of the phase transition in random hypergraphs [6], which builds on

the recent proof strategy of Krivelevich and Sudakov [11] who used the depth-first search algorithm in graphs to give a simple proof of the phase transition in the Erdős-Rényi graph.

References

[1] M. Behrisch, A. Coja-Oghlan, and M. Kang, The order of the giant component of random hypergraphs, *Random Structures Algorithms* **36** (2010), 149–184.
[2] B. Bollobás, The evolution of random graphs, *Trans. Am. Math. Soc.* **286** (1984), 257–274.
[3] B. Bollobás, *Random Graphs*, Cambridge University Press, 2001.
[4] B. Bollobás, C. Borgs, J. Chayes, J. H. Kim, and D. B. Wilson, The scaling window of the 2-SAT transition, *Random Structures Algorithms* **18** (2001), 201–256.
[5] B. Bollobás and O. Riordan, *Percolation*, Cambridge University Press, 2006.
[6] O. Cooley, M. Kang, and Y. Person, *Giant components in random hypergraphs*, manuscript.
[7] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **5** (1960), 17–61.
[8] S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs*, Wiley, New York, 2000.
[9] M. Kang and T. Łuczak, The two critical phase of a random planar graph, *Trans. of AMS* **364** (2012), 4239–4265.
[10] M. Kang, W. Perkins, and J. Spencer, The Bohman-Frieze process near criticality, *Random Structures Algorithms*, DOI: 10.1002/rsa.20437.
[11] M. Krivelevich and B. Sudakov, The phase transition in random graphs: A simple proof, *Random Structures Algorithms*, to appear.
[12] O. Riordan and L. Warnke, Explosive Percolation Is Continuous, *Science* **333** (2011), 322–324.
[13] T. Łuczak, Component behavior near the critical point of the random graph process, *Random Structures Algorithms* **1** (1990), 287–310.

Self-avoiding walks: many questions and a few answers

Hugo Duminil-Copin

The self-avoiding walk (SAW) was introduced independently by Flory and Ott as a model of polymers. It was then studied by mathematicians as a combinatorial model of pristine simplicity in its description, yet of extreme difficulty in its solution; by computer scientists interested in computational complexity; by biologists using it to model properties of DNA and other biological polymers of interest. This talk describes the geometric and combinatorial properties of SAWs.
Definition of the model: The model is defined by assigning equal probability to all paths on a lattice $\mathbb{L}$ (e.g. the hypercubic lattice $\mathbb{Z}^d$ or the hexagonal lattice $\mathbb{H}$) of length n starting from the origin and without self-intersections. More formally, a *walk* of length $n \in \mathbb{N}$ is a map $\gamma : \{0, \dots, n\} \to \mathbb{L}$ such that γ_i and γ_{i+1} are nearest neighbors for each $i \in \{0, \dots, n-1\}$. An injective walk is called *self-avoiding*. Let $\mathbb{P}_n$ denote the uniform law on SAWs of length n starting at the origin (we set $\mathbb{E}_n$ for the corresponding expectation). For future reference, we set c_n for the number of SAWs of length n.

We refer to [1, 5] for overviews on the mathematical and physical aspects of this subject.

Combinatorial aspects of the model: One of the first questions on SAW is to compute the entropy of the system, or equivalently the number of SAWs of length n. While computations for small values of n can be made by hand, they quickly become impossible to perform, due to the fact that c_n grows exponentially fast. Sub-multiplicativity of (c_n) implies that $(c_n^{1/n})$ converges to a constant μ_c called the *connective constant* of the lattice. This connective constant can be estimated for different lattices. For $\mathbb{Z}^2$, the rigorous bounds $\mu_c(\mathbb{Z}^2) \in [2.625\,622, 2.679\,193]$ are known, while for the hexagonal lattice, the following result was proved recently.

Theorem 1 (Duminil-Copin, Smirnov [4])**.** The connective constant of the hexagonal lattice equals $\sqrt{2+\sqrt{2}}$.

The connective constant describes the behavior with the roughest degree of precision. A much more precise behavior is predicted: on $\mathbb{Z}^d$, there should be a constant γ such that $c_n \sim An^{\gamma-1}\mu_c^n$, where $f(n) \sim g(n)$ means $\lim_{n\to\infty} f(n)/g(n) = 1$. The predicted values of the critical exponent γ are:

dimension	1	2	3	≥ 4
γ	1	$\frac{43}{32}$	$1.16\ldots$	1

These results are known for $d = 1$ and for $d \geq 5$ (for a slight modification of the model). For $d \geq 5$, the SAW can be related to the simple random walk, and the scaling limit is Brownian motion. In fact, for $d = 4$, the prediction involves a logarithmic correction and c_n is expected to be equivalent to $A(\log n)^{1/4}\mu_c^n$.

The important point is that the exponent γ has a probabilistic interpretation: γ measures how likely two SAWs are to avoid each other:

$$(\mathbb{P}_n \otimes \mathbb{P}_n)(\omega_1 \cap \omega_2 = \emptyset) = c_{2n}/c_n^2 \approx n^{1-\gamma}.$$

This interpretation suggests that γ depends on the geometric behavior of long SAWs, and that it should therefore be *universal*, meaning that it depends on lattices only through their dimension. This fact motivates the study of the geometric properties of the model, since they are related to the combinatorial ones.

Geometric properties. Originally, Flory was interested in the geometric properties of a random path γ with law $\mathbb{P}_n$. In particular, he focused on the typical distance between the endpoint γ_n and the origin by studying heuristically the following quantity $\mathbb{E}_n[\mathrm{Dist}(\gamma_n, 0)^2]$, called the *mean-square displacement* (Dist refers to the distance with respect to some norm, for instance the Euclidean one). Numerical computations and non-rigorous theory predict the precise behaviour of the mean-squared displacement of the walk's endpoint: $\mathbb{E}_n[\mathrm{Dist}(\gamma_n, 0)^2] = n^{2\nu+o(1)}$ where

dimension	1	2	3	≥ 4
ν	1	$\frac{3}{4}$	$0.59\ldots$	1/2

In dimensions five and above, the conjecture was proved for a version of SAW with weak repulsion in Brydges and Spencer using the celebrated lace-expansion technique. Hara and Slade proved the conjecture for standard SAW.

Until recently, virtually no rigorous non-trivial bounds on the mean-square displacement were known in low dimensions despite the many attempts to understand

properties of the model. In the last few months, two relevant theorems were proved simultaneously. The first one shows that the SAW is not ballistic, meaning that the end-point is not at distance cn from the origin, where $c > 0$ is a positive constant. Precisely, the result of [3] yields that $\mathbb{E}_n[\mathrm{Dist}(\gamma_n, 0)^2] = o(n^2)$ as n tends to infinity. The proof relies on renewal theory and combinatorial tricks based on unfoldings of parts of the walk. This theorem is the first non-trivial upper bound on the mean-square displacement. In the other direction, Madras proved that there exists $c > 0$ such that $\mathbb{E}_n[\mathrm{Dist}(\gamma_n, 0)^2] \geq cn^{4/(3d)}$ for any $n \geq 1$. Let us mention that this result is not straightforward. The fact that a SAW of length n visits n different vertices forces the diameter of the walk to be larger than $\varepsilon n^{1/d}$. Nevertheless, the average distance between the end-point to the origin has a priori no reason to be of the order of the diameter of the walk. In the same spirit, the fact that the probability that a walk ends near the origin was proved to converge to zero in [2]. Let us finish by an open problem.

Open Problem. Prove that in two and three dimensions, there exists $\varepsilon > 0$ such that

$$n^{2/d+\varepsilon} \leq \mathbb{E}_n[\mathrm{Dist}(\gamma_n, 0)^2] \leq n^{2-\varepsilon}$$

for all n large enough.

References

[1] R. Bauerschmidt, H. Duminil-Copin, J. Goodman, and G. Slade. Lectures on self-avoiding walks. In D. Ellwood, C. Newman, V. Sidoravicius, and W. Werner, editors, *Lecture notes, in Probability and Statistical Physics in Two and More Dimensions*. CMI/AMS – Clay Mathematics Institute Proceedings, 2011.

[2] H. Duminil-Copin, A. Glazman, A. Hammond, and A. Manolescu. The endpoint of self-avoiding walk delocalizes. *preprint*, page 30, 2013.

[3] H. Duminil-Copin and A. Hammond. Self-avoiding walk is sub-ballisticity. To appear in CMP, arXiv:1205.0401, 2013.

[4] H. Duminil-Copin and S. Smirnov. The connective constant of the honeycomb lattice equals $\sqrt{2+\sqrt{2}}$. *Annals of Math.*, 175(3):1653–1665, 2012.

[5] N. Madras and G. Slade. *The self-avoiding walk*. Probability and its Applications. Birkhäuser Boston Inc., Boston, MA, 1993.

Random graph coloring

Amin Coja-Oghlan

(joint work with Dan Vilenchik)

Let $G(n, m)$ be the random graph with n vertices and m edges. We make progress on determining the chromatic number $\chi(G(n, m))$, a problem posed by Erdős and Rényi [7].

This question has received considerable attention. Shamir and Spencer [16] proved concentration results. These were enhanced first by Łuczak [12] and then by Alon and Krivelevich [3], who proved that $\chi(G(n, m))$ is concentrated on two consecutive integers if $m \ll n^{3/2}$. In a celebrated paper, Bollobás [5] determined

the exact asymptotics of $\chi(G(n,m))$ in the dense case. This improved prior work by Matula [13]. Łuczak extended the result to sparse random graphs [11].

Achlioptas and Friedgut [1] proved that for any *fixed* $k \geq 3$ there exists a *sharp threshold sequence* $d_{k-\mathrm{col}} = d_{k-\mathrm{col}}(n)$. Furthermore, Achlioptas and Naor [2] proved via the second moment method that

$$d_{k-\mathrm{col}} \geq d_{k,\mathrm{AN}} = 2(k-1)\ln(k-1) = 2k\ln k - 2\ln k - 2 + o_k(1),$$

where the $o_k(1)$ term tends to zero for large k. A "first moment" calculation gives

$$d_{k-\mathrm{col}} \leq d_{k,\mathrm{first}} = 2k\ln k - \ln k.$$

Recently, the "cavity method", a non-rigorous method from statistical physics, has been applied to the problem of finding $\chi(G(n,m))$. This led to the conjecture [10, 14, 15, 17]

$$d_{k-\mathrm{col}} = 2k\ln k - \ln k - 1 + o_k(1).$$

Our main result is as follows.

Theorem 1. The k-colorability threshold satisfies $d_{k-\mathrm{col}} \geq d_{k,\mathrm{cond}} - o_k(1)$, with $d_{k,\mathrm{cond}} = 2k\ln k - \ln k - 2\ln 2$.

The *density* of a measurable set $A \subset \mathbf{R}_+$ is $\lim_{z\to\infty} \frac{1}{z}\int_0^z 1_A$.

Corollary 2. There exists a set $\mathcal{A} \subset \mathbf{R}_+$ of density 1 and a function $F : \mathcal{A} \to \mathbf{Z}_{\geq 0}$ such that for all average degrees $d \in \mathcal{A}$ we have $\chi(G(n,m)) = F(d)$ w.h.p.

Corollary 2 improves a result from [2], where the chromatic number was determined exactly on a set $\mathcal{A}'$ of density $\frac{1}{2}$.

The proof of Theorem 1 is based on a new approach to the second moment method for the random graph coloring problem. The improvement results from harnessing intuition from statistical mechanics on the geometry of the solution space [9, 17]. Indeed, the second moment argument from [2] boils down to an optimization problem over the *Birkhoff polytope*, i.e., the set of doubly-stochastic $k\times k$ matrices. Solving this problem turns out to be a formidable analytic task. However, the statistical mechanics predictions lead to additional constraints. These show that it actually suffices to optimize over a fairly small subset of the Birkhoff polytope. This problem can be solved via local variations arguments.

In addition, we show that the density $d_{k,\mathrm{cond}}$ corresponds in an exact sense to a phase transition. More precisely, following conventions from physics, by a *phase transition* we mean a d_0 where the function

$$\varphi(d) = \lim_{n\to\infty} \mathrm{E}[\sqrt[n]{\#k\text{-colorings of } G(n,m)}] \tag{1}$$

is non-analytic. Because the limit (1) is not currently known to exist for all d, we phrase the result as follows.

Corollary 3. There is $\epsilon_k = o_k(1)$ such that the following is true.

(1) The limit $\varphi(d)$ exists and is analytic for all $d < d_{k,\mathrm{cond}} - \epsilon_k$. Indeed, $\varphi(d) = k(1-1/k)^{d/2}$.

(2) Either $\varphi(d)$ does not exist for some $d \in (d_{k,\mathrm{cond}} - \epsilon_k, d_{k,\mathrm{cond}} + \epsilon_k)$ or $\varphi(d)$ is non-analytic at some point in this interval.

Bayati, Gamarnik and Tetali [4] proved the existence of a related limit, the "free energy". This inspires the following conjecture.

Conjecture 4. For any $k \geq 3$ and any $d > 0$ the limit (1) exists.

References

[1] D. Achlioptas, E. Friedgut: A sharp threshold for k-colorability. Random Structures Algorithms **14** (1999) 63–70.

[2] D. Achlioptas, A. Naor: The two possible values of the chromatic number of a random graph. Ann. of Math. (2) **162** (2005), 1333–1349.

[3] N. Alon, M. Krivelevich: The concentration of the chromatic number of random graphs. Combinatorica **17** (1997) 303–313

[4] M. Bayati, D. Gamarnik, P. Tetali: Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. Proc. 42nd STOC (2010) 105–114.

[5] B. Bollobás: The chromatic number of random graphs. Combinatorica **8** (1988) 49–55

[6] A. Coja-Oghlan, K. Panagiotou, A. Steger: On the chromatic number of random graphs. J. Combin. Theory Ser. B **98** (2008) 980–993.

[7] P. Erdős, A. Rényi: On the evolution of random graphs. Magayar Tud. Akad. Mat. Kutato Int. Kozl. **5** (1960) 17–61.

[8] G. Grimmett, C. McDiarmid: On colouring random graphs. Math. Proc. Cambridge Philos. Soc. **77** (1975) 313–324

[9] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborova: Gibbs states and the set of solutions of random constraint satisfaction problems. Proc. National Academy of Sciences **104** (2007) 10318–10323.

[10] F. Krzakala, A. Pagnani, M. Weigt: Threshold values, stability analysis and high-q asymptotics for the coloring problem on random graphs. Phys. Rev. E **70** (2004) 046705.

[11] T. Łuczak: The chromatic number of random graphs. Combinatorica **11** (1991) 45–54

[12] T. Łuczak: A note on the sharp concentration of the chromatic number of random graphs. Combinatorica **11** (1991) 295–297

[13] D. Matula: Expose-and-merge exploration and the chromatic number of a random graph. Combinatorica **7** (1987) 275–284.

[14] R. Mulet, A. Pagnani, M. Weigt, R. Zecchina: Coloring random graphs. Phys. Rev. Lett. **89** (2002) 268701

[15] J. van Mourik, D. Saad: Random Graph Coloring - a Statistical Physics Approach. Phys. Rev. E **66** (2002) 056120.

[16] E. Shamir, J. Spencer: Sharp concentration of the chromatic number of random graphs $G(n, p)$. Combinatorica **7** (1987) 121–129

[17] L. Zdeborová, F. Krzakala: Phase transition in the coloring of random graphs. Phys. Rev. E **76** (2007) 031131.

Erdős-Hajnal-type theorems in hypergraphs

David Conlon

(joint work with Jacob Fox, Benny Sudakov)

We call a graph H-free if it contains no induced copy of a given graph H. The famous conjecture of Erdős and Hajnal [2] states that if a graph on n vertices is H-free then it must contain either a clique or an independent set of size $n^{\delta(H)}$,

where $\delta(H) > 0$ depends only on the graph H. It is easy to see that this conjecture is true when H is a complete graph. It is also known to be true for all graphs with at most 4 vertices. However, the conjecture remains open even for the cycle C_5.

For general H, Erdős and Hajnal proved that if a graph on n vertices is H-free then it must contain a clique or an independent set of size $e^{c(H)\sqrt{\log n}}$. This is a significant improvement over the bound of $c \log n$ which, by Ramsey's theorem, holds in all graphs, but it is still quite far from the conjecture. However, as observed in [3], their method does allow one to find complete or empty bipartite subgraphs each side of which are of polynomial size. Recently, Fox and Sudakov [5] went one step further by proving that there is either a complete bipartite graph or an independent set of polynomial size.

For 3-uniform hypergraphs, Erdős and Rado [4] proved that, in any 2-colouring of the edges of the complete graph $K_n^{(3)}$ on n vertices, there is a monochromatic clique of size $c \log \log n$. Phrased differently, this says that any 3-uniform hypergraph on n vertices contains either a clique or an independent set of size $c \log \log n$. Given the situation for graphs, it is tempting to conjecture that if a 3-uniform hypergraph on n vertices is $\mathcal{H}$-free, for some given $\mathcal{H}$, then there should be a clique or an independent set of size much larger than $\log \log n$. We feel, but have been unable to prove, that for general $\mathcal{H}$ this may be too much to expect.

Given this state of affairs, we follow a different route, suggested by Rödl and Schacht [7], attempting to extend the bipartite counterpart of the Erdős-Hajnal theorem to tripartite 3-uniform hypergraphs. In any given 3-uniform hypergraph on n vertices, one may always find a complete or empty tripartite subgraph with parts of order at least $c(\log n)^{\frac{1}{2}}$. This follows from a standard extremal result due to Erdős [1]. We improve this result for $\mathcal{H}$-free graphs.

Theorem 1. Let $\mathcal{H}$ be a 3-uniform hypergraph. Then there exists a constant $\delta(\mathcal{H}) > 0$ such that, for n sufficiently large, any $\mathcal{H}$-free 3-uniform hypergraph on n vertices contains a complete or empty tripartite subgraph each part of which has order at least $(\log n)^{\frac{1}{2}+\delta(\mathcal{H})}$.

This improves upon a result of Rödl and Schacht [7]. They used the regularity method for hypergraphs to show that the size of the largest complete or empty tripartite subgraph grows faster than the function $(\log n)^{\frac{1}{2}}$ by a factor tending to infinity. However, because their result depends upon the regularity lemma, it does not provide good bounds on this factor.

Our result, on the other hand, is not far from best possible, since, for many $\mathcal{H}$, one cannot do better than $c \log n$. To see this, consider the random graph on the vertex set $\{1, 2, \ldots, n\}$ where each edge is chosen with probability $\frac{1}{2}$. For c sufficiently large, with high probability, this graph contains no complete or empty bipartite graph with parts of order $c \log n$. Fix such a graph and call it G_n. Let $\mathcal{G}_n$ be the 3-uniform hypergraph on the same vertex set whose edge set consists of all those triples (i_1, i_2, i_3) with $i_1 < i_2 < i_3$ such that (i_1, i_2) is an edge in G_n.

It is easy to see that $\mathcal{G}_n$ contains no complete or empty tripartite subgraph with sets of size $c \log n$. Suppose otherwise and let U, V and W be subsets of size

$c \log n$ which define such a tripartite graph. Without loss of generality, assume that the largest vertex w, in the ordering inherited from the integers, lies in W. Then, by construction, there must be a complete or empty bipartite subgraph in G_n between U and V, a contradiction. Now, for any subset X of the vertices of $\mathcal{G}_n$, let x_1 and x_2 be the two smallest vertices in X. Then, again by construction, for every $x \in X \backslash \{x_1, x_2\}$, either all edges of the form (x_1, x_2, x) are in $\mathcal{G}_n$ or none of them are. Choose a small hypergraph $\mathcal{H}$ containing no vertex pair (x_1, x_2) with this property. For example, one may take $\mathcal{H}$ to be a tight cycle on five vertices, that is, with vertices $\{1, 2, 3, 4, 5\}$ and edge set $\{123, 234, 345, 451, 512\}$. Then $\mathcal{G}_n$ is $\mathcal{H}$-free. Since it also contains no tripartite subgraph of size $c \log n$, this completes our claim.

