Internat. Math. Nachrichten
Nr. 221 (2012), 1–19

The 2012 Abel laureate Endre Szemerédi and his celebrated work

Mihyun Kang
Technische Universität Graz

1 Introduction

The Hungarian-American mathematician Endre Szemerédi received the 2012 Abel Prize from His Majesty King Harald at the award ceremony in Oslo on 22 May 2012, "for his fundamental contributions to discrete mathematics and theoretical computer science, and in recognition of the profound and lasting impact of these contributions on additive number theory and ergodic theory" (a citation by the Abel Committee [1]).

On this occasion, Johannes Wallner, the editor of the International Mathematical News (*Internationale Mathematische Nachrichten*), asked me to write an article on Endre Szemerédi and his work. While I was preparing this article, I met Endre Szemerédi at a conference, during which Michael Drmota, the president of the Austrian Mathematical Society (Österreichische Mathematische Gesellschaft), suggested to me to interview Endre Szemerédi. As a result, this article contains an interview with Endre Szemerédi (Section 2) and a brief overview of his celebrated work (Section 3).

2 Interview with Endre Szemerédi

On behalf of the Austrian Mathematical Society, I interviewed the 2012 Abel laureate Endre Szemerédi. The interview took place on the 26th of June 2012

ISSN 0020-7926

during the Conference on “Perspectives in Discrete Mathematics” in Bellaterra near Barcelona, Spain.

Kang: Professor Szemerédi, I would like to congratulate you on winning the Abel Prize. When and where did you receive the message that you won the Abel Prize?

Szemerédi: That I know exactly. I was at home in Budapest. It was on the 21st of March 2012. One of my friends, Imre Bárány, was there too. It was a kind of a plot. It turned out later that the Norwegian ambassador in Hungary contacted Imre, asking him to make sure I would be at home that day at 10 o’clock in the morning. On Saturday, the 17th of March, Imre called me and said that he wanted to work with me on Wednesday, the 21st of March, on some mathematical problem and that the problem was hard, therefore it would be better to meet early. He insisted on 10 o’clock, even though this time is very early for me. Nevertheless he showed up and because I was not completely awaken yet he was chatting with my wife, Anna. Usually, I do not pick up the phone. Once, for example, when one of my granddaughters took the phone, she told the caller ‘My grandpa says that he is not at home’. At this time, Anna picked up the phone when the phone rang and tried to give it to me. As usual, I told her to say that I was not at home. But she told me that this time I could not play the game, as the call was from Oslo.

Kang: When you heard that the call was from Oslo, did you realize what was going on? Could you describe how you felt at that moment?

Szemerédi: Mathematicians know the day of the Abel Prize announcement. Also they know that the announcement is made in Oslo. I was almost certain that the call was about the Abel Prize.

When I heard the congratulation, first, I could not quite believe it. But I was very happy. Also I was slightly ashamed because I was not sure whether I deserved it while hundreds of others did not get the prize. It was a mixed feeling.

Kang: Among many beautiful theorems that you proved, there are ‘Szemerédi’s theorem’ and the ‘Szemerédi Regularity Lemma’. Let me first ask questions about Szemerédi’s theorem. Could you explain what Szemerédi’s theorem says?

Szemerédi: It says that if you have a positive proportion of the integers from 1 up to N it contains a long arithmetic progression.

Kang: Could you tell us a short mathematical history of Szemerédi’s theorem?

Szemerédi: The history is that van der Waerden proved his famous theorem, stating that if you divide the integers into finitely many classes, then one class contains arbitrarily long arithmetic progressions. Then Erdős-Turán conjectured in 1936 that the important thing is that the set is dense enough: if you have a positive proportion of the integers, then you have already long arithmetic progressions. More precisely speaking, for every positive integer $k \geq 3$ and every $\delta > 0$ there exists a number $S(k,\delta)$ such that for any $N \geq S(k,\delta)$, every subset of the integers $\{1,\ldots,N\}$ of cardinality at least δN contains an arithmetic progression of length k. In 1953 Roth provided a beautiful proof, using harmonic analytic methods, that the conjecture is true for $k = 3$. He proved that among at least $N/\log\log N$ integers there was always an arithmetic progression of length 3. Actually, one of my favourite mathematicians is Roth. When I first went abroad in 1967, he was the mathematician whom I met. I read his proof. But that's not the reason why I started to work on this problem.

Kang: How did you start working on the problem?

Szemerédi: It is a slightly embarrassing story. What happened is that I tried to prove that given a long arithmetic progression it cannot be that a positive percentage of the elements of the arithmetic progression is squares. In order to prove it, I took it for granted that if you have a positive percentage of the integers, then it contains an arithmetic progression of length 4. I proved that if you have an arithmetic progression of length 4, then it cannot be that all of them are squares. If you put these together, you prove what you wanted. I was very proud of 'my result'. I showed the proof to Paul Erdős. Then he told me that there are slight problems with the whole thing. The first one was that I assumed something, which had not been proved yet at that time, namely that any set of a positive percentage of the integers contains an arithmetic progression of length 4. But this was still okay. The second one was really shameful. Erdős told me that the other thing, stating that there are no four squares that form an arithmetic progression, was proved by Euler already 250 years ago. I felt that I must correct this mistake, because Erdős was the God.

Kang: So, Erdős was your God, your supervisor, your mentor, etc.?

