Sparse Pseudorandom Objects

P. Haxell (Waterloo),
V. Rödl (Emory)

May 23–30 2010

1 Overview

It has been known for a long time that many mathematical objects can be naturally decomposed into a 'pseudorandom', chaotic part and/or a highly organized 'periodic' component. Theorems or heuristics of this type have been used in combinatorics, harmonic analysis, dynamical systems and other parts of mathematics for many years, but a number of results related to such 'structural' theorems emerged only in the last decades. A seminal example of such a structural theorem in discrete mathematics is Szeméredi's Regularity Lemma, which was discovered by Szeméredi in the mid-seventies when he proved his famous result on arithmetic progressions in dense subsets of natural numbers. It states that the set of edges of any dense graph can be 'nearly decomposed' into 'pseudorandom' bipartite graphs. The Regularity Lemma has long been recognised as one of the most powerful tools of modern graph theory.

The aim of the meeting was to follow this structural theme and investigate structural results for sparse combinatorial objects. The meeting brought together a number of experts in the area together with several junior researchers and PhD students.

2 Presentations and Discussions

Each presenter described recent developments on a particular topic, outlined some of the main related open problems, and led an interactive discussion on these results and problems. The topics addressed were as follows.

2.1 Extremal problems for random discrete structures (M. Schacht)

We study thresholds for extremal properties of random discrete structures. We determine the threshold for Szemerédi's theorem on arithmetic progressions in random subsets of the integers and its multidimensional extensions and we determine the threshold for Turán-type problems for random graphs and hypergraphs. In particular, we verify a conjecture of Kohayakawa, Łuczak, and Rödl for Turán-type problems in random graphs. Similar results were obtained by Conlon and Gowers.

2.2 Extremal Graph Theory – the Regularity Lemma Revisited (T. Łuczak)

For a graph H and natural numbers k and n let us define the parameter $\nu_{\chi}^{(k)}(H,n)$ [$\nu_{\tau}^{(k)}(H,n)$] as the smallest a such that each H-free graph G with n vertices and the minimum degree $\delta(G) \geq an$ can be homomorphically mapped to K_k [some H-free graph F on k vertices]. The behavior of these two parameters has been studied since the early seventies, for instance, the Andrásfai-Erdős-Sós Theorem determines the asymptotic behavior of $\nu_\chi^{(k)}(K_k, n)$ for large n. Here we propose to concentrate on computing the infimum of the sets $\{\nu_\chi^{(k)}(H, n)\}_k$ and $\{\nu_\tau^{(k)}(H, n)\}_k$ rather than finding their complete characterization. Thus, for a graph H, we study the parameters

$$\nu_\chi(H) = \inf_k \liminf_{n\to\infty} \nu_\chi^{(k)}(H, n)\,,$$

as well as

$$\nu_\tau(H) = \inf_k \liminf_{n\to\infty} \nu_\tau^{(k)}(H, n)\,.$$

and survey some recent results, open problems, and methods which fit the above framework.

In particular, using Szemerédi's Regularity Lemma we reprove the result of Thomassen from 2008 who, answering an old question of Erdős and Simonovits, showed that $\nu_\chi(C_{2k+1}) = 0$ for all $k \geq 2$. We also use the Regularity Lemma together with some local resilience argument to show that $\nu_\tau(K_k) = \frac{2k-5}{2k-3}$ for all $k \geq 3$.

This talk is based on joint work with Stéphan Thomassé.

2.3 Triangle removal lemma (Y. Person)

The theorem of Szemerédi states that for every $\epsilon > 0$ and every $k \in \mathbf{N}$ there exists $n_0 \in \mathbf{N}$ such that every set $A \subseteq [n]$ with $|A| \geq \epsilon n$, $n \geq n_0$, contains an arithmetic progression of length k. A special case of it is the theorem of Roth for arithmetic progressions of length 3. The best known lower bounds on ϵ in terms of n come from Fourier analytic proofs and the currently best lower bound is due to Bourgain, who showed $\epsilon = \frac{C(\log\log n)^2}{(\log n)^{2/3}}$ is enough. On the other hand, Ruzsa and Szemerédi observed more that 30 years ago that the so-called triangle removal lemma yields another, purely combinatorial, proof of Roth's theorem. This lemma states the following.