The proof of Theorem 1 relies upon a new embedding lemma which says that if the edges of $\mathcal{G}$ are fairly well-distributed, in the sense that in any graph containing many triangles a positive proportion of these triangles form edges both of $\mathcal{G}$ and of its complement $\overline{\mathcal{G}}$, then one may embed an induced copy of any particular small hypergraph $\mathcal{H}$. If the hypergraph is not well-distributed in the sense described above, then it turns out that it must contain a complete or empty tripartite subgraph which is much larger than one would normally expect.

Despite this description being a reasonable one for any uniformity, the proof does not extend to the k-uniform case for any $k \geq 4$. Instead, we get more from considering the analogue of the usual Erdős-Hajnal problem in hypergraphs. That is, given an $\mathcal{H}$-free k-uniform hypergraph, how large of a clique or independent set must it contain?

Let $r_k(\ell)$ be the diagonal Ramsey function, that is, the minimum n such that in any 2-colouring of the edges of $K_n^{(k)}$ there is a monochromatic copy of $K_\ell^{(k)}$. The tower function $t_k(x)$ is defined by $t_1(x) = x$ and $t_{i+1}(x) = 2^{t_i(x)}$. A result of Erdős and Rado [4] states that $r_k(\ell) \leq 2^{\binom{r_{k-1}(\ell-1)}{k-1}} + k - 2$. For $k \geq 4$, this implies that

$$r_k(\ell) \leq t_{k-2}((r_3(c_k \ell))^3).$$

On the other hand, a result of Erdős and Hajnal (see [6]), referred to as the stepping-up lemma, allows one, for $k \geq 4$, to take a colouring of the $(k-1)$-uniform hypergraph on n vertices containing no monochromatic cliques of size ℓ and to show that there is a colouring of the k-uniform hypergraph on 2^n vertices containing no monochromatic clique of size $2\ell + k - 5$. In particular, this can be used to show that, for ℓ sufficiently large,

$$r_k(\ell) \geq t_{k-2}(r_3(c'_k \ell)).$$

Therefore, once the asymptotic behaviour of $r_3(\ell)$ is understood, so is that of $r_k(\ell)$. To be more precise, let r_k^{-1} be the inverse function of r_k. Restating the results quoted above and using the fact that $r_3(\ell)$ is at least exponential in ℓ tells us that, for $k \geq 3$ and n sufficiently large depending on k, $r_k^{-1}(n)$ has upper and lower bounds of the form

$$a'_k r_3^{-1}(\log_{(k-3)} n) \leq r_k^{-1}(n) \leq a_k r_3^{-1}(\log_{(k-3)} n),$$

where $\log_{(0)}(x) = x$ and, for $i \geq 1$, $\log_{(i)}(x) = \log(\log_{(i-1)}(x))$ is the iterated logarithm.

For an Erdős-Hajnal-type theorem to hold, we would therefore need that whenever a hypergraph on n vertices is $\mathcal{H}$-free there is a clique or independent set of size much larger than $r_3^{-1}(\log_{(k-3)} n)$. We will disprove this by showing that there are already simple examples of hypergraphs $\mathcal{H}$ which are not contained in step-up colourings. This implies the following theorem.

Theorem 2. For $k \geq 4$, there exists a constant c_k, a k-uniform hypergraph $\mathcal{H}$ and a sequence $\mathcal{G}_n$ of $\mathcal{H}$-free k-uniform hypergraphs with n vertices such that the size of the largest clique or independent set in $\mathcal{G}_n$ is at most $c_k r_k^{-1}(n)$.

References

[1] P. Erdős, On extremal problems of graphs and generalized graphs, *Israel J. Math.* **2** (1964), 183–190.

[2] P. Erdős and A. Hajnal, Ramsey-type theorems, *Discrete Appl. Math.* **25** (1989), 37–52.

[3] P. Erdős, A. Hajnal and J. Pach, A Ramsey-type theorem for bipartite graphs, *Geombinatorics* **10** (2000), 64–68.

[4] P. Erdős and R. Rado, Combinatorial theorems on classifications of subsets of a given set, *Proc. London Math. Soc.* **3** (1952), 417–439.

[5] J. Fox and B. Sudakov, Density theorems for bipartite graphs and related Ramsey-type results, *Combinatorica* **29** (2009), 153–196.

[6] R. L. Graham, B. L. Rothschild, and J. H. Spencer, *Ramsey theory*, 2nd edition, John Wiley & Sons, 1990.

[7] V. Rödl and M. Schacht, Complete partite subgraphs in dense hypergraphs, *Random Structures Algorithms* **41** (2012), 557–573.

Phase transitions and intractability

Mark Jerrum

(joint work with Leslie Ann Goldberg and Colin McQuillan)

"Phase transition" is a term that formally applies to infinite systems. But the effects of a phase transition can be felt in computations on finite problem instances. It is widely appreciated that phase transitions are a barrier to the effective application of certain algorithmic techniques, such as Markov chain Monte Carlo. But in fact one can sometimes exploit the existence of a phase transition to rule out an efficient approximation algorithm of any kind. Occasionally, the point at which a phase transition occurs can be rigorously linked to the exact boundary between tractability and intractability for a computational problem. We explore this phenomenon in the context of the hard-core (independent set) model on a graph, though most of the results generalise to other two-spin systems, such as the Ising model. In the hard core model, the configurations are independent sets in a graph, and each configuration σ is assigned weight $\lambda^{|\sigma|}$, where $\lambda > 0$ is parameter sometimes called "fugacity". The main quantity of interest is the partition function $\sum_\sigma \lambda^{|\sigma|}$.

In retrospect, the title should perhaps have been more specific, to indicate that we are focusing here specifically on counting problems, rather than decision problems (such as random satisfiability problems), which have a rather different flavour. Moreover, it is *approximate* solutions that we seek, since exact solutions are rarely available. A recurring theme in this area is that optimisation problems can be reduced to approximate counting problems. If the optimisation problem is intractable, then so is the counting problem. A simple example, reducing the problem of finding a maximum independent set to that of approximately counting the independent sets of all sizes was presented. (This is in the unweighted case $\lambda = 1$.) Although the reduction appears to require rapidly growing degree, Luby and Vigoda [5] observed that since the optimisation problem is hard to solve within some fixed constant factor, in fact bounded degree graphs suffice. So, the problem of approximating the number of independent sets in a bounded degree graph is NP-hard.

The degree bound that arises from Luby and Vigoda's observation is large. Dyer, Frieze and Jerrum [1] exploited the phase transition for the independent sets model in random regular bipartite graph to provide a bistable gadget, which they then used to encode the Boolean variables in a certain NP-hard optimisation problem. In this way, they were able to reduce the degree bound substantially, to 25. Recently Sly [7], Sly and Sun [8] and Galanis, Ge, Štefankovič, Vigoda, and Yang [2] have carried out a much more delicate analysis of the gadget, thereby obtaining a sharp result. They show there there is a critical fugacity $\lambda_c(d) = (d-1)^{d-1}/(d-2)^d$ such that the independent sets in a random bipartite graph of degree d are unbalanced when $\lambda > \lambda_c$ and balanced when $\lambda < \lambda_c$. (This critical value λ_c is characterised as the "uniqueness threshold" of an infinite d-regular tree.) Thus, the NP-hardness proof works right down to (but not including) λ_c. In the other direction, Weitz [9], Sinclair, Srivastava and Thurley [6] and Li, Lu, and Yin [4] provide an polynomial time algorithm that works for all $\lambda < \lambda_c$. So we are in the remarkable position of knowing the exact threshold for computational tractability for the hard core or independent set model and, even more remarkably, this transition for *general* d-regular graphs is identical to the "uniqueness threshold" for infinite d-regular *trees*. Most of the results quoted above apply to more or less general two-spin systems.

Then on to the specific result of the talk, which is joint work with Leslie Goldberg and Colin McQuillan [3]. We considered the complexity of of the hard core model on a planar graph and showed that the partition function is hard to approximate for sufficiently large λ. A particular obstacle in this case is that there is an efficient approximation algorithm for approximating the *logarithm* of the partition function in a planar graph. This fact suggests that is is unlikely that we can construct a reduction from an optimisation problem that is APX-complete (i.e., one that is difficult even to approximate within constant factor error). So the analysis of the bistable gadgets, which are cylindrical $L \times L$ fragments of $\mathbb{Z}^2$, becomes rather delicate. Basically one needs to show that, conditioned on whether the phase is odd or even, correlations decay quickly with distance.

References

[1] Martin Dyer, Alan Frieze, and Mark Jerrum. On counting independent sets in sparse graphs. *SIAM J. Comput.*, 31(5):1527–1541 (electronic), 2002.

[2] Andreas Galanis, Qi Ge, Daniel Štefankovič, Eric Vigoda, and Linji Yang. Improved inapproximability results for counting independent sets in the hard-core model. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 6845 of *Lecture Notes in Computer Science*, pages 567–578. Springer Berlin / Heidelberg, 2011.

[3] Leslie Ann Goldberg, Mark Jerrum and Colin McQuillan. Approximating the partition function of planar two-state spin systems. arXiv:1208.4987, 2012.

[4] Liang Li, Pinyan Lu, and Yitong Yin. Correlation decay up to uniqueness in spin systems. arXiv:1111.7064, 2011.

[5] Michael Luby and Eric Vigoda. Approximately counting up to four. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 682–687. ACM Press, 1997.

[6] Alistair Sinclair, Piyush Srivastava, and Marc Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, pages 941–953. SIAM, 2012.

[7] Allan Sly. Computational transition at the uniqueness threshold. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS '10)*, pages 287–296, 2010.

[8] A. Sly and N. Sun. The computational hardness of counting in two-spin models on d-regular graphs. arXiv:1203.2602, 2012.

[9] Dror Weitz. Counting independent sets up to the tree threshold. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, STOC '06, pages 140–149, New York, NY, USA, 2006. ACM.

Nimble Algorithms for Cloud Computing

Ravindran Kannan
(joint work with Santosh Vempala)

Cloud computing is a new paradigm where data is stored across multiple servers and the goal is to compute a function of all the data. We consider a simple model where each server uses polynomial time and space, but communication among servers, being more expensive, is ideally bounded by a polylogarithmic function of the input size. We will dub algorithms that satisfy these types of resource bounds as *nimble*.

The main contribution of the paper is to develop nimble algorithms for several areas which involve massive data and for that reason have been extensively studied in the context of Streaming Algorithms. The areas are approximation of Frequency Moments, Counting bipartite homomorphisms (number of copies of a fixed bipartite graph H in a graph G), Rank-k approximation to a matrix, and Clustering. For frequency moments, we will use a new importance sampling technique based on high powers of the frequencies. We reduce the problem of counting homomorphisms to estimating implicitly defined frequency moments. For rank-k

approximations, besides recent results of several authors developed in the Streaming context, we use a new variant of the random projection method. For clustering, we use our rank-k approximation and the small *coreset* of Chen [1] of size at most polynomial in the dimension.

In contrast to our algorithms in the cloud computing model, in the streaming model, known lower bound results for frequency moments and rank-k approximations rule out the existence of algorithms that use polylogarithmic space.

References

[1] K. Chen, *On Coresets for k-Median and k-Means Clustering in Metric and Euclidean Spaces and Their Applications*, SIAM J. Comput. **39** (2009), 923–947.

The evolution of subcritical Achlioptas processes

Lutz Warnke
(joint work with Oliver Riordan)

In the *Erdős–Rényi random graph process*, starting from an empty graph, in each step a new random edge is added to the evolving graph. One of its most interesting features, both mathematically and in terms of applications, is the 'percolation phase transition': as the ratio of the number of edges to vertices increases past a certain critical point, the global structure changes radically, from only small components to a single macroscopic ('giant') component plus small ones.

We consider *Achlioptas processes*, which were introduced by Dimitris Achlioptas around 2000 in an attempt to create random graph processes with potentially different behaviour than the Erdős–Rényi one. Starting from an empty graph these proceed as follows: in each step *two* potential edges e_1 and e_2 are chosen uniformly at random, and using some rule *one* of them is selected and added to the evolving graph. During the last decade, so-called *bounded-size* rules have received considerable attention, see, e.g., [2, 3, 9]. These make their decisions based only on the sizes of the components containing the endvertices of e_1 and e_2, with the restriction that all sizes larger than some constant B are treated in the same way. It turns out that, while bounded size rules can, e.g., delay or accelerate the percolation phase transition, these still share many qualitative similarities with the 'classical' random graph process.

In contrast, few rigorous results are known for the more involved 'unbounded'-size rules, whose choices depend only on the sizes of the four components containing the endvertices of the two offered edges. To illustrate our very limited understanding of these, we mention that in one line of research, stimulated by a conjecture of Achlioptas, D'Souza and Spencer published in *Science* [1] (supported by extensive simulations), it was believed that certain size rules (in particular the so-called *product rule*) could give rise to a phase transition that is particularly radical: more or less as soon as the macroscopic component appears, it is already extremely large; this phenomenon is known as 'explosive percolation'. However, recently it was shown by Riordan and Warnke [4, 5] that this is not the case, by

proving that the phase transition is in fact continuous for all Achlioptas processes. To summarize, the surprises that (unbounded) size rules have shown so far indicate that our intuition for these processes still needs to be developed.

We establish the first rigorous convergence result for Achlioptas processes using unbounded size rules such as the product rule. Intuitively speaking, we prove that certain key statistics are tightly concentrated at least until the susceptibility (the expected size of the component containing a randomly chosen vertex) diverges. To give a more precise statement we need to introduce some notation. Every rule $\mathcal{R}$ defines, for each n, a random sequence $(G_i)_{i\geq 0} = (G_i^{\mathcal{R}})_{i\geq 0}$ of graphs with vertex set $[n]$, where G_i denotes the graph after i steps. Let $N_k(i)$ denote the number of vertices of G_i in components of size k, and define $N_{\geq k}(i)$ similarly. Now we define the susceptibility of G_i as $S(i) = \sum kN_k(i)/n$. As usual, $S(tn) \xrightarrow{\mathrm{p}} \infty$ as $n \to \infty$ means that for any $C > 0$ we have $\Pr(S(tn) \leq C) \to 0$ as $n \to \infty$. With these definitions in hand, we are ready to state (a simplified version) of the main result of [8], which shows that the number of vertices in components of size $k \geq 1$ (and the susceptibility) is tightly concentrated until the susceptibility 'blows up', which happens at a critical time t_{b}.

Theorem 1. Let $\mathcal{R}$ be a size rule. There exist $t_{\mathrm{b}} = t_{\mathrm{b}}^{\mathcal{R}} \in (0, 1]$ and functions $(\rho_k)_{k\geq 1}$ with $\rho_k = \rho_k^{\mathcal{R}} : [0, t_{\mathrm{b}}) \to [0, 1]$ such that the following holds. For every $t \geq t_{\mathrm{b}}$ we have

$$S(tn) \xrightarrow{\mathrm{p}} \infty \tag{1}$$

as $n \to \infty$. For every $t < t_{\mathrm{b}}$ we have $\sum_{k\geq 1} \rho_k(t) = 1$. Also, for every $t < t_{\mathrm{b}}$ there exist $a, A, C > 0$ (depending only on $\mathcal{R}, t$) such that for every $t' \in [0, t]$ we have $\rho_k(t') \leq Ae^{-ak}$ for all $k \geq 1$. In addition, for $n \geq n_0(\mathcal{R}, t)$ the following holds with probability at least $1 - n^{-99}$: for every $0 \leq i \leq tn$ we have

$$|N_k(i) - \rho_k(i/n)n| \leq (\log n)^C n^{1/2} \quad \text{for all } k \geq 1, \tag{2}$$

$$|S(i) - \textstyle\sum_{k\geq 1} k\rho_k(i/n)| \leq (\log n)^C n^{-1/2}, \tag{3}$$

and $N_{\geq k}(i) \leq Ae^{-ak}n$ for all $k \geq 1$.

To interpret this result, we think of the functions $\rho_k(t)$ as describing the 'scaling limit' of the component size distribution at 'time' $t < t_{\mathrm{b}}$, where time is the number of steps divided by n. A key aspect of Theorem 1 is that this limit does not depend on n; in fact, most of our technical work is devoted to establishing this property (to only show concentration around its expectation simpler arguments would suffice). Furthermore, the bound on ρ_k demonstrates that the idealized component size distribution has an exponential tail for $t < t_{\mathrm{b}}$, as one would expect in a strictly subcritical random graph.

The proof of Theorem 1 is based on a variant of the neighbourhood exploration process and relies on branching process (approximation) arguments. This is quite different from other approaches in this area, which are based on differential equations, see, e.g., [7]. Finally, we believe that convergence up to t_{b} is best possible

(see [6]), and also conjecture that t_{b} coincides with the percolation threshold; we refer to [8] for the technical details.

References

[1] D. Achlioptas, R.M. D'Souza, and J. Spencer. Explosive percolation in random networks. *Science* **323** (2009), 1453–1455.

[2] T. Bohman and D. Kravitz. Creating a giant component. *Combinatorics, Probability and Computing* **15** (2006), 489–511.

[3] S. Janson and J. Spencer. Phase transitions for modified Erdős-Rényi processes. *Arkiv för Matematik* **60** (2012), 305–329.

[4] O. Riordan and L. Warnke. Explosive percolation is continuous. *Science* **333** (2011), 322–324.

[5] O. Riordan and L. Warnke. Achlioptas process phase transitions are continuous. *Annals of Applied Probability* **22** (2012), 1450–1464.

[6] O. Riordan and L. Warnke. Achlioptas processes are not always self-averaging. *Physical Review E* **86** (2012), 011129.

[7] O. Riordan and L. Warnke. Convergence of Achlioptas processes via differential equations with unique solutions. *Preprint* (2011). `arXiv:1111.6179`.

[8] O. Riordan and L. Warnke. The evolution of subcritical Achlioptas processes. *Preprint* (2012). `arXiv:1204.5068`.

[9] J. Spencer and N.C. Wormald. Birth control for giants. *Combinatorica* **27** (2007), 587–628.

On the communication complexity of sparse set disjointness and exists-equal problems

Gábor Tardos

(joint work with Mert Sağlam)

The communication complexity of a single equality problem is well known: to achieve ϵ error in the randomized model with a joint random source it is enough to send $\log(1/\epsilon)$ bits in a single round independent of the size of the universe. Here we consider the "exists-equal" problem that is the OR of n instances of the equality problem. We find matching lower and upper bounds for its complexity for any number of rounds. The surprising thing is that to achieve small constant error in a limited number of rounds one needs strictly more communication than n times the communication needed for a single equality problem. In fact, for a single round protocol we prove a lower bound of $\Omega(n \log n)$, meaning that the best that one can do for a constant-error protocol for exists-equal is to combine low-error protocols for the individual equality problems. To our knowledge, this is the first instance where such a super-linear increase in complexity is demonstrated for a single bit combination of many instances of any communication problem.

Observe that exists-equal can be considered a special case of the sparse set disjointness with the players receiving n element subsets of a larger universe and their goal being deciding whether their sets are disjoint or not. Our upper bound protocol works in this more general case. We give a protocol that communicates a total of $O(n \log^{(r)} n)$ bits over r rounds and errs only with probability well below $1/k$. Here we can set the number of rounds $r \leq \log^* n$ arbitrarily and for $r = \log^* n$

we obtain an $O(n)$ total communication $\log^* n$-round protocol, improving on the $O(k)$-bits $\log k$-round protocol of Håstad and Wigderson [1]. In this case the error bound of our protocol is exponentially small, $O(2^{-\sqrt{k}})$, greatly improving on the constant error of the Håstad-Wigderson protocol.

Our main contribution in this paper is a matching lower bound: we show that any r-round randomized protocol for the exists-equal problem with a small constant error probability should have a message of size $\Omega(n \log^{(r)} n)$. Our lower bound holds even for super-constant $r \leq \log^* n$, showing that any $O(n)$ bits exists-equal protocol should have $\log^* n - O(1)$ rounds.