Szemerédi: Yes, he was everything, although, without Paul Turán, I would have never been a mathematician. Because, when I went to university, the course was for teachers of mathematics and physics for the first two years. After that, you could continue to study mathematics and physics for three more years and get a diploma to teach in a high school or you could apply to be among the 20 out of 250 who would be chosen to continue as mathematicians. Turán, in the second year, gave a beautiful lecture on number theory for two semesters. He covered a lot of things. I liked it very much and decided then to become a mathematician. I tried to be among these 20.

Kang: Let us return to the story of Szemerédi's theorem. You said that after Erdős pointing out your mistakes, you were embarrassed. So, you tried to prove your master that you can do something correctly. Was this the proof of the Erdős-Turán conjecture for the case $k = 4$?

Szemerédi: Yes, I wanted to show him that I could prove it correctly and to see whether he can find an error. First I gave an elementary proof for the case $k = 3$. It was a very simple high-school proof. Then I proved it for $k = 4$. In 1967 Paul Erdős arranged for me an invitation to the university of Nottingham. There I was supposed to give a lecture about my proof. My English was practically non-existing. So I just drew some pictures, and Peter Elliot and Edward Wirsing, both number theorists, based on these pictures and my very bad English, wrote down the proof. I am very grateful for their great help. Similarly, when I proved in 1973 the conjecture for the general k, a good friend of mine, András Hajnal, helped me to write up the paper. It would be better to say that he listened to my explanations and then wrote it up. I am very grateful to him, too.

Kang: There are several other proofs of Szemerédi's theorem, for example by Furstenberg, Gowers and Tao amongst others. Could you explain what they proved, with what methods, and how their results are related to your theorem?

Szemerédi: Hillel Furstenberg is a great mathematician and works mainly on ergodic theory. His famous correspondence principle transfers the dense set into a measure space. Then he could use his multiple recurrence theorem. His method is much deeper and much more powerful than my elementary method, and could be generalised into a multi-dimensional setting. He and Yitzhak Katznelson could prove in 1978 a multi-dimensional analogue, and they could finally prove in 1991 the density version of the Hales-Jewett theorem. Their ergodic method is much more complex and powerful.

Timothy Gowers gave a much much better bound than what I had. Even more importantly, he invented many fundamental methods which completely changed the landscape. We cannot overestimate the influence of his paper. Gowers used a higher-order Fourier analysis and introduced his famous Gowers norm which controls the randomness of the set in question. Terence Tao mixes everything. He takes things from Furstenberg and Gowers. These mathematicians are those who really move this field. Without them, my theorem would be just a theorem, nothing more. They strengthened it, invented many directions, found connections between these things and much more. They do simply unbelievable things. Also, for me, the most striking result was the theorem of Ben Green and Terence Tao which states that among the primes there are arbitrarily long arithmetic progressions.

Kang: In literature, one can often read 'An important application of Szemerédi's theorem is the Green-Tao theorem on arithmetic progressions of primes'. Exactly how is your theorem used in the proof of the Green-Tao theorem?

Szemerédi: It would be hard to tell exactly.

Kang: Many influential mathematicians have their own lemma named after them. You also have one, the Szemerédi regularity lemma, which is unarguably the most powerful and commonly used tool in modern extremal graph theory. Could you explain what the Szemerédi regularity lemma says?

Szemerédi: If you don't want to have all these parameters, then it says that the vertex set of a dense graph can be partitioned into relatively small number of disjoint vertex sets, so that if you form bipartite graphs among these vertex sets, then almost every bipartite graph behaves like a random bipartite graph.

Kang: What does a typical extremal graph problem look like?

Szemerédi: That I really don't know. Just to tell you the truth, my mathematical interest has switched into number theory.

Kang: Is it a return to the origin, to Turán's lecture?

Szemerédi: Exactly, that's the idea.

Kang: One of the most important unresolved problems in mathematics, in particular in number theory, is the Riemann conjecture. Are you attacking it?

Szemerédi: No, I am trying some other problems in number theory, which I cannot say.

Kang: What is your favourite result or theorem of your own? Is it Szemerédi's theorem or the Szemerédi regularity lemma?

Szemerédi: My favourite work is the creation of the pseudo-random method together with Miklós Ajtai and János Komlós (it may be that the pseudo-random method existed in some other areas of mathematics). We used this method, when we tried to give a better estimate on the density of an infinite Sidon sequence. Using this method, we also disproved Heilbronn conjecture, which was at that time already 40 years old.

Kang: I would like to learn mathematical wisdom, from the one who has done excellent mathematics during the last at least fifty years. How do you recognise good mathematical problems? What defines a good problem?

Szemerédi: I am afraid I cannot give a really good answer to it. But, in my opinion, a good problem is one that is interesting, is difficult; you have to study a bit to solve the problem. And for the solution we need to invent a method which will be applied to many other problems and areas.

Kang: Like the Szemerédi regularity lemma!

Szemerédi: Actually we have a program to demolish it. Gyárfás, Sárközy, Rödl, Ruciński and myself have already written five papers without the Szemerédi regularity lemma. The Szemerédi regularity lemma is just a philosophy.

Kang: How do you develop a good taste for mathematics?

Szemerédi: It's hard. I am sure everybody has a different taste. You need experience, experiment and also luck.

Kang: Once you have selected a problem to solve, how do you start, proceed and finish the proof?

