

THE WORK OF ENDRE SZEMERÉDI

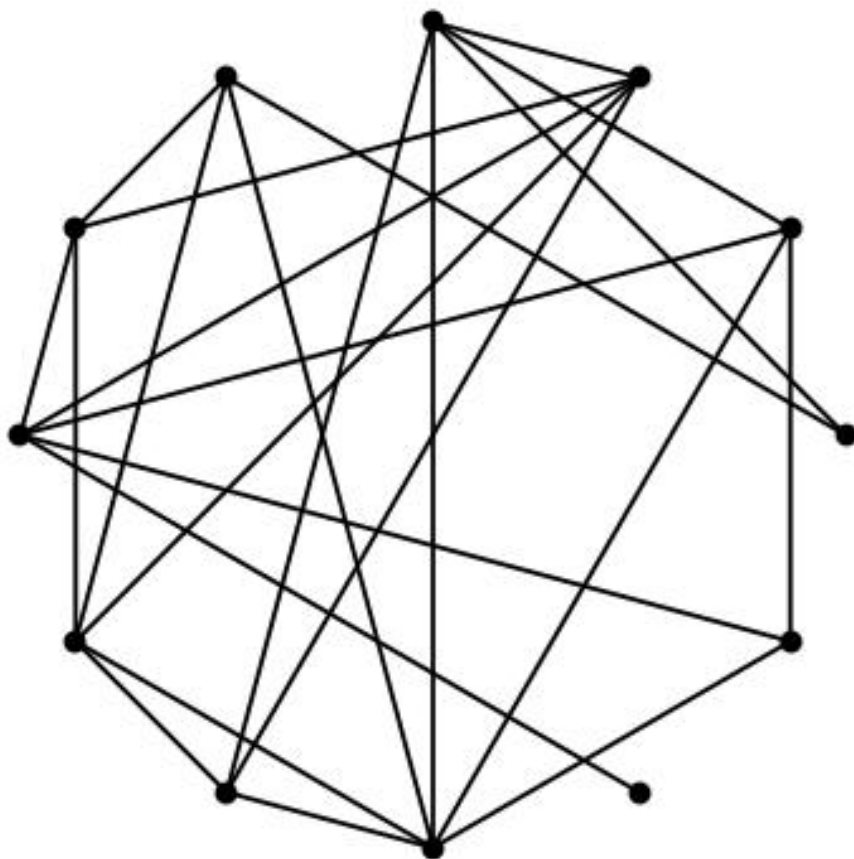
W.T. GOWERS

1. INTRODUCTION

Endre Szemerédi is a towering figure in the area of mathematics known as *combinatorics*, with particularly important contributions to the subarea called *extremal combinatorics*. I will explain what these terms mean in a moment, but first here are a few bald facts about his extraordinary mathematical output. The achievement for which he is best known is his proof, in 1975, of what is now called Szemerédi’s theorem but which at the time was a notorious and decades-old conjecture of Erdős and Turán. This theorem is one of the highlights of twentieth-century mathematics, but it also lies at the heart of a great deal of very recent research. He also gave us Szemerédi’s regularity lemma, a result that originated in the proof of Szemerédi’s theorem but went on to become a major tool in extremal combinatorics. As well as these results, he has published over 200 papers, many of them representing important advances. I shall pick one or two, but it should be understood that they are just a small sample from a huge output that has profoundly influenced many areas of mathematical thought.

What, then, is combinatorics? One possible definition is that it is the study of *discrete structures*. And what are they? Well, the word “discrete” is typically contrasted with the word “continuous”: a structure is continuous if you can move smoothly from one part to another, whereas it is discrete if you have to jump. For example, if you are modelling the flow of a fluid, then the mathematical structures you study will be continuous, since you will specify things like velocities and pressures at various points, and these vary smoothly. By contrast, if you are modelling the inside of a computer, then you will be interested in sequences of 0s and 1s, which is an example of a discrete structure, since to get from one such sequence to another you have to cause at least one 0 to jump to a 1 or vice versa.

Another discrete structure, and perhaps the single most important in combinatorics, is the *graph*. A graph is an object that looks like this.



In other words, it consists of some points, some of which may be joined by lines. The points are called *vertices* and the lines are called *edges*.