Note that the protocol we give errs only with probability that is less than polynomially small and provides guarantees on the total communication for the harder set disjointness problem, whereas our lower bound holds even for a small constant error probability protocols and for the easier exists-equal problem with guarantees on the max-communication. Hence our upper and lower bounds match in a strong sense.

Details can be found in [2].

References

[1] J. Håstad, A. Wigderson, The randomized communication complexity of set disjointness, *Theory of Computing* **3** (2007), 211–219.

[2] M. Sağlam, G. Tardos, On the communication complexity of sparse set disjointness and exists-equal problems, arXiv:1304.1217.

Zero-one laws for minor-closed classes of graphs

Marc Noy

(joint work with Peter Heinig, Tobias Müller, and Anusch Taraz)

Let $\mathcal{G}$ be a class of labelled graphs closed under isomorphism, and let $\mathcal{G}_n$ be the graphs in $\mathcal{G}$ with n vertices. We say that $\mathcal{G}$ has a zero-one law with respect to some logic language L if, for every sentence ϕ in L, the probability that a graph in $\mathcal{G}_n$ satisfies ϕ tends either to 0 or 1, as n goes to infinity. The classical example of a zero-one law is for the class of all graphs in the language of first order (FO) logic, a result due to Glebski et al. and independently to Fagin. The problem has been investigated more generally in the $G(n,p)$ model, where edges are drawn independently with probability p. A zero-one law holds for all constant p and in many other cases. In particular, the celebrated theorem of Shelah and Spencer says that a zero-one law holds if $p = n^{-\alpha}$ and $\alpha > 0$ is an irrational number [7].

In this paper we are interested in different models of random graphs. Consider the class $\mathcal{T}$ of labelled trees with the uniform distribution on trees with n vertices. McColm [4] showed that a zero-one law holds in $\mathcal{T}$ in the monadic second order (MSO) logic, which is FO logic enriched with quantification over sets of vertices. His proof is based on two facts: 1) for each $r > 0$ there exists a rooted tree T_r such that, if two trees A and B both have T_r as a rooted subtree, then Duplicator has a winning strategy in the Ehrenfeucht-Fraissé game with r rounds played on

A and B; and 2) with high probability, a random tree contains T_r as a rooted subtree. In order to cover MSO logic, the Ehrenfeucht-Fraissé game has to be enriched with set moves in addition to vertex moves; Duplicator wins if there is a partial isomorphism between the selected vertices which respects membership in the selected sets.

We show that this approach works much more generally for minor-closed classes that are addable, that is, all the forbidden minors are 2-connected. This includes the class of planar graphs and many others, such as outerplanar, series-parallel, graphs with bounded tree-width, and graphs with given 3-connected components [3]. In all these cases we are able to prove a zero-one law in MSO logic for *connected* graphs under the uniform distribution, using the following result of McDiarmid [6]: if $\mathcal{G}$ is a proper addable minor-closed class and H is a fixed graph in $\mathcal{G}$, then a random graph from $\mathcal{G}$ contains a pendant copy of H, that is, a copy joined to the rest of the graph by a single edge. We remark that if the class is not addable then a zero-one law may not exist (we show this is the case, for instance, for the class of caterpillars).

For arbitrary graphs in an addable minor-closed classes of graphs there cannot be a zero-one law, even in FO logic. The reason is that there are sentences expressible in FO, such as the existence of an isolated vertex, that have a limiting probability strictly between 0 and 1. This is true more generally for the existence of a connected component isomorphic to a given graph in the class. In this situation we prove a *convergence* law, by showing that each sentence in MSO has a limiting probability, possibly different from 0 and 1. The proof is based on the fact that with high probability the largest component has size $n - O(1)$, that the number of components isomorphic to a fixed graph in the class follows asymptotically a Poisson distribution, and that the numbers of components isomorphic to distinct graphs are asymptotically independent.

Let S be a fixed surface and consider the class $\mathcal{G}_S$ of graphs that can be embedded in S. The class $\mathcal{G}_S$ is minor-closed but not addable. In this case we prove a zero-one law in FO for connected graphs in $\mathcal{G}$. The proof is based on adapting an argument from Bender et al. [1] for maps on a surface and on recent results on random graphs of fixed genus [2, 5]. For arbitrary graphs we prove a convergence law in FO. Moreover, we show that the limiting probability of a FO sentence does not depend on the surface and is the same as for planar graphs. In particular, the almost sure theory for connected graphs is independent of the surface. We conjecture that both results, for connected and arbitrary graphs in $\mathcal{G}_S$, extend to the MSO logic. If this is true then the limiting probabilities cannot be independent of the surface, since containing a fixed minor is expressible in MSO logic.

Here is a summary of our main results:

- If $\mathcal{G}$ is a proper addable minor-closed class of graphs, then a *zero-one* law holds in MSO for *connected* graphs in $\mathcal{G}$.
- If $\mathcal{G}$ is a proper addable minor-closed class of graphs, then a *convergence* law holds in MSO for *arbitrary* graphs in $\mathcal{G}$.

- If $\mathcal{G}_S$ is the class of graphs that can be embedded in a fixed surface S, a *zero-one* law holds in FO for *connected* graphs in $\mathcal{G}_S$.
- If $\mathcal{G}_S$ is the class of graphs that can be embedded in a fixed surface S, a *convergence* law holds in FO for *arbitrary* graphs in $\mathcal{G}_S$. Moreover, the limiting probability that a given FO sentence is satisfied is independent of the surface S.

References

[1] E. A. Bender, K. J. Compton, L. B. Richmond. 0-1 laws for maps. *Random Structures Algorithms* **14** (1999), no. 3, 215–237.

[2] G. Chapuy, E. Fusy, O. Giménez, B. Mohar, M. Noy. Asymptotic enumeration and limit laws for graphs of fixed genus. *J. Combin. Theory Ser. A* **118** (2011), 748–777.

[3] O. Giménez, M. Noy, J. Rué. Graph classes with given 3-connected components: asymptotic enumeration and random graphs. *Random Structures Algorithms* (to appear). arXiv:0907.0376

[4] G. L. McColm. MSO zero-one laws on random labelled acyclic graphs. *Discrete Math.* **254** (2002), no. 1-3, 331–347.

[5] C. McDiarmid. Random graphs on surfaces. *J. Combin. Theory Ser. B* **98** (2008), 778–797.

[6] C. McDiarmid. Random graphs from a minor-closed class. *Combin. Probab. Comput.* **18** (2009), 583–599.

[7] J. Spencer. *The Strange Logic of Random Graphs*. Springer (2001).

Independent sets in hypergraphs

József Balogh

(joint work with Robert Morris and Wojciech Samotij)

1. Introduction

Many important theorems and conjectures in combinatorics, such as the theorem of Szemerédi on arithmetic progressions and the Erdős-Stone Theorem in extremal graph theory, can be phrased as statements about families of independent sets in certain uniform hypergraphs. In recent years, an important trend in the area has been to extend such classical results to the so-called 'sparse random setting'. This line of research has recently culminated in the breakthroughs of Conlon and Gowers [2] and of Schacht [8], who developed general tools for solving problems of this type. Although these two papers solved very similar sets of longstanding open problems, the methods used are very different from one another and have different strengths and weaknesses.

We provide a third, completely different approach to proving extremal and structural results in sparse random sets that also yields their natural 'counting' counterparts. We give a structural characterization of the independent sets in a large class of uniform hypergraphs by showing that every independent set is almost contained in one of a small number of relatively sparse sets. We then derive many interesting results as fairly straightforward consequences of this abstract theorem. In particular, we prove the well-known conjecture of Kohayakawa, Łuczak, and

Rödl [4], a probabilistic embedding lemma for sparse graphs, for all 2-balanced graphs. We also give alternative proofs of many of the results of Conlon and Gowers [2] and Schacht [8], such as sparse random versions of Szemerédi's theorem, the Erdős-Stone Theorem and the Erdős-Simonovits Stability Theorem, and obtain their natural 'counting' versions, which in some cases are considerably stronger. We also obtain new results, such as a sparse version of the Erdős-Frankl-Rödl Theorem [3] on the number of H-free graphs and, as a consequence of the KŁR conjecture, we extend a result of Rödl and Ruciński [6] on Ramsey properties in sparse random graphs to the general, non-symmetric setting. Similar results have been discovered independently by Saxton and Thomason [7]. A full version of this extended abstract with a complete reference list is [1].

2. The main result

Our main result gives a structural characterization of the collection of all independent sets in a large class of uniform hypergraphs.

Definition 1. Let $\mathcal{H}$ be a uniform hypergraph with vertex set V, let $\mathcal{F}$ be an increasing family of subsets of V and let $\varepsilon \in (0,1]$. We say that $\mathcal{H}$ is $(\mathcal{F},\varepsilon)$-*dense* if $e(\mathcal{H}[A]) \geq \varepsilon e(\mathcal{H})$ for every $A \in \mathcal{F}$.

Let $\mathcal{F}_\varepsilon = \big\{A \subseteq V(\mathcal{H}) : e(\mathcal{H}[A]) \geq \varepsilon e(\mathcal{H})\big\}$. Given a hypergraph $\mathcal{H}$, for each $T \subseteq V(\mathcal{H})$, we define $\deg_{\mathcal{H}}(T) = |\{e \in \mathcal{H} : T \subseteq e\}|$, and let $\Delta_\ell(\mathcal{H}) = \max\big\{\deg_{\mathcal{H}}(T) : T \subseteq V(\mathcal{H}) \text{ and } |T| = \ell\big\}$.

Theorem 1. For every $k \in \mathbb{N}$ and all positive c, c' and ε, there exists a positive constant C such that the following holds. Let $\mathcal{H}$ be a k-uniform hypergraph and let $\mathcal{F} \subseteq \mathcal{P}(V(\mathcal{H}))$ be an increasing family of sets such that $|A| \geq \varepsilon v(\mathcal{H})$ for all $A \in \mathcal{F}$. Suppose that $\mathcal{H}$ is $(\mathcal{F},\varepsilon)$-dense and $p \in (0,1)$ is such that $p^{k-1}e(\mathcal{H}) \geq c'v(\mathcal{H})$ and for every $\ell \in [k-1]$,

$$\Delta_\ell(\mathcal{H}) \leq c \cdot \min\left\{p^{\ell-k}, p^{\ell-1}\frac{e(\mathcal{H})}{v(\mathcal{H})}\right\}.$$

Then there exists a family $\mathcal{S} \subseteq \binom{V(\mathcal{H})}{\leq Cp\cdot v(\mathcal{H})}$ and functions $f : \mathcal{S} \to \overline{\mathcal{F}}$ and $g : \mathcal{I}(\mathcal{H}) \to \mathcal{S}$ such that for every $I \in \mathcal{I}(\mathcal{H})$, $g(I) \subseteq I$ and $I \setminus g(I) \subseteq f(g(I))$.

3. Some of the applications

Szemerédi type of results. The celebrated theorem of Szemerédi [9] says that for every $k \in \mathbb{N}$, the largest subset of $\{1,\dots,n\}$ that contains no k-term arithmetic progression (AP) has $o(n)$ elements. It immediately follows that there are only $2^{o(n)}$ subsets of $\{1,\dots,n\}$ with no k-term AP. Our first result can be viewed as a sparse analogue of this statement.

Theorem 2. For every positive β and every $k \in \mathbb{N}$, there exist constants C and n_0 such that the following holds: For every $n \in \mathbb{N}$ with $n \geq n_0$, if $m \geq Cn^{1-1/(k-1)}$, then there are at most $\binom{\beta n}{m}$ m-subsets of $\{1,\dots,n\}$ that contain no k-term AP.

The sparse random analogue of Szemerédi's theorem, proved by Schacht [8] and independently by Conlon and Gowers [2], follows as an easy corollary, using the first moment method, of Theorem 2. We say that a set $A \subseteq \mathbb{N}$ is (δ, k)-*Szemerédi* if every subset $B \subseteq A$ with at least $\delta|A|$ elements contains a k-term AP. Let $[n] = \{1, \dots, n\}$ and denote $[n]_p$ the p-random subset of $[n]$.

Corollary 3. For every $\delta \in (0,1)$ and every $k \in \mathbb{N}$, there exists a constant C such that for all sufficiently large n, if $p_n \geq Cn^{-1/(k-1)}$, then

$$\lim_{n\to\infty} \Pr\left([n]_{p_n} \text{ is } (\delta,k)\text{-Szemerédi}\right) = 1.$$

We remark that Theorem 2 and Corollary 3 are both sharp up to the value of the constant C. When we prove Theorem 2, we apply Theorem 1 to the following hypergraph: Let $\mathcal{H}$ be the k-uniform hypergraph of k-term APs in $[n]$, i.e., the hypergraph on the vertex set $[n]$ whose edges are all k-term APs in $[n]$, let $\mathcal{F}$ denote the family of subsets of $[n]$ with at least δn elements, and let $\varepsilon = \varepsilon(\delta, k)$.

The typical structure of H-free graphs. Let H be an arbitrary non-bipartite graph. We say that a graph G is *H-free* if G does not contain H as a subgraph. For an integer n, denote by $f_n(H)$ the number of labeled H-free graphs on the vertex set $[n]$. Since every subgraph of an H-free graph is also H-free, it follows that $f_n(H) \geq 2^{\mathrm{ex}(n,H)}$. Erdős, Frankl, and Rödl [3] proved that this crude lower bound is in a sense tight, namely that

$$f_n(H) = 2^{\mathrm{ex}(n,H)+o(n^2)}. \tag{1}$$

Our next result can be viewed as a 'sparse version' of (1). Such a statement was already considered by Łuczak, who derived it from the KŁR conjecture. For integers n and m with $0 \leq m \leq \binom{n}{2}$, let $f_{n,m}(H)$ be the number of labeled H-free graphs on the vertex set $[n]$ that have exactly m edges. The following theorem refines (1) to n-vertex graphs with m edges.

Theorem 4. Let G be a 2-balanced graph and let δ be a positive constant. There exists a constant C such that for every $n \in \mathbb{N}$, if $m \geq Cn^{2-1/m_2(G)}$, then

$$\binom{\mathrm{ex}(n,G)}{m} \leq f_{n,m}(G) \leq \binom{\mathrm{ex}(n,G)+\delta n^2}{m}.$$

When we prove Theorem 4, we apply Theorem 1 to the following hypergraph: The *hypergraph of copies of G in K_n^t* is the $e(G)$-uniform hypergraph on the vertex set $E(K_n)$ whose edges are the edge sets of all copies of G in K_n.

Note that the KŁR conjecture is also proved by Theorem 1; we skip details here.

Research supported in part by: (JB) NSF CAREER Grant DMS-0745185, UIUC Campus Research Board Grant 11067, and OTKA Grant K76099; (RM) CNPq bolsa de Produtividade em Pesquisa; (WS) ERC Advanced Grant DMMCA and a Trinity College JRF

References

[1] J. Balogh, R. Morris and W. Samotij, *Independent sets in hypergraphs*, submitted.
[2] D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, submitted.
[3] P. Erdős, P. Frankl, and V. Rödl, *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs Combin., **2** (1986), 113–121.
[4] Y. Kohayakawa, T. Luczak, and V. Rödl, *On K_4-free subgraphs of random graphs,* Combinatorica, **17** (1997), 173–213.
[5] T. Łuczak, *On triangle-free random graphs*, Random Structures Algorithms, **16** (2000), 260–276.
[6] V. Rödl and A. Ruciński, *Threshold functions for Ramsey properties*, J. Amer. Math. Soc., **8** (1995), 917–942.
[7] D. Saxton and A. Thomason, Hypergraph containers, submitted.
[8] M. Schacht, *Extremal results for random discrete structures*, submitted.
[9] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.

Small Complete Minors Above the Extremal Edge Density

Asaf Shapira

(joint work with Benny Sudakov)

A graph G has a K_t-minor if G contains t vertex disjoint connected subgraphs $S_1, \ldots, S_t$ and $\binom{t}{2}$ paths $(P_{i,j})_{1 \le i < j \le t}$ such that $P_{i,j}$ connects S_i to S_j, each path $P_{i,j}$ is disjoint from each set S_k with $k \neq i, j$, and the paths $P_{i,j}$ are internally vertex disjoint, that is, $P_{i,j}$ can only intersect with $P_{i,j'}$ or $P_{i',j}$ at its endpoint vertices. This K_t-minor is called topological if all $|S_i| = 1$. The notion of minor is undoubtedly one of the most well studied topics in graph theory. A central result in this area states that a linear number of edges is enough to force the appearance of a K_t-minor. Formally, for every integer $t \ge 3$ define

$$c(t) = \min\{c : d(G) \ge c \text{ implies that } G \text{ has a } K_t\text{-minor}\}\ , \tag{1}$$

where $d(G) = |E(G)|/|V(G)|$. Kostochka [7] and Thomason [12] proved that $c(t) = \Theta(t\sqrt{\log t})$.

Fiorini, Joret, Theis and Wood [3] raised the following problem; how many edges guarantee that a graph contains not only a K_t-minor, but one which has *few vertices*. Observe that graphs with logarithmic girth (which can be constructed by deleting short cycles from a random graph $G(n,p)$ with $p = c/n$, or explicitly see, e.g., [9]) show that for any constant C, there is a graph G with $d(G) \ge C$ and no K_3-minor of order $o(\log n)$. We also need $d(G) \ge c(t) = (\alpha + o(1))t\sqrt{\log t}$ to guarantee *some* K_t-minor. So the question boils down to finding the smallest constant $c > c(t)$ so that any graph G with $d(G) \ge c$ contains a K_t-minor of order $O(\log n)$. Fiorini et al. [3] proved that if $d(G) \ge 2^{t-2} + \epsilon$ then G has a K_t-minor of order $C(\epsilon)\log n$. Note that the average degree here is exponentially larger than the one needed to guarantee a K_t-minor. This motivated Fiorini et al. [3] to conjecture that in fact any graph G with $d(G) \ge c(t) + \epsilon$ contains a K_t-minor of order $C(\epsilon)\log n$. That is, while $c(t)n$ edges are sufficient (and necessary)

to guarantee *some* K_t-minor, adding only $o(n)$ additional edges should force the appearance of the (asymptotically) smallest K_t-minor one can force even with Cn edges, for any constant C. Our main result in this paper comes very close to confirming this conjecture.

Theorem 1. For every $\epsilon > 0$ and integer $t \geq 3$ there exist $n_0 = n_0(\epsilon, t)$ such that every n-vertex graph G with $n \geq n_0$ and $d(G) \geq c(t) + \epsilon$ contains a K_t-minor of order $O(\frac{c(t)t^2}{\epsilon} \log n \log\log n)$.

Let us finally mention an old conjecture of Erdős, stating that a graph with $n^{1+\epsilon}$ edges contains a non-planar subgraph of size $C(\epsilon)$. This conjecture was confirmed (in a very strong sense) by Kostochka and Pyber [6] who proved that any graph with $4^{t^2} n^{1+\epsilon}$ edges contains a topological K_t-minor of size $O(t^2 \log t/\epsilon)$. So the conjecture of [3] that we study here is in some sense a strengthening of the conjecture of Erdős for $\epsilon = 1/\log n$.

We believe that an important aspect of this paper is the proof technique we employ here which relies on the notion of expansion in graphs and might be applicable in other settings. A good perspective on our approach comes from dense graphs, i.e. graphs with cn^2 edges. Probably the most powerful tool one has at his disposal when studying dense graphs is Szemerédi's regularity lemma [11], which asserts that *any* dense graph can be approximated by a graph consisting of a bounded number of quasi-random graphs. Since quasi-random graphs are much easier to work with, this lemma allows one to reduce a problem on arbitrary graphs to the same problem on quasi-random graphs.

When it comes to sparse graphs, there is no analogue of the regularity lemma. In recent years, a parallel paradigm has emerged, the underlying idea of which can be thought of as stating that *any* sparse graph is close to being the disjoint union of expander graphs. While the regularity lemma supplies one notion of approximation/quasi-randomness for all applications involving dense graphs, it seems like for sparse graphs different applications call for different notions of approximation and expansion. We refer the reader to [2, 4, 13] for some examples where this paradigm was applied.

