Mathematisches Forschungsinstitut Oberwolfach

Report No. 23/2009

DOI: 10.4171/OWR/2009/23

Combinatorics and Probability

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

April 26th – May 2nd, 2009

Abstract. The effective application of probabilistic reasoning in the study of problems in diverse areas is one of the most exciting recent developments in Mathematics. Probabilistic methods turned out to be very powerful in Discrete Mathematics, Analysis, Number Theory and Theoretical Computer Science. The meeting was dedicated to recent developments in these areas, focusing on the investigation of combinatorial problems for random sets and probabilistic methods, on the study of of questions in percolation, on the design and analysis of randomized algorithms and derandomization techniques, and on applications of probabilistic ideas in the study of questions in Combinatorial Number Theory and in Combinatorial Geometry.

Mathematics Subject Classification (2000): 05C35, 05C80, 05D05, 05D10, 05D40, 60C05, 68Q25.

Introduction by the Organisers

The conference was organized by Noga Alon (Tel Aviv), Béla Bollobás (Cambridge and Memphis) and Ingo Wegener (Dortmund). The programme consisted of 12 main lectures, supplemented by 17 shorter contributions, and covered many areas in Extremal and Probabilistic Combinatorics as well as in Theoretical Computer Science.

A few months before the meeting, we were devastated to hear that the third organizer of this conference, Professor Ingo Wegener, died in November 26, 2008. Ingo was the friend and colleague of many of us, and his spirit was with us during the meeting. The meeting started with a tribute to his memory and one of the technical lectures focused on recent developments in the investigation of a problem he raised in the previous Oberwolfach meeting we organized with him in 2006.

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

The speakers reported on recent developments in Ramsey theory, including variants that deal with Random Graphs and random structures, on new results in Combinatorial Geometry and on advances in the study of colouring problems for graphs and hypergraphs. Novel results in percolation, extremal graph theory and additive combinatorics have been described as well, combining combinatorial, probabilistic and analytic ideas. Additional active topics discussed included results on hashing, a new efficient algorithm for the local lemma, better algorithms for the random k-set problem, new developments on random walks in random graphs, and a discussion of reachability games.

The workshop focused on the connection and common themes of Combinatorics, Discrete Probability and Theoretical Computer Science, and the lectures, many of which given by young participants, stimulated fruitful discussions. The fact that the participants work in different and yet related topics, and the open problems session held during the meeting, encouraged interesting discussions and collaborations.

Forty nine scientists, including forty from countries other than Germany participated in the meeting. The organizers and participants thank the Mathematisches Forschungsinstitut Oberwolfach for providing an inspiring setting for this conference.

Workshop: Combinatorics and Probability

Table of Contents

Abstracts

Structure and Cardinality of H-free graphs

József Balogh

A graph is called *H-free* if it contains no copy of H. Denote by $f_n(H)$ the number of (labeled) H-free graphs on n vertices. As a natural extension of the Erdős-Stone theorem [13], Erdős conjectured that

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

when H contains a cycle. The lower bound is trivial, as all subgraphs of an extremal H-free graph are H-free.

The conjecture was first shown to be true for cliques by Erdős, Kleitman and Rothschild [12] and later Erdős, Frankl and Rödl [11] proved it for all graphs H with $\chi(H) \geq 3$. We considered three different ways of extending their results:

- Finding more precise estimates for $f_n(H)$, and describing the structure of almost all H-free graphs.
- Proving similar results when H is forbidden as an induced subgraph.
- Trying to prove the conjecture when H is a bipartite graph.

The structure of almost all H-free graphs. In order to state our results, we need some definitions. Denote by I_ν the ν-vertex graph with no edges. Given a graph H, let the *decomposition family* of H, $\mathcal{M} := \mathcal{M}(H)$, be the family of minimal graphs M for which there exist a $t = t_H$ such that $H \subseteq M' \otimes K_{p-1}(t, \ldots, t)$, where $M' = M'(t)$ is the graph obtained by adding t isolated vertices to M. We say that a graph H is *weakly critical* if there is an edge $e \in E(H)$ for which $\chi(H-e) < \chi(H)$. The main result of Balogh, Bollobás and Simonovits [3] is the following.

Theorem 1. *Let H be a graph with chromatic number $r+1 = \chi(H)$. Then there is a constant $c = c(H)$, such that the following holds:*
For almost all H-free graphs G there exists a partition $(A, S_1, \ldots, S_r)$ of $V(G)$, such that

(a) $|A| \leq c$,

(b) $G[S_j]$ is $\mathcal{M}(H)$-free for every $j \in [r]$.

In particular,

$$2^{(1-1/r)\binom{n}{2}} n^{\mathrm{ex}(n/r,\mathcal{M})} \;\leq\; |f_n(H)| \;\leq\; 2^{(1-1/r)\binom{n}{2}} n^{\mathrm{ex}(n,\mathcal{M})+cn}.$$

In [4], Balogh, Bollobás and Simonovits proved more precise results for certain families of graphs. Among others the following was proved:

Theorem 2. *Let r and s be positive integers and H be a weakly critical graph with chromatic number $r+1$. Then almost every sH-free graph G_n on n vertices has a set S of $s-1$ vertices for which $\chi(G_n - S) = r$, where sH is the graph consisting of s vertex disjoint copies of H.*

Theorem 3. *Denote by O_6 the octahedron graph (which is the same as $K(2,2,2)$). The vertices of almost every O_6-free graph can be partitioned into two classes, U_1 and U_2, so that U_1 spans a C_4-free graph and U_2 spans a P_3-free graph.*

Note that Theorem 2 was proved by Prömel and Steger [19] for $s = 1$.

The number of H-free graphs when H is bipartite. For most bipartite H, Conjecture (1) is still wide open, and even the correct order of magnitude of $\log_2 f_n(H)$ is not known. The only nontrivial bipartite graphs, for which an estimate stronger than the trivial bound is known, are cycles. Kleitman and Winston [16] proved that $\log_2 f_n(C_4) \leq 2.16384 \cdot \mathrm{ex}(n, C_4)$, and later Kleitman and Wilson [17] proved $\log_2 f_n(C_6) = \Theta(\mathrm{ex}(n, C_6))$. For a little stronger estimates of the number of graphs with large (even) girth, i.e., graphs with no short (even) cycles, see [17, 18].

Balogh and Samotij [7] and [8] proved that for every $2 \leq s \leq t$, $f_n(K_{s,t}) = 2^{O(n^{2-1/s})}$. This bound is asymptotically sharp for all pairs (s,t) for which the extremal number of $K_{s,t}$ is known. The methods also yield a bound on the number of $K_{s,t}$-free graphs with fixed order and size, extending the result of Füredi [14]. Using this bound, a relaxed version of a conjecture of Haxell, Kohayakawa and Łuczak [15] is proved, and among others, it is showed that almost all $K_{3,3}$-free graphs of order n have more than $1/20 \cdot \mathrm{ex}(n, K_{3,3})$ edges.

The structure of almost all $\mathcal{H}$-free graphs: the induced case. Note that in the induced case it is better to consider the case when not a single graph H but a family of graphs $\mathcal{H}$ is forbidden. Let $f_n^i(\mathcal{H})$ denote the number of graphs with vertex set $[n]$ containing no $H \in \mathcal{H}$ as an induced subgraph. We have to define the *coloring number* of $\mathcal{H}$, $\chi_C(\mathcal{H})$, to be the maximum number $r+1$ for which there exist s and t, with $s + t = r$, such that no G whose vertex set can be partitioned into $U_1, \ldots, U_s, W_1, \ldots, W_t$, where for every i, $G[U_i]$ is a clique and $G[W_i]$ is an independent set, contains any $H \in \mathcal{H}$ as an induced subgraph.

The following result, proved by Alekseev [1], Bollobás and Thomason [9, 10], and Prömel and Steger [20], generalizes the Erdős-Frankl-Rödl theorem.

Theorem 4. *Let $\mathcal{H}$ be a family of graphs, and suppose $\chi_c(\mathcal{H}) = r + 1$. Then*

$$f_n^i(\mathcal{H}) \;=\; 2^{(1-1/r+o(1))n^2/2}.$$

However, their proofs tell us very little about the structure of a *typical* $\mathcal{H}$-free graph G. Their theorem also gives rather weak bounds on the rate of convergence of the entropy as $n \to \infty$. Some fine structural results for the case $r = 1$ were given in [5] and [6].

For each $k \in \mathbb{N}$, the *universal graph* $U(k)$ is the bipartite graph with parts $A \cong [2]^k$ and $B \cong [k]$, and edge set

$$E\big(U(k)\big) \;=\; \big\{ab \,:\, a \in A,\, b \in B \text{ and } b \in a\big\}.$$

A graph G is said to be $U(k)$-*free* if there do not exist disjoint subsets $A, B \subset V(G)$ such that $G[A,B] = U(k)$. The main result of Alon, Balogh, Bollobás and Morris [2] is the following:

Theorem 5. *Let $\mathcal{H}$ be a family of graphs, with colouring number $\chi_c(\mathcal{H}) = r+1$. Then there exist constants $k = k(\mathcal{H}) \in \mathbb{N}$ and $\varepsilon = \varepsilon(\mathcal{H}) > 0$ such that the following holds.*

For almost every $\mathcal{H}$-free graph G there exists a partition $(A, S_1, \ldots, S_r)$ of $V(G)$, such that

(a) $|A| \le n^{1-\varepsilon}$,

(b) $G[S_j]$ *is $U(k)$-free for every $j \in [r]$.*

Moreover

$$2^{(1-1/r)n^2/2} \;\le\; |f_n^i(\mathcal{H})| \;\le\; 2^{(1-1/r)n^2/2 + n^{2-\varepsilon}} \tag{2}$$

for every sufficiently large $n \in \mathbb{N}$.

Note that for $r = 1$ the relation (2) is not far from being best possible.

References

[1] V. E. Alekseev, On the entropy values of hereditary classes of graphs, *Discrete Math. Appl.* **3** (1993), 191–199.

[2] N. Alon, J. Balogh, B. Bollobás and R. Morris, The structure of almost all graphs in a hereditary property, submitted.

[3] J. Balogh, B. Bollobás and M. Simonovits, The typical structure of graphs without given excluded subgraphs, to appear in *Random Structures and Algorithms*.

[4] J. Balogh, B. Bollobás and M. Simonovits, The fine structure of octahedron-free graphs, submitted.

[5] J. Balogh, B. Bollobás, M. Saks and V. T. Sós, On the diversity function of a hereditary graph property, *Journal of Combinatorial Theory A* **99** (2009), 9–19.

[6] J. Balogh, B. Bollobás and D. Weinreich, The penultimate rate of growth for graph properties, *European Journal of Combinatorics* **22** (2001), 277–289.

[7] J. Balogh and W. Samotij, The number of $K_{m,m}$-free graphs, submitted.

[8] J. Balogh and W. Samotij, The number of $K_{s,t}$-free graphs, in preparation.

[9] B. Bollobás and A. Thomason, Projections of bodies and hereditary properties of hypergraphs, *Bull. London Math. Soc.* **27** (1995), 417–424.

[10] B. Bollobás and A. Thomason, Hereditary and monotone properties of graphs, "The mathematics of Paul Erdős, II" (R.L. Graham and J. Nešetřil, Editors), *Alg. and Combin.* Vol. 14, Springer-Verlag, New York/Berlin (1997), 70–78.

[11] 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 and Combin.* **2** (1986), 113–121.

[12] P. Erdős, D. J. Kleitman and B. L. Rothschild, Asymptotic enumeration of K_n-free graphs, in *Colloquio Internazionale sulle Teorie Combinatorie* (Rome, 1973), Vol. II, pp. 19–27. *Atti dei Convegni Lincei* **17**, Accad. Naz. Lincei, Rome, 1976.

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

[14] Z. Füredi, Random Ramsey graphs for the four-cycle, *Discrete Mathematics* **126** (1994), 407–410.

[15] P. Haxell, Y. Kohayakawa and T. Łuczak, Turán's Extremal Problem in Random Graphs: Forbidding Even Cycles, *Journal of Combinatorial Theory B* **64** (1995), 273–287.

[16] D. Kleitman and K. Winston, On the number of graphs without 4-cycles, *Discrete Mathematics* **41** (1982), 167–172.

[17] D. Kleitman and D. B. Wilson, On the number of graphs which lack small cycles, manuscript (1996).

[18] Y. Kohayakawa, B. Kreuter and A. Steger, An extremal problem for random graphs and the number of graphs with large even-girth, *Combinatorica* **18** (1998), 101–120.
[19] H. J. Prömel and A. Steger, The asymptotic number of graphs not containing a fixed color-critical subgraph, *Combinatorica* **12** (1992), 463–473.
[20] H. J. Prömel and A. Steger, Excluding induced subgraphs III., A general asymptotic, *Random Structures and Algorithms* **3** (1992), 19–31.

The early evolution of the H-free process

Tom Bohman
(joint work with Peter Keevash)

In this talk we sketch an analysis of a significant portion of the initial evolution of the H-free process, for some fixed graph H, defined by starting with an empty graph on n vertices and then adding edges one at a time, chosen uniformly at random subject to the constraint that no H subgraph is formed. More formally, we begin with the graph on n vertices with no edges, which we denote $G(0)$. Now suppose $i > 0$ and we have some graph $G(i-1)$. We say that a pair uv of vertices is *open* in $G(i-1)$ if uv is not an edge of $G(i-1)$ and $G(i-1) \cup \{uv\}$ does not contain H as a subgraph. We choose uv uniformly at random among all open pairs in $G(i-1)$ and then $G(i)$ is obtained from $G(i-1)$ by adding the edge $e_i = uv$. The process terminates when there are no open pairs, with some graph G on n vertices that is a maximal H-free graph. Beside being of interest in its own right, our analysis of this process produces new results in Ramsey theory and the theory of Turán problems.

The study of such process was initiated in the late 1980's by Erdős and others. Ruciński and Wormald [9] introduced a differential equations method to analyze the 'maximum degree d' process and thereby resolve a conjecture of Erdős. Erdős, Suen and Winkler [6] obtained results on the triangle-free process and the bipartite process. Their analysis of the triangle-free process led to the best lower bound on the Ramsey number $R(3,t)$ known at that time. The upper bound $R(3,t) = O(t^2/\log t)$ had been obtained by Ajtai, Komlós and Szemerédi [1], but for many years the best known lower bound, due to Erdős [5], was $\Omega(t^2/\log^2 t)$. Spencer conjectured that the triangle-free process is likely to produce a graph that establishes a good lower bound on $R(3,t)$ for t large; the idea being that the triangle-free process admits enough random edges to bring the independence number close to the smallest possible for a triangle-free graph. Finally, Kim [7] determined the order of magnitude, showing that $R(3,t) = \Theta(t^2/\log t)$. His proof made use of a semi-random construction that is motivated (even guided) by the triangle-free process, but the question remained open as to whether the triangle-free process itself gives such a good construction. This was answered by Bohman [3], who showed that with high probability, the graph produced by the

Research by Tom Bohman partially supported by NSF grant DMS-0701183. Research by Peter Keevash partially supported by NSF grant DMS-0555755.

triangle-free process has independence number bounded above by $O(n^{1/2}\log^{1/2} n)$ and minimum degree bounded below by $\Omega(n^{1/2}\log^{1/2} n)$.

The general H-free process was independently studied by Osthus and Taraz [8] and by Bollobás and Riordan [4]. Say that a graph H is *strictly* 2*-balanced* if the number of vertices v_H and edges e_H in H are both at least 3 and

$$\frac{e_H - 1}{v_H - 2} > \frac{e_K - 1}{v_K - 2}$$

for all proper subgraphs K of H with $v_K \geq 3$. Osthus and Taraz showed that if H is strictly 2-balanced then for some $c, C > 0$ with high probability, for the H-free process G has average degree at least $cn^{1-(v_H-2)/(e_H-1)}$ and maximum degree at most $Cn^{1-(v_H-2)/(e_H-1)}(\log n)^{1/(\Delta(H)-1)}$. (In fact they proved the average degree bound under a similar but weaker condition on H.) Wolfovitz [10] showed that if H is strictly 2-balanced and regular then the expected number of edges in G is at least $cn^{2-(v_H-2)/(e_H-1)}(\log\log n)^{1/(e_H-1)}$.

Our first main result gives a lower bound on the number of steps in the H-free process.

Theorem 1. *Suppose that H is a strictly 2-balanced graph with v_H vertices and e_H edges. Then for some $c > 0$ with high probability the minimum degree in the final graph of the H-free process is at least $cn^{1-(v_H-2)/(e_H-1)}(\log n)^{1/(e_H-1)}$. In particular, the Turán number satisfies*

$$\mathrm{ex}(n, H) = \Omega\left(n^{2-(v_H-2)/(e_H-1)}(\log n)^{1/(e_H-1)}\right).$$

We also get an upper bound on the independence number of the graph produced by the H-free process for many choices of H. One consequence is the following new lower bound on the off-diagonal Ramsey numbers.

Theorem 2. *For fixed $s \geq 3$ and $t \to \infty$, the Ramsey number satisfies*

$$R(s, t) = \Omega\left(t^{\frac{s+1}{2}}(\log t)^{\frac{1}{s-2}-\frac{s+1}{2}}\right).$$

Alon, Ben-Shimon and Krivelevich recently gave a construction that, given the graph produced by the K_s-free process as input, produces a *regular* graph that achieves the bound in Theorem 2.

References

[1] M. Ajtai, J. Komlós and E. Szemerédi, A note on Ramsey numbers, *Journal of Combinatorial Theory A* **29** (1980), 354–360.

[2] N. Alon, S. Ben-Shimon and M. Krivelevich, A note on regular Ramsey graphs, `arXiv:0812.2386v1`.

[3] T. Bohman, The triangle-free process, *Advances in Mathematics* **221** (2009), 1653–1677.

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

[5] P. Erdős, Graph theory and probability II, *Canad. J. Math.* **13** (1961), 346–352.

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

[7] J. H. Kim, The Ramsey number $R(3,t)$ has order of magnitude $t^2/\log t$, *Random Structures and Algorithms* **7** (1995), 173–207.
[8] D. Osthus and A. Taraz, Random maximal H-free graphs, *Random Structures and Algorithms* **18** (2001), 61–82.
[9] A. Ruciński and N. Wormald, Random graph processes with degree restrictions, *Combinatorics, Probability and Computing* **1** (1992), 169–180.
[10] 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.

Geometric selection theorems

Boris Bukh
(joint work with Jiří Matoušek and Gabriel Nivasch)

There are several problems in geometric combinatorics that involve approximation of a large point set by a handful of points. The first result of the kind is due to Rado:

Theorem 1 ([7])**.** *For every finite set $X \subset \mathbb{R}^d$ there is a point $p \in \mathbb{R}^d$ such whenever a convex set C does not contain p, C contains at most $\frac{d}{d+1}|X|$ points of X.*

The fraction $\frac{d}{d+1}$ is sharp as witnessed by the vertices of a regular simplex.

More generally, let $X \subset \mathbb{R}^d$ be a finite set, and $\mathcal{F}$ be a family of subsets of $\mathbb{R}^d$. A set $N \subseteq \mathbb{R}^d$ is called a *weak ε-net* with respect to $\mathcal{F}$, where $\varepsilon \in (0,1]$ is a real number, if N intersects every $C \in \mathcal{F}$ with $|X \cap C| \geq \varepsilon |X|$. Of special importance are weak ε-nets with respect to convex sets, which we call simply weak ε-nets. Rado's says that there is a always 1-point weak $\frac{d}{d+1}$-net. Let $f(X,r)$ denote the minimum cardinality of a weak $\frac{1}{r}$-net for X. It is a nontrivial fact, first proved by Alon, Bárány, Füredi, and Kleitman [1], that

$$f(d,r) := \sup\{f(X,r) : X \subset \mathbb{R}^d \text{ finite}\}$$

is finite for every $d \geq 1$ and every $r \geq 1$; that is, for every set X there exist weak ε-nets of size bounded solely in terms of d and ε. The best known upper bound is $f(d,r) = O(r^d \log^{c(d)} r)$.

Consider another approximation problem. A n-point set $X \subset \mathbb{R}^d$ in general position spans the family $\mathcal{S}(X)$ of $\binom{n}{d+1}$ $d+1$-dimensional simplices. Let $g(X) = \frac{1}{\binom{n}{d+1}} \max_{p \in P} \#\{\Delta \in \mathcal{S}(X) : p \in \Delta\}$, and put

$$g(d) = \inf\{g(X) : X \subset \mathbb{R}^d \text{ finite}\}.$$

In a rather surprising result Bárány [2] showed that $g(d)$ is positive. The best lower known bounds are $g(2) \geq 2/9$ due to Boros and Füredi [3] and $g(d) \geq \frac{d^2+1}{(d+1)^{d+1}}$ due to Wagner [8].

There is also a sparse version of the previous problem. For simplicity we consider only dimension $d = 2$. Let $X \subset \mathbb{R}^d$ be an n-point set in general position, but

instead of considering all $\binom{n}{3}$ triangles spanned by X, let $\mathcal{F}$ be a family of $\alpha\binom{n}{3}$ triangles among the triangles spanned by X. Nivasch and Sharir [6] (fixing a proof of Eppstein [5]) showed that in this situation there is always a point in approximately $c\alpha^3\binom{n}{3}$ triangles of $\mathcal{F}$.

In these three approximation problems there have been no known non-trivial constructions known. All the known bounds applied to arbitrary points sets, and the proofs showed that no large point set can be approximated too well by a small point set.

We introduce a class of construction, which establishes the first non-trivial lower bounds for these three problems:

Theorem 2.

(1) *For every d and r there is an $X \subset \mathbb{R}^d$ for which there is no weak $1/r$-net of size $c_d r(\log r)^{d-1}$, i.e., $f(r) = \Omega\big(r(\log r)^{d-1}\big)$.*

(2) *For every d and r there is an $X \subset \mathbb{R}^d$ for which there is no point in $\frac{(d+1)!}{(d+1)^{d+1}}\binom{n}{d+1}$ simplices spanned by X, i.e., $g(d) \le \frac{(d+1)!}{(d+1)^{d+1}}$.*

(3) *For every $\alpha < 1$ there is an n-point set $X \subset \mathbb{R}^d$ in general position and a family $\mathcal{F}$ of $\alpha\binom{n}{3}$ triangles spanned by X such that no point is in more than $\frac{\alpha^2}{\log(1/\alpha)}\binom{n}{3}$ triangles of $\mathcal{F}$.*

The best previously known bounds were $f(d,r) \ge 1/r$, $g(d,r) \le 2^{-d}$ and $\alpha^2\binom{n}{3}$ respectively.

The basic element of the construction is a highly stretched grid $\{x_1,\dots,x_m\} \times \{y_1,\dots,y_m\}$ where the real numbers $x_1,\dots,x_m,y_1,\dots,y_m$ satisfy $1 \ll x_1 \ll \cdots \ll x_m \ll y_1 \ll \cdots y_m$ and $A \ll B$ denotes $\exp(A) \le B$ (any other sufficiently quickly-growing function in place of exp would also work). The crucial property is that the intersections of convex sets with the stretched grid are well approximated by combinatorial objects, called stair-convex sets. Stair-convex sets share most properties of convex sets, but their definition involves no arithmetic operations, which makes them much easier to work with. For full details the reader is referred to [4].

References

[1] N. Alon, I. Bárány, Z. Füredi and D. J. Kleitman, Point selections and weak ε-nets for convex hulls, *Combinatorics, Probability and Computing* **1** (1992), 189–200.

[2] Imre Bárány, A generalization of Carathéodory's theorem, *Discrete Math.* **40** (1982), 141–152.

[3] Endre Boros and Zoltán Füredi, The number of triangles covering the center of an n-set, *Geom. Dedicata* **17** (1984) 69–77.

[4] Boris Bukh, Jiří Matoušek and Gabriel Nivasch, Lower bounds for weak epsilon-nets and stair-convexity, `arXiv:0812.5039` .

[5] David Eppstein, Improved bounds for intersecting triangles and halving planes, *Journal of Combinatorial Theory A* **62** (1993), 176–182. `http://www.ics.uci.edu/~eppstein/pubs/Epp-TR-91-60.pdf` .

[6] Gabriel Nivasch and Micha Sharir, Eppstein's bound on intersecting triangles revisited, *Journal of Combinatorial Theory A*, to appear. `arXiv:0804.4415` .

[7] R. Rado, A theorem on general measure, *J. London Math. Soc.* **21** (1947), 291–300.
[8] Ulrich Wagner. *On k-Sets and Applications*, PhD thesis, ETH Zürich, June 2003, `http://www.inf.ethz.ch/~emo/DoctThesisFiles/wagner03.pdf` .

A Better Algorithm for Random k-SAT

Amin Coja-Oghlan

The k-SAT problem is well known to be NP-hard for $k \geq 3$. But this merely indicates that no algorithm can solve *all* possible inputs efficiently. Therefore, there has been a significant amount of research on *heuristics* for k-SAT, i.e., algorithms that solve "most" inputs efficiently. While some heuristics for k-SAT are very sophisticated, virtually all of them are based on at least one of the following basic paradigms.

Pure literal rule. If a variable x occurs only positively (resp. negatively) in the formula, set it to true (resp. false). Simplify the formula by substituting the newly assigned value for x and repeat.

Unit clause propagation. If the formula contains a clause that consists of only a single literal ("unit clause"), then set the underlying variable so as to satisfy this clause. Simplify and repeat.

Walksat. Initially pick a random assignment. Then repeat the following. While there is an unsatisfied clause, pick one at random, pick a variable occurring in the chosen clause randomly, and flip its value.

Backtracking. Assign a variable x, simplify the formula, and recurse. If the recursion fails to find a satisfying assignment, assign x the opposite value and recurse.

Heuristics based on these paradigms can be surprisingly successful (given that k-SAT is NP-hard) on certain types of inputs, e.g., [7]. However, it remains remarkably simple to generate formulas that elude all known algorithms/heuristics. Indeed, the simplest conceivable type of *random* instances does the trick: let $\mathbf{\Phi}$ denote a k-SAT formula over the variable set $V = \{x_1, \ldots, x_n\}$ that is obtained by choosing m clauses uniformly at random and independently from the set of all $(2n)^k$ possible clauses. Then for a large regime of densities m/n satisfying assignments are known to exist due to non-constructive arguments, but no efficient algorithm is known to find one.

To be precise, keeping k fixed and letting $m = \lceil rn \rceil$ for a fixed $r > 0$, we say that $\mathbf{\Phi}$ has some property *with high probability* ("w.h.p.") if the probability of the property tends to one as $n \to \infty$. Via the (non-algorithmic) second moment method [1, 2] it can be shown that $\mathbf{\Phi}$ has a satisfying assignment w.h.p. if $m/n < (1 - \varepsilon_k)2^k \ln 2$. Here ε_k tends to 0 for large k. On the other hand, a very simple first moment argument shows that no satisfying assignment exists w.h.p. if $m/n > 2^k \ln 2$. In summary, the threshold for $\mathbf{\Phi}$ being satisfiable is asymptotically $2^k \ln 2$.