You might think that a graph is continuous, because you can move continuously along its edges. However, it is just the *picture* that is continuous rather than the graph itself. All we care about with a graph is which pairs of vertices are joined by edges, and this can be specified as a simple list. For example, if the graph looks like the vertices and edges of a square, we can specify it by saying that the vertices are a, b, c and d and listing the edges as ab, bc, cd and da .

2. SZEMERÉDI'S THEOREM

Not all study of discrete structures would be classified as combinatorics. Another characteristic feature of much of the subject is that its problems can be stated in a way that is easy to understand, or at least far easier than it is for the problems in many other areas. Also, the proofs are often elementary, not in the sense that Sherlock Holmes would use

the word but in a rather special mathematical sense. When a mathematician describes a proof as elementary, it means that the argument does not make use of advanced concepts or rely on previously established difficult results. It should not be taken to imply that the proof is easy: one can put together basic ingredients in extremely complicated and sophisticated ways, and some “elementary” proof are amongst the hardest in mathematics. Conversely, some advanced proofs may actually be quite easy when you have invested enough time in understanding the theories on which they depend. (Of course, there are also easy elementary proofs and difficult advanced ones.)

Szemerédi’s theorem is a perfect illustration of what I have just said. It has a very appealing statement, and the proof that Szemerédi gave was both elementary and extremely difficult. Let me begin by explaining what the theorem says.

In order to do this, I need the notion of an *arithmetic progression*. An arithmetic progression is a sequence of numbers that advances in steps of the same size. So the sequence 3, 9, 15, 21, 27, 33, 39 is an arithmetic progression, with steps of size 6, while the sequence 4, 7, 11, 14, 17, 21 is not an arithmetic progression because the differences between successive terms are not all the same (some being 3 and some being 4).

One way of understanding Szemerédi’s theorem is to imagine the following one-player game. You are told a small number, such as 5, and a large number, such as 10,000. Your job is to choose as many integers between 1 and 10,000 as you can, and the one rule that you must obey is that from the integers you choose it should not be possible to create a five-term arithmetic progression. For example, if you were accidentally to choose the numbers 101, 1103, 2105, 3107 and 4109 (amongst others), then you would have lost, because these five numbers form a five-term progression with step size 1002.

Obviously you are destined to lose this game eventually, since, to give an argument that is both elementary and extremely easy, if you keep going for long enough you will eventually have chosen *all* the numbers between 1 and 10,000, which will include many five-term arithmetic progressions. But Szemerédi’s theorem tells us something far more interesting: even if you play with the best possible progression-avoiding strategy, you will lose long before you get anywhere near choosing all the numbers.

To state the result precisely I need just a tiny bit of algebra, though all I mean by that is that I would like to represent the two numbers we start with by letters. Let k stand for the length of the progression we are trying to avoid, and let n stand for the number of numbers we get to choose from. (In the discussion just above, k was 5 and n was 10,000.) Now let us write $S(k, n)$ for the largest number of numbers it is possible to choose while avoiding

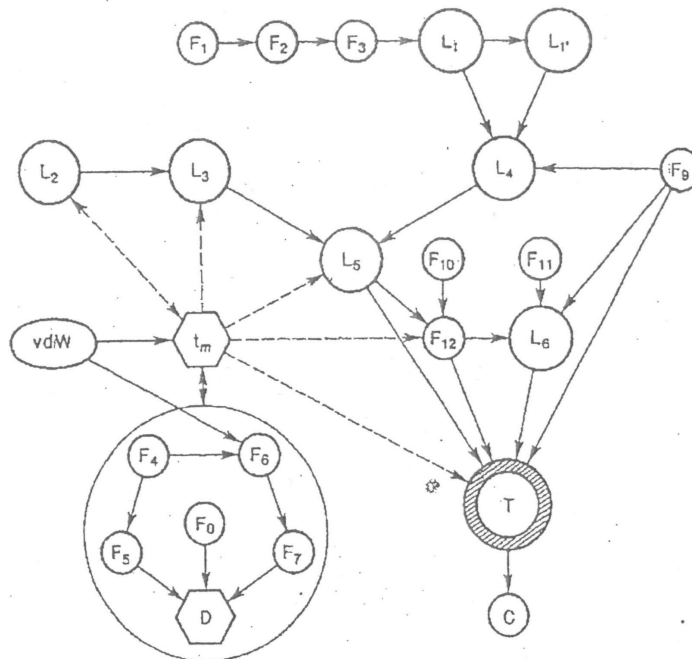
any k -term progression. What Szemerédi showed was that when n is large, $S(k, n)$ is a very small percentage of n . How small? Well, as small as you like – provided only that n is large enough.