Just like graph minors, expansion is one of the most well studied topics in graph theory. There are several known results connecting expansion and existence of K_t-minors in graphs, see, e.g., [1, 10, 5, 8]. In all these papers the goal was to maximize the value of t. Our task here is quite different, we want to minimize the number of vertices in the minor, keeping t fixed.

For the proof of Theorem 1, we will need a very strong notion of expansion. The price will be that we will ask for a very weak notion of approximation, which will turn out to be sufficient for proving Theorem 1. In what follows, for a set of vertices S we use $N(S)$ to denote the neighborhood of S, that is the set of vertices not in S that are connected to at least one vertex in S.

Definition 1 (δ-Expander)**.** An m-vertex graph H is said to be a *δ-expander* if for every integer $0 \leq d \leq \log\log m - 1$ and $S \subseteq V(H)$ of order $|S| \leq m/2^{2^d}$ we have $|N(S)| \geq \frac{\delta 2^d}{\log m (\log\log m)^2} |S|$.

Observe that if G is an m-vertex δ-expander then sets of vertices of size cm have vertex expansion about $1/\log m$ while sets of vertices of size m^c have a nearly constant vertex expansion. The following lemma shows that we can indeed find a δ-expander in any graph with sufficiently many edges.

Lemma 2 (Key Lemma)**.** If G satisfies $d(G) = c$, then for every $0 < \delta \leq \frac{1}{256}$ we can find in G a subgraph H, so that $d(H) \geq (1-\delta)c$ and H is a δ-expander.

The proof of Theorem 1 proceeds by first invoking Lemma 2 on the input graph G thus obtaining a graph H satisfying the expansion properties of Definition 1. We then show how one can find a small K_t-minor inside H, a task which is much easier given the fact that H has strong expansion properties. As we noted above, one can come up with different notions of expansion when studying sparse graphs. And indeed, in order to prove Theorem 1 we will actually have to prove another variant of Lemma 2, which uses a slightly different notion of expansion than the one defined above. It might very well be possible to prove other variants of Lemma 2, suitable for tackling other problems. Lemma 2 can be thought of as a strengthening of Mader's Theorem. Indeed, Mader's Theorem states that any graph G with $d(G) \geq 2k$ has a k-connected subgraph H satisfying $d(H) \geq k$. So Lemma 2 gives a similar conclusion only it replaces the notion of k-connectivity with the stronger notion of vertex expansion.

References

[1] N. Alon, P. Seymour and R. Thomas, A separator theorem for nonplanar graphs, *J. Amer. Math. Soc.* 3 (1990), 801–808.

[2] S. Arora, B. Barak and D. Steurer, Subexponential algorithms for unique games and related problems, *Proc of FOCS 2010*, 563–572.

[3] S. Fiorini, G. Joret, D. O. Theis and D. R. Wood, Small minors in dense graphs, *European J. Combinatorics* 33 (2012), 1226–1245.

[4] O. Goldreich and D. Ron, A sublinear bipartiteness tester for bounded degree graphs, *Combinatorica*, 19 (1999), 335–373.

[5] 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.

[6] A. Kostochka and L. Pyber, Small topological complete subgraphs of "dense" graphs, *Combinatorica*, 8 (1988), 83–86.

[7] A. Kostochka, Lower bound of the Hadwiger number of graphs by their average degree, *Combinatorica*, 4 (1984), 307–316.

[8] M. Krivelevich and B. Sudakov, Minors in expanding graphs, *Geometric and Functional Analysis* 19 (2009), 294–331.

[9] A. Lubotzky, R. Phillips and P. Sarnak, Ramanujan graphs, *Combinatorica* 8 (1988), 261–277.

[10] 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.

[11] E. Szemerédi, Regular partitions of graphs, In: *Proc. Colloque Inter. CNRS* (J. C. Bermond, J. C. Fournier, M. Las Vergnas and D. Sotteau, eds.), 1978, 399–401.

[12] A. Thomason, An extremal function for contractions of graphs, *Math. Proc. Cambridge Philos. Soc.* 95 (1984), 261–265.

[13] L. Trevisan, Approximation algorithms for unique games, *Theory of Computing* 4 (2008), 111–128.

Problem Session

Vera T. Sós

The following open problems were presented in a special problem session chaired by Vera T. Sós.

Nathan Linial

First question. A little background: Erdős has defined the *girth* of a 3-uniform hypergraph to be the smallest integer $v \geq 4$ such that there is a set of v vertices that contains at least $v-2$ hyperedges. He conjectured that there exist Steiner Triple Systems of arbitrarily large girth, but this question is still wide open despite considerable research effort in this direction. On the other hand, Brown, Erdős and Sós long ago observed that for every g there is some $c = c_g > 0$ and arbitrarily large n-vertex 3-uniform hypergraphs with cn^2 hyperedges and girth $\geq g$. In their argument c_g tends to zero as g grows.

Question. Is there an absolute constant $c > 0$ and n-vertex 3-uniform hypergraphs with cn^2 hyperedges and arbitrarily large girth?

Remark. It may be relevant to recall a well-known theorem of Ruzsa and Szemerédi that an n-vertex 3-uniform hypergraph in which no 6 vertices span at least 3 hyperedges, must have only $o(n^2)$ hyperedges.

Second question. The problems in my second set are not due to me and at least one is rather old.

Recall that if A is a real matrix, then its rank is the smallest integer r such that A can be expressed as the sum of r rank-1 matrices. If all entries of A are nonnegative, then it can be expressed as the sum of nonnegative rank-1 matrices. The smallest number of terms in such a sum is called the nonnegative rank of A denoted prank(A).

Question 1 (presumably due to Cohen and Rothblum, early 90's)**.** In principle, the definition of prank(A) may depend on the underlying field. If the entries of A are rational, is it necessarily the case that the rational prank and the real prank coincide?

Question 2. Do there exist nonnegative $n \times n$ matrices of bounded rank whose prank is $\Omega(n)$? In particular, are there matrices of rank 3 and prank n? It is known that rank 3 and prank$= \Omega(\sqrt{n})$ is possible.

These concepts are of substantial current interest in computational complexity and they also have interesting geometric aspects.

Noga Alon

Independence numbers of (sparse) powers of Kneser Graphs

Let K_n^t be the graph whose set of vertices is

$$V(K_n^t) = \{(A_1, A_2, \ldots, A_t) : A_i \subset \{1, 2, \ldots, n\}, |A_i| = n/3 \text{ for all } i\},$$

where the vertices $(A_1, A_2, \ldots, A_t)$ and $(B_1, B_2, \ldots, B_t)$ are adjacent iff there is an index i, $1 \leq i \leq t$, so that $A_j = B_j$ for all $j \neq i$ and $A_i \cap B_i = \emptyset$.

Let $\alpha(K_n^t)$ denote the maximum size of an independent set of K_n^t.

Conjecture. For any $\epsilon > 0$ there are n and t so that

$$\frac{\alpha(K_n^t)}{|V(K_n^t)|} \leq \epsilon.$$

Remark. The t-th *(sparse) power* of an undirected graph $G = (V, E)$, denoted by G^t, is the graph whose vertex set is V^t in which distinct vertices $(x_1 \ldots x_t)$ and $(x'_1 \ldots x'_t)$ are connected iff there exists a single index i such that x_i and x'_i are connected in G and $x_j = x'_j$ for all $j \neq i$. The study of the asymptotic behavior of the independence number of G^t, for a fixed graph G, has been considered in several papers; see [1, 2] and their references. The authors of [2] proved that for every fixed graph G, the limit $\lim_{t \mapsto \infty} \frac{\alpha(G^t)}{|V|^t}$ exists, and is at least $1/\chi(G)$ and at most the reciprocal of the fractional chromatic number of G. Moreover, for every Cayley graph of an Abelian group, this limit is precisely the reciprocal of the fractional chromatic number of G. The conjecture above asserts that when G is the Kneser graph K_n^1, this limit is smaller than ϵ provided n is sufficiently large (the fractional chromatic number of K_n^1 is 3, its chromatic number is $n/3+2$). The motivation for the conjecture is the investigation of a certain hat guessing game considered by various researchers including Feige, Levine, Peres, Tardos, Winkler, Zwick and myself.

References

[1] N. Alon, Graph Powers, in: Contemporary Combinatorics, (B. Bollobás, ed.), Bolyai Society Mathematical Studies, Springer 2002, pp. 11-28.

[2] G. Hahn, P. Hell and S. Poljak, On the ultimate independence ratio of a graph, Europ. J. Combinatorics 16 (1995), 253-261.

Jacob Fox

Problem 1. Is there a $c > 0$ such that every triangle-free graph on $n > 2$ vertices has disjoint vertex subsets of order cn^c and cn with no edges between them?

Problem 2. Call a vertex subset A of a graph ϵ-regular if for all $B \subset A$ with $|B| \geq \epsilon|A|$, the edge density of B differs from the edge density of A by at most ϵ. David Conlon and I proved that for each $\epsilon > 0$ there is $\delta = \delta(\epsilon) > 0$ such that every graph on n vertices has an ϵ-regular subset with at least δn vertices. We showed that δ^{-1} is at least exponential and at most double-exponential in a power of ϵ^{-1}. Close the gap.

Ehud Friedgut

A subset X of S_n is called a t-coset if it is a coset of a stabilizer of t points. For example, the set of all permutations that send 1 to itself and 3 to 7 is a 2-coset. Is it true that any partition of S_n into t-cosets is a refinement of a partition into $(t-1)$-cosets?

Rüdiger Reischuk

Worst inputs for the smoothed complexity of caching

The smoothed complexity of the competitive ratio for the caching problem considers neighbourhoods of an input sequence X generated by small perturbations of X and estimates the ratio between the faults of a given online algorithm and an optimal offline algorithm. For the worst case complexity, input sequences can easily be constructed that yield the maximum competitive ratio k for cache size k. It is an open problem to characterize the sequences that give the worst performance with respect to the smoothed complexity which has shown to be of the order $\log k$. The same problem is open for the maximal length of an increasing subsequence (nonconsecutive) of a sequence of n numbers and the height of a binary search tree generated by such a sequence - in both cases the value is of order $\sqrt{n}$.

Jaroslav Nešetřil

Pentagon Problem

Does there exist ℓ with the following property: For any cubic graph G with girth $\geq \ell$, there exists a homomorphism $G \to C_5$.

Remark 1. This is not true with C_7 instead of C_5 (Hatami).

Remark 2. There exists "target graphs" with odd girth 5.

David Conlon

Suppose that a graph on n vertices contains cn^t copies of K_t, the complete graph on t vertices. A theorem due to Nikiforov [2] then says that there is a blow-up of K_t with each of the t sets in the blow-up having size $c' \log n$, where c' depends only on c and t. This result easily implies a precise version of the Erdős-Stone theorem on the appearance of complete t-partite subgraphs in graphs of density greater than $1 - \frac{1}{t-1}$.

A natural question to ask, discussed at length by Rödl and Schacht [3], is whether this result can be extended to hypergraphs. The simplest open case is to decide whether every 3-uniform hypergraph on n vertices which contains cn^4 copies of $K_4^{(3)}$ also contains a blow-up of $K_4^{(3)}$ with each of the 4 sets of size $c'\sqrt{\log n}$, where c' depends only on c. I would like to reiterate this question because I believe it to be of fundamental importance. Indeed, this result and the hoped-for generalisations would easily imply an optimal extension of the Erdős-Stone theorem to hypergraphs as well as sharp results about almost monochromatic subsets of edge-coloured hypergraphs [1].

References

[1] D. Conlon, J. Fox and B. Sudakov, Large almost monochromatic subsets in hypergraphs, *Israel J. Math.* **181** (2011), 423–432.

[2] V. Nikiforov, Graphs with many r-cliques have large complete r-partite subgraphs, *Bull. London Math. Soc.* **40** (2008), 23–25.

[3] V. Rödl and M. Schacht, Complete partite subgraphs in dense hypergraphs, *Random Structures Algorithms* **41** (2012), 557–573.

János Pach

Homometric sets in graphs

Given a connected graph G and two vertices $x, y \in V(G)$, let $d_G(x, y)$ denote the distance between x and y in G. Two disjoint subsets $A, B \subset V(G)$ are called *homometric* if the multiset of distances between the elements of A is the same as the multiset of distances between the elements of B, that is, if

$$\{d_G(x, y) : x, y \in A\} = \{d_G(x, y) : x, y \in B\},$$

counted with multiplicities.

Let $h(n)$ denote the largest integer h such that every connected graph of n vertices has two disjoint subsets of vertices that are homometric. Is it true that $h(n) = o(n)$?

References

[1] M. O. Albertson, J. Pach, M. E. Young. Disjoint homometric sets in graphs. *Ars Math. Contemp.* **4** (2011), no. 1, 1–4.

Gabor Tardos

(with Mert Sağlam)

If $S_0 \subseteq S$ are finite sets, $1 \leq r \leq n$ are integers, and $H \subseteq S^n$ with $|H| = |S_0|^n$, then

$$\sum_{x \in S^n} \log(|B_r(x) \cap H| + 1) \geq \sum_{x \in S^n} \log(|B_r(x) \cap S_0^n| + 1),$$

where $B_r(x)$ is the Hamming ball of radius r:

$$B_r(x) = \{y \in S^n \mid |\{i \mid x_i \neq y_i\}| \leq r\}.$$

Ravi Kannan

Point process in unit square in $\mathbb{R}^d$

Let $v_1, \ldots, v_k$ be arbitrary points. At each time $t > k$, choose one of $v_1, \ldots, v_{t-1}$ uniformly at random, say v_j, and then pick a point, according to $\mathcal{N}(v_j, r_t^2)$, where $r_t = 1/t^\nu$. This point is labeled v_t.

Conjecture 1 (vague)**.** Phase transition at $\nu = 1/d$.

Jozsef Solymosi

The following conjecture is related to some questions about expanders.

Let $f(x)$ be a polynomial over the integers having degree at least two. Show that for every such f there is a bound B_f such that the range, $f(\mathbb{Z})$, contains no arithmetic progression of length B_f.

This conjecture would follow from various well-known conjectures, most notably from the so-called Bombieri-Lang conjecture which states that the number of rational points on an algebraic curve is uniformly bounded by the genus of the curve (if the genus is at least two). If that was true then there would be a bound on the size of the longest AP in the terms of the degree of the polynomial.

Removal lemmas for Kneser graphs and various product graphs

Ehud Friedgut
(joint work with Oded Regev)

The well known triangle removal lemma of Rusza and Szemeredi [4] states that any graph on n vertices that has less than δn^3 triangles, can be made triangle-free by removing less than ϵn^2 edges. Fox [3] greatly improved the bound of the dependence of ϵ on δ, from $\delta > 1/Tower(poly(\epsilon))$ to $\delta > 1/Tower(\log(1/\epsilon))$.

We prove analogous results for "one level lower", i.e. for certain graphs, where if you have a set of vertices that span few edges they can be turned into an independent set by removing few vertices. One such graph is the Kneser graph, however we will state our main result in the setting of a different graph, $G = K_3^{\otimes n}$. The definition of the graph is as follows. $V(G) = \{0,1,2\}^n$. $E(G) = \{\{u,w\} : u_i \neq w_i \ \forall i\}$. Note that $|V| = 3^n, |E| = 6^n/2$.

Theorem 1. For every $\epsilon > 0$ there exists $\delta > 0$ such that if a set of vertices $W \subset V(K_3^{\otimes n})$ spans less than $\delta 6^n$ edges, then there exists an independent set W such that $|V \setminus W| \leq \epsilon 3^n$.

Furthermore $\delta^{-1} < Tower(\log(1/\epsilon))$.

Despite the tower-like bound on the dependency of delta on epsilon the only examples we have are such that the dependence is polynomial.

Our results extend to a wide variety of product graphs (e.g. when the base graph is regular, or regularizable.) This can also be used to transfer the result to the Kneser graph setting.

For $k < n/2$ define the Kneser graph $K(n,k)$ to be the graph whose vertices are the k-subsets of $[n]$, with an edge between two vertices if the corresponding sets are disjoint. When $k = o(n)$ the independent sets are described precisely (and the sparse sets implicitly) in [1]. However when $k = \Theta(n)$ the following is new.

Theorem 2. For every $\epsilon > 0$ and $\alpha < 2$ there exists $\delta > 0$ such that if a set of vertices $W \subset V(K(n,\alpha n))$ spans less than $\delta\binom{n}{\alpha n}\binom{(1-\alpha)n}{\alpha n}$ edges, then there exists an independent set W such that $|V \setminus W| \leq \epsilon\binom{n}{\alpha n}$.

The motivation for proving these theorems comes from the fact that they enable us to improve the rsults of [2] and [1], and completely characterize independent sets in both settings as sets which may be approximated by "juntas", sets which are determined by a fixed number of coordinates.

References

[1] I. Dinur, E. Friedgut Intersecting families are almost contained in juntas, Combin. Probab. Comput., 18 (2009), Special Issue 1-2, 107–122

[2] I. Dinur, E. Friedgut, O. Regev, Independent sets in graph products are almost contained in juntas, G.A.F.A. 18 (2008), no. 1, 77–97

[3] J. Fox, *A new proof of the graph removal lemma.*, Ann. of Math. (2) 174 (2011), no. 1, 561–579.

[4] I. Z. Ruzsa and E. Szemerdi, *Triple systems with no six points carrying three triangles, Combinatorics*, Vol. II (Amsterdam), Colloq. Math. Soc. János Bolyai, vol. 18, North- Holland, 1978, pp. 939-945.

Chromatic number, clique subdivisions, and the conjectures of Hajós and Erdős-Fajtlowicz

Choongbum Lee

(joint work with Jacob Fox, Benny Sudakov)

A *subdivision* of a graph H is any graph formed by replacing edges of H by internally vertex disjoint paths. This is an important notion in graph theory, e.g., the celebrated theorem of Kuratowski uses it to characterize planar graphs. For a graph G, we let $\sigma(G)$ denote the largest integer p such that G contains a subdivision of a complete graph of order p. Clique subdivisions in graphs have been extensively studied and there are many results which give sufficient conditions for a graph G to have large $\sigma(G)$. For example, Bollobás and Thomason [5], and Komlós and Szemerédi [10] independently proved that every graph of average degree at least d has $\sigma(G) \geq cd^{1/2}$ for some absolute constant c. Motivated by a conjecture of Erdős, in [2] the authors further showed that when $d = \Omega(n)$ in the above subdivision one can choose all paths to have length two. Similar result for subdivisions of general graphs with $O(n)$ edges (a clique of order $O(\sqrt{n})$ clearly satisfies this) was obtained in [9].

For a given graph G, let $\chi(G)$ denote its chromatic number. A famous conjecture made by Hajós in 1961 states that $\sigma(G) \geq \chi(G)$. Dirac [7] proved that this conjecture is true for all $\chi(G) \leq 4$, but in 1979, Catlin [6] disproved the conjecture for all $\chi(G) \geq 7$. Subsequently, several researchers further studied this problem. On the negative side, by considering random graphs, Erdős and Fajtlowicz [8] in 1981 showed that the conjecture actually fails for almost all graphs. On the positive side, recently Kühn and Osthus [11] proved that all graphs of girth at least 186 satisfy Hajós' conjecture. Thomassen [12] studied the relation of Hajós' conjecture to several other problems of graph theory such as Ramsey theory, maximum cut problem, etc., and discovered many interesting connections.

In this talk, we revisit Hajós' conjecture and study to what extent the chromatic number of a graph can exceed the order of its largest clique subdivision. Let $H(n)$ denote the maximum of $\chi(G)/\sigma(G)$ over all n-vertex graphs G. The example of graphs given by Erdős and Fajtlowicz which disprove Hajós' conjecture in fact has $\sigma(G) = \Theta(n^{1/2})$ and $\chi(G) = \Theta(n/\log n)$. Thus it implies that $H(n) = \Omega(n^{1/2}/\log n)$. In [8], Erdős and Fajtlowicz conjectured that this bound is tight up to a constant factor so that $H(n) = O(n^{1/2}/\log n)$. Our first theorem verifies this conjecture.