Szemerédi: Well, first you try to find connections, try to apply different methods, and then try to invent slightly different methods. Most of the time I am not successful. My ratio is very bad. I worked on a lot of problems, but I could not solve many of them. This is not so bad. It comes with the territory. There are mathematicians who have by far better ratio, but I do not mind. The only problem is that if I work on a problem, then I have a hard time to give it up, even when I am having a feeling that I cannot prove it, I do not have new ideas.

Kang: Could you give me a hint what your ratio would be?

Szemerédi: The ratio? It is about 1 to 10.

Kang: It is a good message to young mathematicians: we should try a lot!

Szemerédi: Yes, we should try a lot.

Kang: How or why did you discover or invent the Szemerédi regularity lemma? As far as I know, you have developed a weaker version of the Szemerédi regularity lemma, while proving Szemerédi's theorem.

Szemerédi: Yes, you are right. The paper about arithmetic progressions contains a weaker version of the regularity lemma.
The regularity lemma was needed for a graph-theoretical problem. I listened to a

lecture by Béla Bollobás, which was 1974 in Calgary. He talked about the Erdős-Stone theorem. Bollobás and Erdős wanted to determine the right magnitude and they had a very good bound, the order was $\log n$ and only the constant was missing a little bit. And I decided to work on it, as I liked the problem and Béla presents things very nicely and extremely cleverly. Then I realised that maybe a lemma like the regularity lemma would help considerably. Actually, that was the real reason why the regularity lemma was invented. At least, this is my recollection. I would like to thank Vašek Chvátal for helping me to write up the regularity lemma. Later we together found the exact constant for the Erdős-Stone theorem.

Kang: The title of your talk at the Abel lecture was 'In every chaos there is an order'. Was it your way of thinking when you developed the regularity lemma, or did you hear about this principle from someone else?

Szemerédi: This has already been used by many mathematicians, like in Ramsey theory. If you take a graph and colour the edges, then there is a monochromatic copy of some fixed graph. A graph is chaotic, and if you break it into pieces, then the pieces will have some nice properties. That's just a nice way of saying something. But 'In every chaos there is an order' is not my invention. That's ancient.

Kang: When do 'good' mathematical ideas come to you? While walking in woods, relaxing on a sofa or in bed, working at your office desk, or while discussing with colleagues?

Szemerédi: Working at a desk is out of question, because I never do this. I usually lie down on bed. But when I really concentrate on something, then I walk. It almost always happens that I work on a problem while walking, leave the problem, and next day I go for a walk and continue thinking and so on.

Kang: As both of us are participating in the conference on 'Perspectives in Discrete Mathematics', I would like to ask you what you think about future directions of mathematics, in particular of Discrete Mathematics?

Szemerédi: Well, as others have already explained, there is no such a thing as predicting the future. But during the last 20 years, Discrete Mathematics has developed enormously, because of computers and the interaction between computer scientists and Discrete Mathematicians. Discrete Mathematics is nowadays recognised as a part of mathematics. It was not the case, when I started. Discrete Mathematics was something like just doing it for playing around. But, not any more.

Kang: There are also parts of Discrete Mathematics, where one needs some methods from continuous mathematics.

Szemerédi: That's true. There is an interplay between continuous and discrete mathematics. Well, mathematicians who worked on continuous mathematics believed that this was a one way street and did not care about discrete mathematics. But when you really examine the proof of some of their results, then there are a lot of techniques and theory, which are very hard to understand, but there are some crucial arguments, which they claim to be the bottleneck and the punch line of the proof. Sometimes these crucial arguments are classical combinatorial arguments.

Kang: What kind of a change are you experiencing as a mathematician after having received the Abel Prize?

Szemerédi: That's easy. If it wasn't for the Abel Prize, I would not sit here for the interview.

Kang: Let me ask you one last question. What are you going to do with your prize money, 6,000,000 NOK, around 800,000 EUR?

Szemerédi: I don't know. We have many children and we are also waiting for the tax people to determine how much tax we are supposed to pay.

Kang: I would like to thank you for a pleasant conversation.

3 Szemerédi's Work

Endre Szemerédi has made immense contributions to discrete mathematics, theoretical computer science, ergodic theory, additive and combinatorial number theory and discrete geometry, through the understanding of deep connections between seemingly unrelated fields of mathematics.

Without a doubt, the most celebrated work by Szemerédi is his proof of the Erdős-Turán conjecture (1936), now known as Szemerédi's theorem (1975), which states that every set of integers of positive upper density contains arbitrarily long arithmetic progressions. In order to prove his theorem, Szemerédi introduced the Szemerédi regularity lemma (indeed a weaker version of it), which roughly says that every large dense graph can be partitioned into a bounded number of roughly equally-sized parts so that the graph is random-like between pairs of parts.

Szemerédi's theorem, Szemerédi regularity lemma and subsequent studies related to them have led to much progress in various areas of mathematics, including extremal graph theory, ergodic theory, harmonic analysis, additive number theory, discrete geometry and theoretical computer science.

An important application of Szemerédi's theorem is the Green-Tao theorem, stating that there are arbitrarily long arithmetic progressions of primes [20]. There are several other proofs of Szemerédi's theorem including an ergodic theoretic proof by Furstenberg [11] in 1977, a Fourier analytic proof by Gowers [16] in 2001, and proofs based on hypergraph removal lemma independently by Gowers [17], Tao [32, 34], and Nagle, Rödl, Schacht and Skokan [26, 27]. Some of these results will be discussed in Sections 3.1 and 3.2.