If, for instance, we are trying to avoid progressions of length 23, Szemerédi's theorem tells us that there is some n (which may be huge, but the point is that it exists) such that if we play the game with n numbers, then we cannot choose more than $n/1000$ of those numbers – that is, a mere 0.1% of them – before we lose. And the same is true for any other progression length and any other positive percentage.

How does the proof go? I'm not going to tell you. I'm just going to repeat that is elementary in the technical sense, and attempt to convince you that it is difficult by reproducing a diagram from Szemerédi's original paper.

202

E. Szemerédi



The diagram represents an approximate flow chart for the accompanying proof of Szemerédi's theorem. The various symbols have the following meanings: $F_k \equiv$ Fact k , $L_k \equiv$ Lemma k , $T \equiv$ Theorem, $C \equiv$ Corollary, $D \equiv$ Definitions of B, S, P, a, β , etc., $t_m \equiv$ Definition of t_m , $vdW \equiv$ van der Waerden's theorem, $F_0 \equiv$ "If $f: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ is subadditive then $\lim_{n \rightarrow \infty} \frac{f(n)}{n}$ exists".

3. WHY SHOULD WE CARE ABOUT FINDING ARITHMETIC PROGRESSIONS?

I know of no situation in real life where it is important to be sure that a smallish set of integers is forced to contain an arithmetic progression of length 10. And even if such a situation were to arise, the n you would need for Szemerédi's theorem to tell you anything interesting would be far larger than the number of atoms in the universe, or even the exponential of that number. That puts the theorem well beyond the realms of any practical applications. What is it, then, about Szemerédi's theorem that mathematicians find so fascinating?

There are several answers to this. The most obvious one is the contrast between the simplicity of the statement of Szemerédi's theorem and the difficulty of its proof (and all subsequently discovered proofs). Usually, a simple and natural mathematical statement either has a simple proof or a simple counterexample. From time to time, however, one is surprised: an innocent-seeming question is much harder to resolve than one expects. A significant proportion of such questions turn out to be so hard that nobody believes that they will be solved without ideas that are way beyond anything we have at the moment (an example is the question of whether $e + \pi$ is an irrational number). But some questions are just right: they are simple to ask and very hard to answer, but they have just enough connection with what we do know to make one feel that trying to solve them is not a completely hopeless task. The Erdős-Turán conjecture came into that category.

A second answer is that Szemerédi's theorem has *mathematical* applications even if it does not have practical ones. A particularly notable one is a theorem of Ben Green and Terence Tao, which states that you can find arbitrarily long arithmetic progressions that consist only of prime numbers. This result does not follow directly from Szemerédi's theorem, but Green and Tao found an extremely clever way of converting their problem into a form that allowed them to use Szemerédi's theorem to solve it.

If, however, you insist on practical applications, then all is not lost, provided that you are ready to accept *indirect* applications rather than direct ones. An important and not sufficiently appreciated aspect of mathematics is that the result you prove is often less interesting than the methods you use to prove it. This tendency is especially pronounced in combinatorics, where open problems often become popular not because we are desperate to know the answers to them but because they encapsulate some more general difficulty that we feel is holding us back mathematically. When such a problem is solved, the solution often involves the development of new mathematical tools that go on to be used in many other contexts.

Szemerédi's theorem provides a wonderful illustration of this phenomenon too, as we are about to see.

4. SZEMERÉDI'S REGULARITY LEMMA

I have not yet said what *extremal* combinatorics is, but let me do so now. Here is a question in extremal graph theory: if a graph has n vertices, then how many edges can it have without any three of them forming a triangle? In general, questions in extremal combinatorics ask how big some quantity can be before something else is forced to happen. Szemerédi's theorem itself is another example: it addresses the question of how many numbers you can choose between 1 and n before you are forced to include an arithmetic progression of length k .