Theorem. There exists an absolute constant C such that $H(n) \leq Cn^{1/2}/\log n$ for $n \geq 2$.

The proof shows that we may take $C = 10^{120}$, although we do not try to optimize this constant. For the random graph $G = G(n,p)$ with $0 < p < 1$ fixed, Bollobás and Catlin [4] determined $\sigma(G)$ asymptotically almost surely and later Bollobás [3] determined $\chi(G)$ asymptotically almost surely. These results imply, by picking the optimal choice $p = 1 - e^{-2}$, the lower bound $H(n) \geq (\frac{1}{e\sqrt{2}} - o(1))n^{1/2}/\log n$.

For a graph G, let $\alpha(G)$ denote its independence number. The main theorem actually follows from the study of the relation between $\sigma(G)$ and $\alpha(G)$, which might be of independent interest. Let $f(n,\alpha)$ be the minimum of $\sigma(G)$ over all graphs G on n vertices with $\alpha(G) \leq \alpha$.

Theorem. There exist absolute positive constants c_1 and c_2 such that the following holds.

(1) If $\alpha < 2\log n$, then $f(n,\alpha) \geq c_1 n^{\frac{\alpha}{2\alpha-1}}$, and
(2) if $\alpha = a\log n$ for some $a \geq 2$, then $f(n,\alpha) \geq c_2\sqrt{\frac{n}{a\log a}}$.

Note that for $\alpha = 2\log n$, both bounds from the first and second part gives $f(n,\alpha) \geq \Omega(\sqrt{n})$. Moreover, both parts of this theorem establish the correct order of magnitude of $f(n,\alpha)$ for some range of α. For $\alpha = 2$, it can be shown that in the triangle-free graph constructed by Alon [1], every set of size at least $37n^{2/3}$ contains at least n edges. This implies that the complement of this graph has independence number 2 and the largest clique subdivision of size $t < 37n^{2/3}$. Indeed, if there is a clique subdivision of order $t \geq 37n^{2/3}$, then between each of the at least n pairs of nonadjacent vertices among the t vertices of the subdivided clique, there is at least one additional vertex along the path between them in the subdivision. However, this would require at least $t + n$ vertices in the n-vertex graph, a contradiction. On the other hand, for $\alpha = \Theta(\log n)$, by considering $G(n,p)$ with constant $0 < p < 1$, one can see that the second part of our theorem is tight up to the constant factor. Even for $\alpha = o(\log n)$, by considering the complement of $G(n,p)$ for suitable $p \ll 1$, one can easily verify that there exists an absolute constant c' such that $f(n,\alpha) \leq O(n^{\frac{1}{2}+\frac{c'}{\alpha}})$.

Our theorem can also be viewed as a Ramsey-type theorem which establishes an upper bound on the Ramsey number of a clique subdivision versus an independent set.

References

[1] N. Alon, Explicit Ramsey graphs and orthonormal labelings, *Electron. J. Combin.* **1** (1994), R12.

[2] N. Alon, M. Krivelevich and B. Sudakov, Turan numbers of bipartite graphs and related Ramsey-type questions, *Combin. Probab. Comput.* **12** (2003), 477–494.

[3] B. Bollobás, The chromatic number of random graphs, *Combinatorica* **8** (1988), 49–55.

[4] B. Bollobás and P. A. Catlin, Topological cliques of random graphs, *J. Combin. Theory Ser. B* **30** (1981), 224–227.

[5] B. Bollobás and A. Thomason, Proof of a conjecture of Mader, Erdős and Hajnal on topological complete subgraphs, *European J. Combin.* **19** (1998), 883–887.

[6] P. Catlin, Hajós' graph-coloring conjecture: variations and counterexamples, *J. Combin. Theory Ser. B* **26** (1979), 268–274.

[7] G. Dirac, A property of 4-chromatic graphs and some remarks on critical graphs, *J. London Math. Soc.* **27** (1952), 85–92.

[8] P. Erdős and S. Fajtlowicz, On the conjecture of Hajós, *Combinatorica* **1** (1981), 141–143.

[9] J. Fox and B. Sudakov, Density theorems for bipartite graphs and related Ramsey-type results, *Combinatorica* 29 (2009), 153–196.

[10] J. Komlós and E. Szemerédi, Topological cliques in graphs. II, *Combin. Probab. Comput.* **5** (1996), 79–90.

[11] D. Kühn and D. Osthus, Topological minors in graphs of large girth, *J. Combin. Theory Ser. B* **86** (2002), 364–380.

[12] C. Thomassen, Some remarks on Hajós' conjecture, *J. Combin. Theory Ser. B* **93** (2005), 95–105.

Efficient algorithms for three-dimensional axial and planar random assignment problems

Gregory B. Sorkin

(joint work with Alan Frieze)

An instance of the (two-dimensional) assignment problem is given by an $n \times n$ cost array $M_{i,j}$, and the problem is to select exactly one element from each row and from each column so as to minimize the sum of all selected elements or, formulated as an integer linear program (ILP),

$$\begin{aligned} &\text{minimize } \sum_{i,j} M_{i,j} X_{i,j} \\ &\text{subject to } (\forall i) \sum_j X_{i,j} = 1, \quad (\forall j) \sum_i X_{i,j} = 1, \quad X_{i,j} \in \{0,1\}. \end{aligned}$$

The problem may also be thought of as finding the cheapest perfect matching in the complete bipartite graph with weights given by M, or the cost of assigning jobs i to machines j where each machine can accommodate only one job. The linear relaxation with $X_{i,j} \in [0,1]$ of the ILP above has integer extreme points, so the problem may be solved in polynomial time.

The *random assignment problem*, in its most popular form, is the case when the entries of the cost matrix M are i.i.d. Exp(1) random variables (independent, identically distributed exponential random variables with parameter 1). Since the

problem can be solved in polynomial time, the focus for the random case is on the cost's expectation as a function of n,

$$f(n) = \mathbf{E}\Big[\min_{X_{i,j}} \sum_{i,j} M_{i,j} X_{i,j}\Big]$$

with $X_{i,j}$ subject to the constraints above. This problem received a great deal of study over several decades. It was considered from an operations research perspective by Donath in the 1960s, an asymptotic conjecture $f(n) \to \pi^2/6 = \zeta(2)$ was formulated by statistical physicists Mézard and Parisi in the 1980s based on the mathematically sophisticated but non-rigorous "replica method", an exact conjecture $f(n) = \sum_{i=1}^{n} 1/i^2$ was hazarded by Parisi in the late 1990s, a generalization to partial matchings and non-square matrices was made by Coppersmith and Sorkin, the Mézard–Parisi conjecture was proved by Aldous in a pair of papers in 1992 and 2001, and the Coppersmith–Sorkin conjecture was proved simultaneously in 2004 by two papers using two very different methods, by Nair, Prabhakar and Sharma, and by Linusson and Wästlund. A further generalisation of these conjectures was made by Buck, Chan and Robbins in 2002, and proved by Wästlund in 2005. The study of other aspects of the random assignment problem and related problems is ongoing, for example by Wästlund. See [3] for all citations.

In higher dimensions there are two natural generalizations of the assignment problem. We illustrate in dimension $D = 3$; the generalizations to higher dimensions are clear. Taking as input an $n \times n \times n$ matrix (or "tensor" or "array") M, the Axial assignment problem is minimizing $\sum_{i,j,k} M_{i,j,k} X_{i,j,k}$ where $X_{i,j,k} \in \{0, 1\}$ and there is one selected value per "plane" of the array, of which there are three types, 1-, 2-, and 3-planes, according to which coordinate is fixed:

$$(1) \qquad (\forall i) \sum_{j,k} X_{i,j,k} = 1, \quad (\forall j) \sum_{i,k} X_{i,j,k} = 1, \quad (\forall k) \sum_{i,j} X_{i,j,k} = 1.$$

This is NP-complete; in three dimensions it was one of the original problems listed by Karp [6].

The Planar three-dimensional assignment problem is similar but with one selected value per "line" of the array, with three types of lines:

$$(2) \qquad (\forall i,j) \sum_{k} X_{i,j,k} = 1, \quad (\forall j,k) \sum_{i} X_{i,j,k} = 1, \quad (\forall i,k) \sum_{j} X_{i,j,k} = 1.$$

Again this is NP-complete, as established by Frieze [2].

The *multi-dimensional random assignment problem* we consider here is the case when the entries of the cost matrix are i.i.d. Exp(1) random variables. In this random setting, there are two natural questions. Are there polynomial-time algorithms that find optimal or near-optimal solutions w.h.p.? And what is the expected cost of a minimum assignment? We do not even know how the cost scales with n, although statistical physicists Martin, Mézard and Rivoire [8] conjecture that the Axial problem has an asymptotic expected cost of c/n; see also [7]. We now sketch our results and open problems.

1. Axial assignment

A lower bound of n^{2-D} comes from considering just the first of the three sets of constraints in (1). An upper bound of $n^{2-D}\log n$ comes from a recent non-constructive result on hypergraph factors by Johansson, Kahn and Vu [5], in a multipartite extension verified by us and (for other purposes) by Gerke and McDowell [4]. Our main result for axial assignment is an algorithm `BDAPTA`(d) based on a "bounded depth alternating path tree" of depth d, a generalization of a short alternating path as used to augment a matching. For 3-dimensional assignment it has the following properties.

Theorem 1. Suppose that $1 \leq d \leq \epsilon \log_2 \log n$ where where $0 < \epsilon < 1/2$ is a constant. For a random 3-dimensional axial assignment instance, w.h.p. Algorithm `BDAPTA`(d) runs in time $O(n^3)$ and outputs a solution of cost $O(2^{4d} n^{-1+\eta_d} \log n)$ w.h.p., where $\eta_d = \frac{1}{2^{d+1}-1}$.

An appropriate choice of d yields w.h.p. a solution of expected cost $O(n^{-1-o(1)})$, thus an $n^{o(1)}$ approximation to best possible. Not only is this the first nearly tight upper bound obtained algorithmically, it is the only good bound we are aware of except for the one following from [5].

The algorithm generalizes to dimensions $D \geq 4$, but produces a solution whose expected cost is of order $\Omega(n^{-1})$, just as in the 3-dimensional case, and far from the upper bound of $O(n^{2-D}\log n)$.

2. Planar assignment

For Planar 3-dimensional assignment, a lower bound of $\Omega(n)$ comes from considering just the first of the three sets of constraints in (2). Our second main result is an algorithm with the following properties.

Theorem 2. There is a polynomial-time algorithm that, given a random 3-dimensional planar assignment instance, w.h.p. finds a solution of cost $O(n \log n)$.

Here, the idea is that a 3-dimensional Planar assignment consists of n 2-dimensional assignments, with constraints between them. Our algorithm solves for the 2-dimensional assignments sequentially, respecting the constraints on each from the earlier ones. These 2-dimensional assignment instances have a somewhat complex structure (they are really instances of matching rather than assignment), but a general result of Dyer, Frieze, and McDiarmid [1] is insensitive to the details and yields the upper bound of $O(n \log n)$.

As for the Axial case, our approach to the Planar problem falters for dimensions $D \geq 4$. The natural generalization is to a greedy algorithm that sequentially solves n $(D-1)$-dimensional instances, but even for $D = 4$ such an algorithm can fail, reaching an instance that has no solution, regardless of cost.

3. Open problems

We are left with open questions including these: For $D \geq 3$, what are the true growth rates of the expected costs of optimal D-dimensional axial and planar assignments? For $D \geq 3$, are there asymptotically optimal, polynomial-time algorithms for solving these problems? And for $D > 3$, are there polynomial-time algorithms yielding solutions within logarithmic or $n^{o(1)}$ factors (as we have given for $D = 3$)?

References

[1] M.E. Dyer, A. Frieze, and C.J.H. McDiarmid, *On linear programs with random costs*, Mathematical Programming **35** (1986), 3–16.

[2] A.M. Frieze, *Complexity of a* 3*-dimensional assignment problem*, European J. Oper. Res. **13** (1983), no. 2, 161–164.

[3] A.M. Frieze and G.B. Sorkin, *Efficient algorithms for three-dimensional axial and planar random assignment problems*, Tech. report, 2013, `http://arxiv.org/abs/1008.0390`.

[4] S. Gerke and A. McDowell, *Non-vertex-balanced factors in random graphs*, Tech. report, 2013, `http://arxiv.org/abs/1304.3000`.

[5] A. Johansson, J. Kahn, and V. Vu, *Factors in random graphs*, Random Structures Algorithms **33** (2008), no. 1, 1–28.

[6] R.M. Karp, *Reducibility among combinatorial problems*, Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y., 1972), Plenum, New York, 1972, pp. 85–103.

[7] O.C. Martin, M. Mezard, and O. Rivoire, *A frozen glass phase in the multi-index matching problem.*

[8] O.C. Martin, M. Mézard, and O. Rivoire, *Random multi-index matching problems*, J. Stat. Mech. Theory Exp. (2005), no. 9, P09006, 36 pp. (electronic).

[9] M. Mézard and G. Parisi, *Replicas and optimization*, J. Physique Lettres **46** (1985), 771–778.

[10] ———, *On the solution of the random link matching problems*, J. Physique Lettres **48** (1987), 1451–1459.

Tight Lower Bounds for Greedy Routing in Higher-Dimensional Small-World Grids

Martin Dietzfelbinger

(joint work with Philipp Woelfel)

We consider Kleinberg's celebrated small world graph model [2, 3], in which a D-dimensional grid $\{0, \ldots, n-1\}^D$ is augmented with a constant number of additional unidirectional edges leaving each node. These *long range* edges are determined at random according to a probability distribution (the augmenting distribution), which is the same for each node. Kleinberg suggested using the *inverse D-th power distribution*, in which node v is the long range contact of node u with a probability proportional to $\|u - v\|_2^{-D}$. He showed that such an augmenting distribution allows to route a message efficiently in the resulting random graph: The greedy algorithm, where in each intermediate node the message travels over a link that brings the message closest to the target w.r.t. the Manhattan distance, finds a

path of expected length $O\big((\log n)^2\big)$ between any two nodes. In this paper we prove that greedy routing does not perform asymptotically better for any uniform and isotropic augmenting distribution, i. e., the probability that node u has a particular long range contact v is independent of the labels of u and v and only a function of $\|u - v\|_2$.

The corresponding problem for dimension 1 was solved in [1]. The method does not seem to generalize to higher dimensions.

In order to obtain the result, we introduce a novel proof technique: We define a so-called *budget game*, in which a token travels over a game board, from one end to the other, while the player manages a "probability budget". In each round, the player "bets" part of her remaining probability budget on step sizes. A step size is chosen at random according to a probability distribution of the player's bet. The token then makes progress as determined by the chosen step size, while some of the player's bet is removed from her probability budget. We prove a tight lower bound for such a budget game, and then obtain a lower bound for greedy routing in the D-dimensional grid by a reduction.

References

[1] Martin Dietzfelbinger and Philipp Woelfel, *Tight Lower Bounds for Greedy Routing in Uniform Small World Rings*, 41st ACM Symp. on Theory of Computing, 2009, pp. 591–600.

[2] Jon M. Kleinberg, *Navigation in a Small World*, Nature 406 (2000), p. 845.

[3] Jon M. Kleinberg, *The Small-World Phenomenon: An Algorithmic Perspective*, 32nd ACM Symp. on Theory of Computing, 2000, pp. 163–170.

Random Ramsey

Angelika Steger

(joint work with Rajko Nenadov)

For graphs G and F and a constant $r \in \mathbb{N}$, we denote with

$$G \to (F)^e_r$$

the property that every edge-coloring of G with r colors (we call this *r-coloring*) contains a copy of F with all edges having the same color. Ramsey's theorem then implies that for all graphs F and r we have $K_n \to (F)^e_r$, for n large enough. At first sight it is not immediately clear whether this follows from the density of K_n or its rich structure. As it turns out, studying Ramsey properties of random graphs shows that the later is the case, as random graphs give examples of sparse graphs with the desired Ramsey property.

The study of Random Ramsey Theory was initiated by Łuczak, Ruciński, and Voigt [2] who studied the Ramsey property of random graphs in the vertex-coloring case and also established the threshold for the property $G(n,p) \to (K_3)^e_2$. Thereupon, in a series of papers Rödl and Ruciński [3, 4, 5] determined the threshold of $G(n,p) \to (F)^e_r$, in full generality. Formally, their result reads as follows.

For every graph G we denote by $V(G)$ and $E(G)$ its vertex and edge sets and by v_G and e_G their sizes. For every graph G on at least 3 vertices we set $d_2(G) = (e_G - 1)/(v_G - 2)$. By $m_2(G)$ we denote for every graph G the so-called *2-density*, defined as

$$m_2(G) = \max_{J \subseteq G, v_J \geq 3} d_2(J).$$

If $m_2(G) = d_2(G)$ then we say that a graph G is *2-balanced*, and if in addition $m_2(G) > d_2(J)$ for every subset $J \subset G$ with $v_J \geq 3$, we say that G is *strictly 2-balanced*.

Theorem (Rödl, Ruciński [3, 4, 5])**.** Let $r \geq 2$ and F be a fixed graph that is not a forest of stars or, in the case $r = 2$, paths of length 3. Then there exist positive constants $c = c(F, r)$, and $C = C(F, r)$ such that

$$\lim_{n\to\infty} \Pr[G(n,p) \to (F)^e_r] = \begin{cases} 0 & \text{if } p \leq cn^{-1/m_2(F)} \\ 1 & \text{if } p \geq Cn^{-1/m_2(F)}. \end{cases}$$

For the exceptional case of a star with k rays it is easily seen that the threshold is determined by the appearance of a star with $r(k-1)+1$ rays. For path P_3 of length three the 0-statement only holds for $p \ll n^{-1/m_2(P_3)} = n^{-1}$ since, for example, a C_5 with a pending edge at every vertex has density one and cannot be edge-colored with 2 colors without a monochromatic P_3.

Note that $p = n^{-1/m_2(F)}$ is the density where we expect that every edge is contained in roughly a constant number of copies of F. This observation can be used to provide an intuitive understanding of the bounds of in the theorem. If c is very small, then the number of copies of F is a.a.s. (*asymptotically almost surely*, i.e., with probability $1 - o(1)$ if n tends to infinity) small enough that they are so scattered that a coloring without a monochromatic copy of F can be found. If, on the other hand, C is big then these copies a.a.s. overlap so heavily that every coloring has to induce at least one monochromatic copy of F.

In our talk, we gave a new short proof of the above theorem by Rödl and Ruciński. The proof of the 1-statement is based on the recent beautiful hypergraph container theorems by Saxton, Thomason [6] and Balogh, Morris, Samotij [1]. Our proof of the 0-statement is elementary.

References

[1] Balogh, J., Morris, R., and Samotij, W. Independent sets in hypergraphs. preprint, arXiv:1204.6530, 2012.

[2] Łuczak, T., Ruciński, A., and Voigt, B. Ramsey properties of random graphs. *J. Combin. Theory Ser. B 56*, 1 (1992), 55–68.

[3] Rödl, V., and Ruciński, A. 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, pp. 317–346.

[4] Rödl, V., and Ruciński, A. Random graphs with monochromatic triangles in every edge coloring. *Random Structures Algorithms 5*, 2 (1994), 253–270.

[5] Rödl, V., and Ruciński, A. Threshold functions for Ramsey properties. *J. Amer. Math. Soc. 8*, 4 (1995), 917–942.

[6] Saxton, D., and Thomason, A. Hypergraph containers. preprint, arXiv:1204.6595, 2012.

Steganography as a Game

Rüdiger Reischuk
(joint work with M. Liskiewicz, U. Wölfeld)

The goal of steganography is to hide additional information in arbitrary documents such that even the existence of such information cannot be detected [5]. For this purpose, Alice, the sender or stegoencoder, modifies given covertexts and sends these as stegotexts to Bob, the receiver. The embedding should be *reliable*, meaning that Bob can reconstruct the additional information M with high probability, and secure which in this setting means *undetectable*: A third person called Warden has only a small chance to distinguish stegotexts from covertexts. This can be modeled as a game between Alice and the adversary Warden.