Yet for densities m/n beyond $O(2^k/k)$ no algorithm has been known to find a satisfying assignment in polynomial time with a probability that does not tend to

Algorithm	Density $m/n < \cdots$	Ref., year
Pure Literal	$o(1)$ as $k \to \infty$	[9], 2006
Walksat, rigorous	$\frac{1}{6} \cdot 2^k/k^2$	[6], 2009
Walksat, non-rigorous	$2^k/k$	[11], 2003
Unit Clause	$\frac{1}{2}\left(\frac{k-1}{k-2}\right)^{k-2} \cdot \frac{2^k}{k}$	[4], 1990
Shortest Clause	$\frac{1}{8}\left(\frac{k-1}{k-3}\right)^{k-3}\frac{k-1}{k-2} \cdot \frac{2^k}{k}$	[5], 1992
SC+backtracking ("SCB")	$\sim 1.817 \cdot \frac{2^k}{k}$	[8], 1996
BP+decimation, non-rigorous	$\mathrm{e} \cdot 2^k/k$	[10], 2007

TABLE 1. Algorithms for random k-SAT

zero (Table 1 summarizes previous results). In fact, many algorithms, including Pure Literal, Unit Clause, and DPLL-type algorithms, are known to fail or exhibit an exponential running time beyond $O(2^k/k)$. There is experimental evidence that the same is true of Walksat. Indeed, devising an algorithm to solve random formulas w.h.p. for densities m/n up to $2^k\omega(k)/k$ for *any* (howsoever slowly growing) $\omega(k) \to \infty$ has been a prominent open problem [1, 2, 5, 10], which the following theorem resolves.

Theorem 1. *There is a sequence $\varepsilon_k \to 0$ and a polynomial time algorithm* `Fix` *such that* `Fix` *applied to a random formula $\mathbf{\Phi}$ with $m/n \le (1-\varepsilon_k)2^k \ln(k)/k$ outputs a satisfying assignment w.h.p..*

`Fix` is a deterministic local search algorithm and runs in time $O(n+m)^{3/2}$.

REFERENCES

[1] D. Achlioptas, C. Moore, Random k-SAT: two moments suffice to cross a sharp threshold, *SIAM Journal on Computing* **36** (2006), 740–762.

[2] D. Achlioptas, Y. Peres, The threshold for random k-SAT is $2^k \ln 2 - O(k)$, *Journal of the AMS* **17** (2004), 947–973.

[3] A. Braunstein, M. Mézard, R. Zecchina, Survey propagation: an algorithm for satisfiability, *Random Structures and Algorithms* **27** (2005), 201–226.

[4] M.-T. Chao, J. Franco, Probabilistic analysis of a generalization of the unit-clause literal selection heuristic for the k-satisfiability problem, *Inform. Sci.* **51** (1990), 289–314.

[5] V. Chvátal, B. Reed, Mick gets some (the odds are on his side), *Proc. 33th FOCS* (1992), 620–627.

[6] A. Coja-Oghlan, U. Feige, A. Frieze, M. Krivelevich, D. Vilenchik, On smoothed k-CNF formulas and the Walksat algorithm, *Proc. 20th SODA* (2009), 451–460.

[7] M. Davis, G. Longemann, D. Loveland, A machine program for theorem proving, *Communications of the ACM* **5** (1962), 394–397.

[8] A. Frieze, S. Suen, Analysis of two simple heuristics on a random instance of k-SAT, *Journal of Algorithms* **20** (1996), 312–355.

[9] J. H. Kim, Poisson cloning model for random graph, Preprint (2006).

[10] A. Montanari, F. Ricci-Tersenghi, G. Semerjian, Solving constraint satisfaction problems through Belief Propagation-guided decimation, *Proc. 45th Allerton* (2007).

[11] G. Semerjian, R. Monasson, A study of pure random walk on random satisfiability problems with "physical" methods, *Proc. 6th SAT* (2003), 120–134.

Combinatorial theorems in sparse random sets

David Conlon
(joint work with W. T. Gowers)

Szemerédi's theorem [8] states that for any $k \in \mathbb{N}$ and any $\delta > 0$ there exists n_0 such that if $n \geq n_0$ then any subset of the set $\{1, 2, \ldots, n\}$ of size at least δn contains an arithmetic progression of length k. A natural extension of this definition is to say that a subset I of the integers is (k, δ)-Szemerédi if, for all subsets $J \subset I$ of size $|J| \geq \delta |I|$, J contains an arithmetic progression of length k. For example, the celebrated theorem of Green and Tao [2] says that for any $k \in \mathbb{N}$ and any $\delta > 0$ there is n_0 such that, for any $n \geq n_0$, the set of primes between 1 and n is (k, δ)-Szemerédi.

Given $0 \leq p \leq 1$, define $[n]_p$ to be a random subset of $\{1, 2, \ldots, n\}$ where each integer i is chosen with probability p. Our concern is with the following question: when is the random set $[n]_p$ almost surely (k, δ)-Szemerédi? One simple observation is that when the number of k-term arithmetic progressions in a set is much less than the number of points in the set, the set cannot be (k, δ)-Szemerédi for small δ. In that case we can just remove a point from each of the progressions leaving some points but no progressions. In the random set $[n]_p$, the expected number of k-term arithmetic progressions is at most $n^2 p^k$, while the expected number of points is np. Therefore, the set $[n]_p$ cannot be (k, δ)-Szemerédi if $n^2 p^k \leq (1-\delta)np$. Being a little more rigorous, it is easy to show that if $p \leq cn^{-1/(k-1)}$, $[n]_p$ is almost surely not (k, δ)-Szemerédi. Perhaps surprisingly, Kohayakawa, Łuczak and Rödl [3] showed that if $k = 3$, this is essentially best possible. We extend their result to all values of k.

Theorem 1. *For every $k \in \mathbb{N}$ and $\delta > 0$ there exist constants c and C such that*

$$\lim_{n\to\infty} \mathbb{P}\big[[n]_p \text{ is } (k,\delta)\text{-Szemerédi}\big] = \begin{cases} 0, & \text{if } p < cn^{-1/(k-1)}, \\ 1, & \text{if } p > Cn^{-1/(k-1)}. \end{cases}$$

There are several results and conjectures of this variety in the literature. For example, there is a beautiful theorem of Rödl and Ruciński [5, 6] determining the threshold down to which a random graph $G_{n,p}$ satisfies Ramsey's theorem for a given graph H. Another example is a conjecture of Kohayakawa, Łuczak and Rödl [4] stating a similar threshold for when $G_{n,p}$ almost surely satisfies a version of Turán's theorem. Though some partial results are known (see, for example, [1, 7]), the general conjecture has remained open. Our method seems to be very general and allows us, amongst other things, to reprove (and extend) the result of Rödl and Ruciński and settle the conjecture of Kohayakawa, Łuczak and Rödl.

References

[1] S. Gerke, A. Steger and T. Schickinger, K_5-free subgraphs of random graphs, *Random Structures and Algorithms* **24** (2004), 194–232.

[2] B. Green and T. Tao, The primes contain arbitrarily long arithmetic progressions, *Annals of Math.* **167** (2008), 481–547.

[3] Y. Kohayakawa, T. Łuczak and V. Rödl, Arithmetic progressions of length three in subsets of a random set, *Acta Arithmetica* **LXXV** (1996), 133–163.

[4] Y. Kohayakawa, T. Łuczak and V. Rödl, On K_4-free subgraphs of random graphs, *Combinatorica* **17** (1997), 173–213.

[5] V. Rödl and A. Ruciński, Lower bounds on probability thresholds for Ramsey properties, *Combinatorics, Paul Erdős is Eighty* (Vol.1), Keszthely (Hungary), Bolyai Soc. Math. Studies (1993), 317–346.

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

[7] T. Szabó and V. H. Vu, Turán's theorem in sparse random graphs, *Random Structures and Algorithms* **23** (2003), 225–234.

[8] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199–245.

Testing continuous distributions

Artur Czumaj

(joint work with Michał Adamaszek and Christian Sohler)

We study the task of testing properties of probability distributions and our focus is on understanding the role of continuous distributions in this setting. We consider a scenario in which we have access to independent samples of a distribution $\mathfrak{D}$ over a potentially continuous or uncountable domain (e.g., on the interval $[0,1]$ or on the n-dimensional continuous cube $[0,1]^n$). Our goal is to test whether $\mathfrak{D}$ has a given probability distribution or it is ε-far from it (in the statistical distance).

The topic of testing basic properties of the underlying probability distributions has been extensively studied for many decades. While the standard approach in statistics (and also more modern approaches, e.g., in data mining) have led to the development of many high quality techniques and algorithms, until very recently little attention has been paid to the computational complexity of testing in the situations when the underlying distributions are over very large domains. Motivated by these considerations, a number of new studies have emerged that aim at developing efficient testers for various properties of distributions with the focus on the small number of samples used for testing. In particular, it has been shown that for a number of fundamental properties, such as independence, entropy estimation, and the closeness between distributions, it is possible to test the underlying distribution with the number of samples sublinear in the domain size.

While these studies lead to very efficient testers for various properties for distributions on finite support, they seem to be useless when the underlying distribution is on a continuous or uncountable domain. In this paper, our goal is to study the phenomenon of testability of continuous distributions.

Setting. We assume that there is an underlying probability distribution $\mathfrak{D}$ from which we can draw *independent identically distributed samples* (see, e.g., [4]). We assume that each sample is of infinite precision and we will not consider the issue of representation of the real numbers. The *complexity of the tester* is measured in terms of the *number of samples* required in order to obtain a desired information about the distribution.

We study the similarity and dissimilarity between various probability distributions. We use the total variation distance to measure the similarity between probability distributions. For any two probability distributions $\mathcal{X}$ and $\mathcal{Y}$ over Ω, with density functions $\mathfrak{f}_{\mathcal{X}}$ and $\mathfrak{f}_{\mathcal{Y}}$, respectively, we say $\mathcal{Y}$ *is ε-far from* $\mathcal{X}$ if

$$\frac{1}{2} \cdot \int_{\mathbf{x}\in\Omega} |\mathfrak{f}_{\mathcal{X}}(\mathbf{x}) - \mathfrak{f}_{\mathcal{Y}}(\mathbf{x})| \, d\mathbf{x} \;\geq\; \varepsilon \,.$$

Our goal is to design an algorithm that for a given positive ε and a given underlying probability distribution $\mathfrak{Q}$ and a probability distribution $\mathfrak{D}$ available through random sampling, is able to distinguish between the case when $\mathfrak{Q} = \mathfrak{D}$ and when $\mathfrak{D}$ is ε-far from $\mathfrak{Q}$. The algorithm is allowed to be randomized and can err with probability at most $\frac{1}{4}$.

Continuous distributions are typically not testable. In general, it is infeasible to study interesting properties of continuous distributions without any assumptions on the density function. For example, one can show that for every integer t there is no tester A that distinguishes with at most t samples between uniform distribution $\mathfrak{D}_U$ on $[0, 1]$ and any distribution that is ε-far from uniform (for example, take a uniform distribution on t^3 randomly chosen points from the interval $[0, 1]$; such distribution is discrete and hence it is $\frac{1}{2}$-far from uniform). This result can be generalized for testing a number of natural properties for distributions on continuous or uncountable domains by observing that a small discontinuity of the density function makes testing many natural properties impossible.

One can also derive similar impossibility results from the existing lower bounds for testing properties of discrete distributions. For example, Batu et al. [4] (see also [6]) show that testing if a distribution on the support of size n is uniform requires $\Omega(\sqrt{n}/\varepsilon^2)$ samples. Since any continuous distribution can be seen as a limit of a discrete distribution on a support of size $n \to \infty$, the lower bound due to Batu et al. implies that no algorithm can test if a given distribution on $[0, 1]$ is uniform [4]. This approach yields similar impossibility results for testing if a given distribution is monotone, unimodal, or if two distributions are identical, etc., (see [1, 2, 3, 4, 5, 7, 8, 9] for more examples).

With these negative result, the natural challenge is to investigate if there are any nontrivial continuous distributions that are efficiently testable.

Testing if a monotone high-dimensional distribution is uniform. The main objective of this paper is to investigate if there are any nontrivial continuous distributions that are testable. One of a very few properties of discrete distributions in the CS literature that has only a light dependency on the size of the support (the condition by our arguments above, that seems to be necessary to

obtain a fast tester for continuous distributions) is that of testing if a monotone distribution on a Boolean cube is uniform. Rubinfeld and Servedio [9] consider the following problem: given a *monotone*[1] (discrete) distribution $\mathfrak{D}$ on a Boolean cube $\{0,1\}^n$, test if $\mathfrak{D}$ is uniform. Rubinfeld and Servedio [9] show that without any assumption about the monotonicity of $\mathfrak{D}$, every testing algorithm requires $2^{\Omega(n)}$ samples, however, if $\mathfrak{D}$ is monotone, then one distinguishes between the case when $\mathfrak{D}$ is uniform and when $\mathfrak{D}$ is ε-far from uniform using $\mathcal{O}(n\log(1/\varepsilon)/\varepsilon^2)$ samples. Furthermore, this result is almost optimal in that $\Omega(n/\log^2 n)$ samples are necessary.

Our main contribution is the analysis of this problem in the setting when $\mathfrak{D}$ is a monotone distribution[2] on an n-dimensional (*continuous*) cube $[0,1]^n$. First, we provide a characterization of monotone distributions that are ε-far from uniform:

Theorem 1. *Let $\mathfrak{D}$ be a monotone distribution on $[0,1]^n$ with density function $\mathfrak{f}$. If $\mathfrak{D}$ is ε-far from uniform then*

$$\mathbb{E}_{\mathfrak{f}}[\|\mathbf{x}\|_1] \;=\; \int_{\mathbf{x}} \|\mathbf{x}\|_1 \cdot \mathfrak{f}(\mathbf{x})\, d\mathbf{x} \geq \frac{n}{2} + \frac{\varepsilon}{2}\,.$$

The proof of this theorem can be deduced from the following result (which is the main technical contribution of the paper) by substituting $\mathfrak{g}(\mathbf{x}) = \mathfrak{f}(\mathbf{x}) - 1$.

Theorem 2. *Let $\mathfrak{g}\colon [0,1]^n \to \mathbb{R}$ be a monotone function with $\int_{\mathbf{x}} \mathfrak{g}(\mathbf{x}) = 0$. Then*

$$\int_{\mathbf{x}} \|\mathbf{x}\|_1 \cdot \mathfrak{g}(\mathbf{x})\, d\mathbf{x} \;\geq\; \frac{1}{4}\int_{\mathbf{x}} |\mathfrak{g}(\mathbf{x})|\,.$$

By combining Theorem 1 with the fact that for uniform distribution $\mathfrak{U}$ on $[0,1]^n$ we have $\mathbb{E}_{\mathfrak{U}}[\|\mathbf{x}\|_1] = \frac{n}{2}$, we can show that the following simple algorithm tests if a distribution is uniform or it is ε-far from uniform:

Testing uniformity:

- **Repeat** $r = 20$ times:

  Draw a sample (according to the distribution $\mathfrak{D}$) $S = \langle \mathbf{x}_1, \ldots, \mathbf{x}_s\rangle$ from $[0,1]^n$ with $s = \lceil \frac{40n}{\varepsilon^2} \rceil$

  If $\sum_{i=1}^{s} \|\mathbf{x}_i\|_1 \geq s(\frac{n}{2} + \frac{\varepsilon}{4})$ then **Reject** and exit
- **Accept**

Theorem 3. Testing uniformity *distinguishes between the uniform distribution on $[0,1]^n$ and any monotone distribution over $[0,1]^n$ that is ε-far from uniform. Its sample complexity is $\mathcal{O}(n/\varepsilon^2)$ and it errs with the probability at most $\frac{1}{4}$.*

[1]Distribution $\mathfrak{D}$ *is monotone* if for any $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$, if $x_i \leq y_i$ for every i then $\Pr_{\mathfrak{D}} \mathbf{x} \leq \Pr_{\mathfrak{D}} \mathbf{y}$.

[2]A probability distribution $\mathfrak{D}$ on $[0,1]^n$ with density function $\mathfrak{f}$ is *monotone* if for any $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$, if $x_i \leq y_i$ for every i then $\mathfrak{f}(\mathbf{x}) \leq \mathfrak{f}(\mathbf{y})$.

References

[1] N. Alon, A. Andoni, T. Kaufman, K. Matulef, R. Rubinfeld, and N. Xie, Testing k-wise and almost k-wise Independence, *Proceedings of the 39th Annual ACM Symposium on Theory of Computing* (*STOC*), pp. 496–505, 2007.

[2] T. Batu, S. Dasgupta, R. Kumar, and R. Rubinfeld, The complexity of approximating the entropy, *Proceedings of the 34th Annual ACM Symposium on Theory of Computing* (*STOC*), pp. 678–687, 2002.

[3] T. Batu, E. Fischer, L. Fortnow, R. Kumar, R. Rubinfeld, and P. White, Testing random variables for independence and identity, *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science* (*FOCS*), pp. 442–415, 2001.

[4] T. Batu, L. Fortnow, R. Rubinfeld, W. D. Smith, and P. White, Testing that distributions are close, *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science* (*FOCS*), pp. 259–269, 2000.

[5] T. Batu, R. Kumar, and R. Rubinfeld, Sublinear algorithms for testing monotone and unimodal distributions, *Proceedings of the 36th Annual ACM Symposium on Theory of Computing* (*STOC*), pp. 381–390, 2004.

[6] O. Goldreich and D. Ron, On testing expansion in bounded-degree graphs, *Electronic Colloquium on Computational Complexity* (*ECCC*), Report No. 7, 2000.

[7] S. Raskhodnikova, D. Ron, A. Shpilka, and A. Smith, Strong lower bounds for approximating distribution support size and the distinct elements problem, *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science* (*FOCS*), pp. 559–569, 2007.

[8] R. Rubinfeld, Sublinear time algorithms, *Proceedings of the International Congress of Mathematicians*, Madrid, Spain, August 22–30, 2006.

[9] R. Rubinfeld and R. A. Servedio, Testing monotone high-dimensional distributions, *Proceedings of the 37th Annual ACM Symposium on Theory of Computing* (*STOC*), pp. 147–156, 2005.

On risks of using cuckoo hashing: Strengths and weaknesses of pseudorandom graphs generated by universal hashing

Martin Dietzfelbinger

(joint work with Ulf Schellbach)

Cuckoo hashing, introduced by Pagh and Rodler [10], is a strategy for maintaining hash tables for a set S of keys from a "universe" U, so that lookups take constant time in the worst case. The data structure consists of two tables T_1 and T_2 of size m each, and it uses two hash functions $h_1, h_2 \colon U \to [m]$. Given $S \subseteq U$, of size $n = |S|$, it is required that key x is stored in cell $T_1[h_1(x)]$ or $T_2[h_2(x)]$, and that at most one key of S is stored in each cell. In this case we say that h_1 and h_2 are *suitable for* S. Pagh and Rodler showed that if $m \geq (1+\varepsilon)n$ for a constant $\varepsilon > 0$, and if h_1 and h_2 are $c \log n$-wise independent on S, for some constant $c > 0$, then h_1 and h_2 are suitable for S with high probability, and there is an insertion procedure (the "cuckoo strategy", see [10]), that inserts each key in expected constant time.

The cuckoo hashing data structure immediately induces a bipartite graph, the *cuckoo graph* $G(S, h_1, h_2)$, defined as follows: The left node set and the right node set both are $[m]$, and the edge set is

$$E(S, h_1, h_2) = \{(h_1(x), h_2(x)) \mid x \in S\}.$$

This graph exhibits a certain degree of randomness, depending on the way S and h_1 and h_2 are chosen. Clearly, h_1 and h_2 are suitable for S if and only if no connected component of $G(S, h_1, h_2)$ has more than one cycle.

The talk (which is based on the two papers [6, 7]) addresses the question how strong the hash functions used in the scheme have to be. To achieve $c \log n$-wise independence by means of the hash class, sophisticated constructions must be used [12], which are not really practical. In practice, one is tempted to use cheap and fast hash classes and hope that they will behave almost like random. In experiments, cuckoo hashing works very well with some weaker hash function classes. However, already Pagh and Rodler reported on experimental results that indicate that cuckoo hashing might not work well in combination with the "*multiplicative class*" from [4] and recommended avoiding this class. The multiplicative class $H^{\text{mult}}_{k,\ell}$ consists of functions $h_a \colon [2^k] \to [2^l]$ of the form $h_a(x) = ((a \cdot x) \bmod 2^k) \operatorname{div} 2^{k-l}$, for $0 < a < 2^k$ odd. One of these functions is chosen at random, in the sense of *universal hashing* as introduced by Carter and Wegman [2]. This class is 2-universal in the sense that the collision probability for any two keys is at most $2/m$. Another basic and important class is the "*linear hash class*" $H^{\text{lin}}_{p,m}$, for a prime number $p > n$ and table size $m < p$. The functions in this class have the format $h_{a,b}(x) = ((ax) \bmod p) \bmod m$, for $0 \le a, b < p$ (chosen randomly). This standard class is (almost) two-wise independent.

In 2008, Mitzenmacher and Vadhan [9] proved that if a 2-wise independent hash class is used (or even if the collision probability is $O(1/m)$) and the key set S exhibits a certain kind of partial randomness, and some further technical conditions are fulfilled, then the combination of the key set and a hash function chosen at random from a class of simple functions will behave very close to full randomness, in a technical sense. This setup applies to cuckoo hashing, and it would imply that if the conditions listed in the theorems given by Mitzenmacher and Vadhan are fulfilled, then cuckoo hashing will work.

Our results indicate that care must be taken when using weak universal classes in combination with cuckoo hashing. Under certain circumstances, it may happen that the whole data structure crashes with high probability! In detail, our results are as follows:

Result 1. In cuckoo hashing with the multiplicative class $H^{\text{mult}}_{k,\ell}$ (the table size m is larger than $2n$ and hence way above the threshold sufficient for the standard analysis) *all* function pairs (h_1, h_2) from $H^{\text{mult}}_{k,\ell}$ will work badly with high probability for fully random key sets of size n, if $n/|U| > n^{1-\delta}$ for some constant $\delta > 0$.

Result 1 can be extended to non-dense key sets with $n/|U|$ arbitrarily small, by considering key sets chosen at random from a (structured) sub-universe $U' \subseteq U$.

Result 2. Cuckoo hashing with the standard almost 2-wise independent class $H^{\text{lin}}_{p,m}$ exhibits a similar behaviour as in Result 1, again in the case where the key set is relatively dense in U. This is true even if the two hash functions use different prime moduli.

Result 3. Another construction shows that cuckoo hashing with the multiplicative class does not exhibit the behaviour provable for $c \cdot \log n$-wise independent classes (the failure probability is $\Omega(1)$, while in the $c \cdot \log n$-wise independent case it is $O(1/n)$), for sparse key sets that exhibit a certain grid structure.

The proof techniques for Results 1 and 2 are similar: One shows that the "full" cuckoo graph $G(U, h_1, h_2)$ has a very regular, periodic structure. Further, one shows that it contains constant-size "obstructing subgraphs", meaning connected subgraphs with two cycles. Because of the periodicity, there is a linear number of such subgraphs that do not overlap too much. By a standard argument (conditional expectation inequality or second moment method) one sees that when a relatively dense random set of edges (which corresponds to choosing a random key set S) is chosen, the probability that one of these obstructing subgraphs is chosen completely is $1 - o(1)$, which means that h_1 and h_2 are not suitable for S with probability $1 - o(1)$. For the different hash classes the methods to find the obstructing subgraphs are different.

On the surface, these results seem to contradict the results by Mitzenmacher and Vadhan [9]. However, a closer look reveals that in the case of dense key sets one of the technical assumptions for the main theorems from that paper are not satisfied. Although the key set is fully random, a technical parameter (the "collision probability", closely related to the Renyi entropy) is too large.

On the one hand our results explain the experimental observations by Pagh and Rodler and substantiates their warning against using multiplicative hashing in combination with cuckoo hashing; on the other hand shows that one of the "further technical conditions" of the Mitzenmacher/Vadhan result, namely the requirement that key sets must be relatively sparse in U, is really necessary for that result to be true.

Open problems: 1. It is an open problem to determine how strong universal classes are needed to guarantee that cuckoo hashing works. Our work shows that standard 2-wise independence classes are not enough; with contrived constructions one can show that even bounding the collision probability for up to 5 keys one cannot enforce that cuckoo hashing works ([3], manuscript by Cohen/Kane).
2. Fotakis, Pagh, Sanders, and Spirakis [8] described a generalization of cuckoo hashing that uses more than 2 hash functions, say d many, and one table of size m. It is an open problem to determine the exact threshold density for n/m for their scheme to work, even with fully random hash functions. As described in [5] (Observation 2), the results of Calkin [1] give a one-sided bound. It is unknown (but presumably true) that this bound is not optimal.

References

[1] N. J. Calkin, Dependent sets of constant weight binary vectors, *Combinatorics, Probability and Computing* **6** (1997), 263–271.

[2] L. Carter and M. N. Wegman, Universal classes of hash functions, *J. Comput. Syst. Sci.* **18** (1979), 143–154.

[3] J. Cohen and D. M. Kane, 6.856 Project: Bounds on the Independence Required for Cuckoo Hashing, Manuscript, 2005/2009. `http://www.math.harvard.edu/~dankane/cuchkoohashing.pdf` , last download: 2009/05/31.
[4] M. Dietzfelbinger, T. Hagerup, J. Katajainen, and M. Penttonen, A reliable randomized algorithm for the closest-pair problem, *J. Algorithms* **25** (1997), 19–51.
[5] M. Dietzfelbinger and R. Pagh, Succinct data structures for retrieval and approximate membership, in: 35th ICALP 2008, Part I: Track A: *Algorithms, Automata, Complexity, and Games*, pp. 385–396.
[6] M. Dietzfelbinger and U. Schellbach, On risks of using cuckoo hashing with simple hash functions, in: *Proc. 20th ACM-SIAM Symp. on Discrete Algorithms*, 2009, pp. 795–804.
[7] M. Dietzfelbinger and U. Schellbach, Weaknesses of cuckoo hashing with simple universal hash classes: the case of large universes, in: *Proc. 35th SOFSEM*, 2009, pp. 217–228.
[8] D. Fotakis, R. Pagh, P. Sanders, and P. Spirakis, Space efficient hash tables with worst case constant access time, *Theory of Computing Systems* **38** (2005), 229–248.
[9] M. Mitzenmacher and S. Vadhan, Why simple hash functions work: exploiting the entropy in a data stream, in: *Proc. 19th ACM-SIAM Symp. on Discrete Algorithms*, 2008, pp. 746–755.
[10] R. Pagh and F. F. Rodler, Cuckoo Hashing, *J. Algorithms* **51** (2004), 122–144.
[11] A. Pagh, R. Pagh and M. Ruzic, Linear probing with constant independence, in: *Proc. 39th ACM Symp. on Theory of Computing*, 2007, pp. 318–327.
[12] A. Siegel, On universal classes of extremely random constant-time hash functions, *SIAM J. Comput.* **33** (2004), 505–543.