3.1 Szemerédi's Theorem

Szemerédi's Theorem goes back to the following famous theorem in Ramsey theory.

Theorem 1 (van der Waerden's Theorem)**.** *For every positive integers k and r, there exists an integer $W(k,r)$ such that for any integer $N \geq W(k,r)$, if the positive integers $\{1,2,\ldots,N\}$ are partitioned into r classes, then at least one of the classes contains at least one arithmetic progression $a, a+n, a+2n, \ldots, a+(k-1)n$ of length k, where a, n are positive integers.*

In other words, van der Waerden's Theorem says that for any integer $N \geq W(k,r)$, if we colour the positive integers $\{1,2,\ldots,N\}$ with r colours, then we can find a sequence of k numbers $a, a+n, a+2n, \ldots, a+(k-1)n$, all of which are coloured with a single colour.

Erdős and Turán conjectured in 1936 that the existence of an arithmetic progression would still be guaranteed in a set of integers if the set were dense. Szemerédi gave an affirmative answer to the conjecture by showing that any positive fraction of the positive integers must contain arbitrarily long arithmetic progressions.

Theorem 2 (Szemerédi's Theorem – 'Finite' Version)**.** *For every positive integer k and real number $0 < \delta \leq 1$, there exists an integer $S(k,\delta)$ such that for any integer $N \geq S(k,\delta)$, any subset $A \subset \{1,2,\ldots,N\}$ of cardinality at least δN contains at least one arithmetic progression $a, a+n, a+2n, \ldots, a+(k-1)n$ of length k, where a, n are positive integers.*

Szemerédi's Theorem is usually stated using the concept of the positive upper density as follows.

Theorem 3 (Szemerédi's Theorem)**.** *Let $A \subset \mathbb{N}$ be a subset of the positive integers with positive upper density, i.e.*

$$\bar{\delta}(A) := \limsup_{N \to \infty} \frac{\#\{A \cap [1,N]\}}{N} > 0.$$

Then for any positive integer k, there exist positive integers a, n such that $\{a, a+n, a+2n, \ldots, a+(k-1)n\} \subset A$.

Szemerédi's theorem for the cases $k = 1, 2$ is trivial. Roth [28] proved the theorem for the case $k = 3$ in 1953, using Fourier analysis. Szemerédi proved the theorem first for the case $k = 4$ in 1969 [30], and established the theorem in full generality in 1975 [31], using combinatorial arguments, in particular a weaker version of the Szemerédi regularity lemma.

There are several extensions of Szemerédi's theorem, such as the following strengthened 'infinite' version:

Theorem 4 (Szemerédi's Theorem – 'Infinite' Version). *Let $A \subset \mathbb{N}$ be a subset of the positive integers with positive upper density, i.e. $\bar{\delta}(A) > 0$. Then for any positive integer k, there exist a positive integer n and infinitely many positive integers a such that $\{a, a+n, a+2n, \ldots, a+(k-1)n\} \subset A$.*

Furstenberg [11] provided in 1977 the ergodic theoretic proof of Szemerédi's theorem by showing that Szemerédi's theorem is equivalent to the following.

Theorem 5 (Furstenberg Recurrence Theorem). *Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T : X \to X$ be a measure-preserving bijection on X, i.e. $(X, \mathcal{B}, \mu, T)$ is a measure-preserving system. Then for any $E \in \mathcal{B}$ with positive measure and any positive integer k there exists a positive integer n such that $\mu(E \cap T^n E \cap \ldots \cap T^{(k-1)n} E) > 0$.*

This is equivalent to a stronger theorem, saying that for any set $E \in \mathcal{B}$ with $\mu(E) > 0$ and every positive integer k there exist *infinitely* many positive integers n for which $\mu(E \cap T^n E \cap \ldots \cap T^{(k-1)n} E) > 0$. It corresponds to the 'infinite' version of Szemerédi's theorem. Indeed, even a stronger statement is true:

for any $E \in \mathcal{B}$ with $\mu(E) > 0$,

$$\liminf_{N \to \infty} \sum_{n=1}^{N} \mu(E \cap T^n E \cap \ldots \cap T^{(k-1)n} E) > 0.$$

The case $k = 1$ is trivial, and the case $k = 2$ is the classical Poincaré recurrence theorem.

The ergodic theoretic approach by Furstenberg led to various generalisations of Szemerédi's theorem, including the multidimensional generalisation by Furstenberg and Katznelson [12] and the polynomial generalisation by Bergelson and Leibman [7].

The multidimensional Szemerédi's theorem by Furstenberg and Katznelson was established in 1978 [12], as a consequence of the following extension of the Furstenberg recurrence theorem:

Theorem 6 (Furstenberg-Katznelson Recurrence Theorem)**.** *Let* $(X, \mathcal{B}, \mu)$ *be a probability space and let* $T_1, T_2, \ldots, T_k : X \to X$ *be commuting measure-preserving bijections on* X*. For any* $E \in \mathcal{B}$ *with* $\mu(E) > 0$*,*

$$\liminf_{N\to\infty} \sum_{n=1}^{N} \mu(T_1^n E \cap T_2^n E \ldots \cap T_k^n E) > 0.$$

Theorem 7 (Multidimensional Szemerédi's Theorem)**.** *Let* $d \geq 1$ *and* $A \subset \mathbb{Z}^d$ *with positive upper Banach density, i.e.*

$$\delta^*(A) := \limsup_{N\to\infty} \frac{\#\{A \cap [-N,N]^d\}}{(2N+1)^d} > 0.$$

Then for any $b_1, b_2, \ldots, b_k \in \mathbb{Z}^d$*, there exist a positive integer* n *and infinitely many* $a \in \mathbb{Z}^d$ *such that* $\{a + nb_1, a + nb_2, \ldots, a + nb_k\} \subset A$*.*

In 1996, Bergelson and Leibman [7] generalised the multidimensional Szemerédi's theorem to polynomials.

