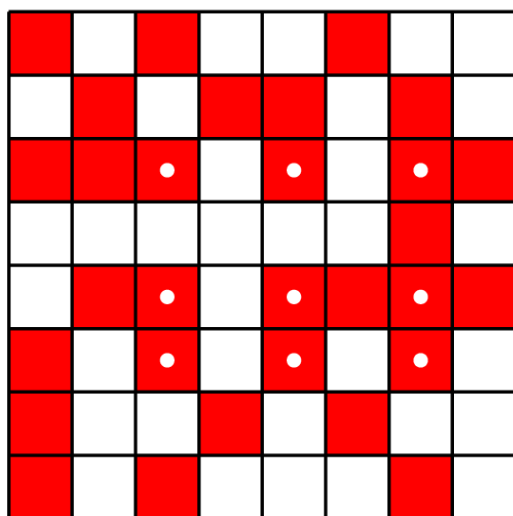


Favourite Open Problems: Prof. Benny Sudakov

Prof. Benny Sudakov is a Professor at ETH Zürich, a leading figure in Combinatorics, and currently a Visiting Professor in the Mathematics Department at LSE. We asked him to tell us about one of his favourite open problems. The post below was written by ***Dr Neil Olver*** and Prof. Benny Sudakov.

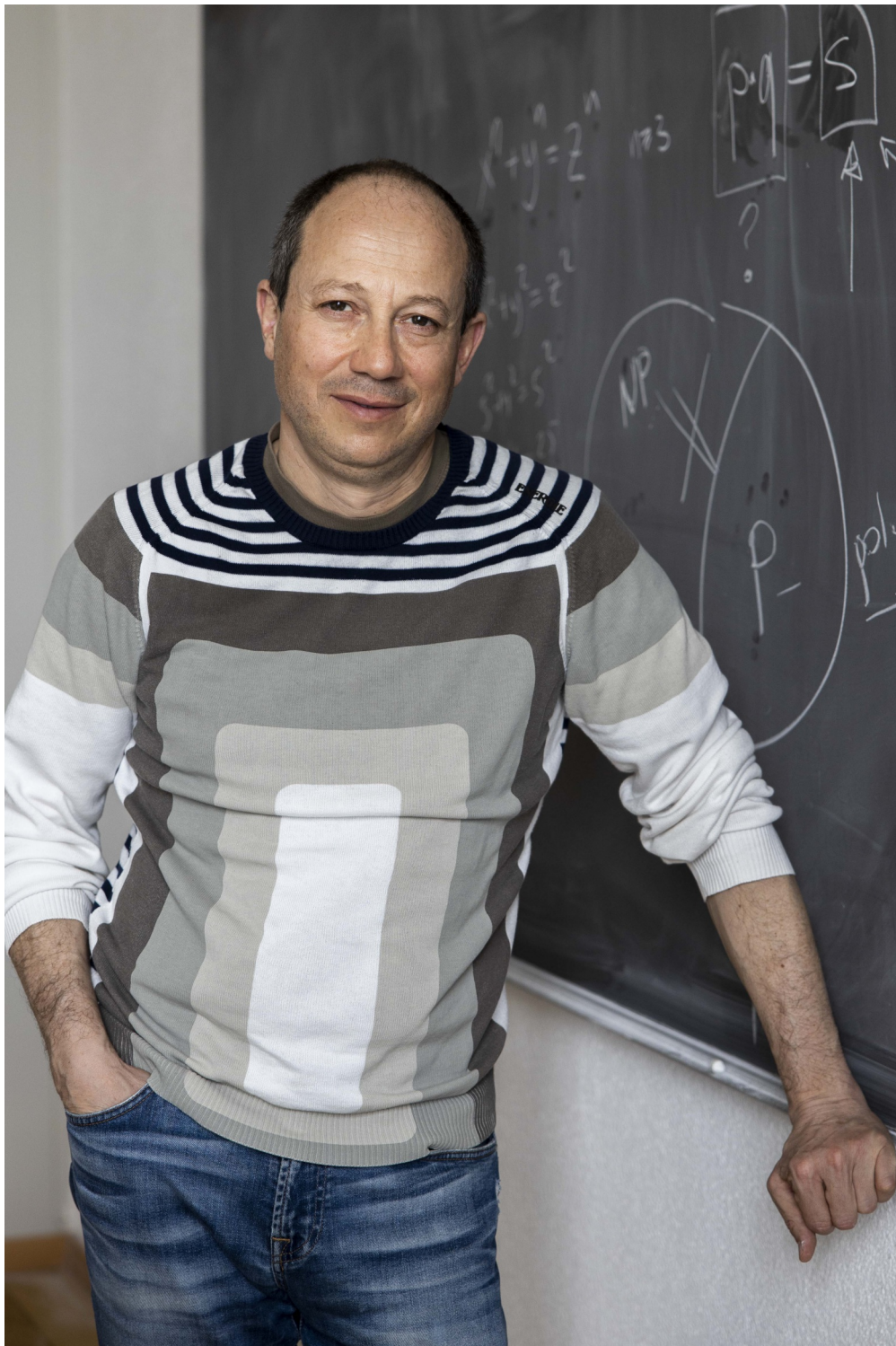
Suppose we have a square grid, with n rows and n columns. Some of the squares are coloured red, and some are uncoloured.