Triangle removal lemma. For every $\epsilon > 0$ there exists $\delta > 0$ such that if G is a graph on n vertices with at most δn^3 triangles, then one can remove at most ϵn^2 edges to make G triangle-free.

Until recently, the only known proof of this lemma was via Szemerédi's regularity lemma and therefore the dependency of δ^{-1} on ϵ is a tower of twos polynomial in ϵ^{-1}. Quite recently, Fox gave a new proof of the triangle removal lemma which avoids the use of the regularity lemma and shows that δ can be taken to be a tower of twos of height $200 \log \epsilon^{-1}$. Still, the dependency of ϵ and n for Roth's theorem is far from the result of Bourgain mentioned above, but Fox's proof suggests new perspectives and it gives better bounds for testing if a graph is triangle-free or is far from it.

In this talk I will discuss the ideas of Fox and present his proof.

2.4 Regular subgraphs (D. Dellamonica)

A paper of Pyber, Rödl and Szemerédi shows: (I) for any k there exists c_k such that any graph with maximum degree Δ and average degree d satisfying $d \geq c_k \log \Delta$ contains a k-regular subgraph. They also show: (II) the existence of graphs with Δ doubly exponential on d which do not contain 3-regular subgraphs. We discuss results and problems related to the following questions.

Question 1: Is $d \geq c_k \log \Delta$ the lowest lower bound possible in (I)?

Question 2: What are the structural properties of graphs which do not contain regular subgraphs? It is possible to show that graphs avoiding regular subgraphs necessarily contain subgraphs which are in some sense similar to the random construction establishing (II). However, in order to transfer properties of the random model used in (II) one would need a finer description of this structure.

2.5 Extremal problems for triple systems (D. Mubayi)

1. Regular substructures: Let $f(n)$ denote the maximum number of edges in a linear triple system on n vertices that contains no 2-regular subsystem. Here 2-regular means that every vertex lies in exactly two edges and linear means that every two edges have at most one point in common. About ten years ago, Verstraëte and I (using an idea of Lovász) proved that $n \log n < f(n) < 4n^{5/3}$. It appears that there have

been no improvements to either of the above bounds. There are absolutely no results on this problem for k-regular subsystems where $k > 2$.

2. Induced substructures: The induced Turán number $exind(n, F)$ of a k-uniform hypergraph is defined as follows. Let H_1 be a collection of k-element subsets of $[n]$ that are regarded as present and H_2 be a collection of k-element subsets of $\binom{[n]}{k} \setminus H_1$ that are considered as absent. Let $H_3 = \binom{[n]}{k} \setminus (H_1 \cup H_2)$. Suppose also that for every subset of k-subsets M of H_3, the k-graph $H_1 \cup M$ contains no induced copy of F. Then $exind(n, F)$ is the maximum size of H_3 subject to the restrictions above. This parameter is crucial to our understanding of the extremal theory of induced structures.

Let G_i be the (induced) 3-graph with four vertices and i edges. A few months ago I proved that $(1/9)n^2 < exind(n, G_1, G_4) < (1/6 - c)n^2$ for some positive c. I conjecture that the lower bound is tight. This is related to recent results of Razborov and Pikhurko about the usual Turán number of the family G_1, G_4. If the conjecture above is true, then one can attempt to prove a much more challenging conjecture (due to Balogh and me) that characterizes the structure of almost all 3-graphs with vertex set $[n]$ that contain no induced copy of G_1 or G_4. Such questions are related to results of Nagle, Rödl and others on counting hypergraphs with forbidden induced substructures.