Such questions fall naturally into two parts. One part is to find examples of structures that avoid what you want to avoid in such a way that the quantity you are interested in is as big as possible. The other part is to show that if the quantity in question reaches a certain size, then you *cannot* avoid what you want to avoid. A profound insight of Erdős was that in many situations a surprisingly good way of carrying out the first part is to choose your structures *randomly*. For example, if your structure is a graph with n vertices, there are some problems for which you get very good answers if for each pair of vertices x and y you simply toss a coin to decide whether to join them by an edge. (If you want your random graph to contain fewer edges, then you can use a biased coin.) It might seem as though there is nothing you can say about a structure that has been defined randomly, and indeed that is more or less true if you are looking for complete certainty. But, and this is the point, there is a huge amount that we can say if we merely ask for *near* certainty. And that is enough: if we can say that a randomly chosen structure almost certainly has the properties we want, then we can also draw the much weaker conclusion that *at least one* structure has the properties we want.

Erdős's insight gave birth to random graph theory, which became a major subarea of combinatorics. As a result, combinatorialists think of random graphs not as unpleasant chaotic objects but as objects that are in many ways *easy to understand*. Just to be clear, I am not saying that random graph theory is an easy area: if you ask sufficiently detailed questions about a random graph, then answering them requires very delicate and difficult probabilistic estimates. But there is nevertheless an important sense in which random graphs are (almost always) predictable and well-behaved.

It is against this background that Szemerédi's regularity lemma should be understood. Before the regularity lemma, graphs were thought of as rather structureless objects. After all, when you are specifying a graph, you decide for each pair of vertices x and y whether to join them by an edge, and there are no constraints whatsoever on your decision. However, Szemerédi realized that once you have lost your fear of randomness you can give a useful structural description to a completely arbitrary graph.

I cannot give a precise statement of the result here, but the rough idea is this. Given any graph, there is a way of dividing up its vertices into a small number of sets in such a way that if you take the edges joining any two of those sets, they look as though they have been chosen randomly. (In fact, even this rough idea is an oversimplification, but it will do for these purposes.) In short, every graph is made out of a small number of random-like graphs.

What this tells us is that we can give a good description of our graph using a very small amount of data: having split the vertices into the sets that the lemma tells us we can find, we just have to say roughly how many edges there are between each pair of those sets, and we know that those edges will be distributed in a way that looks random. That doesn't tell us exactly what the graph is, but for many purposes the difference between two random-like graphs is not important – they are both random-like, so they both have the properties that you expect of a random graph.

Szemerédi's regularity lemma quickly became, and has remained, a central tool in extremal graph theory, and its indirect influence, for example through several modifications and generalizations that have been formulated subsequently, is wider still.

I promised to discuss indirect *practical* applications of the regularity lemma. So let me start with an intellectual endeavour of obvious practical importance and work back to the regularity lemma. Nobody can deny that if one could develop a computer program that was able learn from experience, then it would have innumerable practical applications. The branch of artificial intelligence that attempts to do this is called *machine learning*. A famous abstract model called *PAC learning* was proposed by Leslie Valiant as a good framework for thinking about machine learning. (The letters PAC stand for “probably approximately correct”.) This gave rise to purely mathematical questions that fall under the general heading *property testing*. Roughly speaking, you are presented with a structure and you want to show that either it has a certain property or it is probably very similar to a structure that does not have that property. For example, you might be given a graph and be required to show either that it contains a triangle or that it differs only slightly,

if at all, from a graph with no triangles. What is interesting from the point of view of machine learning is that this can often be done extraordinarily fast: the program can carry out a very small number of simple tests and then form a reliable hypothesis (that is, a hypothesis that is probably approximately correct). And the tool that allows one to prove this? Szemerédi's regularity lemma.

5. SORTING

Szemerédi is also famous for being one third of Miklós Ajtai, János Komlós and Endre Szemerédi, or AKS for short. I shall discuss a couple of highlights of their work, but I must stress that this is a small sample. When I say “highlights” I mean something like the lights on a motorway: they are high, but they are also extremely numerous.

An old topic in computer science is that of *sorting algorithms*. You are given a collection of objects and are able to compare any two of them. Your job is to put them in order using as few comparisons as you can. To think about this, it may help to imagine a collection of rocks that you have to put in order of weight, and all you have to do it is a pair of scales that will tell you, given any two rocks, which is the heavier. You have to pay a dollar every time you use the balance, and you want to spend as little money as possible.