Large induced trees in K_r-free graphs

Jacob Fox
(joint work with Po-Shen Loh and Benny Sudakov)

For a graph G, let $t(G)$ denote the maximum number of vertices in an induced subgraph of G that is a tree. The problem of bounding $t(G)$ in a connected graph G was first introduced twenty years ago by Erdős, Saks, and Sós [4]. Clearly, to get a non-trivial result one must impose some conditions on the graph G, because, for example, the complete graph contains no induced tree with more than 2 vertices. In their paper, Erdős, Saks, and Sós studied the relationship between $t(G)$ and several natural parameters of the graph G. They were able to obtain asymptotically tight bounds on $t(G)$ when either the number of edges or the independence number of G were known.

Erdős, Saks, and Sós also considered the problem of estimating the size of the largest induced tree in graphs with no K_r (complete graph on r vertices). Let $t_r(n)$ be the minimum value of $t(G)$ over all connected K_r-free graphs G on n vertices. In particular, for triangle-free graphs, they proved that

$$\Omega\left(\frac{\log n}{\log\log n}\right) \;\le\; t_3(n) \;\le\; O(\sqrt{n}\log n),$$

Research by Jacob Fox partially supported by an NSF Graduate Research Fellowship and a Princeton Centennial Fellowship. Research by Po-Shen Loh partially supported by a Fannie and John Hertz Foundation Fellowship, an NSF Graduate Research Fellowship, and a Princeton Centennial Fellowship. Research by Benny Sudakov partially supported by NSF CAREER award DMS-0546523, NSF grant DMS-0355497, and a USA-Israeli BSF grant.

and left as an interesting open problem the task of closing the wide gap between these two bounds.

The first significant progress on this question was made only recently by Matoušek and Šámal [14], who actually came to the problem of estimating $t_3(n)$ from a different direction. Pultr had been studying forbidden configurations in Priestley spaces [2], and this led him to ask in [16] how large $t(G)$ could be for connected bipartite graphs G. Let $t_B(n)$ be the minimum value of $t(G)$ over all connected bipartite graphs on n vertices. It is clear that $t_3(n) \leq t_B(n)$, so the result of Erdős, Saks, and Sós immediately gives a lower bound on $t_B(n)$.

Motivated by Pultr's question, Matoušek and Šámal studied $t_B(n)$ and $t_3(n)$. They found the following nice construction which shows that $t_3(n) \leq t_B(n) < 2\sqrt{n} + 1$. Let $m = \sqrt{n}$, and consider the graph with parts V_{-m+1}, V_{-m+2}, ..., V_{m-1}, where $|V_i| = m - |i|$, and each consecutive pair of parts (V_i, V_{i+1}) induces a complete bipartite graph. This graph is clearly bipartite with $m^2 = n$ vertices, and it is easy to see that every induced tree in it has at most $2m - 1$ vertices. On the other hand, Matoušek and Šámal were able to improve the lower bound on $t_B(n)$ and $t_3(n)$, showing that $t_3(n) \geq e^{c\sqrt{\log n}}$ for some constant c. Furthermore, they also proved that if there was even a single value of n_0 for which $t_3(n_0) < \sqrt{n_0}$, then in fact $t_3(n) \leq O(n^\beta)$ for some constant β strictly below $1/2$. The above fact led Matoušek and Šámal to conjecture that the true asymptotic behavior of $t_3(n)$ was some positive power of n which is strictly smaller than $1/2$.

Our first main result essentially solves this problem. It determines that the order of growth of both $t_3(n)$ and $t_B(n)$ is precisely $\Theta(\sqrt{n})$, disproving the conjecture of Matoušek and Šámal.

Theorem 1. *Let G be a connected triangle-free graph on n vertices. Then $t(G) \geq \sqrt{n}$.*

Furthermore, our approach can also be used to give asymptotically tight bounds on the size of the largest induced tree in K_r-free graphs for all remaining values of r. In their original paper, Erdős, Saks, and Sós gave an elegant construction which shows that $t_r(n)$ for $r \geq 4$ has only logarithmic growth. Indeed, let T be a balanced $(r-1)$-regular tree, that is, a rooted tree in which all non-leaf vertices have degree $r - 1$ and the depth of any two leaves differs by at most 1. Then the line graph[1] $L(T)$ is clearly K_r-free, and one can easily check that induced trees in $L(T)$ correspond to induced paths in T, which have only logarithmic length. Optimizing the choice of the parameters in this construction, one can show that $t_r(n) \leq \frac{2\log(n-1)}{\log(r-2)} + 2$. On the other hand, using Ramsey Theory, Erdős, Saks, and Sós also showed that $t_r(n) \geq \frac{c_r \log n}{\log\log n}$, where c_r is a constant factor depending only on r. Our second main result closes the gap between these two bounds as well, and determines the order of growth of $t_r(n)$ up to a small multiplicative constant.

[1]The vertices of $L(T)$ are the edges of T, and two of them are adjacent if they share a vertex in T.

Theorem 2. *Let $r \geq 4$, and let G be a connected graph on n vertices with no clique of size r. Then $t(G) \geq \frac{\log n}{4 \log r}$.*

One can similarly investigate induced forests in K_r-free graphs. Let $f_r(n)$ to be the maximum number such that every K_r-free graph on n vertices contains an induced forest with at least $f_r(n)$ vertices. We trivially have $f_r(n) \geq t_r(n)$. The independence number $\alpha(G)$ of a graph is the size of the largest independent set of vertices in G. The size of the maximum induced forest in any graph is closely related to its independence number. Indeed, since an independent set is a forest and every forest is bipartite, then the number of vertices of the largest induced forest in a graph G is at least $\alpha(G)$ and at most $2\alpha(G)$. Together with the best known upper bound for off-diagonal Ramsey numbers [1], for fixed $r \geq 3$ and all n, we have $f_r(n) \geq c(n \log^{r-2} n)^{\frac{1}{r-1}}$ for some positive constant c. Hence, $f_3(n)$ is a factor on the order of $\sqrt{\log n}$ larger than $t_3(n)$. Furthermore, for fixed $r > 3$, $f_r(n)$ and $t_r(n)$ behave very differently, as $f_r(n)$ is polynomial in n while $t_r(n)$ is only logarithmic in n. This demonstrates that in K_r-free graphs on n vertices the largest guaranteed induced forest is much larger than the largest guaranteed induced tree.

We finish by mentioning some related research. Our work considers the Ramsey-type problem of finding either a clique or a large induced tree. The similar problem of finding an induced copy of a *particular* tree T in a K_r-free graph was independently raised by Gyárfás [8] and Sumner [19]. They conjectured that for any fixed integer r and tree T, any graph with sufficiently large chromatic number (depending on r and T) must contain either an r-clique or an induced copy of T. Note that the essential parameter for the graph G is now the chromatic number and not the number of vertices. Indeed, a complete bipartite graph has no clique of size 3, but contains only stars as induced subtrees. This conjecture is widely open, although some partial results were obtained in [9, 10, 11, 17].

Induced trees were also studied in the context of sparse random graphs. This line of research was started by Erdős and Palka [3], who conjectured that for any constant $c > 1$, the random graph $G(n, c/n)$ would with high probability contain an induced tree of order $\gamma(c)n$. This was solved by Fernandez de la Vega [5], and other variants of this result were obtained in [12, 6, 7, 13, 18]. In another regime, when the edge probability is $p = c \log n/n$, Palka and Ruciński [15] showed that the largest induced tree in $G(n, p)$ has size $\Theta(n \log \log n / \log n)$ with high probability.

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi, A note on Ramsey numbers, *Journal of Combinatorial Theory A* **29** (1980), 354–360.

[2] R. Ball and A. Pultr, Forbidden forests in Priestley spaces, *Cahiers de Topologie et Geométrie Différentielle Catégoriques* **45** (2004), 2–22.

[3] P. Erdős and Z. Palka, Trees in random graphs, *Discrete Math.* **46** (1983), 145–150; Addendum: *Ibid* **48** (1984), 331.

[4] P. Erdős, M. Saks, and V. Sós, Maximum induced trees in graphs, *Journal of Combinatorial Theory B* **41** (1986), 61–79.

[5] W. Fernandez de la Vega, Induced trees in sparse random graphs, *Graphs Combin.* **2** (1986), 227–231.
[6] W. Fernandez de la Vega, The largest induced tree in a sparse random graph, *Random Structures and Algorithms* **9** (1996), 93–97.
[7] A. Frieze and B. Jackson, Large induced trees in sparse random graphs, *Journal of Combinatorial Theory B* **42** (1987), 181–195.
[8] A. Gyárfás, On Ramsey covering numbers, *Coll. Math. Soc. János Bolyai*, in: *Infinite and Finite Sets*, North Holland/American Elsevier, New York (1975), 10.
[9] A. Gyárfás, Problems from the world surrounding perfect graphs, *Zastowania Matematyki Applicationes Mathematicae* **XIX** (1985), 413–441.
[10] A. Gyárfás, E. Szemerédi, and Zs. Tuza, Induced subtrees in graphs of large chromatic number, *Discrete Math.* **30** (1980), 235–244.
[11] H. Kierstead and S. Penrice, Radius two trees specify χ-bounded classes, *Journal of Graph Theory* **18** (1994), 119–129.
[12] L. Kučera and V. Rödl, Large trees in random graphs, *Comment. Math. Univ. Carolin.* **28** (1987), 7–14.
[13] T. Łuczak and Z. Palka, Maximal induced trees in sparse random graphs, *Discrete Math.* **72** (1988), 257–265.
[14] J. Matoušek and R. Šámal, Induced trees in triangle-free graphs, *Electronic Journal of Combinatorics* **15** (2008), R41, 9 pp. (electronic). A conference version appeared in *Electron. Notes in Discrete Math.* **29** (2007), 307–313.
[15] Z. Palka and A. Ruciński, On the order of the largest induced tree in a random graph, *Discrete Applied Mathematics* **15** (1986), 75–83.
[16] A. Pultr, personal communication to J. Matoušek and R. Šámal.
[17] A. Scott, Induced trees in graphs of large chromatic number, *Journal of Graph Theory* **24** (1997), 297–311.
[18] W. Suen, On large induced trees and long induced paths in sparse random graphs, *Journal of Combinatorial Theory B* **56** (1992), 250–262.
[19] D. Sumner, Subtrees of a graph and chromatic number, in: *The Theory and Applications of Graphs*, ed. G. Chartrand, John Wiley & Sons, New York (1981), 557–576.

Coloring simple hypergraphs

Alan Frieze
(joint work with Dhruv Mubayi)

Hypergraph coloring has been studied for almost 50 years, since Erdős' seminal results on the minimum number of edges in uniform hypergraphs that are not 2-colorable. Some of the major tools in combinatorics have been developed to solve problems in this area, for example, the local lemma and the nibble or semi-random method. Consequently, the subject enjoys a prominent place among basic combinatorial questions.

Closely related to coloring problems are questions about the independence number of hypergraphs. An easy extension of Turán's graph theorem shows that a k-uniform hypergraph with n vertices and average degree d has an independent set of size at least $cn/d^{1/(k-1)}$, where c depends only on k. If we impose local constraints on the hypergraph, then this bound can be improved. An i-cycle in a

Research by Alan Frieze partially supported by NSF Grant DMS-0753472. Research by Dhruv Mubayi partially supported by NSF Grant DMS-0653946.

k-uniform hypergraph is a collection of i distinct edges spanned by at most $i(k-1)$ vertices. Say that a k-uniform hypergraph has girth at least g if it contains no i-cycles for $2 \leq i < g$. Call a k-uniform hypergraph simple if it has girth at least 3. In other words, every two edges have at most one vertex in common. Throughout this abstract we will assume that $k \geq 3$ is a fixed positive integer.

Ajtai-Komlós-Pintz-Spencer-Szemerédi [2] proved the following fundamental result that strengthened the bound obtained by Turán's theorem above.

Theorem 1. ([2]) *Let $H = (V, E)$ be a k-uniform hypergraph of girth at least 5 with maximum degree Δ. Then it has an independent set of size at least*

$$c\, n \left(\frac{\log \Delta}{\Delta}\right)^{1/(k-1)}$$

where c depends only on k.

Spencer conjectured that Theorem 1 holds even for simple hypergraphs, and this was later proved by Duke-Lefmann-Rödl [5]. Theorem 1 has proved to be a seminal result in combinatorics, with many applications. Indeed, the result was first proved for $k = 3$ by Komlós-Pintz-Szemerédi [11] to disprove the famous Heilbronn conjecture, that among every set of n points in the unit square, there are three points that form a triangle whose area is at most $O(1/n^2)$.

The goal of this paper is to prove a result that is stronger than Theorem 1 (and also the accompanying result of [5]). Our main result states that not only can one find an independent set of the size guaranteed by Theorem 1, but in fact that the entire vertex set can be partitioned into independent sets with this average size. Recall that the chromatic number $\chi(H)$ of H is the minimum number of colors needed to partition the vertex set so that no edge is monochromatic.

Theorem 2. *Fix $k \geq 3$. Let $H = (V, E)$ be a simple k-uniform hypergraph with maximum degree Δ. Then*

$$\chi(H) < c\left(\frac{\Delta}{\log \Delta}\right)^{\frac{1}{k-1}}$$

where c depends only on k.

It is shown in [4] that Theorem 2 is sharp apart from the constant c. In order to prove Theorem 2 we will first prove the following slightly weaker result. A triangle in a k-uniform hypergraph is a 3-cycle that contains no 2-cycle. In other words, it is a collection of three sets A, B, C such that every two of these sets have intersection of size one, and $A \cap B \cap C = \emptyset$.

Theorem 3. *Fix $k \geq 3$. Let $H = (V, E)$ be a simple triangle-free k-uniform hypergraph with maximum degree Δ. Then*

$$\chi(H) < c\left(\frac{\Delta}{\log \Delta}\right)^{\frac{1}{k-1}},$$

where c depends only on k.

The proof of Theorem 3 rests on several major developments in probabilistic combinatorics over the past 25 years. Our approach is inspired by Johansson's breakthrough result on graph coloring, which proved Theorem 3 for $k = 2$.

The proof technique, which has been termed the semi-random, or nibble method, was first used by Rödl (inspired by earlier work in [2, 11]) to confirm the Erdős-Hanani conjecture about the existence of asymptotically optimal designs. Subsequently, Kim [9] (see also Kahn [8]) proved Johansson's theorem for graphs with girth five and then Johansson proved his result.

The implication Theorem 3 $\rightarrow$ Theorem 2 forms a much shorter (but still non-trivial) part of this paper (See Section 2). Our proof uses a recent concentration result of Kim and Vu [10] together with some additional ideas similar to those from Alon-Krivelevich-Sudakov [3].

Finally, we remark that our proof of Theorem 3 also gives the same upper bound for list chromatic number. On the other hand, we are not able to prove Theorem 2 for list chromatic numbers. We end with a conjecture posed in [6].

Conjecture 1. ([6]) *Let F be a k-graph. There is a constant c_F depending only on F such that every F-free k-graph with maximum degree Δ has chromatic number at most $c_F(\Delta/\log\Delta)^{1/(k-1)}$.*

References

[1] M. Ajtai, P. Erdős, J. Komlós, E. Szemerédi, On Turán's theorem for sparse graphs, *Combinatorica* **1**, 313–317 (1981).

[2] M. Ajtai, J. Komlós, J. Pintz, J. Spencer and E. Szemerédi, Extremal uncrowded hypergraphs, *Journal of Combinatorial Theory A* **32** (1982), no. 3, 321–335.

[3] N. Alon, M. Krivelevich and B. Sudakov, Coloring graphs with sparse neighborhoods, *Journal of Combinatorial Theory B* **77** (1999), 73–82.

[4] T. Bohman, A. Frieze, D. Mubayi, Coloring H-free hypergraphs, submitted. Pre-print available at `http://www.math.cmu.edu/~af1p/Texfiles/Hypchrom.pdf` .

[5] R. Duke, H. Lefmann, V. Rödl, On uncrowded hypergraphs, *Random Structures and Algorithms* **6** (1995), 209–212.

[6] A. Frieze and D. Mubayi, On the chromatic number of simple triangle-free triple systems, to appear in *Electronic Journal of Combinatorics*.

[7] A. Johansson, Asymptotic choice number for triangle free graphs, *DIMACS Technical Report* 91-4, 1196.

[8] J. Kahn, Asymptotically good list-colorings, *Journal of Combinatorial Theory A* **73** (1996), no. 1, 1–59.

[9] J. H. Kim, On Brooks' theorem for sparse graphs, *Combinatorics, Probability and Computing* **4** (1995), no. 2, 97–132.

[10] J. H. Kim and V. Vu, Concentration of multivariate polynomials and its applications, *Combinatorica* **20** (2000), 417–434.

[11] J. Komlós, J. Pintz and E. Szemerédi, A lower bound for Heilbronn's problem, *J. London Math. Soc.* (2) **25** (1982), no. 1, 13–24.

[12] A. Kostochka, D. Mubayi, V. Rödl, P. Tetali, On the chromatic number of set systems, *Random Structures and Algorithms* **19** (2001), no. 2, 87–98.

[13] V. Rödl, On a packing and covering problem, *European Journal of Combinatorics* **6** (1985), no. 1, 69–78.

Bootstrap percolation on $G(n,p)$

Svante Janson
(joint work with Tomasz Łuczak, Tatyana Turova and Thomas Vallier)

Bootstrap percolation on a graph G is defined as the spread of *activation* or *infection* according to the following rule, with a given threshold $r \geq 2$: We start with a set $\mathcal{A}(0) \subseteq V(G)$ of *active* vertices. Each inactive vertex that has at least r active neighbours becomes active. This is repeated until no more vertices become active, i.e., when no inactive vertex has r or more active neighbours.

We are mainly interested in the final size A^* of the active set, and in particular whether eventually all vertices will be active or not. If they are, we say that the initial set $\mathcal{A}(0)$ *percolates*. We will study a sequence of graphs of order $n \to \infty$; we then also say that (a sequence of) $\mathcal{A}(0)$ *almost percolates* if the number of vertices that remain inactive is $o(n)$, i.e., if $A^* = n - o(n)$.

Bootstrap percolation has been studied on various graphs, both deterministic and random; one can study either a random initial set or the deterministic problem of choosing an initial set that is optimal in some sense. For example, a classical folklore problem is to find the minimal percolating set in a two-dimensional grid; see Balogh and Pete [3] and Bollobás [5]. (These references also treat higher-dimensional grids.) Some further references for random initial sets on various graphs are Cerf and Manzo [6], Holroyd [7] (grids); Balogh and Bollobás [1] (hypercube); Balogh, Peres and Pete [2] (infinite trees); Balogh and Pittel [4] (random regular graphs).

We here study bootstrap percolation on the Erdös-Rényi random graph $G_{n,p}$ (which somewhat surprisingly seems to have been neglected so far in this context), with an initial set $\mathcal{A}(0)$ consisting of a given number a vertices chosen at random. This was first studied by Vallier [8]; we here present a simple method that allows us to both simplify the proofs and improve the results.

In order to analyze the bootstrap percolation process on $G_{n,p}$, we change the time scale and consider at each time step the spread of activation from one vertex only. Choose $u_1 \in \mathcal{A}(0)$ and give each of its neighbours a *mark*; we then say that u_1 is *used*, and let $\mathcal{Z}(1) := \{u_1\}$ be the set of used vertices at time 1. We continue recursively: At time $t+1$, choose a vertex $u_{t+1} \in \mathcal{A}(t) \setminus \mathcal{Z}(t)$ (provided this set is non-empty). We give each neighbour of u_{t+1} a new mark. Let $\Delta\mathcal{A}(t+1)$ be the set of inactive vertices with r marks; these now become active and we let $\mathcal{A}(t+1) = \mathcal{A}(t) \cup \Delta\mathcal{A}(t+1)$ be the set of active vertices at time t. We finally set $\mathcal{Z}(t+1) = \mathcal{Z}(t) \cup \{u_{t+1}\} = \{u_i : i \leq t+1\}$, the set of used vertices.

The process stops when $\mathcal{A}(t) \setminus \mathcal{Z}(t) = \emptyset$, i.e., when all active vertices are used. We denote this time by T;

$$T := \min\{t \geq 0 : \mathcal{A}(t) \setminus \mathcal{Z}(t) = \emptyset\}. \tag{1}$$

Thus the final infected set is $\mathcal{A}(T) = \mathcal{Z}(T)$, and its size is

$$A^* := |\mathcal{A}(T)| = |\mathcal{Z}(T)| = T. \tag{2}$$

Hence, the set $\mathcal{A}(0)$ percolates if and only if $T = n$, and $\mathcal{A}(0)$ almost percolates if and only if $T = n - o(n)$.

Since $|\mathcal{Z}(t)| = t$ and $\mathcal{Z}(t) \subseteq \mathcal{A}(t)$ for $t = 0, \dots, T$, we also have, with $A(t) := |\mathcal{A}(t)|$, the number of active vertices at time t,

$$T = \min\{t \geq 0 : A(t) = t\}. \tag{3}$$

We analyze this process by the standard method of revealing the edges of the graph $G_{n,p}$ only on a need-to-know basis. We thus begin by choosing u_1 as above and then reveal its neighbours; we then find u_2 and reveal its neighbours, and so on. Let, for $i \notin \mathcal{Z}(s)$, $I_i(s)$ be the indicator function that there is an edge between the vertices u_s and i. This is also the indicator that i gets a mark at time s, so if $M_i(t)$ is the number of marks i has at time t, then

$$M_i(t) = \sum_{s=1}^{t} I_i(s), \tag{4}$$

at least until i is activated (and what happens later does not matter). Note that if $i \notin \mathcal{A}(0)$, then, for every $t \leq T$, $i \in \mathcal{A}(t)$ if and only if $M_i(t) \geq r$.

The crucial feature of this description of the process, which makes the analysis simple, is that the random variables $I_i(s)$ are i.i.d. $\mathrm{Be}(p)$. We have defined $I_i(s)$ only for $s \leq T$ and $i \notin \mathcal{Z}(s)$, but it is convenient to add further (redundant) variables so that $I_i(s)$ are defined, and i.i.d., for all $i \in V_n$ and all $s \geq 1$.

Define, for $i \in V_n \setminus \mathcal{A}(0)$,

$$Y_i := \min\bigl\{t : M_i(t) \geq r\bigr\}. \tag{5}$$

If $Y_i \leq T$, then Y_i is the time vertex i becomes active, but if $Y_i > T$, then Y_i never becomes active. Thus, for $t \leq T$,

$$\mathcal{A}(t) = \mathcal{A}(0) \cup \{i \notin \mathcal{A}(0) : Y_i \leq t\}.$$

By (4) and (5), each Y_i has a negative binomial distribution $\mathrm{NegBin}(r, p)$;

$$\mathbb{P}(Y_i = k) = \mathbb{P}\bigl(M_i(k-1) = r-1,\, I_i(k) = 1\bigr) = \binom{k-1}{r-1} p^k (1-p)^{r-k};$$

moreover, these random variables Y_i are i.i.d.

We let, for $t = 0, 1, 2, \dots$,

$$S(t) := |\{i \notin \mathcal{A}(0) : Y_i \leq t\}|,$$

so

$$A(t) = S(t) + A(0) = S(t) + a. \tag{6}$$

By (3), (2) and (6), it suffices to study the stochastic process $S(t)$. Note that $S(t)$ is a sum of $n - a$ i.i.d. processes $\mathbf{1}[t \geq Y_i]$, each of which is 0/1-valued and jumps from 0 to 1 at time Y_i. The fact that $S(t)$, and thus $A(t)$, is a sum of i.i.d. processes makes the analysis easy; in particular, for any given t,

$$S(t) \sim \mathrm{Bin}\bigl(n - a, \mathbb{P}(Y_1 \leq t)\bigr).$$

We have, for any given t_0,

$$T \ge t_0 \iff \min_{t<t_0}(A(t)-t) > 0 \iff a + \min_{t<t_0}(S(t)-t) > 0 \iff a > -\min_{t<t_0}(S(t)-t).$$

(Note that is exact; so far no approximation has been done.)

To find the threshold for (almost) percolation, we thus only have to find the minimum $\min_{t<t_0}(S(t)-t)$ for $t_0 = n$ or t_0 close to n. Standard concentration results show that $S(t) \approx \mathbb{E}\, S(t)$, where

$$\mathbb{E}\, S(t) = (n-a)\,\mathbb{P}(Y_1 \le t) = (n-a)\,\mathbb{P}(M_1(t) \ge r),$$

and explicit results are easily found.

For notational simplicity we state the result for $r = 2$ only. In this case, $\mathbb{E}\, S(t) - t$ has a minimum $1/(2np^2)$ at $t = 1/(np^2)$ (asymptotically), and we obtain the following result.

Theorem 1. *Let $r = 2$, and assume $n^{-1} \ll p = p(n) \ll n^{-1/2}$. Then the threshold for (almost) percolation is*

$$a_* := \frac{1}{2np^2}.$$

More precisely, for any fixed $\delta > 0$,

(i) *If $|\mathcal{A}(0)| \le (1-\varepsilon)a_*$, then whp $A^* \le 2|\mathcal{A}(0)|$.*

(ii) *If $|\mathcal{A}(0)| \ge (1+\varepsilon)a_*$, then whp $A^* = n - o(n)$. If further $np \ge \log n + \log\log n + \omega(n)$ for some $\omega(n) \to \infty$, then whp $A^* = n$, so $\mathcal{A}(0)$ percolates completely.*

Moreover, $S(t) - \mathbb{E}\, S(t)$ converges after normalization to a Gaussian process, and it is easy to refine the results above and obtain very precise information on the width of the critical window (which is of the order $\sqrt{a_*}$); we also obtain a Gaussian limit law for the final size A^* in the subcritical case.

Details will appear.

References

[1] J. Balogh and B. Bollobás, Bootstrap percolation on the hypercube, *Probability Theory and Related Fields* **134** (2006), no. 4, 624–648.

[2] J. Balogh, Y. Peres and G. Pete, Bootstrap percolation on infinite trees and non-amenable groups, *Combinatorics, Probability and Computing* **15** (2006), no. 5, 715–730.

[3] J. Balogh and G. Pete, Random disease on the square grid, *Random Structures and Algorithms* **13** (1998), no. 3-4, 409–422.

[4] J. Balogh and B. G. Pittel, Bootstrap percolation on the random regular graph, *Random Structures and Algorithms* **30** (2007), no. 1-2, 257–286.

[5] B. Bollobás, *The Art of Mathematics. Coffee Time in Memphis*. Cambridge University Press, New York, 2006.

[6] R. Cerf and F. Manzo, The threshold regime of finite volume bootstrap percolation, *Stochastic Process. Appl.* **101** (2002), no. 1, 69–82.

[7] A. E. Holroyd, Sharp metastability threshold for two-dimensional bootstrap percolation, *Probabability Theory and Related Fields* **125** (2003), no. 2, 195–224.