Theorem 8 (Polynomial Multidimensional Szemerédi Theorem)**.** *Let* $d \geq 1$ *and* $A \subset \mathbb{Z}^d$ *with positive upper Banach density, i.e.* $\delta^*(A) > 0$*. Then for any polynomials* $P_1, P_2, \ldots, P_k : \mathbb{Z} \to \mathbb{Z}^d$ *with* $P_1(0) = P_2(0) = \ldots = P_k(0) = 0$*, there exist a positive integer* n *and infinitely many* $a \in \mathbb{Z}^d$ *such that* $\{a + P_1(n), a + P_2(n), \ldots, a + P_k(n)\} \subset A$*.*

Szemerédi's Theorem is one of the fundamental results in Ramsey theory, and is the density version of van der Waerden's theorem. Another fundamental result in Ramsey theory is the Hales-Jewett theorem, which is a generalisation of van der Waerden's theorem. In order to state the Hales-Jewett theorem, we need a notion of combinatorial lines, instead of arithmetic progressions. Let W_n be the set of all words of length n over the alphabet $\{1, \ldots, k\}$. Given a word $w(x) \in W_n$ where x occurs at least once. The set $L = \{w(1), \ldots, w(k)\}$, in which $w(i)$ is obtained from $w(x)$ by substituting each x with i, is called a *combinatorial line*.

Theorem 9 (Hales-Jewett Theorem)**.** *For every positive integers* k, r*, there exists a positive number* $H(k,r)$ *such that for any integer* $n \geq H(k,r)$*, every* r*-colouring of* W_n *contains a monochromatic combinatorial line.*

In 1991, Furstenberg and Katznelson [13] proved the density version of the Hales-Jewett theorem, using ergodic theoretic techniques:

Theorem 10 (Density Hales-Jewett Theorem)**.** *For every positive integer* k *and real number* $0 < \delta \leq 1$*, there exists a positive number* $D(k,\delta)$ *such that for any integer* $n \geq D(k,r)$*, every subset* $A \subset W_n$ *of cardinality at least* δk^n *contains a combinatorial line.*

Another approach to proving Szemerédi's theorem is based on hypergraph removal lemmas, for which corresponding hypergraph regularity lemmas have been developed. These will be discussed in the next section.

One of the most important applications of Szemerédi's theorem is:

Theorem 11 (Green-Tao Theorem)**.** *For every integer $k \geq 2$ there exist infinitely many arithmetic progressions of primes of length k.*

We shall conclude this section with what Terence Tao wrote in an article titled "The dichotomy between structure and randomness, arithmetic progressions, and the primes" [35]:

> "A famous theorem of Szemerédi asserts that all subsets of the integers with positive upper density will contain arbitrarily long arithmetic progressions. There are many different proofs of this deep theorem, but they are all based on a fundamental dichotomy between structure and randomness, which in turn leads (roughly speaking) to a decomposition of any object into a structured (low-complexity) component and a random (discorrelated) component. Important examples of these types of decompositions include the Furstenberg structure theorem and the Szemerédi regularity lemma. One recent application of this dichotomy is the result of Green and Tao establishing that the prime numbers contain arbitrarily long arithmetic progressions (despite having density zero in the integers). The power of this dichotomy is evidenced by the fact that the Green-Tao theorem requires surprisingly little technology from analytic number theory, relying instead almost exclusively on manifestations of this dichotomy such as Szemerédi's theorem. In this paper we survey various manifestations of this dichotomy in combinatorics, harmonic analysis, ergodic theory, and number theory. As we hope to emphasise here, the underlying themes in these arguments are remarkably similar even though the contexts are radically different."

3.2 Szemerédi Regularity Lemma

The Szemerédi regularity lemma and its variants belong to the essential tools of extremal graph theory, additive number theory, discrete geometry and theoretical computer science. More remarkably, it was the main tool for proving Szemerédi's theorem.

The Szemerédi regularity lemma roughly says that the vertex set of any large dense graph can be partitioned into a constant number of classes such that almost all of the induced bipartite graphs are "pseudorandom" in the sense that they mimic the behaviour of random bipartite graphs of the same density.

The pseudorandomness plays a fundamental role in structural and algorithmic aspects of many problems. It provides a way to use 'probabilistic intuition' for solving deterministic problems or to extract 'predictable' average structure from a 'mysterious' huge structure. Avi Wigderson said [36]:

> "Pseudorandomness is the study, by mathematicians and computer scientists, of deterministic structures which share some properties of random ones."

In order to state more precisely the Szemerédi regularity lemma we need some definitions. Let $G=(V,E)$ be a graph with vertex set V and edge set $E \subset \binom{V}{2}$, i.e. each edge $e \in E$ is a 2-element subset of V. Given real $\varepsilon > 0$ and disjoint subsets $A, B \subset V$, the pair (A,B) is called ε-*regular* if for all $A' \subset A$ and $B' \subset B$ with $|A'| \geq \varepsilon |A|$ and $|B'| \geq \varepsilon |B|$,

$$\left| d(A,B) - d(A',B') \right| \leq \varepsilon,$$

where $d(A,B) := \frac{E \cap (A \times B)}{|A||B|}$ is the density of edges between A and B.