A beautiful argument shows that if you have n objects, then the number of comparisons you need to make is at least the logarithm (to base 2) of $n!$. (The exclamation mark denotes a factorial, so $n!$ is shorthand for $1 \times 2 \times \cdots \times n$.) The argument is simple enough that I can even give it here. The number of possible orderings of the n objects is $n!$. Each time you ask a question, there are only two possible answers, so you cannot hope to reduce by more than a factor of 2 the number of orderings that are consistent with the answers you have so far received. (It's a bit like playing twenty questions: each question divides what remains of the world into two, and if you're unlucky the answer you get will direct you to the bigger half, so you cannot in general do better than reducing the set of possibilities by 50%.) Therefore, after k steps you cannot reduce the number of possible orderings by a factor of more than 2^k . Therefore in the worst case the number of steps has to be some k such that 2^k is at least $n!$, since otherwise there could well be more than one ordering consistent with the information you have. This is equivalent to saying that k has to be at least the logarithm of $n!$.

The logarithm of $n!$ is roughly $n \log n$ (that is, n times the logarithm of n). Interestingly, sorting algorithms are known that get by with approximately this number of comparisons. That is, there are methods known for deciding which two objects to compare, given previous

answers, that require a number of comparisons that is close to the theoretical minimum. For example, in 1945 John von Neumann invented a method known as Merge Sort that achieves this.

At the end of 1945, Szemerédi was only five. It might therefore appear that the subject of sorting algorithms was itself completely sorted out long before he was mathematically active. However, to think that is to underestimate the capacity of mathematicians to ask good questions.

Let us go back to the rocks that we want to put in order of weight and vary the game slightly. This time you have as many scales as you like, so that you can perform several comparisons simultaneously. So what you do is organize a series of weighings, where in each weighing you can weigh as many pairs of rocks as you like. The one obvious constraint is that the same rock cannot be placed on two different scales, so each rock is compared at most once during each weighing. And your object now is to put the rocks in order using the smallest number of weighings you can get away with.

Since we already know that around $n \log n$ comparisons are needed, and since you can do at most $n/2$ comparisons per weighing, the total number of weighings is obviously going to have to be around $2 \log n$, at the very least. For a long time it was an open problem whether one could get away with this theoretical minimum number, and this is the problem that Ajtai, Komlós and Szemerédi solved. They came up with a brilliantly clever method of sorting that does indeed require on the order of $\log n$ weighings. In the language of computer science, what Ajtai, Komlós and Szemerédi discovered was a *fast parallel algorithm* for sorting.

I cannot describe their method in full here, but I can give some idea about one ingredient of it. Suppose you have 1000 rocks to start with and you divide them into two groups of 500 each. (At this stage you have absolutely no idea how the weights of the rocks compare with each other.) One thing you might do is pair up the rocks in one group with the rocks in the other group, do 500 simultaneous comparisons, and put all the lighter rocks into a group marked L and the heavier ones into a group marked H. We would be very happy if all the rocks in L were lighter than all the rocks in H, but so far there is absolutely no reason to believe that this is true. However, we can repeat the process. Of course, to get more information, we now want to pair the rocks up in a different way.

Taking our cue from Erdős, we might think, correctly as it happens, that a good thing to do now is *randomly* pair up the rocks in L with the rocks in R and repeat the process.

That is, each time a rock in L is heavier than the rock we pair it up with in H, we swap them over, and each time it is lighter we leave it where it is.

If we keep randomly pairing up the rocks in L with the rocks in H, then there is a general tendency for lighter rocks to end up in L and heavier rocks to end up in H. What's more, as the process continues, this tendency gets more and more pronounced: if the rocks in L are already mostly the lighter ones, then it is unlikely that they will get moved up into H or that rocks in H will get moved down into L. Ajtai, Komlós and Szemerédi proved that after a constant number of comparisons (that is, a number that does not get bigger with n), the vast majority of the rocks in L are lighter than the vast majority of the rocks in H.

If *all* the rocks in L were lighter than *all* the rocks in H, then we could simply repeat this process within each group and after about $\log n$ rounds we would be finished. However, all we know is that this is true for *almost* all the rocks, so although Ajtai, Komlós and Szemerédi did indeed reach this conclusion, they had to work hard to do so.

One final word about this algorithm: as I described it above it was a *randomized* algorithm, because of all the random pairings. However, there is a very useful kind of graph known as an *expander* that can often be used as a substitute for a random graph. Here, instead of tossing a coin, one can use the edges of an expander to help one decide how to pair up the rocks, and the algorithm still works. This is an example of *derandomization*, which is one of the fundamental ideas in theoretical computer science, and it means that the algorithm can be performed in a fully deterministic way.