[8] T. Vallier, *Random graph models and their applications*, Ph. D. thesis, Lund University, 2007.

A Probability inequality using typical moments and Concentration Results

Ravi Kannan

The well-known Höffding-Azuma (H-A) inequality is widely used to prove concentration results. For real-valued random variables $X_1, X_2, \ldots X_n$ satisfying two conditions:

$\lvert X_i\rvert \leq 1$	Absolute Bounds
$E(X_i \mid X_1, X_2, \ldots X_{i-1}) = 0$	Martingale Difference Sequence,

the inequality gives "Gaussian" tail bounds on the probability that $\sum_{i=1}^n X_i$ deviates from 0:

$$\Pr\left(\left|\sum_{i=1}^n X_i\right| \geq t\right) \leq c_1 e^{-c_2 t^2/n},$$

for some constants c_1, c_2. The main aim of this paper is to weaken the assumption of an absolute bound, while retaining the essential strength of the conclusion. We present two theorems which do this, (of which only the simpler Theorem 1 is stated here) both upper bounding $E(\sum_{i=1}^n X_i)^p$ (the pth moment of $\sum_{i=1}^n X_i$) for some even integer p; from this it is simple to get tail bounds. Our Theorem 1 is simply stated and proved. But both H-A inequality and Chernoff bounds are very special cases of it. Also, several hard concentration results made easy by Talagrand's celebrated inequality – like the (geometric) Travelling Salesperson problem, Minimum weight spanning trees, Longest Increasing Sequence and bin-packing – are also simply tackled by Theorem 1 which yields similar (within constants) Gaussian tail bounds. We are also able to get other results – for example, chromatic number of sparse random graphs, and random projections. Theorem 1 also weakens the Martingale difference condition to a condition we call Strong Negative Correlation; this weakening has several uses too as we will see.

The absolute bound of H-A is weakened in Theorem 1 to bounds on (even) moments of X_i conditioned on (any) value of $X_1 + X_2 + \cdots + X_{i-1}$. A further weakening is obtained in our Main Theorem – Theorem 2 whose proof is more complicated, but still elementary. In Theorem 2, we use information on conditional moments of X_i conditioned on "typical values" of $X_1 + X_2 + \cdots + X_{i-1}$ as well as the "worst-case" values. This is very useful in many contexts as we show. Using Theorem 2, we settle the (discrete case of the) stochastic bin-packing problem studied by Rhee and Talagrand and others by proving concentration results which we show are best possible. We also give exponential tail bounds on the number of triangles in sparse random graphs and indicate several other applications.

Theorem 1. *Let $X_1, X_2, \ldots X_n$ be real valued random variables and p an even positive integer satisfying*

$$E\, X_i(X_1 + X_2 + \ldots X_{i-1})^l \leq 0, \qquad l < p, \ \textit{odd;}\ i = 2, 3, \ldots n.$$

$$E(X_i^l \mid X_1 + X_2 + \ldots X_{i-1}) \leq \left(\tfrac{n}{p}\right)^{(l-2)/2} l!, \quad l \leq p, \ \textit{even;}\ i = 1, 2, 3, \ldots n.$$

Then we have $E\,(\sum_{i=1}^{n} X_i)^p \leq (24np)^{p/2}$.

We will apply the theorem usually with $p \leq n$. With this, it is clear that the bounds on conditional moments we assume are weaker than the absolute bound of 1 in H-A. In fact, the $l!$ will help in many cases, for the reason that it is the l th moment of the Poisson distribution with mean 1.

References

[1] D. Achlioptas, Database friendly random projections, *Proc. Principles of Database systems (PODS)* (2001), 274–281.

[2] D. Achlioptas and A. Naor, The two possible values of the chromatic number of a random graph, *Ann. of Math.* **162** (2005), no. 3, 1335–1351.

[3] B. Bollobás, *Random Graphs*, Cambridge Studies in advanced mathematics, 73 (2001).

[4] B. Bollobás, Martingales, isoperimetric inequalities and random graphs, in *Combinatorics*, Proceedings Eger., (1987) (Hajnal, et. al. Eds) *Colloq. Math. Soc. Janós Bolyai*, **52** North-Holland, Amsterdam, pp 113–139.

[5] S. Boucheron, O. Bousquet, G. Lugosi and P. Massart, Moment inequalities for functions of independent random variables, *The Annals of Probability* **33** (2005), no. 2, 514–560.

[6] Fan Chung and Linyuan Lu, Concentration Inequalities and Martingale Inequalities: A Survey, *Internet Mathematics* **3**, no. 1, 79–127.

[7] E. G. Coffman Jr. and G. S. Lueker, *Probabilistic Analysis of Packing and Partitioning Algorithms*, Wiley & Sons, 1991.

[8] S. Dasgupta and A. Gupta, An elementary proof of the Johnosn-Lindenstrauss Lemma, *International Computer Science Institute*, TR-99-006, (1999).

[9] D. Dubhashi and D. Ranjan, Balls and bins: A study in negative dependence, *Random Structures and Algorithms* **13** (1998), 99–124.

[10] D. Dubhashi, J. Jonasson and D. Ranjan, Positive influence and negative dependence, *Combinatorics, Probability and Computing* **16** (2007), no. 1, 29–41.

[11] P. Hitczenko, Best constants in martingale version of Rosenthal's inequality, *The Annals of Probability* **18** (1990), no. 4, 1656–1668.

[12] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley- Interscience Series in Discrete Mathematics and Optimization (2000).

[13] S. Janson, K. Oleszkiewicz and A. Ruciński, Upper tails for subgraph counts in random grphs, *Israeli Journal of Mathematics* **142** (2004), 61–92.

[14] K. Joag-Dev and F. Proschan, Negative-association of random variables with applications, *Annals of Statistics* **11** (1983), 286–295.

[15] S. Janson, A. Ruciński, The deletion method for upper tail estimates, *Combinatorica* **24** (2004), no. 4, 615–640.

[16] J. Kahn, Asympotically good list-colorings, *Journal of Combinatorial Theory A* **73** (1996), 1–59.

[17] C. McDiarmid, "Concentration." In *Probabilistic Methods for Algorithmic Discrete Mathematics*, edited by M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, pp. 195–248, Algorithms and Combinatorics 16. Berlin: Springer, 1998

[18] J. Csirik, D. S. Johnson, C. Kenyon, J. B. Orlin, P. W. Shor, and R. R. Weber, On the Sum-of-Squares Algorithm for Bin Packing, *Thirty-Second Annual ACM Symposium on Theory of Computing (STOC)* 208–217, 2000. Journal version in *JACM* **53** (2006), no. 1, 1–65.

[19] J. H. Kim and V. Vu, Concentration of multivariate polynomials and its applications, *Combinatorica* **20** (2000), 417–434.

[20] A. Srinivasan, Distributions on level sets with applications to approximation algorithms, in *the Proc. of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)* 2001.

[21] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, *Publications mathématiques de l'I.H.É.S.* **81** (1995), 73–205.
[22] W. Rhee, Inequalities for the bin packing problem III, *Optimization* **29** (1994), 381–385.
[23] W. Rhee and M. Talagrand, A sharp deviation for the stochastic Traveling salesman problem, *Annals of Probability* **17** (1989), 1–8.
[24] E. Shamir and J. Spencer, Sharp concentration of the chromatic number of random graphs, *Combinatorica* **7** (1987), no. 1, 121–129.
[25] J. M. Steele, Probability Theory and Combinatorial Optimization, *CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM* (1997)
[26] S. Vempala, The random projection method, *DIMACS Series in Discrete Mathemtics and TCS*, volume 65 (2000).

Hamiltonicity problems in random graphs

Michael Krivelevich

Hamiltonicity has always been one of the most central and attractive subjects in the theory of random graphs, with a variety of impressive results achieved and ingenious arguments discovered during the last five decades. Its most important achievements include:

- If $p(n) = \frac{\ln n + \ln\ln n + \omega(n)}{n}$ for any function $\omega(n) \to \infty$, then a random graph $G(n,p)$ is almost surely Hamiltonian (Komlós and Szemerédi [9], Bollobás [4]);
- almost every random graph process $\tilde{G} = (G_i)_{i=0}^{\binom{n}{2}}$ is such that exactly at the very moment i the last vertex of degree 1 disappears, the graph G_i is Hamiltonian (Ajtai, Komlós and Szemerédi [1], Bollobás [4]);
- for every fixed $d \geq 3$, a random d-regular graph $G_{n,d}$ is almost surely Hamiltonian (Robinson and Wormald [12], [13]);
- if $p(n) = \frac{\ln n + O(\ln\ln n)}{n}$, then $G \sim G(n,p)$ contains almost surely $\lfloor \frac{\delta(G)}{2} \rfloor$ edge disjoint Hamilton cycles (Bollobás and Frieze [5]).

Recent years brought a surge of renewed interest in the subject, and quite a few interesting results and approaches appeared. In this survey talk I will discuss some of them, and will indicate briefly main ideas of the proofs. The results to be discussed include:

Hamiltonicity in (n,d,λ)-graphs

A graph $G = (V,E)$ is called an (n,d,λ)-graph if it has n vertices, is d-regular and all of its eigenvalues but the largest one are at most λ in their absolute values. The setting of (n,d,λ)-graphs supplies a convenient vehicle to study properties of regular – and typically pseudo-random – graphs. It has been proven that if G is an (n,d,λ)-graph and

$$\lambda \leq \frac{(\ln\ln n)^2}{1000 \ln n \cdot \ln\ln\ln n}\, d\,,$$

then G is Hamiltonian for large enough n (Krivelevich and Sudakov [11]).

Packing edge disjoint Hamilton cycles

- If $p = const$, then a random graph $G(n,p)$ contains almost surely $(1 - o(1))np/2$ edge disjoint Hamilton cycles (Frieze and Krivelevich [6]);
- If $p = \frac{(1+o(1))\ln n}{n}$, then $G \sim G(n,p)$ contains almost surely $\lfloor \frac{\delta(G)}{2} \rfloor$ edge disjoint Hamilton cycles (Frieze and Krivelevich [7]).

Local resilience w.r.t. Hamiltonicity

This is a relatively new concept introduced by Sudakov and Vu [14]. It is aimed to measure quantitatively the strength with which a graph G possesses certain property P. Applied to Hamiltonicity, the problem reads as follows: given a Hamiltonian graph G, what is the maximal $\Delta = \Delta(G)$ such that for every subgraph $H \subset G$ with $\Delta(H) \leq \Delta$, the subgraph $G - H$ is still Hamiltonian? The above parameter Δ is called then the *local resilience of G with respect to Hamiltonicity*. It is easy to see that for a nearly d-regular graph G, the local resilience of G is at most $(1 + o(1))d/2$, thus providing an easy upper bound of $np/2$ for a random graph $G(n,p)$. The following results have been obtained recently:

- if $p > \ln^4 n/n$, then almost surely the local resilience of $G(n,p)$ with respect to Hamiltonicity is $(1/2 - o(1))np$ (Sudakov and Vu [14]);
- There exist constants $C > 0$ and $\epsilon > 0$ such that if a graph G is generated according to $G(n,p)$ with $p \geq C \ln n/n$, then almost surely the local resilience of G with respect to Hamiltonicity is at least ϵCnp (Frieze and Krivelevich [7]);
- For every $\epsilon > 0$ and large enough constant d, the local resilience of a random d-regular graph $G_{n,d}$ with respect to Hamiltonicity is at least $(1/6 - \epsilon)d$ (Ben-Shimon, Krivelevich and Sudakov [3]).

Hamiltonicity in Achlioptas processes

Consider the following online random process with parameter $r = r(n)$. Start with an empty graph G_0 on n vertices. At round i, $i \geq 1$, r edges $e_{i1}, \ldots, e_{ir}$ are chosen uniformly at random from the set of currently missing edges $E(K_n) - E(G_{i-1})$ and are shown to an algorithm. The algorithm, based only on the edges of the current round and edges currently on the board, decides to keep one of the current edges e_{ij} and returns the rest to the pool. The goal is to choose edges so as to advance or alternatively to delay a given graph property P. The setting, clearly generalizing the usual random graph process (case $r = 1$) is meant to reflect and to investigate the power of multiple choices in online algorithms on random inputs, similarly to the classical "power of two choices" result of Azar, Broder, Karlin and Upfal [2] about the balls-into-bins question. The model was suggested by Dimitris Achlioptas, who asked whether for $r = 2$ there exists an online algorithm that almost surely avoids creating a giant component for sizably longer than $0.5n$ rounds of the usual random graph process.

For the case of advancing Hamiltonicity the following trivial lower bounds apply: (1) the number of rounds required is at least $((1 + o(1))n \ln n)/(2r)$ (to see a Hamilton cycle in the union of edges presented at the first r rounds; (2) the

number of rounds required is at least n, the number of edges in a Hamilton cycle. Krivelevich, Lubetzky and Sudakov proved [10] the following matching upper bounds:

- the sublogarithmic case $r = o(\log n)$: there is an algorithm constructing a Hamilton cycle almost surely in $((1 + o(1))n \ln n)/(2r)$ rounds;
- the superlogarithmic case: $r \gg \log n$: there is an algorithm constructing a Hamilton cycle almost surely in $n + o(n)$ rounds;
- the logarithmic case $r = \gamma \ln n$: at least $(1+1/(2\gamma))n$ rounds are almost surely required; there is an algorithm that almost surely produced a Hamilton cycle in at most $(3 + 1/\gamma)n$ rounds.

Maker-Breaker Hamiltonicity games

These are games between two players, called Maker and Breaker, who take turns in claiming unoccupied edges of a graph G, one edge at at turn. Maker wins iff he possesses a Hamilton cycle of his edges by the end of the game. Consider now the case where the board of the game G is generated as a random graph $G(n, p)$. Clearly, when $p(n)$ is below the threshold for Hamiltonicity, Maker typically loses, as G has no Hamilton cycle to begin with. In a joint work with Hefetz, Stojakovič and Szabó [8] we proved that if $p(n) \geq \frac{\ln n + (\ln \ln n)^c}{n}$ for some constant $c > 0$, then for almost every board $G \sim G(n, p)$ Maker wins the Hamiltonicity game. This shows that quite close to the Hamiltonicity threshold not only $G(n, p)$ is typically Hamiltonian, it is also robustly Hamiltonian – it is possible to construct a Hamilton cycle even when playing against an adversary.

References

[1] M. Ajtai, J. Komlós and E. Szemerédi, First occurence of Hamilton cycles in random graphs, in: *Cycles in Graphs*, North-Holland Math. Studies **115**, North-Holland, Amsterdam 1985, pp. 173–178.

[2] Y. Azar, A. Broder, A. Karlin and E. Upfal, Balanced allocations, *SIAM Journal on Computing* **29** (1999), 180–200.

[3] S. Ben-Shimon, M. Krivelevich and B. Sudakov, Local resilience of random and pseudo-random regular graphs, in preparation.

[4] B. Bollobás, The evolution of sparse graphs, in: *Graph Theory and Combinatorics*, Academic Press, London, 1984, pp. 35–57.

[5] B. Bollobás and A. Frieze, On matchings and Hamiltonian cycles in random graphs, in: *Random Graphs'83*, North-Holland Math. Studies **118**, North-Holland, Amsterdam 1985, pp. 23–46.

[6] A. Frieze and M. Krivelevich, On packing Hamilton cycles in ϵ-regular graphs, *Journal of Combinatorial Theory Series B* **94** (2005), 159–172.

[7] A. Frieze and M. Krivelevich, On two Hamilton cycle problems in random graphs, *Israel Journal of Mathematics* **166** (2008), 221-234.

[8] D. Hefetz, M. Krivelevich, M. Stojaković and T. Szabó, A sharp threshold for the Hamilton cycle Maker-Breaker game, *Random Structures and Algorithms* **34** (2009), 112–122.

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

[10] M. Krivelevich, E. Lubetzky and B. Sudakov, Hamiltonocity thresholds in Achlioptas processes, *Random Structures and Algorithms*, to appear.

[11] M. Krivelevich and B. Sudakov, Sparse pseudo-random graphs are Hamiltonian, *Journal of Graph Theory* **42** (2003), 17–33.
[12] R. Robinson and N. Wormald, Almost all cubic graphs are Hamiltonian, *Random Structures and Algorithms* **3** (1992), 117–125.
[13] R. Robinson and N. Wormald, Almost all regular graphs are Hamiltonian, *Random Structures and Algorithms* **5** (1994), 363–374.
[14] B. Sudakov and V. Vu, Local resilience of graphs, *Random Structures and Algorithms* **33** (2008), 409–433.

The probabilistic method in topology

Nati Linial

The probabilistic method has revolutionized all of discrete mathematics. It has had a great impact on other areas as well, e.g., on the study of normed spaces. Our hope is that similar ideas can as well be very beneficial to topology. The most obvious contact point is the theory of simplicial complexes. A *simplicial complex* is a very natural object for combinatorics and for geometry alike. From the combinatorial perspective, a simplicial complex $\mathcal{F}$ is a finite family of sets that is closed under taking subsets. Namely, if $A \in \mathcal{F}$ and $B \subseteq A$, then $B \in \mathcal{F}$ as well. Members of $\mathcal{F}$ are called *faces* or *simplices* and the *dimension* of the face $A \in \mathcal{F}$ is defined as $\dim(A) := |A| - 1$. The dimension of $\mathcal{F}$ is defined as the largest dimension of a face in $\mathcal{F}$. We say that $\mathcal{F} \subseteq 2^{[n]}$ has a *full k-dimensional skeleton* if every $A \subseteq [n]$ of cardinality $\leq k+1$ belongs to $\mathcal{F}$. A simplicial complex can be *realized* geometrically e.g., by associating to every face $A \in \mathcal{F}$ a (geometric or topological) simplex of dimension $\dim A$. In this way simplicial complexes provide a useful way of depicting geometric objects.

It is often mentioned that a graph is nothing but a one-dimensional simplicial complex. This simple observation was the starting point to our joint work with Roy Meshulam [3]. We interpret the $G(n,p)$ model of random graphs as a one-dimensional simplicial complex whose one-dimensional faces (= edges) are selected at random independently and with probability p. We have introduced a higher-dimensional analogue of this construction, $X_d(n,p)$, a model of a random d-dimensional complex on vertex set $[n]$. Such a complex has a full $(d-1)$-dimensional skeleton and every d-dimensional face is included in the complex independently and with probability p. It should be clear that $X_1(n,p)$ is nothing but $G(n,p)$. We hope to develop a theory that determines the typical properties of complexes in $X_d(n,p)$. The questions we investigate are, to a large extent inspired by the main findings in the $G(n,p)$ theory. Thus, the higher-dimensional analogue of graph connectivity is the vanishing of the $(d-1)$-st homology of a random complex from $X_d(n,p)$. The corresponding threshold was determined in [3]. Subsequent work by Meshulam and Wallach [5] has extended these results to homology with coefficients from any finite abelian group. The same problem remains open for homology with integer coefficients. The threshold for simple connectivity in the 2-dimensional case has been almost exactly determined by Babson, Hoffman and Kahle in [1].

This perspective leads to numerous interesting problems where fundamental facts and problems in graph theory suggest new problems in higher-dimensional complexes. A good illustration is provided by the quest for a higher-dimensional analogue of Cayley's formula for the number of labeled trees. One realizes that a spanning tree can be viewed as a column basis for the inclusion matrix of $[n] \times \binom{[n]}{2}$ matrix. The same question for the $\binom{[n]}{d} \times \binom{[n]}{d+1}$ inclusion matrix is at present widely open. In this context one should mention Kalai's work [2] which provides a beautiful *weighted* higher-dimensional Cayley's formula. In recent work with Meshulam and Rosenthal [4] we introduce some interesting new constructions of higher-dimensional analogues of spanning trees (so called $\mathbb{Q}$-hypertrees) and determine exactly when they are collapsible.

References

[1] Eric Babson, Christopher Hoffman, Matthew Kahle, The fundamental group of random 2-complexes, `http://arxiv.org/abs/0711.2704` .

[2] Gil Kalai, Enumeration of Q-acyclic simplicial complexes, *Israel J. Math.* **45** (1983), no. 4, 337–351.

[3] Nathan Linial, Roy Meshulam, Homological connectivity of random 2-complexes, *Combinatorica* **26** (2006), no. 4, 475–487.

[4] Nathan Linial, Roy Meshulam, Mishael Rosenthal, Sum complexes – a new family of hypertrees, `http://arxiv.org/abs/0903.1359` .

[5] R. Meshulam, N. Wallach, Homological connectivity of random k-dimensional complexes, `http://arxiv.org/abs/math.CO/0609773` .

Cutoff phenomena for random walks on random regular graphs

Eyal Lubetzky

(joint work with Allan Sly)

A finite ergodic Markov chain is said to exhibit *cutoff* if its distance from the stationary measure drops abruptly, over a negligible time period known as the *cutoff window*, from near its maximum to near 0. That is, one has to run the Markov chain until the cutoff point in order for it to even slightly mix, and yet running it any further would be essentially redundant.

Let (X_t) be an aperiodic irreducible Markov chain on a finite state space Ω with transition kernel $P(x, y)$ and stationary distribution π. The worst-case total-variation distance to stationarity at time t is defined by

$$d(t) \triangleq \max_{x \in \Omega} \| \mathbb{P}_x(X_t \in \cdot) - \pi \|_{\mathrm{TV}},$$

where $\mathbb{P}_x$ denotes the probability given $X_0 = x$, and where $\|\mu - \nu\|_{\mathrm{TV}}$, the *total-variation distance* of two distributions μ, ν on Ω, is given by

$$\|\mu - \nu\|_{\mathrm{TV}} \triangleq \sup_{A \subset \Omega} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

We define $t_{\text{MIX}}(\varepsilon)$, the total-variation *mixing-time* of (X_t) for $0 < \varepsilon < 1$, as

$$t_{\text{MIX}}(\varepsilon) \triangleq \min\{t : d(t) < \varepsilon\}.$$

Next, let $(X_t^{(n)})$ be a family of such chains, each with its corresponding worst-case total-variation distance from stationarity $d_n(t)$, its mixing-times $t_{\text{MIX}}^{(n)}$, etc. We say that this family of chains exhibits *cutoff* at time $t_{\text{MIX}}^{(n)}(\frac{1}{4})$ iff the following sharp transition in its convergence to stationarity occurs:

$$\lim_{n\to\infty} t_{\text{MIX}}^{(n)}(\varepsilon)/t_{\text{MIX}}^{(n)}(1-\varepsilon) = 1 \quad \text{for any } 0 < \varepsilon < 1. \tag{1}$$

The rate of convergence in (1) is addressed by the following: A sequence $w_n = o\big(t_{\text{MIX}}^{(n)}(\frac{1}{4})\big)$ is called a *cutoff window* for the family of chains $(X_t^{(n)})$ if for any $\varepsilon > 0$ there exists some $c_\varepsilon > 0$ such that for all n,

$$t_{\text{MIX}}^{(n)}(\varepsilon) - t_{\text{MIX}}^{(n)}(1-\varepsilon) \le c_\varepsilon w_n. \tag{2}$$

That is, there is cutoff at time $t_n = t_{\text{MIX}}^{(n)}(\frac{1}{4})$ with window w_n if and only if

$$t_{\text{MIX}}^{(n)}(s) = (1 + O(w_n))\, t_n = (1 + o(1))t_n \quad \text{for any fixed } 0 < s < 1,$$

or equivalently, cutoff at time t_n with window w_n occurs if and only if

$$\begin{cases} \lim_{\lambda\to\infty} \liminf_{n\to\infty} \; d_n(t_n - \lambda w_n) = 1, \\ \lim_{\lambda\to\infty} \limsup_{n\to\infty} d_n(t_n + \lambda w_n) = 0. \end{cases}$$

Although many natural families of chains are believed to exhibit cutoff, determining that cutoff occurs proves to be an extremely challenging task even for fairly simple chains, as it often requires the full understanding of the delicate behavior of these chains around the mixing threshold. Before reviewing some of the related work in this area, as well as the conjectures that our work addresses, we state a few of our main results.

The focus of this work is on random walks on a random regular graph, namely on $G \sim \mathcal{G}(n,d)$, a graph uniformly distributed over the set of all d-regular graphs on n vertices, for $d \ge 3$ and n large. This important class of random graphs has been extensively studied, among other reasons due to the remarkable expansion properties of its typical instance. One useful implication of these expansion properties is the rapid mixing of the corresponding *simple random walk* (SRW), the chain whose states are the vertices of the graph, and moves at each step to a uniformly chosen neighbor. Namely, the SRW on such a graph has a mixing time of $O(\log n)$ *with high probability* (**whp**), that is, with probability tending to 1 as $n \to \infty$.

Our first result establishes both cutoff and its optimal window for the SRW on a typical instance of $\mathcal{G}(n,d)$ for any $d \ge 3$ fixed. This settles conjectures of Durrett [2, Conjecture 6.3.5] and Peres [4] in the affirmative.

Theorem 1. *Let $G \sim \mathcal{G}(n,d)$ be a random regular graph for $d \ge 3$ fixed. Then* **whp**, *the* simple *random walk on G exhibits cutoff at $\frac{d}{d-2}\log_{d-1} n$ with a window*

of order $\sqrt{\log n}$. *Furthermore, for any fixed* $0 < s < 1$, *the worst case total-variation mixing time* **whp** *satisfies*

$$t_{\text{MIX}}(s) = \frac{d}{d-2}\log_{d-1} n - (\Lambda + o(1))\Phi^{-1}(s)\sqrt{\log_{d-1} n},$$

where $\Lambda = \frac{2\sqrt{d(d-1)}}{(d-2)^{3/2}}$ *and* Φ *is the c.d.f. of the standard normal.*

The essence of the cutoff for the SRW on a typical $G \sim \mathcal{G}(n,d)$ lies in the behavior of its counterpart, the non-backtracking random walk (NBRW), that does not traverse the same edge twice in a row (formally defined soon). Curiously, this chain also exhibits cutoff on $\mathcal{G}(n,d)$ **whp**, only this time the cutoff window is *constant*: (2) holds for $w_n = 1$ and c_ε logarithmic in $1/\varepsilon$:

Theorem 2. *Let* $G \sim \mathcal{G}(n,d)$ *be a random regular graph for* $d \geq 3$ *fixed. Then* **whp**, *the* non-backtracking *random walk on* G *has cutoff at* $\log_{d-1}(dn)$ *with a constant-size window. More precisely, for any fixed* $\varepsilon > 0$, *the worst case total-variation mixing time* **whp** *satisfies*

$$t_{\text{MIX}}(1-\varepsilon) \geq \lceil \log_{d-1}(dn) \rceil - \lceil \log_{d-1}(1/\varepsilon) \rceil,$$
$$t_{\text{MIX}}(\varepsilon) \leq \lceil \log_{d-1}(dn) \rceil + 3\lceil \log_{d-1}(1/\varepsilon) \rceil + 4.$$

Establishing the above theorems requires a careful analysis of the local geometry around typical pairs of vertices, via a Poissonization argument. Namely, we show that the number of edges between certain neighborhoods of two prescribed vertices is roughly Poisson. Similar arguments then allow us to formulate analogous results for the case of regular graphs of high degree, that is, $\mathcal{G}(n,d)$ where d is allowed to tend to ∞ with n, up to $n^{o(1)}$. In particular, this resolves a conjecture of Hildebrand [3] in a strong sense (from worst-case starting point rather than average one, and after replacing the $o(1)$ error-term by an additive 2).