Security in the area of cryptography is well understood: An adversary with bounded resources cannot decipher a secret message. If a cryptosystem is not secure then there exists such an adversary with a significant advantage over random guessing. A precise definition for security becomes much more challenging for steganography. The distributions of covertexts, also called *channels* play an important role. Security of a stegosystem is intrinsically related to knowledge about the channel. Therefore, it is important to set up the game between the stegoencoder and the adversary appropriately, and in particular to determine the level of influence that both players have in choosing the covertext channel.

We show that a stegosystem that is not secure according to the definition used so far might still not be detectable by an adversary. So far, a stegosystem is defined as *insecure* if the strongest possible adversary can detect the use of steganography. It suffices that this is true for a single channel chosen among all possible channels. A stegosystems secure in this strong sense has been proposed in [2] and was coined rejection sampling. However, a "useful" stegosystem should in addition be *efficient* (i.e., the time, space and oracle query complexities should be polynomial in the length of the hidden message) and achieve a good *transmission rate* for the hidden messages (i.e., the ratio between message bits per covertext and covertext entropy should not be too small).

However, rejection sampling embeds only 1 bit of hidden information in a document regardless of its size. Dedić et al. [1] have analysed generalizations of this idea where $b > 1$ bits are embedded in a covertext in order to achieve a better transmission rate. For this and a general class of such stegosystems they have shown that the query complexity has to grow exponentially in b. Thus, such stegosystem either have a very bad rate, or are very inefficient. In fact, it has been shown that in the common black-box setting this exponential sampling complexity holds for all stegosystems that fulfill this strong security condition. In this model the stegoencoder has no knowledge whatsoever about the covertext channel (except its min-entropy) and can only access it via a sampling oracle while the adversary is supposed to know everything about the channel. In particular, this leads to the strong conclusion that all schemes used in practice are insecure if security is defined based on this extreme setting.

In [3] we have discussed this problem and provided a new model to resolve the highly unbalanced knowledge about the covertext channel by the adversary and the encoder. In the proposed grey-box model of steganography the encoder starts with some partial knowledge – at least about the type of covertext channel.

To get a finer differentiation, we now introduce another person, Chang, who selects the channel, and consider a 3-person game between Alice, Chang and Warden. Different coalitions describe different situations of knowledge about the channel and give rise to different security measures for steganography. It is argued that the most fair and realistic situation is obtained when Chang plays randomly.

We show that one of these measures obtained by this game theoretic setting called *undetectable on average* provides much more appropriate results for real systems. For this purpose, two families of channels are constructed based on random sets and pseudorandom permutations. They are instances of what is called *flat h-channels* in [1]. In practice, one family is easy to detect while for the other it should be impossible if suitable pseudorandom permutations are used. It is shown that among all measures considered so far, only *undetectable on average* is able to classify these families correctly [4].

Summarizing, it is shown that the detectability of a stegosystem can be based on the difficulty to learn the covertext distribution. There is a tight analytical relationship between these two tasks. We leave as an open problem whether the pseudorandom functions used in our constructions can be replaced by cryptographic functions.

References

[1] N. Dedić, G. Itkis, L. Reyzin, S. Russell, *Upper and lower bounds on black-box steganography*, Journal of Cryptology **22(3)** (2009), 365–394.

[2] N. Hopper, J. Langford, L. von Ahn, *Provably secure steganography*, Advances in Cryptology – CRYPTO 2002, LNCS **2442**, Springer (2002), 77–92.

[3] M. Liśkiewicz, R. Reischuk, U. Wölfel, *Grey-Box Steganography*, Theoretical Computer Science (2012) online, to appear in print 2013.

[4] M. Liśkiewicz, R. Reischuk, U. Wölfel, *Security Levels in Steganography – Insecurity does not Imply Detectability*, TR ITCS, Univ. Lübeck (2012).

[5] H. Wang, S. Wang, *Cyber Warfare: Steganography vs. Steganalysis*, C. ACM **47**, 2004, 76–82.

Sperner's Theorem and a problem of Erdős-Katona-Kleitman

Benny Sudakov

(joint work with Shagnik Das, Wenying Gan)

Extremal set theory is one of the most rapidly developing areas in combinatorics, having applications to other branches of mathematics and computer science including discrete geometry, functional analysis, number theory and complexity. The typical extremal problem has the following form: how large can a structure be without containing some forbidden configuration? A classical example, considered by many to be the starting point of extremal set theory, is a theorem of Sperner

[8]. An *antichain* is a family of subsets of $[n]$ that does not contain sets $F_1 \subset F_2$. Sperner's Theorem states that the largest antichain has $\binom{n}{\lfloor n/2 \rfloor}$ sets, a bound that is easily seen to be tight by considering the family of sets of size $\lfloor \frac{n}{2} \rfloor$. This celebrated result enjoys numerous applications and has many extensions, many of which are discussed in Engel's book [1]. One particular extension, due to Erdős [2], shows that the size of the largest set family without a *k-chain*, that is, k-sets $F_1 \subset F_2 \subset \ldots \subset F_k$, is the sum of the $k-1$ largest binomial coefficients, $M_{k-1} = \sum_{i=\lceil \frac{n-k+2}{2} \rceil}^{\lceil \frac{n+k-2}{2} \rceil} \binom{n}{i}$. When $k = 2$, we recover Sperner's Theorem.

We study an Erdős-Rademacher-type extension of Erdős' theorem, a name we now explain. Arguably the most well-known result in extremal combinatorics is a theorem of Mantel [7] from 1907, which states that an n-vertex triangle-free graph can have at most $\lfloor \frac{n^2}{4} \rfloor$ edges. In an unpublished result, Rademacher strengthened this theorem by showing that any graph with $\lfloor \frac{n^2}{4} \rfloor + 1$ edges must contain at least $\lfloor \frac{n}{2} \rfloor$ triangles. Erdős [3] then extended this to graphs with a linear number of extra edges, and in [4] studied the problem for larger cliques. More generally, for any extremal problem, the corresponding Erdős-Rademacher problem asks how many copies of the forbidden configuration must appear in a structure larger than the extremal bound.

In the context of Sperner's Theorem, this problem was first considered by Erdős and Katona, who conjectured that a family with $\binom{n}{\lfloor n/2 \rfloor} + t$ sets must contain at least $t \lceil \frac{n+1}{2} \rceil$ 2-chains. Kleitman [6] confirmed the conjecture, and, in a far-reaching generalization, showed the minimum number of 2-chains in a family of any fixed size is obtained by choosing sets of size as close to $\frac{n}{2}$ as possible. He then conjectured (see [5, 6]) that the same families minimize the number of k-chains, a problem that has remained open for nearly fifty years.

Conjecture 1. The number of k-chains in a family is minimized by choosing sets of sizes as close to $\frac{n}{2}$ as possible.

We obtain Erdős-Rademacher-type extensions of the theorems of Sperner and Erdős. We first rediscovered a result of Kleitman on the minimum number of 2-chains in a family of any number of sets. However, through slightly more careful calculations, and by introducing an additional argument, we are able to characterize all extremal families, as given below.

Theorem 2. Let $\mathcal{F}$ be a family of subsets of $[n]$, with $|\mathcal{F}| = s \geq \binom{n}{\lfloor n/2 \rfloor}$. Let $r \in \frac{1}{2}\mathbb{N}$ be the unique half-integer such that $\sum_{i=\frac{n}{2}-r+1}^{\frac{n}{2}+r-1} \binom{n}{i} < s \leq \sum_{i=\frac{n}{2}-r}^{\frac{n}{2}+r} \binom{n}{i}$. Then $\mathcal{F}$ minimizes the number of 2-chains if and only if the following conditions are satisfied:

(1) For every $F \in \mathcal{F}$, $\frac{n}{2} - r \leq |F| \leq \frac{n}{2} + r$.
(2) For any $A \subset [n]$ with $\frac{n}{2} - r + 1 \leq |A| \leq \frac{n}{2} + r - 1$, we have $A \in \mathcal{F}$.
(3) If $s \leq \sum_{i=\frac{n}{2}-r}^{\frac{n}{2}+r-1} \binom{n}{i}$, then $\{F \in \mathcal{F} : |F| = \frac{n}{2} \pm r\}$ forms an antichain.
(4) If $s \geq \sum_{i=\frac{n}{2}-r}^{\frac{n}{2}+r-1} \binom{n}{i}$, then $\{F \notin \mathcal{F} : |F| = \frac{n}{2} \pm r\}$ forms an antichain.

Our main results verify Conjecture 1 for families of certain sizes. To begin with, recall that Erdős showed the largest family without k-chains consists of the $k-1$ middle levels of the hypercube, whose size we denote by M_{k-1}. If we were to add one set to this family, the best we could do would be to add it to the kth level, in which case we would create $\binom{\lfloor (n+k)/2 \rfloor}{k-1}(k-1)!$ k-chains. Indeed, we show that every additional set must contribute at least this many new k-chains, and the above construction shows this is tight when our extremal family is contained within the k middle levels.

Theorem 3. If $\mathcal{F}$ is a set family over $[n]$ of size $s = M_{k-1} + t$, then $\mathcal{F}$ contains at least $t\binom{\lfloor (n+k)/2 \rfloor}{k-1}(k-1)!$ k-chains.

We are then able to extend our argument to work for larger set families, obtaining a result that is tight when the extremal family is contained within the $k+1$ middle levels.

Theorem 4. Provided $n \geq 15$ and $k \leq n-6$, if $\mathcal{F}$ is a set family over $[n]$ of size $s = M_k + t$, then the number of k-chains in $\mathcal{F}$ is at least

$$\binom{n}{\lceil (n-k)/2 \rceil}\binom{\lfloor (n+k)/2 \rfloor}{k-1}(k-1)! + \\ + t\left(\binom{\lceil (n+k)/2 \rceil}{k-1} + \binom{\lceil (n+k)/2 \rceil}{k}\binom{k}{2}\right)(k-1)!.$$

In both cases, we actually obtain stronger results, providing stability versions of the above theorems, showing that if a family has close to the minimum number of k-chains, it must be close in structure to the extremal example. These stability results are of interest even in the case $k = 2$, as one does not obtain any stability from the Kleitman proof for 2-chains. We then use the stability results to show that when the above bounds are tight, the extremal families are exactly as in Theorem 2.

References

[1] K. Engel, **Sperner Theory**, Cambridge University Press, 1997.

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

[3] P. Erdős, On a theorem of Rademacher-Turán, *Illinois J. Math.* **6** (1962), 122 – 127.

[4] P. Erdős, On the number of complete subgraphs contained in certain graphs, *Magy. Tud. Acad. Mat. Kut. Int. Közl.* **7** (1962), 459 – 474.

[5] P. Erdős and D. Kleitman, Extremal problems among subsets of a set, *Discrete Math.* **8** (1974), 281 – 294.

[6] D. Kleitman, A conjecture of Erdős-Katona on commensurable pairs among subsets of an n-set, *Theory of Graphs, Proc. Colloq., Tihany* (1966), 215 – 218.

[7] W. Mantel, Problem 28, *Winkundige Opgaven* **10** (1907), 60 – 61.

[8] E. Sperner, Ein Satz über Untermengen einer endlichen Menge, *Math. Z.* **27** (1928), 544 – 548.

Large deviations in random graphs

Yufei Zhao
(joint work with Eyal Lubetzky)

The goal of this talk is to answer the following question of Chatterjee and Varadhan [3].

Question 1. Fix $0 < p < r < 1$. Let G_n be an instance of the Erdős-Rényi random graph $G(n,p)$ conditioned on the rare event of having at least $\binom{n}{3}r^3$ triangles. For large n, is G_n close in cut distance to a typical $G(n,r)$?

By cut distance here we mean the quantity

$$\delta_\square(G_n, r) = \max_{A,B\subseteq V(G_n)} \frac{1}{n^2}|e(A,B) - r|A||B||.$$

The problem of large deviations in random graphs has a long history. A representative example that drew much interest is the upper tail problem of estimating the probability that $G(n,p)$ has at least $(1+\eta)\binom{n}{3}p^3$ triangles, where η is fixed and p is allowed to vary with n. Janson, Oleszkiewicz, and Ruciński [5] and Kim and Vu [6] developed powerful techniques for proving concentration bounds for this problem. In recent breakthroughs independently by Chatterjee [1] and DeMarco and Kahn [4] they showed that the upper tail probability for triangle counts is $e^{-\Theta_\eta(n^2p^2\log(1/p))}$ when $p \geq \log n/n$, thereby determining the correct order in the exponent.

We consider the case of constant p, with the goal of determining the constant in the exponent of the upper tail probability. More precisely we are interested in the quantity

$$\text{Rate} := -\lim_{n\to\infty} \frac{1}{\binom{n}{2}} \log \mathbb{P}(G(n,p) \text{ has at least } \tbinom{n}{3}r^3 \text{ triangles}).$$

Observe that if the number of edges in $G(n,p)$ deviates to $\binom{n}{2}r$, with the edges uniformly distributed, then one has the desired triangle count deviation. This gives an upper bound on the rate

$$\begin{aligned}\text{Rate} \leq h_p(r) &:= r\log\frac{r}{p} + (1-r)\log\frac{1-r}{1-p} \\ &= -\lim_{N\to\infty}\frac{1}{N}\log\mathbb{P}(\text{Binom}(N,p)\geq Nr).\end{aligned}$$

However, there could be other "reasons" for generating too many triangles. One has the following dichotomy, with names borrowed from statistical physics, for the question: *what is the most likely reason for having too many triangles?*

Replica symmetric phase: Too many edges, uniformly distributed.
In this case Rate $= h_p(r)$ and the answer to Question 1 is YES.

Symmetry breaking phase: Some other configuration (e.g., a large clique)
In this case Rate $< h_p(r)$ and the answer to Question 1 is NO.

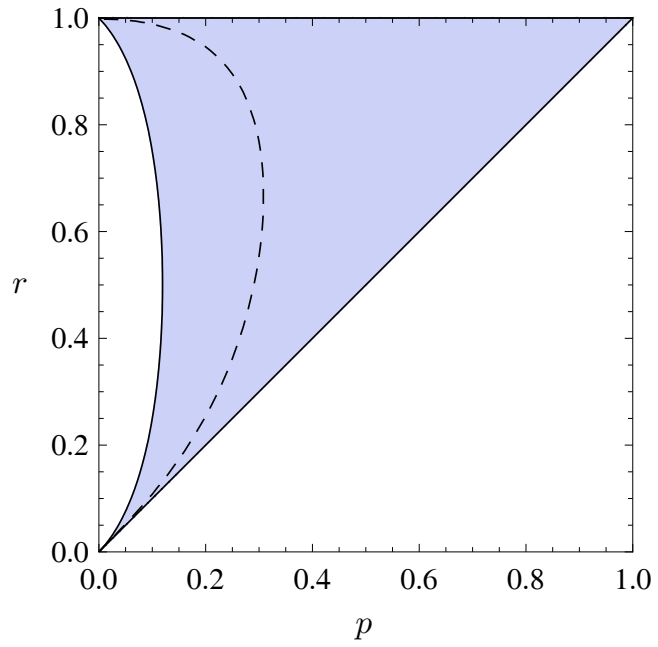

FIGURE 1. Phase diagram for the upper tail of triangle counts. Shaded region is the replica symmetric phase; the region to its left is the symmetry breaking phase. Previous results [2, 3] established replica symmetry to the right of the dashed curve.

The answer turns out to depend on (p, r). Chatterjee and Dey [2] showed, using Stein's method, that the region to the right of the dashed curve in Figure 1 belongs to the replica symmetric phase. Chatterjee and Varadhan [3] developed a new framework using graph limits and rediscovered the same replica symmetric region as [2], and furthermore they showed that for each r, one must enter the symmetry breaking phase for sufficiently small p.

Applying the framework of Chatterjee and Varadhan, which reduced the determination of the large deviation rate to an extremal problem for graphons, we identified the complete replica symmetric phase, plotted as the shaded region in Figure 1.

Theorem 2. [7] The replica symmetric phase for upper tail large deviations in triangle counts is given by $\{(p, r) : (1 + (r^{-1} - 1)^{1/(1-2r)})^{-1} \leq p < r\}$.

Furthermore, we identified the replica symmetric phase when the triangle is replaced by any d-regular graph. It turns out that the phase diagram depends only on d. The boundary curves are shown in Figure 2. We also considered the upper tail problem of having largest eigenvalue of $G(n, p)$ being at least nr (it's typically concentrated near np). The resulting phase diagram coincides with triangle count large deviations.

In the talk I explained the Chatterjee-Varadhan framework, our solution to the problem, as well as many open problems that remain.

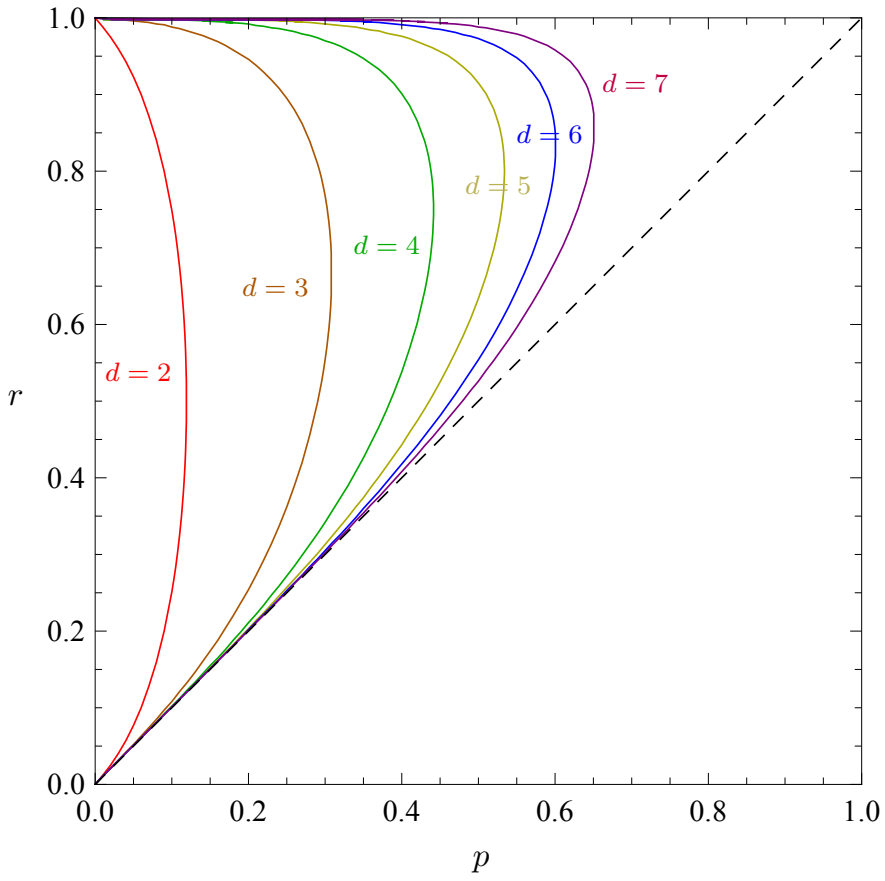

FIGURE 2. The phase boundary for counts of d-regular fixed subgraphs in $G(n,p)$.

REFERENCES

[1] S. Chatterjee. The missing log in large deviations for triangle counts. *Random Structures Algorithms*, 40(4):437–451, 2012.

[2] S. Chatterjee and P. S. Dey. Applications of Stein's method for concentration inequalities. *Ann. Probab.*, 38(6):2443–2485, 2010.

[3] S. Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.*, 32(7):1000–1017, 2011.

[4] B. DeMarco and J. Kahn. Upper tails for triangles. *Random Structures Algorithms*, 40(4):452–459, 2012.

[5] S. Janson, K. Oleszkiewicz, and A. Ruciński. Upper tails for subgraph counts in random graphs. *Israel J. Math.*, 142:61–92, 2004.

[6] J. H. Kim and V. H. Vu. Divide and conquer martingales and the number of triangles in a random graph. *Random Structures Algorithms*, 24(2):166–174, 2004.