2.6 Counting induced substructures (B. Nagle)

Let F be a k-graph on f vertices, and let H be k-graph on n vertices. In this talk, we consider the algorithmic problem of computing the number $\#(F, H)$ of (labeled, or unlabeled) induced copies of F in H. (The greedy algorithm can do this in time $O(n^f)$.) In 1986, Nešetřil and Poljak gave an algorithm for graphs for computing $\#(F, H)$ in time $O(n^e)$, where the exponent $e = \omega\lfloor f/3 \rfloor + r$ for remainder $r = f \pmod 3$. In 2005, Yuster studied the problem of computing $\#(F, H)$ for hypergraphs with $k \geq 3$, and conjectured that this quantity may be computed in time $o(n^f)$. In this talk, we present such an algorithm with running time $O(n^f / \log n)$. We formalize the problem that this running time should be reducible to $O(n^{f-\epsilon})$, for an absolute constant $\epsilon > 0$.

We also discuss a few approximation algorithms for estimating $\#(F, H)$. In 1992, Duke, Lefmann and Rödl showed that $\#(F, H)/n^f$ can be determined asympototically in time $O(n^2)$. In 2005, Haxell, Nagle and Rödl showed that $\#(F, H)/n^f$, for 3-graphs, can be determined asymptotically in time $O(n^6)$. Very recently, these results were extended to linear k-uniform hypergraphs by Nagle, Schacht and their graduate students.

3 Scientific Progress Made

All participants of the meeting worked together on three group projects. A question about what structure of a graph is forced when one knows an upper bound on the number of copies of a fixed tree was proposed by Schacht. The question of finding better bounds on $f(n)$, the maximum number of edges in a linear triple system on n vertices that contains no 2-regular subsystem, was proposed by Mubayi (see 2.5). The problem of estimating the extremal function $exind(n, G_1, G_4)$ (see 2.5) was also proposed by Mubayi. Here we describe in detail only the first of these projects, as progress on the others is still ongoing.

3.1 Trees force almost regular graphs

Sidorenko's conjecture in extremal graph theory due to Erdős and Simonovits [7] and Sidorenko [5, 6] asserts the following for every bipartite graph F and every $p > 0$. If an n-vertex graph G contains at least $p\binom{n}{2}$ edges, then the number of labeled copies of F in G is at least $(1 - o(1))p^{e_F} n^{v_F}$, where $o(1)$ tends to 0 as $n \to \infty$. This conjecture is known to be true for several classes of graphs including forests, even cycles, and complete bipartite graphs [6], Boolean cubes [4] and bipartite graphs F which contain a vertex that is connected to every vertex in the other vertex class [3]. A related and somewhat stronger conjecture was stated by Skokan and Thoma [8]. Those authors asked if every bipartite graph F which contains at least one cycle *forces* a graph G to be *quasi-random* if the number of labeled copies of F in an n-vertex graph G with at least $p\binom{n}{2}$ edges is at most $(1 + o(1))p^{e_F} n^{v_F}$. Here a graph G is quasi-random if it satisfies the properties considered in the Chung-Graham-Wilson theorem [2]. Since Sidorenko's conjecture is known to be true for trees and the Chung-Graham-Wilson theorem asserts a matching lower bound for the number of any graph F in a quasi-random graph G, a resolution of the *forcing conjecture* would yield a proof of Sidorenko's conjecture.

We studied a similar question when F is a tree. In view of the forcing conjecture for bipartite graphs F which contain a cycle, one may ask which structure or how much control over a graph G we can force by an upper bound on the number of labeled copies of a given tree T. We say a graph $G = (V, E)$ (more precisely a sequence of graphs $(G_n)_{n\in\mathbf{N}}$) is *nearly p-regular* for some $p > 0$, if

$$\sum_{v\in V} \big| \deg(v) - p|V| \big| = o(|V|^2).$$

It is easy to see for every fixed $\ell \in \mathbf{N}$ that if an n-vertex graph G is nearly p-regular, then for any tree F with ℓ edges the number $N_F(G)$of labeled copies of F in G satisfies

$$N_F(G) = (1 \pm o(1))p^\ell n^{\ell+1}\,.$$

At the workshop we obtained the opposite implication, stating that trees force the property of being nearly regular, i.e., every graph G with $N_F(G)$ being close to the minimal value must be nearly regular. As a consequence we obtain the following characterization of nearly regular graphs.