Remarks:

- We have established the cutoff phenomenon for SRWs and NBRWs on almost every d-regular graph on n vertices, where $3 \leq d \leq n^{o(1)}$ (beyond which the mixing time is $O(1)$ and we cannot have cutoff). For both walks, we obtained the precise cutoff location and window:
  1. For the SRW, the cutoff point is **whp** at $\frac{d}{d-2}\log_{d-1} n$, and in fact, we obtained the *two* leading order terms of $t_{\text{MIX}}(s)$ for any fixed s.
  2. For the NBRW, cutoff occurs at $\log_{d-1}(dn)$ **whp** ($\frac{d}{d-2}$ times faster than the SRW) with an $O(1)$ window. Moreover, for large d, the entire mixing transition takes place within a 2-step cutoff window.
- In addition, we provided a randomized algorithm to approximate $t_{\text{MIX}}(s)$ of the SRW on an input d-regular graph, with a runtime of $\widetilde{O}(n \cdot t_{\text{MIX}}(s))$. One may thus test (in nearly linear time for typical graphs) whether the SRW on a given d-regular graph indeed exhibits the above mentioned sharp transition in its mixing.

- Finally, we provided explicit constructions of d-regular graphs (for any $d \geq 3$ fixed) where the SRW has cutoff at prescribed locations.
- It would be interesting to extend our results to any arbitrary family of expanders. While one may design such graphs where the SRW has no cutoff, such constructions seem highly asymmetric, and the following conjecture seems plausible (see also [1, Question 5.2]):

  Conjecture 1. *The* SRW *on any family of vertex-transitive expander graphs exhibits cutoff.*

- Similarly, recalling the above comparison of t_{MIX} of the SRW and the NBRW on random regular graphs, it would be interesting to extend this result to any family of vertex-transitive expander graphs.

References

[1] Jian Ding, Eyal Lubetzky, Yuval Peres, Total-variation cutoff in birth-and-death chains, *Probabability Theory and Related Fields*, to appear.

[2] Rick Durrett, *Random graph dynamics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2007 pp. x+212.

[3] Martin Hildebrand, Random walks on random simple graphs, *Random Structures and Algorithms* **8** (1996), no. 4, 301–318.

[4] Yuval Peres, American Institute of Mathematics (AIM) research workshop *"Sharp Thresholds for Mixing Times"*, (Palo Alto, December 2004). Summary available at `http://www.aimath.org/WWN/mixingtimes` .

The evolution of random planar graphs

Tomasz Łuczak

(joint work with Mihyun Kang)

We are concerned with the evolution of the random planar graph $G_{\text{pl}}(n, M)$, i.e., a graph chosen uniformly at random from the family $\mathcal{G}_{\text{pl}}(n, M)$ of all planar graphs with vertex set $\{1, 2, \ldots, n\}$ and M edges.

The studies of random graphs and maps on 2-dimensional surfaces were initiated in sixties by Tutte [6]. The enumerative and structural problems concerning random maps, i.e., *embedded* graphs on a surface, are nowadays well settled. In particular, the number of all planar maps was found already by Tutte [7]. Recently, the topological structure of scaling limits of random maps was determined and studied; see, for instance, [1, 2, 5].

On the other hand, similar problems for random graphs which are *embeddable* on a surface are still widely open. The asymptotic formula for the number of all planar graphs has been found only recently by Giménez and Noy [3]. They also determined the asymptotic size of the family $\mathcal{G}_{\text{pl}}(n, M)$ for $M = \alpha n$, where $\alpha \in (1, 3)$. For $\alpha < 1/2$ the size of $\mathcal{G}_{\text{pl}}(n, M)$ can be easily deduced from the behavior of the uniform random graph $G(n, M)$, but in the range $\alpha \in [1/2, 1]$ the asymptotic formula for $|\mathcal{G}_{\text{pl}}(n, M)|$ has not been known. As for the structure of the random planar graph the main reference here is a beautiful paper of McDiarmid,

Steger, and Welsh [4], who deduced many properties of the random planar graph from the asymptotic behavior of a certain Markov chain.

We combine the generating functions and combinatorial methods to estimate the number of planar random graphs $|\mathcal{G}_{\rm pl}(n, M)|$ for $n/2 \le M \le n + n^{2/3}$. We also identify two periods of the critical behavior in the evolution of $G_{\rm pl}(n, M)$. The first critical phenomenon, similar to the phase transition observed in the evolution of random forests, occurs when $M = n/2 + O(n^{2/3})$ and the dominant component emerges in $G_{\rm pl}(n, M)$. The other critical phase can be observed when $M = n+O(n^{3/5})$. At that time of the evolution of $G_{\rm pl}(n, M)$ the size of the largest component reaches $n - o(n)$ and, consequently, the rate of its growth significantly slows down.

References

[1] P. Chassaing and G. Schaeffer, Random planar lattices and integrated superBrownian excursion, *Probability Theory and Related Fields* **128** (2004), 161–212.

[2] J. F. Le Gall, The topological structure of scaling limits of large planar maps, *Invent. Math.* **169** (2007), 621–670.

[3] O. Giménez and M. Noy, Asymptotic enumeration and limit laws of planar graphs, *J. Amer. Math. Soc.* **22** (2009), 309–329.

[4] C. McDiarmid, A. Steger and D. Welsh, Random planar graphs, *Journal of Combinatorial Theory B* **93** (2005), 187–205.

[5] O. Schramm, Conformally invariant scaling limits: an overview and a collection of problems. In *International Congress of Mathematicians*, Madrid, 2006, Vol. I, pages 513-543. *Eur. Math. Soc.*, Zürich, 2007.

[6] W. T. Tutte, A census of planar triangulations, *Canad. J. Math.* **14** (1962), 21–38.

[7] W. T. Tutte, A census of planar maps, *Canad. J. Math.* **15** (1963), 249–271.

Assigning Papers to Referees

Kurt Mehlhorn

Refereed conferences require every submission to be reviewed by members of a program committee (PC) in charge of selecting the conference program. A main responsibility of the PC chair is to organize the review process, in particular, to decide which papers are assigned to which member of the PC. The PC chair typically bases her decision on input from the PC, her knowledge of submissions and PC members, or scores that are computed automatically from keywords provided by authors and PC members. From now on, we call PC members reviewers or referees. There are many software systems available that support the PC chair in her task; for example, EasyChair [8], HotCRP [7], Softconf [2], Linklings [1], CMT [4], and Websubrev [6]. Used in more than 1300 conferences in 2008 alone [9], EasyChair is currently the most popular conference management software. The system asks the reviewers to declare conflicts of interests and to rank the papers (for which the reviewer has no conflict of interest) into three classes: high interest, medium interest, and low interest. This process is called bidding. Based on this information, the system automatically computes an assignment that the PC chair can later review and modify accordingly. Creating an assignment from scratch by

hand is normally not feasible since many conferences get in excess of 500 submissions [3]. The talk will be based on the paper Assigning Papers to Referees [5] by Naveen Garg, Telikepalli Kavitha, Amit Kumar, Kurt Mehlhorn, and Julián Mestre. In this paper, we propose to optimize a number of criteria that aim at achieving fairness among referees/papers. Some of these variants can be solved optimally in polynomial time, while others are NP-hard, in which case we design approximation algorithms. Experimental results strongly suggest that the assignments computed by our algorithms are considerably better than those computed by popular conference management software.

References

[1] Linklings. `http://www.linklings.com/` .

[2] Sofconf. `http://www.softconf.com/` .

[3] P. Apers, Acceptance rates major database conferences, `http://wwwhome.cs.utwente.nl/~apers/rates.html` .

[4] S. Chaudhuri. Microsoft's academic conference management service, `http://cmt.research.microsoft.com/cmt/` .

[5] N. Garg, T. Kavitha, A. Kumar, K. Mehlhorn, and J. Mestre, Assigning Papers to Referees, `http://www.mpi-inf.mpg.de/~mehlhorn/ftp/RefereeAssignment.pdf` , 2008.

[6] S. Halevi. Websubrev. `http://people.csail.mit.edu/shaih/websubrev/` .

[7] E. Kohler. HotCRP. `http://www.cs.ucla.edu/~kohler/hotcrp/` .

[8] A. Voronkov, EasyChair. `http://www.easychair.org` .

[9] A. Voronkov. EasyChair – users. `http://www.easychair.org/users.cgi` .

Some recent results and some open problems concerning solving infinite duration games

Peter Bro Miltersen

Dante in Purgatory – a riddle: *There are seven terraces in Purgatory, indexed* $1, 2, 3, 4, 5, 6, 7$. *Dante enters Purgatory at terrace* 1. *Each day, if Dante finds himself at some terrace* $i \in \{1, 2, \ldots, 7\}$, *he must play a game of matching pennies against Lucifer: Lucifer hides a penny, and Dante must try to guess if it is heads up or tails up. If Dante guesses correctly, he proceeds to terrace* $i + 1$ *the next morning – if* $i + 1$ *is* 8, *he enters Paradise and the game ends. If, on the other hand, Dante guesses incorrectly, there are two cases. If he incorrectly guesses "heads", he goes back to terrace* 1 *the next morning. If he incorrectly guesses "tails" the game ends and Dante forever loses the opportunity of visiting Paradise. How can Dante ensure ending up in Paradise with probability at least* $3/4$*? How long should he expect to stay in Purgatory before the game ends in order to achieve this?*

The somewhat striking answer to this riddle is that it *is* possible for Dante to go to Paradise with probability at least $3/4$, but any strategy achieving this guarantee has the downside that it allows Lucifer to confine Dante to Purgatory

Research supported by Center for Algorithmic Game Theory, funded by the Carlsberg Foundation.

for roughly 10^{25} years (In comparison, the current age of the universe is less than 10^{11} years so even playing one move per nanosecond would not help Dante much.) Other strategies guarantee Dante to go to Paradise with probability at least 99% or 99.9999%, but he would have to be even more patient to play these. Details can be found in [5] which is one of three papers we surveyed in this talk, the others being [1] and [4]. All papers are concerned with solving *two-player*, *zero-sum*, *terminal-payoff*, *finite-state* games of *potentially infinite duration*, a class of games that has been the subject of formal study since Zermelo [6]. We consider algorithmically *solving* such games when they are given explicitly as graphs by which we mean computing optimal strategies (or near-optimal strategies, when optimal ones fail to exist). The three papers consider three subclasses of these games. In the talk, we described each and mentioned interesting unsolved open problems. The most important are repeated below.

- Andersson *et al.* [1] study *deterministic graphical games*, or "Awari-like games". These are turn-based games of perfect information with no stochasticity in the rules of the game. The paper shows that such games can solved in almost linear time. A main open question is whether such games can be solved in linear time, by a comparison based algorithm.

- Miltersen and Gurvich [4] study *simple stochastic games* or "Backgammon-like" games[1]. These can be viewed as deterministic graphical games augmented with coin or dice throws. The paper shows that solving these games are in fact equivalent to solving various other classes of two-player, zero-sum games, in particular mean-payoff and discounted-payoff games. A main open question, first stated by Condon [2] twenty years ago is whether simple stochastic games can be solved in polynomial time.

- Hansen, Koucky and Miltersen [5] study *deterministic graphical games* or "Poker-tournament-like" games. These add to the previous models simultaneous moves, i.e., the games are no longer turn-based. Everett [3] showed that while such games in general do not possess optimal strategies, they possess ϵ-optimal mixed strategies for any $\epsilon > 0$. These are the objects we consider computing. Hansen, Koucky and Miltersen exhibit a game, PURGATORY$_n$ (Dante's Purgatory described above is PURGATORY$_7$) with the property that representing ϵ-optimal strategies requires exponential space when standard (say, fraction) representation of probabilities is used. It is an open problem if some non-standard representation can be devised enabling representation and computation of ϵ-optimal strategies in polynomial space.

[1]Oliver Riordan pointed out to the author during the workshop that with the doubling option, backgammon is in fact not a finite-state game, so we mean backgammon without the doubling rule.

References

[1] Daniel Andersson, Kristoffer Arnsfelt Hansen, Peter Bro Miltersen, and Troels Bjerre Sørensen. Deterministic graphical games revisited, In Arnold Beckmann, Costas Dimitracopoulos, and Benedikt Löwe, editors, *CiE*, volume 5028 of *Lecture Notes in Computer Science*, pages 1–10. Springer, 2008.

[2] Anne Condon, The complexity of stochastic games, *Information and Computation* **96** (1992), 203–224

[3] H. Everett. Recursive games, In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games Vol. III*, volume 39 of *Annals of Mathematical Studies*. Princeton University Press, 1957.

[4] Vladimir Gurvich and Peter Bro Miltersen, On the computational complexity of solving stochastic mean-payoff games, *CoRR*, `abs/0812.0486` , 2008.

[5] Kristoffer Arnsfelt Hansen, Michal Koucky, and Peter Bro Miltersen, Winning concurrent reachability games requires doubly-exponential patience, In *Proceedings of LICS'09*, to appear.

[6] Ernst Zermelo, Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels, In *Proceedings of the Fifth International Congress of Mathematicians*, pages 501–504. Cambridge University Press, 1913.

Critical thresholds in bootstrap percolation

Robert Morris

Let G be a graph, $r \in \mathbb{N}$ be a positive integer, and $A \subset V(G)$ be a set of 'infected' sites. Bootstrap percolation (or r-neighbour bootstrap percolation) is the following deterministic process: at each time-step we infect sites with at least r already infected neighbours, and infected sites remain infected forever. To be precise, let $A_0 = A$, and

$$A_{t+1} \;=\; A_t \,\cup\, \big\{v \in V(G) \,:\, |\Gamma(v) \cap A_t| \geq r\big\}$$

for each $t \geq 0$. We write $[A] = \bigcup A_t$, and say A *percolates* if $[A] = V(G)$.

The bootstrap process is closely related to the Ising model, and was introduced around thirty years ago in the context of statistical physics [11]. The process has been most commonly studied on the torus $[n]^d$, with the elements of the set A chosen independently at random, with probability p, say. The main question is to determine the critical threshold for percolation,

$$p_c(G,r) \;:=\; \inf\big\{p \,:\, \mathbb{P}(A \text{ percolates}) \geq 1/2\big\}.$$

The groundwork in the area was laid by Aizenman and Lebowitz [1] and by Schonmann [24], and important breakthroughs were made by Cerf and Cirillo [9], and by Holroyd [17]. Building on the ideas in these papers, we have recently proved the following results, amongst others (see [3, 4, 5, 6, 15, 20, 21]).

Theorem 1 (M.)**.** *There exist constants $C > c > 0$ such that*

$$\frac{\pi^2}{18\log n} - \frac{C}{(\log n)^{3/2}} \;\leq\; p_c([n]^2, 2) \;\leq\; \frac{\pi^2}{18\log n} - \frac{c}{(\log n)^{3/2}}.$$

Let $\log_{(r)}$ denote an r-times iterated logarithm.

Theorem 2 (Balogh, Bollobás and M. (for $d = r = 3$), and Balogh, Bollobás, Duminil-Copin and M. (for all $d \geq r \geq 2$))**.** *Let $d, r \in \mathbb{N}$, with $d \geq r \geq 2$. Then there exists a constant $\lambda(d,r) > 0$ such that*

$$p_c([n]^d, r) \;=\; \left(\frac{\lambda(d,r) + o(1)}{\log_{(r-1)} n}\right)^{d-r+1}$$

as $n \to \infty$. Moreover, we can determine $\lambda(d,r)$ exactly.

The constant $\lambda(d,r)$ is expressed as an integral which we cannot solve. However, Holroyd [17] showed that $\lambda(2,2) = \pi^2/18$, and in [4] we show that $\lambda(d,2) = \frac{d-1}{2} + o(1)$, and $d\lambda(d,d) \to \frac{\pi^2}{6}$ as $d \to \infty$.

Balogh and Bollobás [2] were the first to consider the bootstrap process on the hypercube, and determined p_c up to a constant factor when $r = 2$. We have recently improved this result as follows. Let $x \approx 1.16577$ be the smallest positive root of the equation

$$\sum_{k=0}^{\infty} \frac{(-1)^k x^k}{2^{k^2-k} k!} \;=\; 0.$$

Theorem 3 (Balogh, Bollobás and M.)**.** *There exist constants $C > c > 0$ such that*

$$\left(\frac{16x}{n^2} + \frac{c \log n}{n^{5/2}}\right) 2^{-2\sqrt{n}} \;\leq\; p_c(Q_n, 2) \;\leq\; \left(\frac{16x}{n^2} + \frac{C(\log n)^2}{n^{5/2}}\right) 2^{-2\sqrt{n}}.$$

We showed moreover that

$$p_c([n]^d, 2) \;=\; \frac{4x + o(1)}{d^2}\, 2^{-2\sqrt{d \log_2 n}}$$

whenever $d \gg \log n$. We also studied the problem for majority percolation, and proved the following theorem.

Theorem 4 (Balogh, Bollobás and M.)**.**

$$p_c\big([2]^n, n/2\big) \;=\; \frac{1}{2} \;-\; \frac{1}{2}\sqrt{\frac{\log n}{n}} \;\pm\; O\left(\frac{\log\log n}{\sqrt{n \log n}}\right).$$

Moreover, if $d \geq (\log\log n)^{2+\varepsilon}$ then $p_c([n]^d, d) = \frac{1}{2} + o(1)$ as $n, d \to \infty$.

Let $p_c(\mathbb{Z}^d)$ denote the critical probability for (zero-temperature) Glauber dynamics on $\mathbb{Z}^d$, i.e., the smallest bias ($p > 1/2$) in the initial density of + spins leads to fixation (all vertices +) with probability 1. It is a folklore conjecture that $p_c(\mathbb{Z}^d) = 1/2$ for every $2 \leq d \in \mathbb{N}$. Using techniques from the proof of Theorem 4 (see [3]), I proved the following result.

Theorem 5 (M.)**.** $p_c(\mathbb{Z}^d) \to \frac{1}{2}$ *as $d \to \infty$.*

There are many open questions in the area. For example:

Question 1. *We know that*

$$p_c([n]^d, d) \;=\; \begin{cases} o(1), & \text{if } n \text{ is larger than a tower of 2s of height } d, \\ 1/2 + o(1), & \text{if } n \le 2^{2^{\sqrt{d}}}. \end{cases}$$

What can we say about the transition between these two extremes?

Question 2. *What is the critical threshold for other graphs? The bootstrap process has been studied on*

- *trees, by Balogh, Pete and Peres* [7]*, and by Fontes and Schonmann* [12]*,*
- G_d*, the random regular graph, by Balogh and Pittel* [8]*,*
- *other lattices, by Holroyd, Liggett and Romik* [19]*, and by Holroyd and Duminil-Copin* [18]*,*
- *the percolation cluster in* $\mathbb{Z}^2$*, by Gravner and McDonald* [16]*,*
- *a wide range of 'locally tree-like' regular graphs, by Balogh, Bollobás and Morris* [3]*.*

Theorem 3 gives sharp bounds on $p_c([n]^d, 2)$ whenever $d \gg \log n$. However, we can say almost nothing about the high dimensional case when $r = 3$, even for the hypercube.

Problem 1. *Determine* $p_c([2]^n, 3)$*.*

Solving the following problem would be a very useful step. It is also an interesting extremal question in its own right.

Problem 2. *It is easy to show that*

$$n \;\le\; m(n) \;:=\; \min\big\{|A| \,:\, A \subset [2]^n \text{ percolates with } r = 3\big\} \;\le\; \frac{1}{3}\binom{n}{2} + n.$$

Determine $m(n)$*. In particular, is it true that* $m(n) = \Theta(n^2)$*?*

Finally, we return to the Ising model.

Problem 3. *The following are the best-known bounds for Glauber dynamics on* $\mathbb{Z}^2$ *(due to Fontes, Schonmann and Sidoravicius* [13]*):*

$$\frac{1}{2} \;\le\; p_c(\mathbb{Z}^2) \;<\; 1 - \frac{1}{10^{10}}.$$

Improve the upper bound.

References

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

[2] J. Balogh and B. Bollobás, Bootstrap percolation on the hypercube, *Probability Theory and Related Fields* **134** (2006), 624–648.

[3] J. Balogh, B. Bollobás and R. Morris, Majority bootstrap percolation on the hypercube, *Combinatorics, Probability and Computing* **18** (2009), 17–51.

[4] J. Balogh, B. Bollobás and R. Morris, Bootstrap percolation in three dimensions, to appear in *Annals of Probability*.

[5] J. Balogh, B. Bollobás, H. Duminil-Copin and R. Morris, The sharp threshold for r-neighbour bootstrap percolation, in preparation.
[6] J. Balogh, B. Bollobás and R. Morris, Bootstrap percolation on the hypercube, in preparation.
[7] J. Balogh, Y. Peres and G. Pete, Bootstrap percolation on infinite trees and non-amenable groups, *Combinatorics, Probability and Computing* **15** (2006), 715–730.
[8] J. Balogh and B. Pittel, Bootstrap percolation on random regular graphs, *Random Structures and Algorithms* **30** (2007), 257–286.
[9] R. Cerf and E. N. M. Cirillo, Finite size scaling in three-dimensional bootstrap percolation, *Ann. Prob.* **27** (1999), 1837–1850.
[10] R. Cerf and F. Manzo, The threshold regime of finite volume bootstrap percolation, *Stochastic Proc. Appl.* **101** (2002), 69–82.
[11] J. Chalupa, P. L. Leath and G. R. Reich, Bootstrap percolation on a Bethe latice, *J. Phys.* C **12** (1979), L31–L35.
[12] L. R. Fontes and R.H. Schonmann, Bootstrap percolation on homogeneous trees has 2 phase transitions, *J. Statist. Phys.* **132** (2008), 839–861.
[13] L. R. Fontes, R. H. Schonmann and V. Sidoravicius, Stretched Exponential Fixation in Stochastic Ising Models at Zero Temperature, *Commun. Math. Phys.* **228** (2002), 495–518.
[14] J. Gravner and A. E. Holroyd, Slow convergence in bootstrap percolation, *Ann. Appl. Prob.* **18** (2008), 909–928.
[15] J. Gravner, A. E. Holroyd and R. Morris, Beyond metastability: the critical threshold for two-dimensionsal bootstrap percolation, in preparation.
[16] J. Gravner and E. McDonald, Bootstrap percolation in a polluted environment, *J. Statist. Phys.* **87** (1997), 915–927.
[17] A. E. Holroyd, Sharp Metastability Threshold for Two-Dimensional Bootstrap Percolation, *Probability Theory and Related Fields* **125** (2003), 195–224.
[18] A. E. Holroyd and H. Duminil-Copin, personal communication.
[19] A. E. Holroyd, T. M. Liggett and D. Romik, Integrals, partitions, and cellular automata, *Trans. Amer. Math. Soc.* **356** (2004), 3349–3368 (electronic).
[20] R. Morris, Glauber dynamics in high dimensions, submitted.
[21] R. Morris, The phase transition for bootstrap percolation in two dimensions, in preparation.
[22] S. Nanda, C. M. Newman and D. Stein, Dynamics of Ising spin systems at zero temperature, In *On Dobrushin's way* (*From Probability Theory to Statistical Mechanics*), eds. R. Minlos, S. Shlosman and Y. Suhov, *Am. Math. Soc. Transl.* (2) **198** (2000), 183–194.
[23] C. M. Newman and D. Stein, Zero-temperature dynamics of Ising spin systems following a deep quench: results and open problems, *Physica* A **279** (2000), 159–168.
[24] R. H. Schonmann, On the behaviour of some cellular automata related to bootstrap percolation, *Ann. Prob.* **20** (1992), 174–193.

A constructive proof of the general Lovász Local Lemma

Robin Moser
(joint work with Gábor Tardos)

The Lovász Local Lemma [1] is a powerful tool to non-constructively prove the existence of combinatorial objects meeting a prescribed collection of criteria that do not interleave to much. It is usually formulated in terms of probability theory in the following fashion.

Theorem 1. *Let $\mathcal{A} = \{A_1, A_2, \ldots, A_m\}$ be a finite collection of events in some probability space. For each i, let $\Gamma(A_i) \subseteq \mathcal{A}$ be a subset of events such that A_i*

is independent of the collection $\mathcal{A}\backslash(\Gamma(A_i) \cup \{A_i\})$. *If there exists an assignment* $x\colon \mathcal{A} \to (0,1)$ *of reals to the events such that for each* i,

$$\Pr[A_i] \leq x(A_i) \cdot \prod_{B \in \Gamma(A_i)} x(B)$$

then the probability that all events in $\mathcal{A}$ *are avoided is positive.*

Since the classical proof of this statement is non-constructive and the probability of avoiding all the events may be very small, it has for a long time remained an open question whether it is possible to efficiently construct an outcome of the experiment that does. There is a big number of applications of the lemma, among which a classical one is hypergraph 2-colouring: in the simplest case, if a hypergraph is k-uniform and each edge intersects at most $2^k/e - 1$ other edges, then there exists a 2-colouring of the hypergraph's vertices such that no edge becomes monochromatic. In 1991, Beck achieved a first breakthrough in the search question by providing an algorithm that non-monochromatically 2-colours a k-uniform hypergraph where each edge intersects at most $2^{k/48}$ other edges. Alon improved this to neighbourhoods of size $2^{k/8}$. Later, Srinivasan [4] found a method that copes with $2^{k/4}$ dependencies and we gave such methods for $2^{k/2}$ in [5] and $\mathcal{O}(2^k)$ in [6].

We now finally prove that the search problem is polynomial in the case of almost all known applications of the lemma. We try to be as general as possible. However, in order to make the problem algorithmically accessible at all, we need to make a natural assumption about the structure of a given application by introducing the notion of variables and slightly weakening the requirements for dependency. We prove the following statement.