[7] E. Lubetzky and Y. Zhao. On replica symmetry of large deviations in random graphs. Preprint `http://arxiv.org/abs/1210.7013`arXiv:1210.7013.

The triangle-free process and $R(3,k)$

Robert Morris
(joint work with Gonzalo Fiz Pontiveros and Simon Griffiths)

One of the central problems in Ramsey Theory is that of bounding the so-called *Ramsey numbers*:

$$R(k,\ell) \,=\, \min\big\{n\in\mathbb{N} \,:\, \text{every red-blue colouring of } E(K_n) \text{ contains either a red copy of } K_k \text{ or a blue copy of } K_\ell\big\}.$$

The order of magnitude of $R(k,\ell)$ is only known in one (non-trivial) case, due to the following theorem of Ajtai, Komlós and Szemerédi [1] and Kim [8].

Theorem 1 (Ajtai, Komlós and Szemerédi (1981), Kim (1995))**.**

$$R(3,k) \,=\, \Theta\left(\frac{k^2}{\log k}\right).$$

The method of [1] was refined by Shearer [10], who improved the constant in the upper bound. However, a factor of roughly 100 remained between his result and Kim's lower bound.

The triangle-free process. For each $n\in\mathbb{N}$, consider the following random graph process $(G_m)_{m\in\mathbb{N}}$ on vertex set $[n]=\{1,\dots,n\}$. Let G_0 be the empty graph and, for each $m\in\mathbb{N}$, let G_m be obtained from G_{m-1} by adding a single edge, chosen uniformly from those non-edges of G_{m-1} which do not create a triangle. The process ends when we reach a (random) maximal triangle-free graph $G_{n,\triangle}$.

This process, now known as the *triangle-free process*, was first suggested by Bollobás and Erdős at the "Quo Vadis, Graph Theory?" conference in 1990 as a possible method of obtaining random graphs with good Ramsey-type properties, and was first studied rigorously by Erdős, Suen and Winkler [6], who showed that, with high probability[1], $G_{n,\triangle}$ has at least $cn^{3/2}$ edges for some constant $c>0$. Determining the order of magnitude of $e(G_{n,\triangle})$ remained an open problem for nearly 20 years until the breakthrough paper of Bohman [2], who followed the triangle-free process for a constant proportion of its lifespan, and hence proved that

$$e\big(G_{n,\triangle}\big) \,=\, \Theta\big(n^{3/2}\sqrt{\log n}\big).$$

He also showed that G_m has various pseudo-random properties, and as a consequence was able to give a second proof of Kim's lower bound on $R(3,k)$, with a similar constant. We remark that a corresponding result in the more general setting of the H-free process (defined analogously) was obtained by Bohman and Keevash [3], improving on earlier results of Bollobás and Riordan [5] and Osthus and Taraz [9].

The main result of this talk, proved in [7], is the following sharp version of Bohman's Theorem.

[1]We write 'with high probability' to mean with probability tending to 1 as $n\to\infty$.

Theorem 2.

$$e\big(G_{n,\triangle}\big) \;=\; \left(\frac{1}{2\sqrt{2}} + o(1)\right) n^{3/2}\sqrt{\log n},$$

with high probability as $n \to \infty$.

We moreover control various parameters associated with the graph process, showing that they take the values one would expect in a random graph of the same density. Using these, it is possible (with substantially more work) to obtain the following improvement of Kim's lower bound on $R(3,k)$.

Theorem 3.

$$\left(\frac{1}{4} - o(1)\right)\frac{k^2}{\log k} \;\leqslant\; R(3,k) \;\leqslant\; \big(1 \pm o(1)\big)\frac{k^2}{\log k}$$

as $k \to \infty$.

We remark that very similar results have recently been obtained independently by Bohman and Keevash [4] using related methods. We also repeat, for emphasis, that the upper bound in Theorem 3 was proved by Shearer [10] over 25 years ago.

An outline of the proof. A technique which has proved extremely useful in the study of random graph processes is the so-called 'differential equations method' whose application in Combinatorics was pioneered by Wormald, see [11]. The basic idea is to 'track' a large (but finite) collection of graph parameters, such that the (expected) rate of change of each depends only on some subset of the others. We show that, for each of these parameters, the probability that it is the *first* parameter to go astray (that is, to have normalized error larger than 1) is extremely small.

The most basic parameter we need to track is the number of *open edges* in G_m, where

$$O(G_m) \;=\; \big\{e \in E(K_n) \setminus E(G_m) \,:\, e \not\subseteq N_{G_m}(v) \text{ for every } v \in V(G_m)\big\}.$$

We shall write $Q(m) = |O(G_m)|$. Observe that the open edges of G_m are exactly those which can be added to the graph at step $m+1$. In order to control $Q(m)$, we need to track the parameters $X_e(m)$ (the number of open-open pairs on edge e) and $Y_e(m)$ (the number of edge-open edge pairs on e) for each $e \in O(G_m)$; these variables control the number of edges which are closed at each step of the process, and (together with $Q(m)$) they also control their respective derivatives.

In order to prove Theorem 2, we need to control these parameters up to an error term which *decreases super-exponentially quickly* in $t = m \cdot n^{-3/2}$. In order to obtain such a tiny error, we exploit the self-correcting nature of the triangle-free process; doing so requires three separate steps, each of which relies crucially on the other two.

First, we show that $Q(m)$ evolves (randomly) with $\overline{X}(m)$ and $\overline{Y}(m)$ (the *averages* over all open edges of G_m of the variables $X_e(m)$ and $Y_e(m)$, respectively) according to a 'whirlpool-like' structure. Using a suitably chosen Lyapunov function, we are able to show that this three-dimensional system is self-correcting, even though $Q(m)$ itself is not.

Second, for each $k \in \mathbb{N}$ and $e \in O(G_m)$, we track a variable $V_e^{(k)}(m)$, which is (roughly speaking) the k^{th} derivative of $Y_e(m)$. To define this variable, consider for each $m \in \mathbb{N}$ the graph (the 'Y-graph' of G_m) with vertex set $O(G_m)$, and an edge between each pair $\{f, f'\}$ such that $f' \in Y_f(m)$; then $V_e^{(k)}(m)$ is the average of $Y_f(m)$ over the edges $f \in O(G_m)$ at walk-distance k from e in the Y-graph. Crucially, our error bounds on these variables decrease exponentially quickly in k, and using this fact we are able to prove self-correction. A vital ingredient in this calculation amounts to showing that a random walk on the Y-graph mixes in constant time, and the proof of this property of the Y-graph uses the fact that we can track certain 'ladder-like' graph structures in G_m.

Finally, in order to control the number of 'ladder-like' structures, we in fact track the number of copies of *every* graph structure F which occurs in G_m (at a given 'root'), up to the point at which it is likely to disappear, and after this time we bound the number of copies up to a polylog-factor. Such a general result is not only interesting in its own right; it is necessary for our proof to work, because (for our martingale bounds) we need to track the maximum possible number of copies of each structure which are created or destroyed in a single step of the triangle-free process, which depends on (the number of copies of) several other structures, some of which may be tracking, and others not.

References

[1] M. Ajtai, J. Komlós and E. Szemerédi, A dense infinite Sidon sequence, *Europ. J. Combinatorics*, **2** (1981), 1–11.

[2] T. Bohman, The triangle-free process, *Adv. Math.*, **221** (2009), 1653–1677.

[3] T. Bohman and P. Keevash, The early evolution of the H-free process, *Inventiones Math.*, **181** (2010), 291–336.

[4] T. Bohman and P. Keevash, Dynamic Concentration of the Triangle-Free Process, submitted.

[5] B. Bollobás and O. Riordan, Constrained graph processes, *Electronic J. Combinatorics*, **7** (2000), R18.

[6] P. Erdős, S. Suen and P. Winkler, On the size of a random maximal graph, *Random Structures Algorithms*, **6** (1995), 309–318.

[7] G. Fiz Pontiveros, S. Griffiths and R. Morris, The triangle-free process and $R(3, k)$, submitted.

[8] J.H. Kim, The Ramsey number $R(3, t)$ has order of magnitude $t^2 / \log t$, *Random Structures Algorithms*, **7** (1995), 173–207.

[9] D. Osthus and A. Taraz, Random maximal H-free graphs, *Random Structures Algorithms*, **18** (2001), 61–82.

[10] J.B. Shearer, A note on the independence number of triangle-free graphs, *Discrete Math.*, **46** (1983), 83–87.

[11] N.C. Wormald, The differential equation method for random graph processes and greedy algorithms, in: Lectures on Approximation and Randomized Algorithms, PWN, Warsaw, 1999, 73–155.

The combinatorics and probability of simplicial complexes

Nati Linial

(joint work with Lior Aronshtam, Tomasz Luczak, Roy Meshulam, Yuval Peled)

This is part of an ongoing research effort to investigate higher-dimensional counterparts of some basic combinatorial constructs. By now there are studies concerning high-dimensional permutations (e.g. [5]), tournaments [4] and more. One of the most obvious domains in which this program suggests itself quite naturally is the combinatorial study of simplicial complexes. This is a particularly natural line of study, since graphs are, from the geometric perspective, *just* one-dimensional simplicial complexes. In particular it is of interest to consider *random simplicial complexes* in light of what is already known concerning random graphs. One of our main hopes is that the probabilistic method which is so central to modern combinatorics can be just as beneficial in the geometric/topological realm.

In [3] we introduced a model of random d-dimensional simplicial complexes which coincides with $G(n,p)$ for $d = 1$. In that paper and in subsequent work of Meshulam and Wallach we established a natural analog to the fact that the threshold for graph connectivity is $p = \frac{\log n}{n}$.

More recently we have focused our attention on seeking an analog to the phase transition of $G(n,p)$ that occurs at $p = \frac{1}{n}$. This famous set of results can be viewed as answering the following question: What is the critical p at which graphs from $G(n,p)$ a.s. cease to be forests. As it turns out, there are at least two natural high-dimensional analogs of being a forest. Namely, collapsibility and the vanishing of the d-th homology. In [1, 2] we investigate the thresholds for these two events.

Finally I reported on ongoing work with Yuval Peled on the nature of $\mathbb{Q}$-hypertrees as defined by Kalai (1983).

References

[1] Collapsibility and vanishing of top homology in random simplicial complexes Lior Aronshtam, Nati Linial, Thomasz Luczak and Roy Meshulam, Discrete and Computational Geometry, to appear.

[2] When does the top homology of a random simplicial complex vanish?, Lior Aronshtam and Nati Linial, Random Structures and Algorithms, to appear.

[3] Homological connectivity of random 2-dimensional complexes, N. Linial, and R. Meshulam, Combinatorica, 26(2006) 475–487.

[4] On high-dimensional acyclic tournaments, Nati Linial and Avraham Morgenstern, manuscript.

[5] An upper bound on the number of high-dimensional permutations, Nati Linial and Zur Luria, Discrete and Computational Geometry, to appear.

Turan's Theorem for random graphs

Jeff Kahn
(joint work with Robert DeMarco)

Write $t_r(G)$ (resp. $b_r(G)$) for the maximum size (i.e. number of edges) of a K_r-free (resp. $(r-1)$-partite) subgraph of a graph G. Of course $t_r(G) \geq b_r(G)$, and Turán's Theorem [5] says that equality holds when $G = K_n$. The question addressed here is, roughly, when is equality likely to hold for the random graph $G_{n,p}$; that is, for what $p = p(n)$ is it true that

$$t_r(G_{n,p}) = b_r(G_{n,p}) \text{ w.h.p.?} \tag{1}$$

(As usual an event holds *with high probability* (w.h.p.) if its probability tends to 1 as $n \to \infty$.) Note that (1) holds for small enough p, for the silly reason that $G_{n,p}$ is itself likely to be $(r-1)$-partite; but we are thinking of more interesting values of p.

The problem seems to have first been considered by Babai, Simonovits and Spencer [1], who showed that for $r = 3$ (in which case Turán's Theorem is actually Mantel's [4]), (1) holds when $p > 1/2$ (more precisely, when $p > 1/2 - \varepsilon$ for some fixed $\varepsilon > 0$), and asked whether this could be extended to $p > n^{-c}$ for some fixed positive c. This was accomplished (with $c = 1/250$) by Brightwell, Panagiotou and Steger [2], who actually proved the corresponding result for every (fixed) r:

Theorem 1 ([2])**.** For each r there is a $c > 0$ such that if $p = p(n) > n^{-c}$ then w.h.p. every maximum K_r-free subgraph of $G_{n,p}$ is $(r-1)$-partite.

It was also suggested in [2] that when $r = 3$, $p > n^{-1/2+\varepsilon}$ might suffice for (1); the precise answer in this case ((1) holds for p at least $Cn^{-1/2}\log^{1/2} n$) was proved in [3]. Here we settle the problem for every r:

Theorem 2. For each fixed r there is a C such that if

$$p > Cn^{-\frac{2}{r+1}} \log^{\frac{2}{(r+1)(r-2)}} n,$$

then w.h.p. every maximum K_r-free subgraph of $G_{n,p}$ is $(r-1)$-partite.

This is best possible (apart from the value of C), basically because for smaller p there are usually edges not lying in K_r's, and while these are automatically in all maximum K_r-free subgraphs, there's no reason to expect that they will all be contained in every (or any) maximum $(r-1)$-partite subgraph.

References

[1] L. Babai, M. Simonovits and J. Spencer, Extremal subgraphs of random graphs, *J. Graph Th.* **14** (1990), 599-622.
[2] G. Brightwell, K. Panagiotou and A. Steger, Extremal subgraphs of random graphs, *Random Structures Algorithms*, to appear. http://arxiv.org/abs/0908.3778 (Extended abstract: pp. 477-485 in *SODA '07 (Proc. 18th ACM-SIAM Symp. Discrete Algorithms)*.
[3] R. DeMarco and J. Kahn, Mantel's Theorem for random graphs, submitted.
[4] W. Mantel, Problem 28, *Wiskundige Opgaven,* 10:60-61, 1907.

[5] P. Turán, Eine Extremalaufgabe aus der Graphentheorie, *Mat. Fiz Lapook* **48** (1941), 436-452.

Extremal results for string graphs and graph drawings

Jacob Fox
(joint work with János Pach)

A graph $G = (V, E)$ is called a *string graph* if it is the intersection graph of curves in the plane, i.e., if there is a collection of curves ("strings") γ_v in the plane, one curve for each vertex $v \in V$, such that two curves γ_u and γ_v intersect if and only if u and v are adjacent in G.

A *separator* in a graph $G = (V, E)$ is a subset S of the vertex set V such that no connected component of $G \setminus S$ has more than $\frac{2}{3}|V|$ vertices. Equivalently, S is a separator of G if there is a partition $V = S \cup V_1 \cup V_2$ with $|V_1|, |V_2| \leq \frac{2}{3}|V|$ such that no vertex in V_1 is adjacent to any vertex in V_2.

In [6], we proved that every string graph G with m edges has a separator of size $O(m^{3/4}\sqrt{\log m})$, and conjectured that this bound can be improved to $O(\sqrt{m})$. This result, if true, would be be best possible. In [5], we proved our conjecture in the special case where the vertices of G can be represented by curves in the plane with the property that every pair of them intersect in at most a bounded number of points. The starting point of our investigations was a recent paper of Matoušek [11], in which he ingeniously adapted some powerful techniques developed by Feige, Hajiaghayi, and Lee [4], using the framework of multicommodity flows to design efficient approximation algorithms for finding small separators in general graphs. (See [2], for a similar application.) Matoušek [11] proved our above conjecture up to a logarithmic factor.

Lemma 1. [11] Every string graph with m edges has a separator of size at most $d\sqrt{m}\log m$, where d is an absolute constant.

The aim of this note is to combine Lemma 1 with some previous results of the authors to substantially improve the best known estimates for various important parameters of string graphs.

Our first result provides an upper bound on the chromatic number of string graphs with no complete subgraph of size t, which is polylogarithmic in the number of vertices.

Theorem 2. There is an absolute constant C such that every K_t-free string graph on n vertices has chromatic number at most $(\log n)^{C \log t}$.

Previously, it was not even known if the chromatic number of every triangle-free string graph on n vertices is at most $n^{o(1)}$. In the other direction, solving an old problem of Erdős, for every n, Pawlik, Kozik, Krawczyk, Lasoń, Micek, Trotter, and Walczak [14] constructed a triangle-free intersection graph of n *segments* in the plane with chromatic number at least $\log\log n$. In particular, it follows that

the chromatic number of triangle-free string graphs cannot be bounded from above by a constant.

A *topological graph* is a graph drawn in the plane so that its vertices are represented by points and its edges are represented by (possibly crossing) curves connecting the corresponding point pairs. We also assume that no edge passes through any point representing a vertex other than its endpoints. For any integer $t \geq 2$, we say that a topological graph is *t-quasi-planar* if it has no t pairwise crossing edges. According to an old conjecture made independently by several people (see, e.g., Problem 6 in [12]), for any integer $t \geq 2$, there is a constant c_t such that every t-quasi-planar topological graph on n vertices has at most $c_t n$ edges. Theorem 2 immediately implies the following result, originally established in [7] by a more complicated argument.

Corollary 3. [7] Every t-quasi-planar topological graph on $n > 2$ vertices has at most $n(\log n)^{c \log t}$ edges, for an appropriate constant c.

A family of graphs is called *hereditary* if it is closed under induced subgraphs. The Erdős-Hajnal conjecture [3] states that for every hereditary family F of graphs which is not the family of all graphs, there is a constant $c = c_F$ such that every graph in F on n vertices contains a clique or independent set of size n^c. A weaker estimate, with $e^{c\sqrt{\log n}}$ instead of n^c, was established by Erdős and Hajnal. It is not known whether the Erdős-Hajnal conjecture holds for string graphs. Our next theorem, which follows from Theorem 2, represents the first progress on this problem.

Theorem 4. For every $\varepsilon > 0$, there is a constant $c = c_\varepsilon > 0$ such that the following holds. Every string graph on $n > 2$ vertices contains a complete subgraph of order $n^{c/\log\log n}$ or an independent set of order $n^{1-\varepsilon}$. That is, every collection of $n > 2$ curves in the plane contains a subcollection of at least $n^{c/\log\log n}$ pairwise intersecting curves or a subcollection of at least $n^{1-\varepsilon}$ pairwise disjoint curves.

The classical Kővári-Sós-Turán theorem [9] states that any $K_{t,t}$-free graph with n vertices has at most $n^{2-1/t} + tn/2$ edges. Pach and Sharir [13] conjectured that, for string graphs, this upper bound can be replaced by a bound linear in n. That is, every $K_{t,t}$-free string graph on n vertices has at most $c_t n$ edges. They verified this conjecture up to a polylogarithmic factor in n. In [6], it was proved that the conjecture is true with $c_t \leq t^{c \log\log t}$. The authors further conjectured that the statement also holds with $c_t = ct \log t$, which would be best possible. We get close to this conjecture, proving the upper bound $c_t \leq t(\log t)^{O(1)}$.

Theorem 5. There is a constant c such that for any positive integers t and n, every $K_{t,t}$-free string graph with n vertices has at most $t(\log t)^c n$ edges.

The celebrated crossing lemma of Ajtai, Chvátal, Newborn, Szemerédi [1] and, independently, Leighton [10] states that in every drawing of a graph with n vertices and $m \geq 4n$ edges, there are at least $\Omega(\frac{m^3}{n^2})$ pairs of crossing edges. This is easily seen to be equivalent to the existence of *one* edge that crosses $\Omega(\frac{m^2}{n^2})$ other edges.

Indeed, by the crossing lemma, the average number of edges a single edge crosses is $\Omega(\frac{m^2}{n^2})$. In the other direction, by repeatedly pulling out one edge at a time that crosses $\Omega(\frac{m^2}{n^2})$ of the remaining edges, a total of $\Omega(m)$ edges are pulled out that each cross $\Omega(\frac{m^2}{n^2})$ other edges. This gives $\Omega(m \cdot \frac{m^2}{n^2})$ pairs of crossing edges, and hence implies the crossing lemma.