Theorem. Let $p > 0$ and let $(G_n = (V_n, E_n))_{n\in\mathbf{N}}$ be a sequence of graphs with $|E_n| \geq p\binom{|V_n|}{2}$. The sequence $(G_n)_{n\in\mathbf{N}}$ is nearly p-regular if and only if there exists some tree F with $\ell \geq 2$ edges such that $N_F(G_n) \leq (1 + o(1))p^\ell|V_n|^{\ell+1}$.

This theorem follows from a more precise lower bound estimate on $N_F(G)$ in terms of the degrees of G. For that we verify a counting formula for a graph $G = (V, E)$ of the following form

$$\hat{N}_F(G) \geq 2|E| \left(\prod_{v\in V} \deg(v)^{\frac{\deg(v)}{2|E|}} \right)^{\ell-1}, \tag{1}$$

where $\hat{N}_F(G)$ denotes the number of homomorphisms from F to G. Note that for dense graphs G we have

$$N_F(G) \leq \hat{N}_F(G) \leq N_F(G) + o(n^{v_F})\,.$$

It is easy to show that (1) is minimized when G is a regular graph and in this case we obtain $N_F(G) \geq (1 - o(1))p^\ell|V|^{\ell+1}$. Moreover, one can show that a "matching" upper bound on $N_F(G)$ forces the graph G to be nearly regular.

For the proof of (1) we extend the ideas of Alon, Hoory and Linial [1] who obtained the same formula for paths.

4 Outcomes of the Meeting

We expect several papers to result from the group projects outlined in the previous section. In addition, all the participants derived great benefit from being together in Banff and able to focus on fundamental problems related to spearse pseudo-random objects. We believe the meeting was of particular benefit to the young researchers in the group, and we close with some representative comments from one of the more junior participants.

"My research experience at the BIRS workshop 10frg131, Sparse pseudorandom objects, was extremely positive. This workshop brought together 8 researchers with sharp interests in pseudorandom graph and hypergraph theory for 6 full days of rich discussion and problem-solving. The workshop used an excellent mix of senior and junior researchers. As I consider myself still a junior researcher, I benefited tremendously from time spent with such talented and knowledgable experts, and also appreciated tremendously making collaborative ties with other young researchers whom I didn't previously know. I believe the insights I gained from my mentors, and the relationships I grew with my peers, will assist me invaluably in my career."

References

[1] N. Alon, S. Hoory, and N. Linial, The Moore Bound for Irregular Graphs, *Graphs Combin.* **18** (2002), no. 1, 53–57.

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

[3] D. Conlon, J. Fox, and B. Sudakov, An approximate version of Sidorenko's conjecture, preprint.

[4] H. Hatami, Graph norms and Sidorenko's conjecture, *Israel J. Math.*, to appear.

[5] A. F. Sidorenko, Inequalities for functionals generated by bipartite graphs, *Diskret. Mat.* **3** (1991), no. 3, 50–65.

[6] A. F. Sidorenko, A correlation inequality for bipartite graphs, *Graphs Combin.* **9** (1993), no. 2, 201–204.

[7] M. Simonovits, Extremal graph problems, degenerate extremal problems and super-saturated graphs, in: *Progress in graph theory* (Waterloo, Ont., 1982), Academic Press, Toronto, ON, 1984, 419–437.

[8] J. Skokan and L. Thoma, Bipartite subgraphs and quasi-randomness, *Graphs Combin.* **20** (2004), no. 2, 255–262.