Theorem 2. *Let* $\mathcal{P} = \{P_1, P_2, \ldots, P_n\}$ *be a finite set of independent random variables over some probability space. Let* $\mathcal{A} = \{A_1, A_2, \ldots, A_m\}$ *be a collection of events which are determined by* $\mathcal{P}$ *and let* $\mathrm{vbl}(A_i) \subseteq \mathcal{A}$ *denote the unique minimal set of variables by which* A_i *is determined. For each* i, *let* $\Gamma(A_i) := \{A_j \in A \mid A_j \neq A_i,\ \mathrm{vbl}(A_i) \cap \mathrm{vbl}(A_j) \neq \emptyset\}$. *Suppose there exists an algorithm that samples, for a given event* A_i, *new random values for all variables in* $\mathrm{vbl}(A_i)$, *we will say that it* resamples A_i, *and another one that checks whether an event occurs for given evaluations. Consider the randomized procedure that samples all variables/events at random once and then repeatedly picks any occurring event and resamples it until no events occur anymore. If there exists an assignment* $x\colon \mathcal{A} \to (0,1)$ *of reals to the events such that for each* i,

$$\Pr[A_i] \leq x(A_i) \cdot \prod_{B \in \Gamma(A_i)} x(B),$$

then that procedure finds an evaluation of the variables such that all events are avoided, resampling each event A_i *at most* $x(A_i)/(1-x(A_i))$ *times in expectation.*

The proof is mainly a matter of accurate bookkeeping and becomes, once the right notions have been chosen, very straightforward. We introduce the concept of

witness trees: for each step the algorithm performs, a "justification" can be given in the form of a tree representing all steps that have been executed before and which are "responsible" for the resampled event to occur at the time in question. From the information encoded in such a witness tree, large parts of the random input used can be reconstructed making each single witness tree unlikely to appear. Juxtaposing their small probability with the number of ways in which witness trees can be constructed yields an effective bound on the expected number of valid trees that can witness the resampling of each given event and therefore on the expected number of times each given event can be resampled.

We remark that it is possible to give a parallel version of the algorithm that runs in basically $\mathcal{O}(\log^2(m))$ time when attaching a processor to each event and making mild additional assumptions. Moreover, if we restrict ourselves to cases where the neighbourhood sizes are bounded by some constant, then we can derandomize the algorithm by listing all possible witness trees up to some size and then tailoring a sequence of evaluations of the variables which will prevent any of these trees from occurring when running the usual procedure.

References

[1] Paul Erdős and László Lovász, Problems and results on 3-chromatic hypergraphs and some related questions, In A. Hajnal, R. Rado and V.T. Sós, editors, *Infinite and Finite Sets* (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), volume II, pages 609–627. North-Holland, 1975.

[2] József Beck, An Algorithmic Approach to the Lovász Local Lemma, *Random Structures and Algorithms* **2** (1991), no. 4, 343–365.

[3] Noga Alon, A parallel algorithmic version of the local lemma, *Random Structures and Algorithms* **2** (1991), no. 4, 367–378.

[4] Aravind Srinivasan, Improved algorithmic versions of the Lovász Local Lemma, *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms (SODA)*, San Francisco, California, pp. 611–620, 2008.

[5] Robin A. Moser, Derandomizing the Lovász Local Lemma more Effectively, Eprint, `arXiv:0807.2120v2` , 2008.

[6] Robin A. Moser, A constructive proof of the Lovász Local Lemma, *Proceedings of the 41st annual ACM Symposium on Theory of Computing (STOC 2009)*, to appear. Available at `arXiv:0810.4812v2` , 2008.

Mean-field conditions for percolation on finite graphs

Asaf Nachmias

Let $\{G_n\}$ be a sequence of finite transitive graphs with vertex degree $d = d(n)$ and $|G_n| = n$. Denote by $\mathbf{p}^t(v, v)$ the return probability after t steps of the *non-backtracking* random walk on G_n. We show [3] that if $\mathbf{p}^t(v, v)$ has *quasi-random* properties, then critical bond-percolation on G_n behaves as it would on a random graph. More precisely, if

$$\limsup_n n^{1/3} \sum_{t=1}^{n^{1/3}} t\mathbf{p}^t(v, v) < \infty \,,$$

then the size of the largest component in p-bond-percolation with $p = \frac{1+O(n^{-1/3})}{d-1}$ is roughly $n^{2/3}$. In Physics jargon, this condition implies that there exists a scaling window with a *mean-field* width of $n^{-1/3}$ around the critical probability $p_c = \frac{1}{d-1}$.

A consequence of our theorems is that if $\{G_n\}$ is a transitive expander family with girth at least $(\frac{2}{3}+\epsilon)\log_{d-1} n$ then $\{G_n\}$ has the above scaling window around $p_c = \frac{1}{d-1}$. In particular, bond-percolation on the celebrated Ramanujan graph constructed by Lubotzky, Phillips and Sarnak [2] has the above scaling window. This provides the first examples of *quasi-random* graphs [1] behaving like random graphs with respect to critical bond-percolation.

References

[1] F. R. K. Chung, R. L. Graham and R. M. Wilson, Quasi-random graphs, *Combinatorica* **9** (1989), no. 4, 345–362.

[2] A. Lubotzky, R. Phillips, P. Sarnak, Ramanujan graphs, *Combinatorica* **8** (1988), no. 3, 261–277.

[3] A. Nachmias, Mean-field conditions for percolation on finite graphs, to appear in *Geometric and Functional Analysis* (GAFA) (2007).

Game-theoretic Analysis of Steganographic Security

Rüdiger Reischuk

1. Introduction

The aim of steganography is to hide secret messages in unsuspicious covertexts in such a way that the mere existence of a hidden message is concealed. The basic scenario assumes two communicating parties Alice (sender) and Bob (receiver) as well as an adversary Eve who is often also called a "warden" due to Simmons' [1] motivation of the setting as secret communication among prisoners. Eve wants to find out whether Alice and Bob are exchanging hidden messages among their covertext communication. The stegosystem has to satisfy two conditions – (a) *reliability*, i.e., the ability of Alice to effectively transmit secret information to Bob and (b) *security*, i.e., the ability to prevent Eve from distinguishing between original covertexts and modified stegotexts.

In the last years some advances have been made in the analysis of steganographic systems (see for example [2]). Using notions from cryptography such as *indistinguishability* and adapting them to a steganography scenario, Hopper et al. have shown that it is possible to construct stegosystems that are provably secure against passive and active attacks [3]. However, their construction has several drawbacks in terms of practicality, in particular a very low transmission rate. Dedić et al. have analysed a generalisation of the scheme to a larger number of bits per document [4]. They have shown that for a reliable and secure *black-box stegosystem* (i.e., one in which Alice has no knowledge whatsoever of the covertext channel), the number of sample documents drawn from the covertext channel grows exponentially in the number of bits embedded per document.

Here, we report on recent results on setting up a framework for a detailed analysis of stegosystems [5, 6]. Steganography can be modelled as a game between the stegoencoder and the warden. Alice selects an embedding strategy and Eve a procedure for steganalysis with the goal to maximize, resp. minimize the security- The inequality in knowledge between the encoder and the adversary about the covertext distribution assumed so far is not adequate to model typical situations when steganography is used in practice. In reality, Alice neither has zero nor full knowledge about the covertext channel, but rather something in between, since she has the option to choose which kind of covertext channels (pictures, texts, music, ...) to be used, and the warden has to cope with this choice. Therefore, we propose a more realistic model of steganography, called *grey-box steganography*, in which the encoder has *partial knowledge* of the covertext channel, and investigate the influence of different levels of knowledge.

2. Basic Notation and Definitions

Symbols u taken from an alphabet Σ are called *documents*. A finite concatenation of such documents $u_1||u_2||\ldots||u_\ell$ is a *communication sequence* or *covertext*. Typically, the document models a piece of data (e.g., a digital image or fragment of the image) while the communication sequence models the complete message sent to the receiver in a single communication exchange.

Definition (Channel)**.** *A channel $\mathcal{C}$ is a function that takes a history $\mathcal{H} \in \Sigma^*$ as input and produces a probability distribution $D_{\mathcal{H}}$ on Σ. A history $\mathcal{H} = s_1 s_2 \ldots s_m$ is legal if each subsequent symbol is obtainable given the previous ones, i.e., if $\Pr_{D_{s_1 s_2 \ldots s_{i-1}}}[s_i] > 0$ for all $i \leq m$. The min-entropy of the channel $\mathcal{C}$ is the value $\min_{\mathcal{H}} H_\infty(D_{\mathcal{H}})$ where the minimum is taken over all legal histories $\mathcal{H}$.*

A steganographic information transmission is thought of taking a covertext $C = c_1 \ldots c_\ell \in \Sigma^\ell$ and modifying it to a stegotext $S = s_1 \ldots s_\ell \in \Sigma^\ell$ such that S additionally encodes an independent message M. Let b denote the message encoding rate, i.e., (on average) a single stegodocument s_j encodes b bits of M. For this purpose we require the channel to be sufficiently random. We will assume that the covertext channel distribution has a sufficiently large min-entropy h, that is larger than b.

Definition (Stegosystem)**.** *In the following, let $n = \ell \cdot b$ denote the length of the messages to be embeded into covertexts. A stegosystem $\mathcal{S}$ for the message space $\{0,1\}^n$ is a triple of probabilistic algorithms $[SK, SE, SD]$ with the following functionality:*

- *SK is the key generation procedure that on input 1^n outputs a key K of length κ, where κ is a security parameter that may depend on n;*
- *SE is the encoding algorithm that takes a key $K \in \{0,1\}^\kappa$, a message $M \in \{0,1\}^n$, accesses the sampling oracle $EX_{\mathcal{C}}()$ of a given covertext channel $\mathcal{C}$ and returns a stegotext $S \in \Sigma^{n/b}$;*
- *SD is the decoding algorithm that takes K and S and returns a message M'.*

$\mathcal{S}$ is called a black-box stegosystem *if the algorithms SE, SD have no a priori knowledge about the distribution of the covertext channel and can obtain information about it only by querying the sampling oracle. The unreliability of $\mathcal{S}$ with respect to the covertext channel $\mathcal{C}$ is given by*

$$\mathtt{UnRel}_{\mathcal{C}}^{\mathcal{S}} := \max_{M\in\{0,1\}^n} \Pr[K \leftarrow SK(1^n),\ EX_{\mathcal{C}} : SD(K, SE(K, M)) \neq M] .$$

The time complexities of the algorithms SK, SE, SD are measured with respect to n, κ, and σ, where an oracle query is charged as one unit step. A stegosystem is *efficient* if its time complexites are polynomially bounded.

To measure the security of a stegosystem we have to estimate how likely an adversary, the warden W, can discover that the covertext channel is used for transmitting additional information? If we put no algorithmic restrictions on W (i.e., information-theoretic security) it is necessary that (1) the stegotext S lies in the support of the covertext channel, otherwise W could test S for membership in $\mathrm{supp}(\mathcal{C})$, and (2) the probability of producing a stegotext S equals the probability of drawing S according to $\mathcal{C}$. To simplify the analysis, we will assume that the distribution on the support is uniform. Thus, we concentrate on the problem how the encoder can learn the support of the channel and then uniformly generate stegotexts. Learning complex distributions will be another issue.

For security analyses with complexity theoretic restrictions these assumptions can be relaxed in such a way that W is assumed to be polynomially time-bounded. Thus, Alice has to make sure that a adversary cannot detect deviations from the two conditions above in polynomial time.

Definition (Chosen Hiddentext Attack)**.**
Let $\mathcal{S} = [SK, SE, SD]$ be a stegosystem and $\mathcal{C}$ be a covertext channel. In a chosen hiddentext attack for parameters (n, κ), W has access to two oracles:

- *a* reference *oracle $EX_{\mathcal{C}}()$ that he can query for samples from the covertext channel $\mathcal{C}$ and*
- *a* challenge *oracle CH that is randomly selected being either*
  - *OS – the oracle that for a randomly chosen key K of length κ and a message M of length n provided by W outputs the stegotext $S = SE(K, M)$, or*
  - *OC – the oracle that for a randomly chosen key K of length κ and the message M draws a covertext of length $|SE(K, M)|$ from the covertext channel $\mathcal{C}$.*

The task of W is to determine the nature of CH. We define his advantage *over random guessing for a given covertext channel $\mathcal{C}$ as $\mathtt{Adv}_{\mathcal{C}}^{\mathcal{S}}(W) : |\Pr[W^{OS} = 1] - \Pr[W^{OC} = 1]|$, where W^{OS} (resp. W^{OC}) means that the challenge oracle is the oracle OS (resp. OC) and $W^{OS} = 1$ means that W decides on "stegotext".*

W may have additional information, for example about the channel $\mathcal{C}$, that can help him to distinguish the random selection of the oracle. In the most favourite case, W possesses a complete specification of $\mathcal{C}$, in which case he does not need the reference oracle. Let d be a bound on the maximal amount of such information and call this the *description size* of W.

Definition (Steganographic Security against CHA)**.**
The insecurity *of a stegosystem $\mathcal{S}$ with respect to a covertext channel $\mathcal{C}$ and complexity bounds d, t, q, l is defined by* $\mathtt{InSec}^{\mathcal{S}}_{\mathcal{C}}(d,t,q,l) := \max_W\{\mathtt{Adv}^{\mathcal{S}}_{\mathcal{C}}(W)\}$*, where the maximum is taken over all adversaries W of description length d working in time at most t and making at most q queries of total length l bits to the challenge oracle CH.*

Insecurity *and* unreliability *of a stegosystem $\mathcal{S}$ with respect to a channel family $\mathcal{F}$ are given by* $\mathtt{InSec}^{S}_{\mathcal{F}}(d,t,q,l) := \max_{C\in\mathcal{F}} \mathtt{InSec}^{\mathcal{S}}_{\mathcal{C}}(d,t,q,l)$ *and* $\mathtt{UnRel}^{\mathcal{S}}_{\mathcal{F}} := \max_{\mathcal{C}\in\mathcal{F}} \mathtt{UnRel}^{\mathcal{S}}_{\mathcal{C}}$ *.*

3. Results

In [5] it is shown that depending on the learning complexity of the channel family, the complexity of membership tests and the complexity of the construction process, efficient and secure steganography is possible. Explicit constructions are given for the classes of monomials, decision lists and DNF-formulae.

The security level of a stegosystem derived from the notion of insecurity as given above is quite strong since the worst case channel is measured. In [6] we give examples of families of channels and corresponding stegosystems that have high insecurity, but still cannot be broken by an adversary almost everywhere. We develop alternative notions and investigate their relations. In particular, the following definition turns out to be useful.

Definition. *The* detectability on average *of a stegosystem $\mathcal{S}$ with respect to the channel family $\mathcal{F}$ is given as follows, where the maximum is taken over all (d,t,q,λ)-wardens W:*

$$\mathtt{AvgDetect}^{\mathcal{S}}_{\mathcal{F}}(d,t,q,\lambda) := \max_W \mathbf{E}_{\mathcal{C}\in\mathcal{F}}[\mathtt{Adv}^{\mathcal{S}}_{\mathcal{C}}(W)] \ .$$

We construct families of so called flat h-channels that are random or pseudorandom subsets of size 2^h of the document space Σ [4, 7], and show that detectability on average gives the right measure of security that one would expect intuitively in practice.

References

[1] G. Simmons, The Prisoners' Problem and the Subliminal Channel, In Chaum, D., ed.: *Advances in Cryptology*: Proc. CRYPTO'1983, New York, Plenum Press, 1984, 51–67.

[2] P. Moulin, J. O'Sullivan, Information-theoretic Analysis of Information Hiding, *IEEE Tr. Information Theory* **49** (2003), 563–593.

[3] N. Hopper, J. Langford, L. von Ahn, Provably Secure Steganography, *IEEE Tr. Computers* **58** (2009), 662–676.

[4] N. Dedić, G. Itkis, L. Reyzin, S. Russell, Upper and Lower Bounds on Black-box Steganography, *J. Cryptology* **22** (2009), 365–394.

[5] M. Liśkiewicz, R. Reischuk, U. Wölfel, Grey-box Steganography, Technical Report, Institut für Theoretische Informatik, Universität zu Lübeck, 2009.

[6] M. Liśkiewicz, R. Reischuk, U. Wölfel, Security Levels in Steganography, Technical Report, Institut für Theoretische Informatik, Universität zu Lübeck, 2009.

[7] O. Goldreich, S. Goldwasser, A. Nussboim, On the Implementation of Huge Random Objects, *Proc. 44. IEEE Symp. on Foundations of Computer Science* (FOCS'03), 2003, 68–79.

Susceptibility in inhomogeneous random graphs

Oliver Riordan
(joint work with Svante Janson)

Given a graph G with n vertices, its *susceptibility* $\chi(G)$ may be defined as the mean size of the component containing a random vertex:

$$\chi(G) = \frac{1}{n} \sum_{v \in V(G)} |\mathcal{C}(v)|, \tag{1}$$

where $\mathcal{C}(v)$ denotes the component of G containing the vertex v, and $|H|$ the number of vertices in a graph H. Equivalently, listing the components of G as $\mathcal{C}_i = \mathcal{C}_i(G)$, $i = 1, \dots, K$, we have

$$\chi(G) = \sum_{i=1}^{K} \frac{|\mathcal{C}_i|}{n} |\mathcal{C}_i| = \frac{1}{n} \sum_{i=1}^{K} |\mathcal{C}_i|^2. \tag{2}$$

Listing the components in decreasing order of size, breaking ties in any way, the *modified susceptibility* $\widehat{\chi}(G)$ is defined by

$$\widehat{\chi}(G) = \frac{1}{n} \sum_{i=2}^{K} |\mathcal{C}_i|^2, \tag{3}$$

where the largest component has been omitted from the sum. If the graph G contains a 'giant' component, containing a constant fraction of the vertices, then the sum defining $\chi(G)$ is typically dominated by this component, and it turns out to be more informative to study $\widehat{\chi}(G)$. In what follows the graph G will itself be random, so both $\chi(G)$ and $\widehat{\chi}(G)$ are random variables.

The quantities defined above are closely related to ones appearing in statistical physics, in particular in percolation: $\chi(G)$, or rather its expectation, is analogous to the expected size $\mathbb{E}(|C_0|)$ of the open cluster containing a given (or random) vertex, and $\widehat{\chi}(G)$ to $\mathbb{E}(|C_0|; |C_0| < \infty)$. These latter quantities have been extensively studied; see, for example, [5]. In contrast, not much rigorous work has been done for finite random graphs.

In both contexts, the behaviour of χ or $\widehat{\chi}$ is most interesting near the *phase transition* at which a giant component emerges, where these functions have singularities. For results for $G(n,p)$, see, for example, Durrett [8], or the very detailed results of Janson and Luczak [9]. One of the main motivations for studying susceptibility is that in more complicated models it can be used to give information about the phase transition, while being much simpler to calculate than the size of the largest component. For an example of this approach see Spencer and Wormald [10].

In the present work we study the susceptibility of the random graphs $G_n = G(n, \kappa)$ produced by the general inhomogeneous model of Bollobás, Janson and Riordan [2] and its generalizations in [3, 4]; these models generalize many sparse random graph models introduced earlier, and form a natural setting for the study of phase transitions in random graphs.

Without going into full details of the models, a key feature in all cases is that the model is defined using a *kernel* κ, i.e., a symmetric measurable function from $[0,1]^2$ to $\mathbb{R}$. In the model of [2], each vertex i has a *type* $x_i \in [0,1]$. These types, which must satisfy a certain uniform distribution assumption, may be deterministic or random. Given the types, G_n is simply the random graph in which edges are independent, and the probability of the edge ij is $\min\{\kappa(x_i,x_j)/n, 1\}$.

As in [2], we aim to relate the behaviour of G_n to that of the Poisson Galton–Watson branching process $\mathfrak{X}_\kappa$ associated to κ; this is the multi-type process in which the types of the children of a particle of type x form a Poisson process on $[0,1]$ with intensity measure $\kappa(x,y)\,\mathrm{d}y$. Writing $|\mathfrak{X}_\kappa|$ for the total number of particles in $\mathfrak{X}_\kappa$, the natural analogues of χ and $\widehat{\chi}$ in this context are

$$\chi(\kappa) = \mathbb{E}\,|\mathfrak{X}_\kappa|$$

and

$$\widehat{\chi}(\kappa) = \mathbb{E}\big(|\mathfrak{X}_\kappa| 1_{|\mathfrak{X}_\kappa|<\infty}\big).$$

Note that both expectations may be infinite.

One of our main results is that under either of two additional assumptions, that κ is bounded, or that the types x_i are independent and identically distributed, we then have $\chi(G_n) \xrightarrow{\mathrm{p}} \chi(\kappa)$ and $\widehat{\chi}(G_n) \xrightarrow{\mathrm{p}} \widehat{\chi}(\kappa)$. In the bounded case, what we prove is more general: we consider random graphs G_n with independence between the edges, where the matrices A_n specifying the edge probabilities are uniformly bounded and converge in a certain weak sense to κ. The sense of convergence here involves the *cut metric* of Borgs, Chayes, Lovász, Sós and Vesztergombi [6]; sequences G_n of this form were studied by Bollobás, Borgs, Chayes and Riordan [1].

Fixing a kernel κ, and considering $G_n = G(n, \lambda\kappa)$ for a real parameter $\lambda > 0$, as shown in [2] there is a phase transition at a certain value of λ, above which a giant component appears. Here we study the behaviour of $\chi(\lambda\kappa)$ and $\widehat{\chi}(\lambda\kappa)$ as functions of the parameter $\lambda \in (0,\infty)$, and in particular the behaviour at the threshold for existence of a giant component. We show that $\chi(\lambda\kappa)$, which may found by solving a certain linear equation, may be used to find the critical value of λ.

Finally, we illustrate our results by considering several specific models, including the CHKNS model of Callaway, Hopcroft, Kleinberg, Newman and Strogatz [7] and the closely related graphs introduced by Dubins.

Acknowledgements. Part of this work was carried out during the programme "Combinatorics and Statistical Mechanics" at the Isaac Newton Institute, Cambridge, 2008, where SJ was supported by a Microsoft fellowship, and part during a visit of both authors to the programme "Discrete Probability" at Institut Mittag-Leffler, Djursholm, Sweden, 2009.

References

[1] B. Bollobás, C. Borgs, J. Chayes and O. Riordan, Percolation on dense graph sequences, *Annals of Probability*, to appear. `arXiv:0701346` .

[2] B. Bollobás, S. Janson and O. Riordan, The phase transition in inhomogeneous random graphs, *Random Structures and Algorithms* **31** (2007), 3–122.

[3] B. Bollobás, S. Janson and O. Riordan, Sparse random graphs with clustering, Preprint (2008). arXiv:0807:2040 .
[4] B. Bollobás, S. Janson and O. Riordan, The cut metric, random graphs, and branching processes, Preprint (2009). arXiv:0901:2091 .
[5] B. Bollobás and O. Riordan, *Percolation*. Cambridge University Press, Cambridge, 2006, x + 323 pp.
[6] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós and K. Vesztergombi, Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing, *Advances in Math.* **219** (2008), 1801–1851.
[7] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman and S. H. Strogatz, Are randomly grown graphs really random? *Phys. Rev. E* **64** (2001), 041902.
[8] R. Durrett, *Random Graph Dynamics*. Cambridge University Press, Cambridge, 2007.
[9] S. Janson and M. Luczak, Susceptibility in subcritical random graphs, *J. Math. Phys.*, to appear. arXiv:0806.0252
[10] J. Spencer and N. Wormald, Birth control for giants, *Combinatorica* **27** (2007), 587–628.

On the Number of Triangles in a Random Graph

Alex Scott
(joint work with Atsushi Tateno)

For fixed $p \in (0,1)$, let $G \in \mathcal{G}(n,p)$ be a random graph and let $X(G)$ be the number of triangles in G. Then $\mu := \mathbb{E}\, X(G) = \Theta(n^3)$ and $\sigma^2 := \mathrm{Var} X(G) = \Theta(n^4)$. It is well known that $\widetilde{X} := (X-\mu)/\sigma$ converges in distribution to a standard normal distribution. In particular, a special case of results of Barbour, Karoński and Ruciński [1] shows that

$$\sup_{x\in\mathbb{R}} \big|\, \mathbb{P}[X < \mu + x\sigma] - \Phi(x) \big| = O(n^{-1/2}), \tag{1}$$

where Φ is a standard normal distribution function. While this gives a good global picture of the distribution of X, it does not tell us much about the probability that X takes particular values. For instance, it does not tell us anything about the probability that X is even.

Loebl, Matoušek and Pangrác [2] looked at the properties of X modulo q and proved the following.

Theorem 1 ([2])**.** *There are constants q_0, $C > 0$ such that if $q \in [q_0, C\log n]$ is prime then*

$$\max_k \big|\, \mathbb{P}(X \equiv k \bmod q) - 1/q \big| = o(1/q).$$

How far could such a result be extended? If $q = \Omega(\sigma) = \Omega(n^2)$, it follows from (1) that X cannot be asymptotically uniform (a standard normal distribution is not asymptotically uniform modulo r unless $r = o(1)$). Thus we must have $q = o(n^2)$. We prove the following.

Theorem 2. *Let $p \in (0,1)$ be fixed, and let $G \in \mathcal{G}(n,p)$ be a random graph. If $q(n) = o(n^2/\log n)$ is integer-valued, then*

$$\max_k \big|\, \mathbb{P}[X(G) \equiv k \bmod q] - 1/q \big| = o(1/q).$$

Note that we do not require q to be prime.

Our methods also yield more detailed local information: indeed, we are able to prove a local limit theorem for X.

Theorem 3. *Let $p \in (0,1)$ be fixed, and let $G \in \mathcal{G}(n,p)$ be a random graph. Then*

$$\max_k \left| \mathbb{P}[X(G) = k] - \phi_n(k) \right| = o(n^{-2}).$$

Here,

$$\phi_n(k) = \int_{(k-\mu-1/2)/\sigma}^{(k-\mu+1/2)/\sigma} \phi(x)\,dx$$

is the natural guess for $\mathbb{P}[X = k]$ given the central limit theorem for X.

Both results follow from a "smoothness theorem." Roughly speaking, this smoothness theorem asserts that X is (in a certain sense) very close to a sum of several independent random variables with nice distributions. In particular, over a large range, the distribution of X is extremely 'flat'. This gives us enough local information to prove the two theorems above.

We conjecture that analogous results will hold for the number of copies of any fixed subgraph H.

References

[1] A. D. Barbour, M. Karoński and A. Ruciński, A central limit theorem for decomposable random variables with applications to random graphs, *Journal of Combinatorial Theory B* **47** (1989), 125–145.

[2] M. Loebl, J. Matoušek and O. Pangrác, Triangles in random graphs, *Discrete Mathematics* **289** (2004), 181–185.

A Proof of Green's Conjecture Regarding the Removal Properties of Sets of Linear Equations

Asaf Shapira

A system of ℓ linear equations in p unknowns $Mx = b$ is said to have the *removal property* if every set $S \subseteq \{1,\ldots,n\}$ which contains $o(n^{p-\ell})$ solutions of $Mx = b$ can be turned into a set S' containing no solution of $Mx = b$, by the removal of $o(n)$ elements. Green [GAFA 2005] proved that a single homogenous linear equation always has the removal property, and conjectured that every set of homogenous linear equations has the removal property. In this paper we confirm Green's conjecture by showing that every set of linear equations (even non-homogenous) has the removal property. We also discuss some applications of our result in theoretical computer science, and in particular, use it to resolve a conjecture of Bhattacharyya, Chen, Sudan and Xie [3] related to algorithms for testing properties of boolean functions.