6. THE RAMSEY NUMBER $R(3, k)$

A *triangle* in a graph is what you might expect: three vertices all joined to each other. An *independent set* is a collection of vertices none of which are joined. Here is a simple argument to show that a graph with k^2 vertices must contain either a triangle or an independent set of size k . Equivalently – and this is what I shall actually prove – if a graph does *not* contain a triangle or an independent set of size k , then it must have fewer than k^2 vertices.

The argument goes as follows. Let v be any vertex in the graph. No two vertices that are joined to v can be joined to each other, since then we would have a triangle. Thus, the vertices that are joined to v form an independent set. It follows that v cannot be joined to more than $k - 1$ other vertices (since we are assuming that the graph does not contain an independent set of size k).

Now imagine that we are trying to find a large independent set. One way we might go about it is simply to pick a sequence of vertices v_1, v_2, v_3, \dots , making sure that each new vertex we pick is not joined to any of the previous vertices. Can we do this? Well, suppose that so far we have chosen the vertices v_1, v_2, \dots, v_j . Each one is joined to at most $k - 1$ vertices, so between them they are joined to at most $(k - 1)j$ vertices. Therefore, when it comes to picking a new vertex, there are at most $j + (k - 1)j = kj$ vertices that we must avoid (the first j being the vertices v_1, \dots, v_j). So as long as the number of vertices is greater than kj , we can extend the sequence. If we want to get j all the way up to k , then we need the number of vertices to be greater than $k(k - 1)$. Since k^2 is greater than this, if we have k^2 vertices, we will be able to continue our sequence up to at least v_k , and that will give us an independent set of size k .

Do we really need that many vertices to guarantee a triangle or an independent set of size k ? Well, the argument gives a slightly stronger result than k^2 : it shows that $k(k - 1) + 1$ vertices will suffice. But are there graphs with $k(k - 1)$ vertices that do *not* contain triangles or independent sets of size k ? If not, how many vertices can such a graph have? The largest possible number is called the *Ramsey number* $R(3, k)$.

The argument above is (by the standards of research mathematics) very simple, but deciding whether it can be improved turns out to be much harder. This was the problem that Ajtai, Komlós and Szemerédi solved: they proved that $R(3, k)$ is at most $k^2 / \log k$, to within a constant factor. If you really want to appreciate this achievement, you should spend a while thinking about the problem for yourself, but if you do not have the time for that, then you can at least bear in mind that the problem was decades old, and that fifteen years later Jeong Han Kim, in another famous paper, proved that the result was best possible: that is, $R(3, k)$ is not merely *at most* $k^2 / \log k$ but actually *equal* to $k^2 / \log k$ (again, up to a constant factor).

In order to prove their result, Ajtai, Komlós and Szemerédi had somehow to do better than the easy argument sketched above. Very roughly, what they did was to run the same argument, but to do so in a much more careful way. Recall that the basic idea is to pick vertices v_1, v_2, \dots and avoid the neighbours of vertices you have already chosen. We can describe this process as follows. We start with the entire graph. We pick a vertex v_1 and then we throw away all its neighbours. Then from the remaining part of the graph we pick a vertex v_2 and throw away all *its* neighbours. And we carry on like this. Because no vertex has more than $k - 1$ neighbours, we don't throw away too many vertices at each stage, so we can continue for quite a long time.

Now it is clear that if at any stage of this process we could find a vertex with considerably fewer than $k - 1$ neighbours, we would be very pleased: we could choose that vertex, and then we wouldn't have to throw away so much of the graph, so the process could go on for longer. But why should we be able to find such a vertex? To begin with, there is no reason. But we do have some control over the process: when we choose a vertex v_i , we throw away its neighbours. If these neighbours themselves have many neighbours, then when we throw them away, we will also have thrown away a large number of edges. So the basic idea is to choose the sequence v_1, v_2, \dots in such a way that the number of edges in the part of the graph that you have not thrown away goes down as much as possible. Ajtai, Komlós and Szemerédi showed that it was possible to pursue this idea and end up prolonging the process so that it can be continued for roughly $\log k$ times as long as the simple argument would suggest.

The argument I have just hinted at was not in fact the first proof that Ajtai, Komlós and Szemerédi came up with, but it is easier to describe. Their earlier argument is also important, because it was an early example of the so-called *semirandom method*, which I shall not discuss here, except to say that it is another general technique that has had many applications.

7. CONCLUSION

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 theorems. 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.