Theorem 12 (Szemerédi Regularity Lemma)**.** *For every positive real* ε *and every positive integer* t_0*, there exist positive integers* $T(\varepsilon, t_0)$ *and* $N(\varepsilon, t_0)$ *such that for every graph* $G=(V,E)$ *on* $|V| \geq N(\varepsilon, t_0)$ *vertices there exists a partition* $P = \{V_1, V_2, \ldots, V_t\}$ *of* V *with* $t_0 \leq t \leq T(\varepsilon, t_0)$ *satisfying the following properties:*

- $|V_1| \leq |V_2| \leq \ldots \leq |V_t| \leq |V_1| + 1$
- *all but at most* εt^2 *pairs* (V_i, V_j) *with* $1 \leq i < j \leq t$ *are* ε*-regular.*

Applying the Szemerédi regularity lemma, Ruzsa and Szemerédi [29] established in 1976 the triangle removal lemma, which roughly says that every graph which does not contain many triangles can be made triangle-free by removing few edges.

Theorem 13 (Triangle Removal Lemma)**.** *For every* $\delta > 0$ *there exist a positive real* γ *and a positive integer* N *such that every graph* G *on* $n \geq N$ *vertices containing at most* γn^3 *triangles can be made triangle-free by removing at most* $\delta \binom{n}{2}$ *edges.*

Using this lemma, Ruzsa and Szemerédi [29] gave a short alternative proof of Roth's theorem (i.e. Szemerédi's theorem for $k=3$). Frankl and Rödl [14] showed that Szemerédi's theorem for $k+1$ follows from a removal lemma for k-uniform hypergraphs (where a k-uniform hypergraph $H=(V,E)$ is a pair of vertex set V

and edge set $E \subset \binom{V}{k}$, i.e. each edge $e \in E$ is a k-element subset of V). Recently various generalisations of the regularity lemma and the removal lemma for hypergraphs were obtained independently by Gowers [17], Tao [32], Nagle, Rödl, Schacht and Skokan [26, 27]. The main applications of these generalisations include alternative combinatorial proofs of Szemerédi's Theorem.

Theorem 14 (Hypergraph Removal Lemma)**.** *Let F be a fixed k-uniform hypergraph on f vertices and let $\delta > 0$ be given. Then there exist a positive real γ and a positive integer N such that every k-uniform hypergraph on $n \geq N$ vertices containing at most γn^f copies of F can be made F-free by removing at most $\delta \binom{n}{k}$ edges.*

The Szemerédi regularity lemma works very well for *dense* graphs, where the number of edges is quadratic in the number of vertices, since the error term in the lemma is quadratic in the number of vertices with an arbitrary small multiplicative constant. In order to deal with sparse cases, some variants of the regularity lemmas are developed independently by Kohayakawa [23] and Rödl (unpublished) for sparse graphs and by Alon, Coja-Oghlan, Hàn, Kang, Rödl and Schacht [3] for sparse graphs with general degree distribution. Another variant of Szemerédi regularity lemma and removal lemma has been obtained by Green [19] for abelian groups.

As discussed in the previous section, Szemerédi's theorem has enjoyed its strong connections to ergodic theory. Likewise, the Szemerédi regularity lemma has strong links to probability theory and analysis. For example, Tao [33] gave a probabilistic and information theoretic version of the regularity lemma. Lovász and Szegedy [24] obtained 'analytic versions' of the regularity lemma and showed that some versions of the regularity lemma can be interpreted as approximation of elements in Hilbert spaces or as the compactness of an important metric space of two-variable functions.

In order to state an analytic version of the regularity lemma by Lovász and Szegedy [24], we need some notation. Let $\mathcal{F}$ denote the set of all bounded symmetric measurable functions $F : [0,1]^2 \to \mathbb{R}$ and let $\mathcal{F}_0$ denote the set of all symmetric measurable functions $F : [0,1]^2 \to [0,1]^2$. Given $F \in \mathcal{F}$ and a partition $\mathcal{P} = \{P_1, P_2, \ldots, P_t\}$ of $[0,1]$, let $F_{\mathcal{P}} : [0,1]^2 \to \mathbb{R}$ denote the step-function obtained from F by substituting its value at $(x,y) \in P_i \times P_j$ by $\int_{P_i \times P_j} F(x,y)\, dxdy$.

Theorem 15 (Strong Analytic Regularity Lemma)**.** *For every positive real ε, there exists a positive integer $T(\varepsilon)$ such that for every function $F \in \mathcal{F}_0$, there is a partition $\mathcal{P} = \{P_1, P_2, \ldots, P_t\}$ of $[0,1]$ into $t \leq T(\varepsilon)$ sets of equal measures so that for every set $S \subset [0,1]^2$ which is the union of at most t^2 rectangles,*

$$\left| \int_S (F - F_{\mathcal{P}})\, dxdy \right| \leq \varepsilon.$$

Lovász and Szegedy [24] extended the regularity lemma to a general setting of Hilbert spaces.