1. Introduction

The (triangle) removal lemma of Ruzsa and Szemerédi [15], which is by now a cornerstone result in combinatorics, states that a graph on n vertices that contains only $o(n^3)$ triangles can be made triangle free by the removal of only $o(n^2)$ edges. Or in other words, if a graph has asymptomatically few triangles then it is asymptotically close to being triangle free. While the lemma was proved in [15] for triangles, an analogous result for any fixed graph can be obtained using the same proof idea. Actually, the main tool for obtaining the removal lemma is Szemerédi's regularity lemma for graphs [17], another landmark result in combinatorics. The removal lemma has many applications in different areas like extremal graph theory, additive number theory and theoretical computer science. Perhaps its most well known application appears already in [15] where it is shown that an ingenious application of it gives a very short and elegant proof of Roth's Theorem [14], which states that every $S \subseteq [n] = \{1, \ldots, n\}$ of positive density contains a 3-term arithmetic progression.

Recall that an r-uniform hypergraph $H = (V, E)$ has a set of vertices V and a set of edges E, where each edge $e \in E$ contains r distinct vertices from V. So a graph is a 2-uniform hypergraph. Szemeredi's famous theorem [16] extends Roth's theorem by showing that every $S \subseteq [n]$ of positive density actually contains arbitrarily long arithmetic progressions (when n is large enough). Motivated by the fact that a removal lemma for graphs can be used to prove Roth's theorem, Frankl and Rödl [4] showed that a removal lemma for r-uniform hypergraphs could be used to prove Szemeredi's theorem on $(r+1)$-term arithmetic progressions. They further developed a regularity lemma, as well as a corresponding removal lemma, for 3-uniform hypergraphs thus obtaining a new proof of Szemeredi's theorem for 4-term arithmetic progressions. In recent years there have been many exciting results in this area, in particular the results of Gowers [6] and of Nagle, Rödl Schacht and Skokan [12, 13], who independently obtained regularity lemmas and removal lemmas for r-uniform hypergraph, thus providing alternative combinatorial proofs of Szemeredi's Theorem [16] and some of it generalizations, notably those of Furstenberg and Katznelson [5]. Tao [18] and Ishigami [9] later obtained another proof of the hypergraph removal lemma and of its many corollaries mentioned above. For more details see [7].

In this paper we will use the above mentioned hypergraph removal lemma in order to resolve a conjecture of Green [8] regarding the removal properties of sets of linear equations. Let $Mx = b$ be a set of linear equations, and let us say that a set of integers S is (M, b)-free if it contains no solution to $Mx = b$, that is, if there is no vector x, whose entries all belong to S, which satisfies $Mx = b$. Just like the removal lemma for graphs states that a graph that has few copies of H should be close to being H-free, a removal lemma for sets of linear equations $Mx = b$ should say that a subset of the integers $[n]$ that contains few solutions to $Mx = b$, should be close to being (M, b)-free. Let us start be defining this notion precisely.

Definition (Removal Property)**.** *Let M be an $\ell \times p$ matrix of integers and let $b \in \mathbb{N}^\ell$. The set of linear equations $Mx = b$ has the* removal property *if for every $\delta > 0$ there is an $\epsilon = \epsilon(\delta, M, b) > 0$ with the following property: if $S \subseteq [n]$ is such that there are at most $\epsilon n^{p-\ell}$ vectors $x \in S^p$ satisfying $Mx = b$, then one can remove from S at most δn elements to obtain an (M, b)-free set.*

Green [8] has initiated the study of the removal properties of sets of linear equations. His main result was the following:

Theorem 1 (Green [8])**.** *Any single homogenous linear equation has the removal property.*

The main result of Green actually holds over any abelian group. To prove this result, Green developed a regularity lemma for abelian groups, which is somewhat analogous to Szemerédi's regularity lemma for graphs [17]. Although the application of the group regularity lemma for proving Theorem 1 was similar to the derivation of the graph removal lemma from the graph regularity lemma, the proof of the group regularity lemma was far from trivial. One of the main conjectures raised in [8] is that a natural generalization of Theorem 1 should also hold (Conjecture 9.4 in [8]).

Conjecture 1 (Green [8])**.** *Any system of homogenous linear equations $Mx = 0$ has the removal property.*

Very recently, Král', Serra and Vena [10] gave a surprisingly simple proof of Theorem 1, which completely avoided the use of Green's regularity lemma for groups. In fact, their proof is an elegant and simple application the removal lemma for directed graphs [1], which is a simple variant of the graph removal lemma that we have previously discussed. The proof given in [10] actually extends Theorem 1 to any single non-homogenous linear equation over arbitrary groups. Král', Serra and Vena [10] also show that Conjecture 1 holds when M is a 0/1 matrix, which satisfies certain conditions. But these conditions are not satisfied even by all 0/1 matrices.

In this paper we confirm Green's conjecture for every homogenous set of linear equations. In fact, we prove the following more general result.

Theorem 2 (Main Result)**.** *Any set of linear equations $Mx = b$ has the removal property.*

After our paper appeared on the Arxiv we learned that independently of our work, Král', Serra and Vena managed to improve upon their results in [10, 11] and obtain a proof of Conjecture 1.

References

[1] N. Alon and A. Shapira, Testing Subgraphs in Directed Graphs, *Journal of Computer and System Sciences* **69** (2004), 354–382.

[2] T. Austin and T. Tao, On the testability and repair of hereditary hypergraph properties, manuscript, 2008.

[3] A. Bhattacharyya, V. Chen, M. Sudan and N. Xie, Testing linear-invariant non-linear properties, manuscript, 2008.
[4] P. Frankl and V. Rödl, Extremal problems on set systems, *Random Structures and Algorithms* **20** (2002), 131–164.
[5] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for commuting transformations, *J. Analyse Math.* **34** (1978), 275–291.
[6] T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, *Ann. of Math.* **166**, no. 3, (2007), 897–946.
[7] T. Gowers, Quasirandomness, counting and regularity for 3-uniform hypergraphs, *Combinatorics, Probability and Computing* **15** (2006), 143–184.
[8] B. Green, A Szemerédi-type regularity lemma in abelian groups, *GAFA* **15** (2005), 340–376.
[9] Y. Ishigami, A simple regularization of hypergraphs, `http://arxiv.org/abs/math/0612838` .
[10] D. Král', O. Serra and L. Vena, A combinatorial proof of the removal lemma for groups, `arXiv:0804.4847v1` .
[11] D. Král', O. Serra and L. Vena, A removal lemma for linear systems over finite fields, *Jornadas de Matematica Discreta y algortimica* 2008.
[12] B. Nagle, V. Rödl and M. Schacht, The counting lemma for regular k-uniform hypergraphs, *Random Structures and Algorithms* **28** (2006), 113–179.
[13] V. Rödl and J. Skokan, Regularity lemma for k-uniform hypergraphs, *Random Structures and Algorithms* **25** (2004), 1–42.
[14] K. F. Roth, On certain sets of integers, *J. London Math. Soc.* **28** (1953), 104–109.
[15] I. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, in *Combinatorics* (Keszthely, 1976), *Coll. Math. Soc. J. Bolyai* **18**, Volume II, 939–945.
[16] E. Szemerédi, Integer sets containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 299–345.
[17] 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.
[18] T. Tao, A variant of the hypergraph removal lemma, *Journal of Combinatorial Theory A* **113** (2006), 1257–1280.

Average-case analyses of Vickrey costs

Gregory Sorkin
(joint work with Prasad Chebolu, Alan Frieze and Páll Melsted)

1. The VCG auction mechanism

Suppose that in a graph, each edge is provided by an independent, selfish agent who incurs a cost for supplying it (or for allowing us to drive over it, transmit data over it, or whatever). This "private" cost, the price point at which the agent is neutral between selling the edge or not, is known only to herself. We wish to buy some structure, for example a path between two particular points, or a spanning tree, as cheaply as possible. An obvious "mechanism" to do this is to ask each agent the cost of her edge, find the cheapest structure, and pay each agent accordingly. The problem with this and many other mechanisms is that agents have an incentive to lie: by inflating her claimed cost, an agent may get more money.

Research by Alan Frieze partially supported by NSF grant DMS 6721878.

A Vickrey-Clarke-Groves (VCG) auction [23, 5, 11] is a cleverly designed "truthful" mechanism: assuming that the agents act without collusion, in a VCG auction it is in each agent's best interest to name her true cost. Under the same assumption, a VCG auction also maximizes "social welfare": the structure selected is the one that is genuinely cheapest (and so the least possible resource is consumed in road maintenance, data-server support, or whatever).

In a VCG auction, an "auctioneer" first finds a cheapest structure S^*, according to the edge costs $c(e)$ declared by the agents. (This might be a cheapest path, for example; VCG was first explicitly applied to the shortest-path problem in [20, 21].) For each edge $e \in S^*$ in this structure, the auctioneer pays the corresponding agent not the stated cost $c(e)$ of the edge, but a measure of the benefit it provides, namely the difference between what a cheapest structure would have cost if the edge were not present or had infinite cost, call it $c(S_e^\infty)$, and what the cheapest structure would have cost if the edge were free, call it $c(S_e^0)$. It is clear that neither of these terms depends on $c(e)$. An agent whose edge is not used, $e \notin S^*$, is not paid anything. It is well known and easily verified that if an edge e is used the amount paid for it is at least $c(e)$, that the auction is truthful, and (using the truthfulness property) that it maximizes social welfare.

2. Average-case analysis

Naturally, the VCG mechanism pays more than the cost of the cheapest structure, and unfortunately the overpayment can be arbitrarily large. In [3, 4] it is shown that any truthful mechanism has bad worst-case s–t path overpayment. One alternative to this pessimistic worst-case analysis is through real-world measurements of the VCG overpayment, and such a study appears in [9]. Another alternative, and the one we adopt here, is to compare the VCG cost with the minimum cost in an average-case setting. This was done for shortest paths in certain graphs in [18, 6, 13, 8]. We consider the expected VCG overpayment in three settings, in each of which the expected minimum cost is a classical result in the analysis of random structures.

2.1. **Shortest paths.** We first consider shortest paths in the complete graph K_n, or complete digraph $\vec{K}_n$, with i.i.d. exponential(1) edge weights, where exponential(1) denotes the exponential distribution with mean 1. (We use the terms edge weight, cost, or length interchangeably, and a shortest path is a cheapest path.) Janson [12] has shown that **whp** the distance between two vertices, say 1 and n, in this model is $(1+o(1))\log n/n$. We prove that the asymptotic expected Vickrey cost is twice as large.

Theorem 1. *Suppose that the edges of the complete graph K_n (respectively, digraph $\vec{K}_n$) have i.i.d. exponential mean-1 edge weights. Let* $\mathbf{E}(\mathrm{SP})$ *be the expected cost of a shortest path from* 1 *to* n*. Then*

$$\mathbf{E}(\mathrm{VCG}) \sim 2\,\mathbf{E}(\mathrm{SP}).$$

2.2. **Minimum Spanning Tree.** We next consider a minimum spanning tree of K_n with uniform $[0,1]$ edge weights. It was shown by Frieze [10] that the expected cost $\mathbf{E}(\mathrm{MST})$ of a minimum spanning tree on K_n satisfies $\lim_{n\to\infty} \mathbf{E}(\mathrm{MST}) = \zeta(3)$. Even though there is no nice expression for the exact expectation for finite n, we prove that the expected VCG cost is *exactly* (not just asymptotically) twice as large.

Theorem 2. *Suppose that the edges of the complete graph K_n have i.i.d. uniform $[0,1]$ edge weights. Let $\mathbf{E}(\mathrm{MST})$ be the expected cost of a minimum spanning tree. Then*

$$\mathbf{E}(\mathrm{VCG}) = 2\,\mathbf{E}(\mathrm{MST}).$$

2.3. **Assignment.** Finally, we consider the VCG cost of a perfect matching in a complete bipartite graph with random edge weights, known as the "random assignment problem". When the edge weights are i.i.d. exponential(1) random variables, Mézard and Parisi [15, 16, 17] gave a sophisticated mathematical physics argument, using the "replica method", that the minimum cost AP satisfies $\lim_{n\to\infty} \mathbf{E}(\mathrm{AP}) = \zeta(2) = \pi^2/6$. Aldous [1, 2] made this mathematically rigorous through reasoning about a "Poisson weighted infinite tree". For finite values of n, Parisi [22] conjectured the expected cost to be $\sum_{i=1}^{n} i^{-2}$, Coppersmith and Sorkin [7] extended the conjecture to cheapest cardinality-k assignments in $K_{m,n}$, and these results were proved simultaneously, by different methods, by Linusson and Wästlund [14] and Nair, Prabhakar and Sharma [19]. A beautiful short proof was later found by Wästlund [24].

As in the previous cases, we find that the expected VCG cost is twice the minimum cost asymptotically.

Theorem 3. *Suppose that the edges of the complete bipartite graph $K_{n,n}$ have i.i.d. exponential mean-1 edge weights. Let $\mathbf{E}(\mathrm{AP})$ be the expected cost of a minimum weight perfect matching. Then*

$$\begin{aligned}\mathbf{E}(\mathrm{VCG}) = \mathbf{E}(\mathrm{AP}) + n\Bigg(&\frac{1}{n-1} + \sum_{l=1}^{n-1} \frac{1}{l}\frac{n-l}{n} \\ &- \sum_{l=2}^{n-1} \frac{1}{l(l-1)} \sum_{i=0}^{l-1} \frac{n-i}{n} \prod_{j=i+1}^{l} \frac{(n-j)j}{(n-j+1)j-1}\Bigg) \\ \sim\ & 2\,\mathbf{E}(\mathrm{AP}).\end{aligned}$$

3. Open questions

The same question can be raised for any combinatorial optimization problem with random weights. Natural candidates for consideration include a minimum spanning arborescence (rooted tree with all arcs oriented away from the root) in the complete digraph $\vec{K}_n$ (the directed analogue of MST result), a minimum-weight perfect matching in K_n (the non-bipartite analogue of Random Assignment), and the symmetric or asymmetric Traveling Salesman Problem (in K_n or $\vec{K}_n$ respectively). The natural edge weight distributions are i.i.d. uniform $[0,1]$ or i.i.d.

exponential(1); asymptotically these will be equivalent, but one or other may be more convenient, and with luck one might give an exact (non-asymptotic) result like the one given here for MST. We conjecture that in these cases too the expected Vickrey cost is twice the expected minimum cost, asymptotically. Assuming this pattern holds, it would be most interesting to understand why it is so.

References

[1] David J. Aldous, Asymptotics in the random assignment problem, *Probability Theory and Related Fields* **93** (1992), 507–534.

[2] David J. Aldous, The $\zeta(2)$ limit in the random assignment problem, *Random Structures and Algorithms* **18** (2001), no. 4, 381–418.

[3] Aaron Archer and Éva Tardos, Frugal path mechanisms, *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (San Francisco, California), January 06-08 2002, pp. 991–999.

[4] Aaron Archer and Éva Tardos, Frugal path mechanisms, *ACM Trans. Algorithms* **3** (2007), no. 1, Art. 3, 22.

[5] E. H. Clarke, Multipart pricing of public goods, *Public Choice* **8** (1971), 17–33.

[6] Artur Czumaj and Amir Ronen, On the expected payment of mechanisms for task allocation (extended abstract), *Proceedings of the Fifth ACM Conference on Electronic Commerce*, 2004, pp. 252–253.

[7] Don Coppersmith and Gregory B. Sorkin, Constructive bounds and exact expectations for the random assignment problem, *Random Structures and Algorithms* **15** (1999), no. 2, 113–144.

[8] Abraham Flaxman, David Gamarnik, and Gregory B. Sorkin, First-passage percolation on a width-2 strip and the path cost in a VCG auction, *Proceedings of the Second International Workshop on Internet and Network Economics*, WINE 2006, Patras, Greece, December 15-17, 2006 (Paul G. Spirakis, Marios Mavronicolas, and Spyros C. Kontogiannis, eds.), LNCS, vol. 4286, Springer, 2006, pp. 99–111.

[9] Joan Feigenbaum, Christos Papadimitriou, Rahul Sami, and Scott Shenker, A BGP-based mechanism for lowest-cost routing, *PODC'02: Proceedings of the Twenty-First Annual Symposium on Principles of Distributed Computing* (New York, NY, USA), ACM Press, 2002, pp. 173–182.

[10] Alan Frieze, On the value of a random minimum spanning tree problem, *Discrete Applied Mathematics* **10** (1985), 47–56.

[11] Theodore Groves, Incentives in teams, *Econometrica* **41** (1973), no. 4, 617–631.

[12] Svante Janson, One, two, three $\log n/n$ for paths in a complete graph with random weights, *Combinatorics, Probability and Computing* **8** (1999), 347–361.

[13] David Karger and Evdokia Nikolova, Brief announcement: on the expected overpayment of VCG mechanisms in large networks, *PODC'05: Proceedings of the 24th Annual ACM Symposium on Principles of Distributed Computing* (New York, NY, USA), ACM Press, 2005, pp. 126–126.

[14] Svante Linusson and Johan Wästlund, A proof of Parisi's conjecture on the random assignment problem, *Probability Theory and Related Fields* **128** (2004), 419–440.

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

[16] M. Mézard and G. Parisi, Mean-field equations for the matching and the travelling salesman problems, *Europhys. Lett.* **2** (1986), 913–918.

[17] M. Mézard and G. Parisi, On the solution of the random link matching problems, *J. Physique Lettres* **48** (1987), 1451–1459.

[18] Milena Mihail, Christos Papadimitriou, and Amin Saberi, On certain connectivity properties of the Internet topology, *FOCS'03: Proceedings of the 44th Annual IEEE Symposium on*

Foundations of Computer Science (Washington, DC, USA), IEEE Computer Society, 2003, p. 28.

[19] Chandra Nair, Balaji Prabhakar, and Mayank Sharma, Proofs of the Parisi and Coppersmith Sorkin random assignment conjectures, *Random Structures and Algorithms* **27** (2005), no. 4, 413–444.

[20] Noam Nisan and Amir Ronen, Algorithmic mechanism design (extended abstract), *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing* (Atlanta, Georgia, United States), 1999, pp. 129–140.

[21] Noam Nisan and Amir Ronen, Algorithmic mechanism design, *Games Econom. Behav.* **35** (2001), no. 1-2, 166–196, Economics and artificial intelligence.

[22] Giorgio Parisi, A conjecture on random bipartite matching, *Physics e-Print archive*, `http://xxx.lanl.gov/ps/cond-mat/9801176` , January 1998.

[23] William Vickrey, Counterspeculation, auctions and competitive sealed tenders, *Journal of Finance* **16** (1961), 8–37.

[24] Johan Wästlund, An easy proof of the zeta(2) limit in the random assignment problem, *Electronic Journal of Probability*, to appear.

Synchrony and Asynchrony in Neural Networks

Angelika Steger

(joint work with Fabian Kuhn, Konstantinos Panagiotou and Joel Spencer)

The dynamics of large networks is an important and fascinating problem. Key examples are the Internet, social networks, and the human brain. In this paper we consider a model introduced by DeVille and Peskin [1] for a stochastic pulse-coupled neural network. The key property of their model, which they studied experimentally and by some non-rigorous estimates of the expected behavior, is that the network can exhibit both *synchronous* and *asynchronous* behavior, a property which is omnipresent in the human brain. Synchrony is achieved when there are massive interactions between the neurons, i.e., huge bursts that include a substantial fraction of the neurons. Alternatively, the network is in an asynchronous state when there are only few interactions between the neurons, and only small bursts occur.

The system of DeVille and Peskin consists of n identical neurons, which have k different levels of internal potential. Whenever a neuron has the largest possible potential, it *fires*, having the effect that the potential of each other neuron increases with some small probability p (typically, p will be such that in expectation only constantly many neurons are affected). Note that this may be the beginning of a huge chain reaction: after the first firing, other neurons may have reached the highest potential, and hence will also fire, which in turn could stimulate other neurons to fire. This chain reaction, which we will call a *burst*, ends at the first moment in which there is no neuron with a potential that is large enough. Then, the potential of all neurons that fired is reset to zero. On the other hand, if no neuron fires, a randomly selected neuron increases its potential by some fixed quantity. In this paper we analyze this model rigorously, thereby achieving two goals. Firstly, we are able to answer the questions of DeVille and Peskin about the actual parameter settings and thresholds for which changes between synchronous

and asynchronous behavior occur. In addition to that, we give a very precise picture for the dynamics of the system, depending on the values of the parameters p and k. Secondly, we kind of disprove an observation of DeVille and Peskin that one should expect spontaneous transitions between these two extreme states. In fact we show, that once the system is in any of the two configurations, it stays there with very high probability.

Before we state our results in detail we first try to convey some intuition. Assume for the moment (the unrealistic setting) that $k = 1$. Then each neuron that is promoted one level in the interburst mode starts to fire immediately. The following burst mode can then be viewed as follows. Think of a random graph $G_{n,p}$ (in the Erdős-Rényi sense, with n nodes/neurons and edge probability p). Then the neuron that fires is in some connected component of $G_{n,p}$, and the other neurons that will fire in this burst phase are exactly the other nodes in that component. That is, whether we will see a "big" burst (linear in n) or only "small" bursts (of size $o(n)$) depends on the relation of p and n. More precisely, by applying well-known results from random graph theory, see e.g. [2], we will see only small bursts if $p \ll 1/n$, while in the case $p \gg 1/n$ *huge* bursts also happen.

If $k > 1$ things get much more interesting. Clearly, we can still think of the neurons at the highest level $k-1$ forming a random graph with edge probability p, and we know that all neurons in the component that contains a firing neuron will also fire. However, due to the presence of the smaller levels also other neurons may fire. Assume, for example, that a neuron from level $k-2$ is promoted during a burst phase to level $k-1$. It then has to be integrated into the random graph – and by that it may combine two connected components into a larger one.

At this point the following should be plausible.

Theorem 1. *Suppose that $p = \beta k/n$, where $\beta > 1$. Then, for sufficiently large k, regardless of the starting configuration, we will observe with high probability after finitely many time steps a big burst in which $\Theta(n)$ neurons fired.*

Now let us consider the case $\beta < 1$. If we start with a configuration in which all levels contain roughly n/k neurons then nothing exciting is going to happen: we will probably never experience any big bursts and the system stays in the state where all levels contain roughly the same number of neurons. If on the other hand we start in a configuration in which all neurons are in the same level, say at level 0, then we show that also this type of state is *preserved*. That is, the neurons move "simultaneously" up towards level $k-1$, and only "tiny" bursts are observed. Once this level contains a sufficient number of neurons, a big burst starts – and it brings most neurons back to level 0; and a new cycle starts from the beginning. More precisely, we show the following.

Theorem 2. *There exists a $c > 0$ such that the following is true. Suppose that $p = \beta k/n$, where*

$$\beta \geq \frac{2^{5/4}}{(k \ln k)^{1/4}} \cdot \left(1 + \frac{c}{\sqrt{\ln k}}\right).$$

Then for k sufficiently large, when starting with all neurons in level 0, the system converges to a stable state where bursts of size at least $(1 - o(1)) \cdot n/2$ occur at least every kn time units.

In addition to this, we provide a very precise characterization of the stable state, and determine the asymptotic fraction of neurons that are involved in a big burst. Note that the above theorem only applies if $\beta < 1$ is not too small. The next results says that the lower bound from Theorem 2 is essentially sharp.

Theorem 3. *There exists a $c > 0$ such that the following is true. Suppose that $p = \beta k/n$, where*

$$\beta \leq \frac{2^{5/4}}{(k \ln k)^{1/4}} \cdot \left(1 - \frac{c}{\sqrt{\ln k}}\right).$$

Then, for k sufficiently large, the following holds with high probability. If the system starts with all neurons in level 0, then there will be a finite number of big busts, and after that the system will remain in an asynchronous state.

We prove these theorems by showing that the typical behavior of the system is actually very close to the trajectories described by an associated mean-field model. In particular, we prove that the trajectory of our system will deviate significantly from some deterministic trajectory with only exponentially small probability. A major difficulty that we have to overcome here is that the system jumps between two inherently different states: the burst and the interburst mode. While the system is in either of these two modes, by applying the so-called 'Differential Equation Method' (see Wormald [4] and Seierstad [3]) we obtain after some technical work that with probability very close to one the actual behavior is not far from a solution of a set of differential equations. However, handling the 'transitions' between the two modes is a challenging task, which required the development of new techniques.

References

[1] R. E. L. DeVille and C. S. Peskin, Synchrony and asynchrony in a fully stochastic neural network, *Bulletin of Mathematical Biology* **70**(6):1608–1633.

[2] S. Janson, T. Łuczak and A. Rucinski, *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[3] T. G. Seierstad. Stronger large deviation bounds for Wormald's differential equation method, Submitted, 2008.

[4] N. Wormald. The differential equation method for random graph processes and greedy algorithms, In M. Karonski and H.J. Proemel, editors, *Lectures on Approximation and Randomized Algorithms*, pages 73–155. 1999.

Hypergraph Ramsey problem

Benny Sudakov
(joint work with David Conlon and Jacob Fox)

The *Ramsey number* $r(s,n)$ is the least integer N such that every red-blue coloring of the edges of the complete graph K_N on N vertices contains a red K_s (i.e., a complete subgraph all of whose edges are colored red) or a blue K_n. Ramsey's theorem states that $r(s,n)$ exists for all s and n. Determining or estimating Ramsey numbers is one of the central problems in combinatorics. A classical result of Erdős and Szekeres, which is a quantitative version of Ramsey's theorem, implies that $r(n,n) \leq 2^{2n}$ for all n. Erdős showed using probabilistic arguments that $r(n,n) > 2^{n/2}$ for $n > 2$. Despite efforts by various researchers, the best known constant factors in the above exponents so far remain the same.

Off-diagonal Ramsey numbers, i.e., $r(s,n)$ with $s \neq n$, have also been intensively studied. For example, after several successive improvements, it is known that there are constants $c_1, \ldots, c_4$ such that $c_1 \frac{n^2}{\log n} \leq r(3,n) \leq c_2 \frac{n^2}{\log n}$, and for fixed $s > 3$,

$$c_3 \left(\frac{n}{\log n}\right)^{(s+1)/2} \leq r(s,n) \leq c_4 \frac{n^{s-1}}{\log^{s-2} n}. \tag{1}$$

Although already for graph Ramsey numbers there are significant gaps between the lower and upper bounds, our knowledge of hypergraph Ramsey numbers is even weaker. The Ramsey number $r_k(s,n)$ is the minimum N such that every red-blue coloring of the unordered k-tuples of an N-element set contains a red set of size s or a blue set of size n, where a set is called red (blue) if all k-tuples from this set are red (blue). Erdős, Hajnal, and Rado [5] showed that there are positive constants c and c' such that

$$2^{cn^2} < r_3(n,n) < 2^{2^{c'n}}.$$

They also conjectured that $r_3(n,n) > 2^{2^{cn}}$ for some constant $c > 0$ and Erdős offered a \$500 reward for a proof. Similarly, for $k \geq 4$, there is a difference of one exponential between the known upper and lower bounds for $r_k(n,n)$, i.e., $t_{k-1}(cn^2) \leq r_k(n,n) \leq t_k(c'n)$, where the tower function $t_k(x)$ is defined by $t_1(x) = x$ and $t_{i+1}(x) = 2^{t_i(x)}$.