We say that we have a *hidden square of size k* if we can remove $n - k$ columns and $n - k$ rows so that what remains is a $k \times k$ grid where every square is red. The figure below shows an example of containing a hidden square of size 3, but no hidden square of size 4.



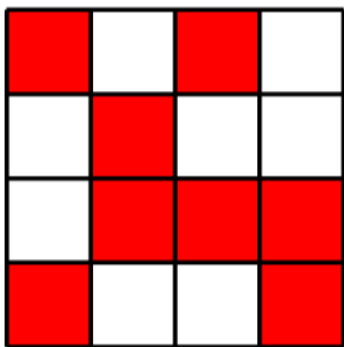
A warm-up question

Can you colour a 4×4 grid in such a way that at least half of the squares are red, and yet there is no hidden square of size 2? What about a 5×5 grid? You might want to try to answer this for yourself – the solution will be discussed below.



Benny Sudakov

For a 4×4 grid, the answer is yes, as demonstrated by the following example (other solutions are possible).



For a 5×5 grid however, the answer is no – there will always be a hidden square of size 2. Here's a (crude) argument as to why this is the case. First, if we had a column with all entries red, then any other column with at least two coloured entries would provide a hidden 2×2 square; and if all other columns had just one red entry, that's only 9 red squares, whereas we must have at least 13. If we had a column with four red entries, we can argue similarly; if any other column had 3 red entries, that would give us a hidden square, so all other columns have at most 2 red entries, but that gives us only $4 + 2 \cdot 4 = 12$ red squares.

So we can restrict our attention to the case that each column has at most 3 coloured squares. There must be at least three columns with 3 coloured squares (if we only had 2, we would have at most $2 \cdot 3 + 3 \cdot 2 = 12$ coloured squares). But then it's not too difficult to convince yourself that we cannot organise these 3 columns such that each pair has at most one coloured row in common.

What about if we only require 10% of the squares to be coloured? Of course, one can now colour a 5×5 grid without a hidden 2×2 square, but what about larger grids — can you colour an arbitrarily large grid with 10% of the squares coloured and avoid a hidden square of size 2? Or colouring only some α fraction of the entries, for some very small $\alpha > 0$? What about hidden squares of larger size?

There is a very nice answer to this question. Fix any k and $\alpha > 0$. Then for big enough n ("big enough" depends on k and α), any colouring of an $n \times n$ grid with at least an α

fraction of entries coloured will have a $k \times k$ hidden square. We will say more about the result, its context, and history shortly.

The open question

But here we come to the open question: *how big is “big enough”?*

For any k and $n \geq k$, let $\alpha(n,k)$ be the smallest value α for which any colouring of an $n \times n$ grid with more than α fraction of the entries coloured *must* have a hidden $k \times k$ square. It was shown by Kővári, Sós and Turán in 1954 that there are constants c_1, c_2, \dots such that $\alpha(n,k) \leq c_k n^{-1/k}$ for all n and k . That is, for fixed k , $\alpha(n,k)$ doesn't grow faster than $n^{-1/k}$ as n gets large. But is this the correct answer? Are there constants d_1, d_2, \dots so that $\alpha(n,k) \geq d_k n^{-1/k}$ for all n and k ?

This is true for $k = 2$ and $k = 3$, but open for $k \geq 4$.

This is an intriguing question, but why is it one of Prof. Sudakov's favourite open problems? To get a sense of this, we will discuss the question again in its original context, namely *extremal graph theory*.