Can the crossing lemma be strengthened to show that every graph drawn with n vertices and $m \geq 4n$ edges contains two sets E_1, E_2 of edges, each of size $\Omega(\frac{m^2}{n^2})$, such that every edge in E_1 crosses every edge in E_2? In [8], the authors and Cs. Tóth proved that, although the answer is no, the statement is true up to a polylogarithmic factor. It is also an easy consequence of Theorem 5.

Corollary 6. [8] In every topological graph G with n vertices and $m \geq 4n$ edges, there are two disjoint edge sets E_1, E_2 with $|E_1|, |E_2| \geq \frac{m^2}{n^2(\log \frac{m}{n})^c}$, such that every edge in E_1 crosses every edge in E_2. Here $c > 0$ is a suitable absolute constant.

At least one logarithmic factor is needed in Corollary 6. In [8], we constructed topological graphs with n vertices and $m \geq 4n$ edges for which the largest pair of crossing sets has cardinality $O(\frac{m^2}{n^2 \log \frac{m}{n}})$.

References

[1] M. Ajtai, V. Chvátal, M. Newborn, and E. Szemerédi, *Crossing-free subgraphs*, in: Theory and Practice of Combinatorics, Math. Stud. **60**, North-Holland, Amsterdam, 1982, pp. 9–12.

[2] P. Biswal, J. R. Lee, and S. Rao, *Eigenvalue bounds, spectral partitioning, and metrical deformations via flows*, J. ACM **57** (2010), no. 3, Art. 13, 23 pp.

[3] P. Erdős and A. Hajnal, *Ramsey-type theorems*, Discrete Appl. Math. **25** (1989), 37–52.

[4] U. Feige, M. Hajiaghayi, and J. R. Lee, *Improved approximation algorithms for minimum weight vertex separators*, SIAM J. Comput. **38** (2008), 629–657.

[5] J. Fox and J. Pach, *Separator theorems and Turán-type results for planar intersection graphs*, Adv. Math. **219** (2008), 1070–1080.

[6] J. Fox and J. Pach, *A separator theorem for string graphs and its applications*, Combin. Probab. Comput. **19** (2010), 371–390.

[7] J. Fox and J. Pach, *Coloring K_k-free intersection graphs of geometric objects in the plane*, European J. Combin. **33** (2012), 853–866.

[8] J. Fox, J. Pach, and Cs. D. Tóth, *A bipartite strengthening of the crossing lemma*, J. Combin. Theory Ser. B **100** (2010), 23–35.

[9] T. Kővári, V. Sós, and P. Turán, *On a problem of K. Zarankiewicz*, Colloq. Math. **3** (1954), 50–57.

[10] T. Leighton, *New lower bound techniques for VLSI*, Math. Systems Theory **17** (1984), 47–70.

[11] J. Matoušek, *Near-optimal separators in string graphs*, preprint.

[12] J. Pach, *Notes on geometric graph theory*, Discrete and Computational Geometry (J.E. Goodman et al, eds.), DIMACS Series **6**, Amer. Math. Soc., Providence, 1991, 273–285.

[13] J. Pach and M. Sharir, *On planar intersection graphs with forbidden subgraphs*, J. Graph Theory **59** (2009), 205–214.

[14] A. Pawlik, J. Kozik, T. Krawczyk, M. Lasoń, P. Micek, W. T. Trotter, and B. Walczak, *Triangle-free intersection graphs of line segments with large chromatic number*, arXiv:1209.1595.

Property testing for hyperfinite properties

Christian Sohler
(joint work with Ilan Newman)

In this talk, I will summarise some of the recent developments in the area of property testing in bounded degree graphs. The area of graph property testing deals with the question of whether a given graph has a predetermined property or is ε-far away from every graph that has this property. If the maximum degree of a graph is bounded by d, being ε-far from a property means that one has to insert and/or delete more than εdn edges to obtain a graph that has the tested property. The goal is to fulfill this task by random sampling.

It turns out that hyperfinite graphs, *i.e.* graphs that can be decomposed into components of constant size by the removal of at most εdn edges, play an important role in the recent development in this area. It was shown that every bounded degree hyperfinite graph is already determined up to εdn edges by its local structure, *i.e.* the distribution of the rooted subgraphs induced by the vertices with a constant distance from the root.

So, informally, every bounded degree hyperfinite graph has a constant size description that determines its structure up to the insertion and/or deletion of εdn edges.

Given this result, every hyperfinite property is testable.

On the structure of dense graphs with small clique number

Mathias Schacht
(joint work with Heiner Oberkampf)

Chromatic thresholds. We are interested in structural properties of large and dense graphs $G = (V, E)$ that do not contain a copy of a fixed small graph F. Here *density* will be given a condition on the minimum degree of G and *structural properties* are captured by studying homomorphic images of G.

For example, if $F = K_r$ is a clique and $\delta(G)$ is sufficiently high, then Turán's theorem [12] asserts that G is $(r-1)$-partite, in particular, the chromatic number $\chi(G)$ is bounded by a constant independent of the number of vertices of G. More generally, Andrásfai [2] raised the following problem: *Given a graph F and an integer t, determine the smallest function $f_{F,t}(n)$ such that the minimum degree condition $\delta(G) \geq f_{F,t}(n)$ for any n-vertex, F-free graph G yields $\chi(G) \leq t$.*

We are interested in the case for which minimum degree condition implies some bound on $\chi(G)$ independent from the size of the graph G itself. This leads to the so-called *chromatic threshold*. For a graph F we define

$$\delta_{F,\chi} = \inf\{\alpha \in [0,1] \colon \text{there is } t_\alpha \in \mathbb{N} \text{ such that every } F\text{-free graph } G \text{ with } \delta(G) > \alpha|V(G)| \text{ satisfies } \chi(G) \leq t_\alpha\}\,.$$

If $F' \subseteq F$, then obviously $\delta_{F',\chi} \leq \delta_{F,\chi}$. Moreover, it follows from the Erdős–Stone theorem [5] that $\delta_{F,\chi} \leq \frac{\chi(F)-2}{\chi(F)-1}$ for every graph F with at least one edge.

For $F = K_3$ it was shown in [4] that $\delta_{K_3,\chi} \geq 1/3$ and in the other direction Thomassen [10] obtained a matching upper bound, i.e.,

$$\delta_{K_3,\chi} = \frac{1}{3}\,. \tag{1}$$

In fact, Erdős and Simonovits [4] asked whether all triangle-free graphs G with $\delta(G) \geq (1/3 + o(1))|V(G)|$ are 3-colorable. This was answered negatively by Häggkvist [7], but recently Brandt and Thomassé [3] showed that the chromatic number of such graphs is bounded by 4.

Nikiforov [9] extended these results from triangles to r-cliques and showed that

$$\delta_{K_r,\chi} = \frac{2r-5}{2r-3}$$

and, moreover, he showed that $\chi(G) \leq r+1$ for every K_r-free graph G with $\delta(G) > \frac{2r-5}{2r-3}|V(G)|$. In the case when F is an odd cycle of length at least five it was shown by Thomassen [11] that the chromatic threshold is zero and recently, Allen et al. [1] determined the threshold for every graph F.

Homomorphism thresholds. What can be said if we move away from bounding the chromatic number for dense, F-free graphs G and require instead that there is a homomorphic image H of G, i.e., $G \xrightarrow{\text{hom}} H$, which itself is F-free and of bounded size. More precisely, in [10] Thomassen posed the following question: *Given a fixed constant c, does there exist a finite family of triangle-free graphs such that every triangle-free graph on n vertices with minimum degree greater than cn is homomorphic to some graph of this family?* We formalize this question by defining

$$\begin{aligned}\delta_{F,\text{hom}} = \inf\{\alpha \in [0,1]\colon\ &\text{there is } t_\alpha \in \mathbb{N} \text{ such that for all } F\text{-free graphs } G\\ &\text{with } \delta(G) > \alpha|V(G)| \text{ there exists an } F\text{-free graph } H\\ &\text{satisfying } G \xrightarrow{\text{hom}} H \text{ and } |V(H)| \leq t_\alpha\}\,.\end{aligned}$$

Thomassen asked to determine $\delta_{K_3,\text{hom}}$. The existence of a (F-free) graph H of order t and of a homomorphism $G \xrightarrow{\text{hom}} H$, clearly implies $\chi(G) \leq t$. Consequently, for all graphs F we have

$$\delta_{F,\text{hom}} \geq \delta_{F,\chi}\,.$$

Łuczak [8] answered Thomassen's question and proved that $\delta^h_{K_3,\text{hom}} = 1/3$. Hence, for K_3 the homomorphism threshold and the chromatic threshold equal. Goddard and Lyle [6] extended Łuczak's result showing, that K_r-free graphs with minimum degree bigger than $\frac{2r-5}{2r-3}$ are homomorphic to the join $K_{r-3} \vee H$, where H is a triangle-free graph with $\delta(H) > |V(H)|/3$. Thus we have the following theorem.

Theorem 1 (Goddard & Lyle)**.** For every integer $r \geq 3$ we have

$$\delta_{K_r,\text{hom}} = \delta_{K_r,\chi} = \frac{2r-5}{2r-3}\,.$$

We discuss a new proof of Theorem 1 based on a simple probabilistic argument. Roughly speaking, the proof proceeds as follows. Given an K_r-free graph $G = (V, E)$ with $\delta(G) \geq (\frac{2r-5}{2r-3} + \varepsilon)|V|$. Consider a random set X of size polynomial in $1/\varepsilon$ and r. In particular, the size of X is bounded by a function depending on ε and r only. With high probability all but at most $\varepsilon|V|/10$ vertices in G have at least $(\frac{2r-5}{2r-3} + \varepsilon/2)|X|$ neighbours in X. Let $V = U \cup Z$, where Z is the set of the exceptional vertices (i.e., $|Z| \leq \varepsilon|V|/10$). Next we consider a partition $U_1 \cup \ldots \cup U_s$ of U according the neighbourhood in X, i.e., two vertices u and u' belong to the same class U_i if and only if $N_G(u) \cap X = N_G(u') \cap X$. In particular, the number of classes $s \leq 2^{|X|}$ is bounded. It is easy to check that the classes U_i are independent in G. Moreover, one can show that $G[U]$ is isomorphic to a blow-up of an K_r-free graph on s vertices and, hence, $G[U]$ has a K_r-free homomorphic image on s vertices. Finally, we deal with the vertices in Z. Since, $|Z| \leq \varepsilon|V|/10$ every vertex $z \in Z$ has at least $(\frac{2r-5}{2r-3} + \varepsilon/2)|U|$ neighbours in U. Now we partition the vertices in Z according to their neighbourhood pattern among $U_1, \ldots, U_s$. More precisely, we put z and z' into the same class Z_j if and only if

$$\{i \in [s] \colon N(z) \cap U_i \neq \emptyset\} = \{i \in [s] \colon N(z') \cap U_i \neq \emptyset\}\,.$$

Therefore, we partition Z into at most $2^s \leq 2^{2^{|X|}}$ classes. Then we show similarly as before that every set Z_j induces an independent set in G and the induced bipartite graphs $G[U_i, Z_j]$ and $G[Z_j, Z_k]$ are either complete or empty. In other words, G is a blow-up of a graph H with at most $2^{|X|} + 2^{2^{|X|}}$ vertices. Since G is K_r-free the graph H must be also K_r-free and the theorem follows.

References

[1] P. Allen, J. Böttcher, S. Griffiths, Y. Kohayakawa, and R. Morris, *The chromatic thresholds of graphs*, manuscript.

[2] B. Andrásfai, *Über ein Extremalproblem der Graphentheorie*, Acta Math. Acad. Sci. Hungar. **13** (1962), 443–455.

[3] St. Brandt and St. Thomassé, *Dense triangle-free graphs are four-colorable: A solution to the Erdős-Simonovits problem*, J. Combin. Theory Ser. B, to appear.

[4] P. Erdős and M. Simonovits, *On a valence problem in extremal graph theory*, Discrete Math. **5** (1973), 323–334.

[5] P. Erdős and A. H. Stone, *On the structure of linear graphs*, Bull. Amer. Math. Soc. **52** (1946), 1087–1091.

[6] W. Goddard and J. Lyle, *Dense graphs with small clique number*, J. Graph Theory **66** (2011), no. 4, 319–331.

[7] R. Häggkvist, *Odd cycles of specified length in nonbipartite graphs*, Graph theory (Cambridge, 1981), North-Holland Math. Stud., vol. 62, North-Holland, Amsterdam, 1982, pp. 89–99.

[8] T. Łuczak, *On the structure of triangle-free graphs of large minimum degree*, Combinatorica **26** (2006), no. 4, 489–493.

[9] V. Nikiforov, *Chromatic number and minimum degree of K_r-free graphs*, submitted.

[10] C. Thomassen, *On the chromatic number of triangle-free graphs of large minimum degree*, Combinatorica **22** (2002), no. 4, 591–596.

[11] ———, *On the chromatic number of pentagon-free graphs of large minimum degree*, Combinatorica **27** (2007), no. 2, 241–243.

[12] P. Turán, *Eine Extremalaufgabe aus der Graphentheorie*, Mat. Fiz. Lapok **48** (1941), 436–452.

Reporter: Karen Gunderson

Participants

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

Dr. Paul Balister
Department of Mathematical Sciences
The University of Memphis
345 Dunn Hall
Memphis TN 38152-3240
UNITED STATES

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

Prof. Dr. Bela Bollobas
Department of Pure Mathematics
University of Cambridge
Wilberforce Road
Cambridge CB3 0WB
UNITED KINGDOM

Prof. Dr. Amin Coja-Oghlan
Institut für Mathematik
Goethe-Universität Frankfurt
Robert-Mayer-Str. 6-10
60325 Frankfurt am Main
GERMANY

Dr. David Conlon
Wadham College
Parks Road
Oxford OX1 3PN
UNITED KINGDOM

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

Prof. Dr. Hugo Duminil-Copin
Département de Mathématiques
Université de Geneve
Case Postale 64
2-4 rue du Lievre
1211 Geneve 4
SWITZERLAND

Dr. Jacob Fox
Department of Mathematics
Massachusetts Institute of
Technology
77 Massachusetts Avenue
Cambridge, MA 02139-4307
UNITED STATES

Prof. Dr. Ehud Friedgut
Department of Mathematics
The Weizmann Institute of Science
P. O. Box 26
Rehovot 76 100
ISRAEL

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

Prof. Dr. Zoltan Furedi
Department of Mathematics
University of Illinois at
Urbana-Champaign
1409 West Green Street
Urbana IL 61801
UNITED STATES

Prof. Dr. David Gamarnik
Massachusetts Institute of Technology
Sloan School of Management
E 53 - BS 7
77 Massachusetts Avenue
Cambridge MA 02139-4307
UNITED STATES

Heidi Gebauer
Department of Computer Science
ETH Zürich
Universitätstr. 6
8092 Zürich
SWITZERLAND

Dr. Karen Gunderson
Department of Mathematics
University of Bristol
University Walk
Bristol BS8 1TW
UNITED KINGDOM

Prof. Dr. Svante Janson
Matematiska Institutionen
Uppsala Universitet
Box 480
751 06 Uppsala
SWEDEN

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

Prof. Dr. Jeff Kahn
Department of Mathematics
Rutgers University
Piscataway NJ 08854-8019
UNITED STATES

Prof. Dr. Mihyun Kang
Institut für Optimierung und Diskrete Mathematik
Technische Universität Graz
Steyrergasse 30
8010 Graz
AUSTRIA

Prof. Dr. Ravindran Kannan
Microsoft Research
SVC-6
1288 Pear Avenue
Mountain View, CA 94043
UNITED STATES

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

Prof. Dr. Jeong Han Kim
School of Computational Sciences
Korea Institute for Advanced Study (KIAS)
85 Hoegiro, Dongdaemun-gu
Seoul 130-722
KOREA, REPUBLIC OF

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

Choongbum Lee
Department of Mathematics
University of California at
Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90095-1555
UNITED STATES

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

Prof. Dr. Tomasz Luczak
Zaklad Matematyki Dyskretnej
Wydzial Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza
61-614 Poznan
POLAND

Prof. Dr. Robert Morris
IMPA
Instituto Matematica Pura e Aplicada
Estrada Dona Castorina, 110
Jardim Botanico
Rio de Janeiro 22460-320
BRAZIL

Prof. Dr. Jaroslav Nesetril
Department of Applied Mathematics
Charles University
Malostranske nam. 25
118 00 Praha 1
CZECH REPUBLIC

Prof. Dr. Marc Noy
Departament de Matematica Aplicada II
Universitat Politecnica de Catalunya
c/Jordi Girona 1-3
08034 Barcelona
SPAIN

Prof. Dr. Janos Pach
École Polytechnique Fédérale de
Lausanne
SB MATHGEOM
Station 8
1015 Lausanne
SWITZERLAND

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

Prof. Dr. Oliver M. Riordan
Mathematical Institute
Oxford University
24-29 St. Giles
Oxford OX1 3LB
UNITED KINGDOM

Dr. Wojciech Samotij
Trinity College
Cambridge CB2 1TQ
UNITED KINGDOM

Prof. Dr. Mathias Schacht
Fachbereich Mathematik
Universität Hamburg
Bundesstr. 55
20146 Hamburg
GERMANY

Prof. Dr. Uwe Schöning
Institut für Theoretische Informatik
Fak. für Ingenieurwissenschaften & Inf.
Universität Ulm
89069 Ulm
GERMANY

Prof. Dr. Alex Scott
Merton College
Oxford University
Oxford OX1 4JD
UNITED KINGDOM

Prof. Dr. Asaf Shapira
Department of Mathematics
School of Mathematical Sciences
Tel Aviv University
Ramat Aviv, Tel Aviv 69978
ISRAEL

Paul Smith
Instituto de Matematica Pura e Aplicada - IMPA
Jardim Botanico
Estrada Dona Castorina, 110
22460 Rio de Janeiro, RJ 320
BRAZIL

Prof. Dr. Christian Sohler
Heinz-Nixdorf Institut &
Institut für Informatik
Universität Paderborn
Fürstenallee 11
33102 Paderborn
GERMANY

Prof. Dr. Jozsef Solymosi
Department of Mathematics
University of British Columbia
1984 Mathematics Road
Vancouver BC V6T 1Z2
CANADA

Prof. Dr. Gregory B. Sorkin
Department of Mathematics
London School of Economics
Houghton Street
London WC2A 2AE
UNITED KINGDOM

Prof. Dr. Vera T. Sós
Alfred Renyi Institute of Mathematics
Hungarian Academy of Sciences
P.O.Box 127
1364 Budapest
HUNGARY

Prof. Dr. Angelika Steger
Institut für Theoretische Informatik
ETH Zürich
ETH Zentrum
8092 Zürich
SWITZERLAND

Prof. Dr. Benjamin Sudakov
UCLA
Department of Mathematics
Los Angeles, CA 90095-1555
UNITED STATES

Prof. Dr. Tibor Szabo
Fachbereich Mathematik & Informatik
Freie Universität Berlin
Arnimallee 6
14195 Berlin
GERMANY

Prof. Dr. Gabor Tardos
Alfred Renyi Institute of Mathematics
Hungarian Academy of Sciences
Realtanoda u. 13-15
1053 Budapest
HUNGARY

Prof. Dr. Prasad Tetali
School of Mathematics
Georgia Institute of Technology
686 Cherry Street
Atlanta, GA 30332-0160
UNITED STATES

Prof. Dr. Andrew Thomason
Dept. of Pure Mathematics and
Mathematical Statistics
University of Cambridge
Wilberforce Road
Cambridge CB3 0WB
UNITED KINGDOM

Dr. Lutz Warnke
Dept. of Pure Mathematics and
Mathematical Statistics
University of Cambridge
Wilberforce Road
Cambridge CB3 0WB
UNITED KINGDOM

Prof. Dr. Emo Welzl
Theoretische Informatik
ETH Zürich
8092 Zürich
SWITZERLAND

Yufei Zhao
Department of Mathematics
MIT
77 Massachusetts Ave.
Cambridge, MA 02139
UNITED STATES