The study of 3-uniform hypergraphs is particularly important for our understanding of hypergraph Ramsey numbers. This is because of an ingenious construction called the stepping-up lemma due to Erdős and Hajnal. Their method allows one to construct lower bound colorings for uniformity $k+1$ from colorings for uniformity k, effectively gaining an extra exponential each time it is applied. Unfortunately, the smallest k for which it works is $k = 3$. Therefore, proving that $r_3(n,n)$ has doubly exponential growth will allow one to close the gap between the

Research by David Conlon supported by a Junior Research Fellowship at St John's College, Cambridge. Research by Jacob Fox supported by an NSF Graduate Research Fellowship and a Princeton Centennial Fellowship. Research by Benny Sudakov partially supported by NSF CAREER award DMS-0812005 and by USA-Israeli BSF grant.

upper and lower bounds for $r_k(n,n)$ for all uniformities k. There is some evidence that the growth rate of $r_3(n,n)$ is closer to the upper bound, namely, that with four colors instead of two this is known to be true. Erdős and Hajnal constructed a 4-coloring of the triples of a set of size $2^{2^{cn}}$ which does not contain a monochromatic subset of size n. This is sharp up to the constant c. It also shows that the number of colors matters a lot in this problem and leads to the question of what happens in the intermediate case when we use three colors. The 3-color Ramsey number $r_3(n,n,n)$ is the minimum N such that every 3-coloring of the triples of an N-element set contains a monochromatic set of size n. In this case, Erdős and Hajnal have made some improvement on the lower bound 2^{cn^2}, showing that $r_3(n,n,n) \geq 2^{cn^2 \log^2 n}$. Here, we substantially improve this bound, extending the above mentioned stepping-up lemma of these two authors to show

Theorem 1. *There is a constant $c > 0$ such that $r_3(n,n,n) \geq 2^{n^{c \log n}}$.*

For off-diagonal Ramsey numbers, a classical argument of Erdős and Rado [6] from 1952 demonstrates that

$$r_k(s,n) \leq 2^{\binom{r_{k-1}(s-1,n-1)}{k-1}}.$$

Together with the upper bound in (1) it gives, for fixed s, that

$$r_3(s,n) \leq 2^{\binom{r_2(s-1,n-1)}{2}} \leq 2^{c\frac{n^{2s-4}}{\log^{2s-6} n}}.$$

Our next result improves the exponent of this upper bound by a factor of $\frac{n^{s-2}}{\operatorname{polylog} n}$.

Theorem 2. *For fixed $s \geq 4$ and sufficiently large n,*

$$\log r_3(s,n) \leq \Big(\tfrac{(s-3)}{(s-2)!} + o(1)\Big) n^{s-2} \log n.$$

Erdős and Hajnal [4] showed that $\log r_3(4,n) > cn$ using the following simple construction. They consider a random tournament on $[N] = \{1, \ldots, N\}$ and color the triples from $[N]$ red if they form a cyclic triangle and blue otherwise. Since it is well known and easy to show that every tournament on four vertices contains at most two cyclic triangles and a random tournament on N vertices with high probability does not contain a transitive subtournament of size $c' \log N$, the resulting coloring neither has a red set of size 4 nor a blue set of size $c' \log N$. In the same paper from 1972, they suggested that $\frac{\log r_3(4,n)}{n} \to \infty$. Here we prove the following new lower bound which implies this conjecture.

Theorem 3. *There is a constant $c > 0$ such that $\log r_3(s,n) \geq c\, sn \log\left(\frac{n}{s} + 1\right)$ for all $4 \leq s \leq n$.*

Despite the fact that Erdős [2] believed $r_3(n)$ is closer to $2^{2^{cn}}$, together with Hajnal [1], he discovered the following interesting fact about hypergraphs which maybe indicates the opposite. They proved that there are $c, \epsilon > 0$ such that every 2-coloring of the triples of an N-element set contains a subset S of size $s > c(\log N)^{1/2}$ such that at least $(1/2 + \epsilon)\binom{s}{3}$ triples of S have the same color. That is, this subset deviates from having density $1/2$ in each color by at least some

fixed positive constant. Erdős [3] further remarks that he would begin to doubt that $r_3(n)$ is double-exponential in n if one can prove that any 2-coloring of the triples of an N-set contains some set of size $s = c(\epsilon)(\log N)^\delta$ for which at least $(1-\epsilon)\binom{s}{3}$ triples have the same color, where $\delta > 0$ is an absolute constant. Erdos and Hajnal proposed [1] that such a statement may even be true with $\delta = 1/2$. Our first result shows that this is indeed the case.

Theorem 4. *For each $\epsilon > 0$ and ℓ, there is $c = c(\ell, \epsilon) > 0$ such that every ℓ-coloring of the triples of an N-element set contains a subset S of size $s = c\sqrt{\log N}$ such that at least $(1-\epsilon)\binom{s}{3}$ triples of S have the same color.*

By considering random ℓ-coloring of the triples, it is easy to see that this theorem is tight up to the constant factor c. Our result also demonstrates (at least for $\ell \geq 3$) that the maximum almost monochromatic subset that an ℓ-coloring of the triples must contain is much larger than the corresponding monochromatic subset. This is in striking contrast with graphs, where these two quantities have the same order of magnitude, as demonstrated by a random ℓ-coloring of the edges of K_N.

Another open problem from the 1989 paper of Erdős and Hajnal [1] asks whether one can exhibit a fixed hypergraph of density larger than $1/2 + \epsilon$ on $c\sqrt{\log N}$ vertices that occurs monochromatically. That is, can we find dense hypergraphs with small Ramsey numbers? We show that this is indeed the case by obtaining a new upper bound on the ℓ-color Ramsey number of a complete multipartite 3-uniform hypergraph. A *hypergraph* $H = (V, E)$ consists of a vertex set V and an edge set E, which is a collection of subsets of V. A hypergraph is *k-uniform* if each edge has exactly k vertices. For a k-uniform hypergraph H, the Ramsey number $r(H; \ell)$ is the minimum N such that every ℓ-coloring of the k-tuples of an N-element set contains a monochromatic copy of H. The *complete d-partite k-uniform hypergraph* $K_d^k(n)$ is the k-uniform hypergraph whose vertex set consists of d parts of size n and whose edges are all k-tuples that have their vertices in some k different parts. The number of vertices of $K_d^3(n)$ is dn and the number of edges in $K_d^3(n)$ is $\binom{d}{3}n^3 > (1-\frac{3}{d})\binom{dn}{3}$, i.e., it has edge density more than $1-\frac{3}{d}$. In particular, as d increases, the edge density of $K_d^3(n)$ tends to 1. Therefore, Theorem 4 is an immediate corollary of the following theorem.

Theorem 5. *The ℓ-color Ramsey number of the complete d-partite hypergraph $K_d^3(n)$ satisfies*

$$r(K_d^3(n); \ell) \leq 2^{\ell^{2r} n^2},$$

where $r = r_2(d-1; \ell)$ is the ℓ-color Ramsey number of the complete graph on $d-1$ vertices.

References

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

[2] P. Erdős, Problems and results on graphs and hypergraphs: similarities and differences, in Mathematics of Ramsey theory, *Algorithms Combin.* Vol. **5** (J. Nešetřil and V. Rödl, eds.) 12–28. Berlin: Springer-Verlag, 1990.

[3] P. Erdős, Problems and results in discrete mathematics, *Discrete Math.* **136** (1994), 53–73.

[4] P. Erdős and A. Hajnal, On Ramsey like theorems, Problems and results, Combinatorics (*Proc. Conf. Combinatorial Math.*, Math. Inst., Oxford, 1972) pp. 123–140, Inst. Math. Appl., Southend-on-Sea, 1972.
[5] P. Erdős, A. Hajnal and R. Rado, Partition relations for cardinal numbers, *Acta Math. Acad. Sci. Hungar.* **16** (1965), 93–196.
[6] P. Erdős and R. Rado, Combinatorial theorems on classifications of subsets of a given set, *Proc. London Math. Soc.* **3** (1952), 417–439.

Uncoordinated Two-Sided Matching Markets

Berthold Vöcking
(joint work with Heiner Ackermann, Paul W. Goldberg, Vahab S. Mirrokni and Heiko Röglin)

A matching is *stable* if it does not contain a *blocking pair*, that is, a pair of agents from different sides who can deviate from this matching and both benefit. Gale and Shapley [2] showed that stable matchings always exist and can be found in polynomial time. Besides their theoretical appeal, two-sided matching models have proved useful in the empirical study of many labor markets such as the National Resident Matching Program (NRMP). Since the seminal work of Gale and Shapley, there has been a significant amount of work in studying two-sided markets. See for example, the book by Knuth [4], the book by Gusfield and Irving [3], or the book by Roth and Sotomayor [5].

In many real-life markets, there is no central authority to match agents, and agents are self-interested entities. This motivates the study of *uncoordinated two-sided markets*, first proposed by Knuth [4]. Uncoordinated two-sided markets can be modeled as a game among agents of one side, which we call the *active* side. The strategy of each active agent is to choose one agent from the *passive* side, and stable matchings correspond to Nash equilibria of the corresponding games. In order to understand the behavior of the agents in these uncoordinated markets, it is interesting to analyze the dynamics that arise when agents play repeatedly better or best responses to the strategies of the other agents.

1. Better Response Dynamics

To the best of our knowledge, Donald Knuth was the first who suggested to consider Nash dynamics in two-sided markets. He showed that the better response dynamics can cycle [4]. This means, it is possible to start with a matching M_1 and to resolve some blocking pairs, leading to a sequence of matchings $M_1, M_2, \ldots, M_k$ with $M_k = M_1$. Hence, in the worst case the better response dynamics never stabilizes. This, however, assumes that blocking pairs are resolved in a certain order, which is not realistic in an uncoordinated environment. Hence, he suggested to analyze the random better response dynamics. Roth and Vande Vate [6] proved that for every matching M, there exists a polynomial sequence of blocking pairs that lead to a stable matching when resolved consecutively. Hence, the random

better response dynamics reaches a stable matching in a finite number of steps with probability one.

This leaves open the question of how long it takes to stabilize. We believe that this is a crucial question as it corresponds to the question of how long an uncoordinated market needs to stabilize. In [1], we resolved this question and proved that there exists a family of two-sided markets such that the random better response dynamics takes a number of steps that is exponential in the size of the graph. This result indicates that coordination is necessary as there exist uncoordinated markets that need with high probability exponential time to stabilize.

2. Best Response Dynamics

Both Knuth's cycle [4], and Roth and Vande Vate's proof [6] hold only for the better response dynamics, and not for the *best response dynamics*, where women are activated and choose their best blocking pair. We extended these results to best responses. That is, we showed that also the best response dynamics can cycle and that starting from any matching, there exists a short sequence of best responses to a stable matching. As a corollary of the proof of the latter result, we obtain that every sequence of best responses starting with the empty matching reaches a stable matching after a polynomial number of steps. Hence, when starting with the empty matching, no central coordination is needed to reach a stable matching quickly if agents play only best responses.

In contrast to this, we showed that the result for random better responses can be extended to the random best response dynamics when arbitrary starting configurations are allowed. Hence, even if agents play only best responses, coordination is necessary if arbitrary initial matchings are allowed.

References

[1] H. Ackermann, P. W. Goldberg, V. S. Mirrokni, H. Röglin, and B. Vöcking, Uncoordinated two-sided matching markets, In *Proceedings of the 9th ACM Conference on Electronic Commerce* (EC) (2008), 256–263.

[2] D. Gale, and L. S. Shapley, College admissions and the stability of marriage, *American Mathematical Monthly* **69** (1962), 9–15.

[3] D. Gusfield, and R. W. Irving, *The stable Marriage Problem: Structure and Algorithms*, MIT Press (1989).

[4] D. E. Knuth, *Marriage Stables et leurs relations avec d'autres problèmes Combinatoires*, Les Presses de l'Université de Montréal, 1976.

[5] A. E. Roth, and M. A. O. Sotomayor, *Two-sided Matching: A study in game-theoretic modeling and analysis*, Cambridge University Press, 1990.

[6] A. E. Roth, and J. H. V. Vate, Random paths to stability in two-sided matching, *Econometrica* **58** (1990), no. 6, 1475–1480.

Tight Lower Bounds for Greedy Routing in Uniform Small World Rings

Philip Woelfel
(joint work with Martin Dietzfelbinger)

ABSTRACT

Motivated by Kleinberg's Small World Graph model and packet routing strategies in peer-to-peer networks, greedy routing algorithms on augmented networks have been investigated thoroughly. We prove tight lower bounds for one- and two-sided greedy routing on augmented rings.

1. Overview

An augmented network consists of a *base network* $G = (V, E)$ together with a probability distribution that for each vertex $v \in V$ defines probabilities of additional links from v to vertices $u \in V - \{v\}$ to be present. These additional links are called *long range contacts*. Often, the base network has a natural notion of "distance", and the probability distribution for the long range contacts is *uniform* in the sense that the probability for the existence of a long range contact from u to v does not depend on the labels of u and v, but only on the distance from u to v. Such uniform augmented networks were first considered by Watts and Strogatz [16] as a mathematical model to understand the "small-world phenomenon" occurring in social networks and the web. (The most prominent example of the phenomenon is the short chains of acquaintances, or "six degrees of separation" between any two individuals in the United States, as observed in Milgram's famous experiment [14].) Kleinberg [10, 11] started studying algorithmic aspects of such networks, in particular the cost of finding short paths greedily, i.e., by always following the edge that minimizes the distance to the destination. He proved that for uniform augmented meshes with a carefully chosen distribution for long range contacts, the path found by this greedy method has expected length $O((\log n)^2)$.

Since then, greedy routing and uniform augmented graph models have found a tremendous amount of interest (see the surveys [5, 12]). We focus on the one-dimensional case, where the underlying network is a one-dimensional grid or a one-dimensional ring. Here, deterministic constructions were given for $\ell = \Theta(\log n)$ long range contacts (Chord [6, 15]). Probabilistic constructions for $\ell \in [\Omega(1), O(\log n)]$ were proposed by Kleinberg [10, 11], in Symphony [13], and in Randomized-Chord [8, 17]. In all these constructions, the expected number of steps for greedy routing is $O((\log n)^2/\ell)$ (see also [1]). One distinguishes the "one-sided case" in which a token cannot change its direction on the ring and the "two-sided case" where this is allowed. Regarding upper bounds, no difference between these cases is known. It is a natural question to ask whether there are distributions for the long range contacts that allow faster greedy routing than the known upper bound $O((\log n)^2/\ell)$.

Lower bounds were obtained either for exactly ℓ or for an expected number of ℓ long range contacts per node. Aspnes, Diamadi, and Shah [1] offered a bound of $\Omega((\log n)^2/(\ell \log\log n))$ for the one-sided case, which was improved to $\Omega((\log n)^2/(\ell \cdot a^{\log^* n}))$, for some constant $a > 1$, by Giakkoupis and Hadzilacos [7]. Although not explicitly stated, a lower bound technique for black-box optimization in [2] implies a tight lower bound of $\Omega((\log n)^2)$ for one-sided greedy routing on the uniform augmented ring with $\ell = 1$ long range contact per node.

For the two-sided case, in [1] a lower bound of $\Omega((\log n)^2/(\ell^2 \log\log n))$ was shown, for distributions μ that obeyed certain, quite strict, monotonicity conditions. Under similar assumptions, Flammini et al. [4] showed an optimal lower bound of $\Omega((\log n)^2)$ for the diameter of undirected paths, where each node has exactly one long-range contact. (Note that the diameter is potentially larger than the expected greedy routing time.)

Aspnes, Diamadi and Shah conjectured a lower bound of $\Omega((\log n)^2/(\ell \log\log n))$ for both, the one- and the two-sided case; Giakkoupis and Hadzilacos [7] conjectured a lower bound of $\Omega((\log n)^2/\ell)$, which matches the known upper bounds for $\ell = O(\log n)$. Our contribution is a proof of the latter conjecture for both the two-sided and the one-sided case, with no restriction on the distribution μ at all, except that a node has an expected number of $\ell \in [\Omega(1), O((\log n)^2)]$ long range contacts.

2. The Main Results

In order to prove lower bounds for one- and two-sided greedy routing on the ring, we consider a random process in which a token is moved over a board with $2n+1$ cells, labelled from left to right with the numbers $-n, \ldots, n$. At the beginning, the token is placed in a location X_0 chosen uniformly at random from $\{1, \ldots, n\}$, and then it is moved in a series of steps towards its *target*, which is cell 0. In each step, a set $D \subseteq \{-n, \ldots, -1, 1, \ldots, n\}$ of *step sizes* is chosen at random according to a probability distribution μ on the set of all such step size sets. (The distribution remains the same throughout the game.) Then the token is moved from its current position X to position $X - d$, where d is the step size in D that minimizes the distance $|X - d|$ of the token to its target. If $|X - d| \geq |X|$ for all $d \in D$, then the token stays put.[1] Let T_2 denote the number of steps needed until the target is reached, and let $L = \mathbf{E}_\mu(|D|)$ be the average number of possible step sizes. We establish the following lower bound on the expected number $\mathbf{E}_\mu(T_2)$ of steps until the token reaches its destination. (This is asymptotically tight for $L = O(\log n)$.)

Theorem 1. *Let μ be an arbitrary probability distribution over the subsets of $\{-n, \ldots, -1, 1, \ldots, n\}$ and let $L = \mathbb{E}_\mu(|D|) = O((\log n)^2)$. Then $\mathbb{E}_\mu(T_2) = \Omega((\log n)^2/L)$.*

[1] In the corresponding graph routing problem, only sets D that contain $\{-1, 1\}$ are considered (to capture the links of the base network). Thus, the token moves in every step until the target has been reached. Adding a constant number of elements to D does not change our asymptotic lower bounds.

For $L = \omega(\log n \cdot \log\log n)$, the lower bound $\Omega(\log n/\log L)$, which is larger than $\Omega((\log n)^2/L)$, can be proved rather easily by other means (see, e.g., Theorem 3 in [1]).

Applying Theorem 1 with $L = \ell + 2$ for $\Omega(1) \le \ell \le O((\log n)^2)$ we obtain a lower bound of $\Omega((\log n)^2/\ell)$ for two-sided greedy routing on a ring of size n, with uniform long range contacts chosen at random according to an arbitrary distribution μ, with an expected number of ℓ long range contacts at each vertex.

We state a second, analogous, theorem for the one-sided case. The process now starts on a randomly chosen cell X_0 in $\{1,\dots,n\}$. In each step, as long as the current cell X is not 0, a set $D \subseteq \{1,\dots,n\}$ is chosen, according to a distribution μ on all subsets of $\{1,\dots,n\}$, and one moves to cell $X-d$ for the largest $d \in D \cup \{0\}$ with $d \le X$. Let T_1 denote the number of steps it takes in the one-sided process until the token reaches cell 0.

Theorem 2. *Let μ be an arbitrary probability distribution over the subsets of $\{1,\dots,n\}$ and let $L = \mathbf{E}_\mu(|D|) = O((\log n)^2)$. Then $\mathbf{E}_\mu(T_1) = \Omega((\log n)^2/L)$.*

3. Our Proof Technique

As already observed in [1] (and in [2]), instead of studying the "token process" directly, it is advantageous to delay the decision on the starting point, and consider a new Markov chain $S_0, S_1, S_2, \dots$ that has intervals as states. An interval $S_t = \{a,\dots,b\}$ represents the information that the token is in one of the points of the interval, with equal probability. (For example, $S_0 = \{1,\dots,n\}$, because in the beginning the token is placed uniformly at random on a point in this set.) The new Markov chain is equivalent to the token process in the sense that the expected time to reach $\{0\}$ is the same as $\mathbf{E}_\mu(T_2)$.

A state (interval) S has "weight" $|S|$, its size. A very coarse measure for the progress made by a step from S to S' is the quotient $|S|/|S'| > 1$. In a run $S_0, S_1, S_2, \dots, S_{T_2}$ of the interval process the total progress, obtained by multiplying these quotients, must be n. Each step of the interval process, from state S to state S', say, that uses set D of step sizes, is *associated* with one $d \in D$ in some clever way. We use an "accounting method" approach as it is often used in amortized analysis. With each step we associate an abstract "cost" caused by this step, which takes S, S', D, and the associated step size d into account.

We then show that the *expected* cost of one step, starting in an arbitrary state S, is $O(1)$. Further, we show by a convexity argument that if in a run $S_0, S_1, S_2, \dots, S_T$ of the interval process much of the naive progress (shrinking interval lengths) is achieved, then the total cost (sum) of these steps in this run is $\Omega((\log n)^2/L)$. Overall we get that in every run the total cost is $\Omega((\log n)^2/L)$, and the expected cost incurred in one step is $O(1)$. Then it is intuitively clear, and it can be proved by a variant of Wald's equation from [9], that the expected number of steps must be $\Omega((\log n)^2/L)$.

For details of the proof, we refer to [3].

References

[1] J. Aspnes, Z. Diamadi, and G. Shah, Fault-tolerant routing in peer-to-peer systems, *Proc. 21st PODC*, pp. 223–232, 2002.

[2] M. Dietzfelbinger, J. E. Rowe, I. Wegener, and P. Woelfel, Tight bounds for blind search on the integers, *Proc. 25th STACS*, pp. 241–252, 2008.

[3] M. Dietzfelbinger, and P. Woelfel, Tight Lower Bounds for Greedy Routing in Uniform Small World Rings, *Proc. 41st STOC*, 2009. To appear.

[4] M. Flammini, L. Moscardelli, A. Navarra, and S. Pérennes, Asymptotically optimal solutions for small world graphs, *Proc. 19th DISC*, pp. 414–428, 2005.

[5] P. Fraigniaud, Small worlds as navigable augmented networks: Model, analysis, and validation, *Proc. 15th ESA*, pp. 2–11, 2007.

[6] P. Ganesan and G. S. Manku, Optimal routing in chord, *Proc. 15th SODA*, pp. 176–185, 2004.

[7] G. Giakkoupis and V. Hadzilacos. On the complexity of greedy routing in ring-based peer-to-peer networks, *Proc. 26th PODC*, pp. 99–108, 2007.

[8] P. K. Gummadi, R. Gummadi, S. D. Gribble, S. Ratnasamy, S. Shenker, and I. Stoica, The impact of DHT routing geometry on resilience and proximity, *Proc. SIGCOMM*, pp. 381–394, 2003.

[9] J. Jägersküpper, Algorithmic analysis of a basic evolutionary algorithm for continuous optimization, *Theoretical Computer Science*, 279:329–347, 2007.

[10] J. M. Kleinberg, Navigation in a small world, *Nature*, p. 845, 2000.

[11] J. M. Kleinberg, The small-world phenomenon: An algorithmic perspective, *Proc. 32nd STOC*, pp. 163–170, 2000.

[12] J. M. Kleinberg, Complex networks and decentralized search algorithms, *Proc. International Congress of Mathematicians*, Vol. III, pp. 1019–1044, 2006.

[13] G. S. Manku, M. Bawa, P. Raghavan, and V. Inc, Symphony: Distributed hashing in a small world, *Proc. 4th USENIX Symp. on Internet Technologies and Systems*, pp. 127–140, 2003.

[14] S. Milgram, The small-world problem, *Psychology Today*, 67(1):60–67, 1967.

[15] I. Stoica, R. Morris, D. Liben-Nowell, D. R. Karger, M. F. Kaashoek, F. Dabek, and H. Balakrishnan, Chord: a scalable peer-to-peer lookup protocol for internet applications, *IEEE/ACM Transactions on Networking*, 11:17–32, 2003.

[16] D. Watts and S. Strogatz, Collective dynamics of small-world networks, *Nature*, 393:440–442, 1998.

[17] J. Xu, On the fundamental tradeoffs between routing table size and network diameter in peer-to-peer networks, *Proc. 22nd INFOCOM*, 2003.

Reporter: Paul Balister

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
USA

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

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

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

Boris Bukh
Department of Mathematics
Princeton University
Fine Hall
Washington Road
Princeton , NJ 08544
USA

Dr. Amin Coja-Oghlan
University of Edinburgh
Informatics Forum
10 Crichton Street
GB-Edinburgh EH8 9AB

David Conlon
St. John's College
Department of Mathematics
GB-Cambridge CB2 1TP

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

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

Jacob Fox
Department of Mathematics
Princeton University
Fine Hall
Washington Road
Princeton , NJ 08544
USA

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

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

Dr. Stefanie Gerke
Department of Mathematics
Royal Holloway College
University of London
Egham
GB-Surrey TW 20 OEX

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

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

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

Prof. Dr. Jeff Kahn
Department of Mathematics
Rutgers University
New Brunswick NJ 08903-2101
USA

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

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

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

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

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

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

Dr. Eyal Lubetzky
Microsoft Research
One Microsoft Way
Redmond , WA 98052-6399
USA

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

Prof. Dr. Yuri Matiyasevich
St. Petersburg Branch of Steklov
Mathematical Institute of
Russian Academy of Science
Fontanka 27
191023 St. Petersburg
RUSSIA

Prof. Dr. Kurt Mehlhorn
Max-Planck-Institut
für Informatik
Stuhlsatzenhausweg 85
66123 Saarbrücken

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

Prof. Dr. Robert Morris
Dept. of Pure Mathematics and
Mathematical Statistics
University of Cambridge
Wilberforce Road
GB-Cambridge CB3 0WB

Robin Moser
Institut für theoretische
Informatik
ETH-Zentrum
Universitätstr.6
CH-8092 Zürich

Dr. Asaf Nachmias
Theory Group
Microsoft Research
One Microsoft Way
Redmond , WA 98052
USA

Angelica Pachon Pinzon
Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld

Prof. Dr. Hans Jürgen Prömel
Technische Universität Darmstadt
Rundeturmstr. 12
64283 Darmstadt

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

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

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

Prof. Dr. Alex Scott
Merton College
Oxford University
GB-Oxford OX1 4JD

Prof. Dr. Asaf Shapira
School of Mathematics
Georgia Institute of Technology
686 Cherry Street
Atlanta , GA 30332-0160
USA

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

Prof. Dr. Vera T. Sos
Alfred Renyi Institute of Mathematics
Hungarian Academy of Sciences
P.O.Box 127
H-1364 Budapest

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

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

Prof. Dr. Endre Szemeredi
Department of Computer Sciences
Rutgers University
Piscataway , NJ 08855
USA

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

Prof. Dr. Gabor Tardos
School of Computing Science
Simon Fraser University
8888 University Drive
Burnaby , B.C. V5A 1S6
CANADA

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

Prof. Dr. Berthold Vöcking
Lehrstuhl für Informatik I
RWTH Aachen
Ahornstr. 55
52074 Aachen

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

Prof. Dr. Philipp Woelfel
University of Calgary
Department of Computer Science
ICT 602
2500 University Drive NW
Calgary , AB T2N 1N4
CANADA