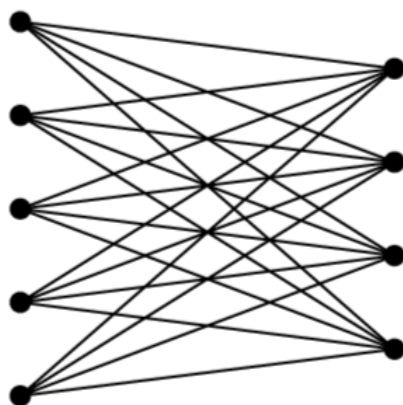
A brief history of extremal graph theory

Willem Mantel was an officer in the Dutch army around the start of the 20th century, but also a mathematician. In those days, a form of recreational maths journal was popular; you would send in a problem, as well as your solution. Your problem (without solution) would be published, and the solution revealed in a later issue. Mantel proposed (and answered) the following problem in 1907.

How many edges can a graph on n vertices have, if it does not contain any triangles?

For example, if $n = 4$, then one can easily get 4 edges; but with 5 edges there will always be a triangle.

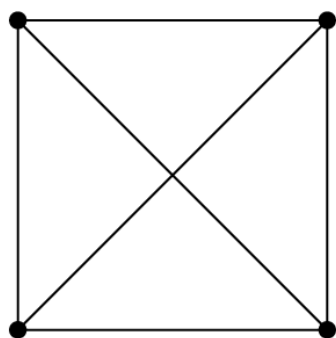
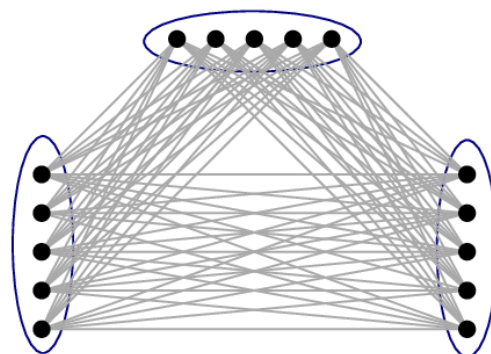
Mantel proved that the answer is roughly $n^2/4$ (more precisely, it is $n^2/4$ if n is even, and $(n^2-1)/4$ if n is odd). The construction is quite simple: divide the nodes of the graph into two equal parts (as equal as possible), and put in all possible edges between these two parts, but no edges within a part. The figure below shows this for $n = 9$; we will denote this graph by $K_{5,4}$, representing that it consists of all possible edges between a “left” side of size 5 and a right side of size 4. It’s easy to convince yourself that this graph has no triangle; indeed, it has no cycle of odd length. The more difficult thing is to show that this is the best possible construction.



The graph $K_{5,4}$.

Turán generalised this theorem in 1941. Instead of disallowing *triangles*, what if we disallow a clique of size $r + 1$, for some $r \geq 2$? A clique of size $r + 1$, denoted by K_{r+1} , is simply the graph on $r + 1$ nodes where every single pair of nodes is connected by an edge; so a clique of size 3 is precisely a triangle. How many edges can one have without containing a K_{r+1} , for some given r ?

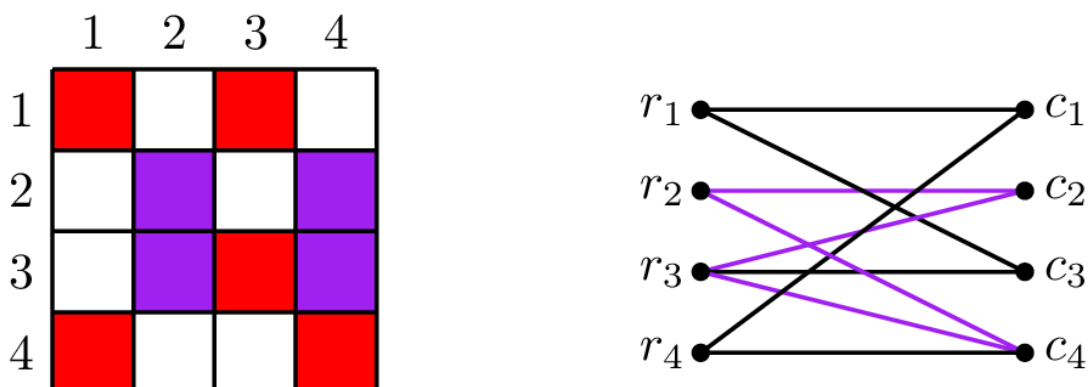
The answer turns out to be what you might guess. Divide the nodes of the graph into r parts, as equally as possible. Now add an edge between every pair of nodes in distinct parts. This graph will have (roughly) $(r - 1)/2r \cdot n^2$ edges, and clearly contains no K_{r+1} , since for any subset of size $r + 1$, two nodes will have to be in the same part, and hence will not be connected by an edge. Again, the hard part is showing that this is the best possible construction.