Theorem 16 (Regularity Lemma in Hilbert Space)**.** *Let $S_1, S_2, \ldots$ be arbitrary subsets of a Hilbert space $\mathcal{H}$. Then for every positive real ε and $f \in \mathcal{H}$, there exist a positive integer $t \leq 1/\varepsilon^2$, $f_i \in S_i$ for $1 \leq i \leq t$ and $a_1, a_2, \ldots, a_t \in \mathbb{R}$ such that for every $g \in S_{t+1}$,*

$$\Big|\ \langle g, f - (a_1 f_1 + a_2 f_2 + \ldots + a_t f_t) \rangle\ \Big| \ \leq\ \varepsilon ||g|| \cdot ||f||.$$

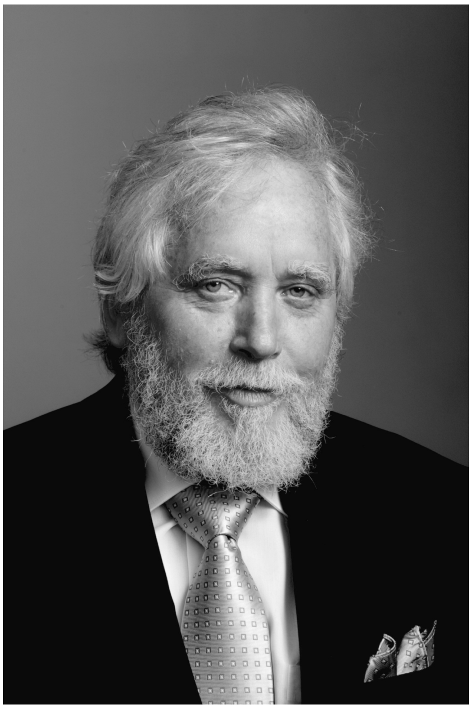
This work is along the lines of the study of connections between the regularity lemma and graph limits by Lovász and Szegedy [25], by Borgs, Chayes, Lovász, Sós, Szegedy and Vesztergombi [9], and by Bollobás, Janson and Riordan [8].

Last but not least, we shall discuss important applications of the Szemerédi regularity lemma in computer science. The regularity lemma guarantees the *existence* of a regular partition that approximates a graph by a constant number of pseudorandom graphs. The question of whether we can *compute in polynomial time* such a regular partition is algorithmically very important, from the viewpoint of theoretical computer science. A number of NP-hard problems can be solved in polynomial time on graphs that admit regular partitions in polynomial time. Algorithmic regularity lemmas that compute a regular partition in polynomial time were obtained by Alon, Duke, Lefmann, Rödl, and Yuster [4] and by Frieze and Kannan [10] for dense graphs, by Alon, Coja-Oghlan, Hàn, Kang, Rödl and Schacht [3] for sparse graphs, and by Haxell, Nagle and Rödl [21] for hypergraphs. An important application of algorithmic regularity lemmas on graphs is a polynomial time approximation scheme for the MAX-CUT problem.

Another very important application of the regularity lemma in computer science is the property testing [2, 5, 15]. Alon and Shapira [6] showed that every monotone property is testable. Roughly speaking, a property is said to be *testable* if there exists a randomised algorithm with constant running time which distinguishes between instances with the property and those which are far from it. To be more precise, let us call a graph property *monotone* if it is closed under removal of edges and vertices. Given a positive integer n and real ε, a graph G on n vertices is said to be ε-*far* from satisfying a graph property P if at least εn^2 edges should be added or deleted from G in order to make the resulting graph satisfy P. A *tester* for P is a randomised algorithm that distinguishes with high probability (e.g. 2/3) between the case of G satisfying P and the case of G being ε-far from satisfying P. A tester is said to have *one-sided error* if whenever G satisfies P, the algorithm declares that this is the case with probability one.

Theorem 17 (Testability of Monotone Properties)**.** *For every monotone graph property P there is a tester with one-sided error.*

4 An Irregular Mind



“Szemerédi has *an irregular mind*, his brain is wired differently than most mathematicians. Many of us admire his unique way of thinking, his extraordinary vision.” [22]

Endre Szemerédi was born in 1940 in Budapest, Hungary. He studied at Eötvös Loránd University and received his Ph.D. from Moscow State University. He is Research Fellow at the Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences and Professor of Computer Science at Rutgers University.

Endre Szemerédi has received many awards and honours for his essential contributions to mathematics and computer science including the Abel Prize (2012), the Leroy P. Steele Prize (2008), the Rolf Schock Prize (2008), the Prize of the Hungarian Academy of Sciences (1979), the Pólya Prize (1975), the Rényi Prize (1973), the Grünwald Prize (1968, 1967).

Endre Szemerédi and his work are best described by Timothy Gowers [18]:

> “Some mathematicians are famous for one or two major theorems. Others are famous for a huge and important body of high-class results. Very occasionally, there is a mathematician who is famous for both. No account of Szemerédi’s work would be complete without a discussion of Szemerédi’s theorem and Szemerédi’s regularity lemma. However, there is much more to Szemerédi than just these two the-

orems. He has published over 200 papers, as I mentioned at the beginning, and at the age of 71 he shows no signs of slowing down. It is extremely fitting that he should receive an award of the magnitude of the Abel Prize. I hope that the small sample of his work that I have described gives at least some idea of why, even if I have barely scratched the surface of what he has done."

References

[1] The Abel Prize Laureate 2012, *http://www.abelprize.no/c54147/seksjon/vis.html?tid=54148*

[2] N. Alon, Testing subgraphs in large graphs, *Random Structures and Algorithms* **21** (2002), 359-370.

[3] N. Alon, A. Coja-Oghlan, H. Hàn, M. Kang, V. Rödl and M. Schacht, Quasi-randomness and algorithmic regularity for graphs with general degree distributions, *SIAM J. Comput.* **39** (2010), 2336-2362.

[4] N. Alon, R.A. Duke, H. Lefmann, V. Rödl, and R. Yuster, The algorithmic aspects of the regularity lemma, *J. of Algorithms* **16** (1994), 80-109.

[5] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, *Combinatorica* **20** (2000), 451-476.

[6] N. Alon and A. Shapira, Every monotone graph property is testable, in: *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pp. 128–137, ACM Press (2005).

[7] V. Bergelson and A. Leibman, Polynomial extensions of van der Waerden's and Szemerédi's theorems, *J. Amer. Math. Soc.* **9** (1996), 725-753.

[8] B. Bollobás, S. Janson and O. Riordan, Monotone graph limits and quasimonotone graphs, *Internet Mathematics* **8** (2012), 187-231.

[9] C. Borgs, J. Chayes, L. Lovász, V. Sós, B. Szegedy and K. Vesztergombi, Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing, *Advances in Math.* **219** (2008), 1801–1851.

[10] A. Frieze and R. Kannan, Quick approximation to matrices and applications, *Combinatorica* **19** (1999), 175-200.

[11] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Analyse Math.* **31** (1977), 204-256.

[12] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for commuting transformations, *J. Analyse Math.* **34** (1978), 275-291.

[13] H. Furstenberg and Y. Katznelson, A density version of the Hales-Jewett theorem, *J. Analyse Math.* **57** (1991), 64-119.

[14] P. Frankl and V. Rödl, Extremal problems on set systems, *Random Structures Algorithms* **20** (2002), 131-164.

[15] O. Goldreich, S. Goldwasser and D. Ron, Property testing and its connection to learning and approximation, *Journal of the ACM* **45** (1998), 653-750.

[16] T. Gowers, A new proof of Szemerédi's theorem, *Geom. Func. Anal.* **11** (2001), 465-588.

[17] T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, *Annals of Mathematics* **166** (2007),897-946.

[18] T. Gowers, The work of Endre Szemerédi, *http://gowers.files.wordpress.com/2012/03/talktalk2.pdf*

[19] B. Green, A Szemerédi-type regularity lemma in abelian groups, *Geom. Funct. Anal.* **15** (2005), 340-376.

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

[21] P. Haxell, B. Nagle, and V. Rödl, An algorithmic version of the hypergraph regularity method, in: *Proceedings of the fourty-sixth annual IEEE Symposium on Foundations of Computer Science*, pp. 439-448, IEEE Computer Society, 2005.

[22] *An irregular mind*, Szemerédi is 70. Edited by Imre Bŕány and József Solymosi. Bolyai Society Mathematical Studies, 21. János Bolyai Mathematical Society, Budapest; Springer, Berlin, 2010, 758 pp.

[23] Y. Kohayakawa, Szemerédi's regularity lemma for sparse graphs, *Foundations of computational mathematics* (1997), 216-230.

[24] L. Lovász, B. Szegedy, Szemerédi's regularity lemma for the analyst, *Geom. Funct. Anal.* **17** (2007), 252-270.

[25] L. Lovász, B. Szegedy, Limits of dense graph sequences, *J. Combin. Theory Ser. B* **96** (2006), 933-957.

[26] B. Nagle, V. Rödl, and M. Schacht, The counting lemma for regular k-uniform hypergraphs, *Random Structures & Algorithms* **28** (2006), no. 2, 113-179.

[27] V. Rödl and J. Skokan, Regularity lemma for k-uniform hypergraphs, *Random Structures & Algorithms* **25** (2004), no. 1, 1-42.

[28] K. Roth, On certain sets of integers, *J. London Math. Soc.* **28** (1953), 245-252.

[29] I. Z. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, in Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), *Coll. Math. Soc. J. Bolyai* **18**, Vol. II, pp. 939-945,

[30] E. Szemerédi, On sets of integers containing no four elements in arithmetic progression, *Acta Math. Acad. Sci. Hungar.* **20** (1969), 89-104.

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

[32] T. Tao, A variant of the hypergraph removal lemma, *J. Combin. Theory Ser. A* **113** (2006), 1257-1280.

[33] T. Tao, Szemerédi's regularity lemma revisited, *Contrib. Discrete Math.* **1** (2006), 8-28.

[34] T. Tao, A correspondence principle between (hyper)graph theory and probability theory, and the (hyper)graph removal lemma, *J. Anal. Math.* **103** (2007), 1-45.

[35] T. Tao, The dichotomy between structure and randomness, arithmetic progressions, and the primes, *http://arxiv.org/abs/math/0512114v2*

[36] A. Wigderson, Randomness and pseudorandomness, Abstract of the Abel Lecture 2012, *http://www.abelprize.no/nyheter/vis.html?tid=54530*

The photo of Endre Szemerédi was taken by Knut Falch, and is from the photo archives of the Norwegian Academy of Science and Letters. It is used here with kind permission. Many thanks to Anne-Marie Astad.

Author's address:

Mihyun Kang
Technische Universität Graz
Institut für Optimierung und Diskrete Mathematik
Steyrergasse 30, 8010 Graz, Austria
e-mail kang@math.tugraz.at