 K_4 15 node graph with no K_4

We can further generalise the question. Instead of looking for graphs without cliques, what about graphs missing some other arbitrary graph H ? It turns out we understand the answer very well for most graphs H . There is one important class of exceptions however; we do not know the answer for *bipartite* graphs, and this is where our hidden square problem comes in.

The connection with the hidden square problem

Suppose we are given a $n \times n$ grid with some red squares. We can construct an associated graph as follows. It will have n nodes corresponding to the n rows (call these nodes r_1, r_2, \dots, r_n), and n nodes corresponding to the n columns (call these c_1, c_2, \dots, c_n). Then for every grid position (i, j) which is coloured, we include the edge between r_i and c_j ; if the grid entry is not coloured, then we do not include the edge.



In this view, a hidden square of size 2 has a clear interpretation: it is nothing more than a collection of 2 column nodes and 2 row nodes such that each of the column nodes is connected by an edge to each of the row nodes; in other words, a copy of $K_{2,2}$. (In the example above, $\{r_2, r_3, c_2, c_4\}$ is a copy of $K_{2,2}$.) Similarly, if we ask the question for a hidden square of size k , then we are asking whether the corresponding graph has a copy of $K_{k,k}$.

Our open question, phrased in terms of this graph viewpoint, is the following.

Let $\alpha(n,k)$ be the smallest value of α for which every subgraph of $K_{n,n}$ with at least αn^2 edges must contain a copy of $K_{k,k}$. Then is it true that there are constants d_1, d_2, \dots so that $\alpha(n,k) \geq d_k n^{-1/k}$ for all n and k ?

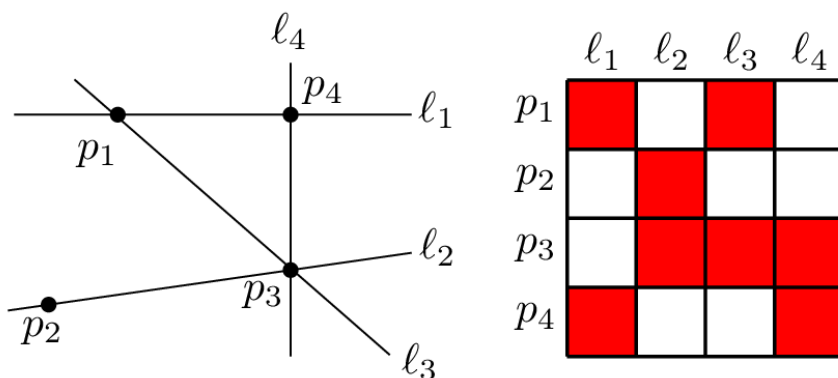
So the question fits squarely in the domain of extremal graph theory. This is a very active area of research, and the Department of Mathematics at LSE has a number of experts on the topic ([Prof. Peter Allen](#), [Prof. Julia Böttcher](#) and [Prof. Jozef Skokan](#)).

An application in geometry

Here is an example of how results of this type can be useful for a problem in a slightly different area of mathematics. Suppose we have a set P of n distinct points on the plane, as well as a set L of n distinct lines (a line extends infinitely in each direction). An *incidence* is a pair (p, ℓ) , with $p \in P$ and $\ell \in L$, where p lies on ℓ . The question is: what is the largest possible number of incidences possible, as a function of n ?

The obvious trivial upper bound is n^2 , if we could arrange somehow that every single point in P lies on every line of L , but it's easy to see that this is nowhere near possible. Can we do better?

We can! Observe that if we have two points $p_1, p_2 \in P$, then there cannot be two distinct lines $\ell_1, \ell_2 \in L$ so that (p_1, ℓ_1) , (p_1, ℓ_2) , (p_2, ℓ_1) and (p_2, ℓ_2) are all incidences. So given any choice of P and L , construct an $n \times n$ grid where the rows are labelled by the points in P , the columns are labelled by the points in L , and a square at position (p, ℓ) is coloured red precisely if (p, ℓ) is an incidence. Then this colouring has no hidden 2×2 square. Hence our upper bound on the number of red squares without a hidden square of size 2 applies, and so there can be at most $cn^{3/2}$ incidences, for some constant c .



This is not the correct answer (the correct exponent is $4/3$, not $3/2$), but it's very easy to obtain. And there are a number of other problems in geometry where these results from extremal graph theory are very useful.