Combinatorial geometry and graph colorings: from algebra to probability

A.M. Raigorodskii, with the help of V. Bulankina and A. Gusev

1 Definitions and notation

One of the most famous and fascinating objects of combinatorial geometry is *the chromatic number of a space*. Before we introduce it, we remind that the space $\mathbb{R}^n$, which is called *the n-dimensional Euclidean space*, is just the set of all "points" $\mathbf{x}$, each of which is a sequence consisting of n real numbers: $\mathbf{x} = (x_1, \ldots, x_n)$. Moreover, between any two points $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$, one can find the distance using the formula

$$|\mathbf{x} - \mathbf{y}| = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}.$$

In particular, for $n = 1$, we get the usual line, for $n = 2$ — the usual plane, for $n = 3$ — the usual space.

The chromatic number of $\mathbb{R}^n$ is the quantity denoted by $\chi(\mathbb{R}^n)$ and equal to the minimum number of colors needed to color all the points of the space $\mathbb{R}^n$, so that the distance between any two points of the same color is not 1.

We will start from the simplest facts, which are widely known, and we will finally come to advanced results obtained just few months before the Summer Conference. Moreover, the methods, which we shall study, will be very different and nontrivial varying from linear algebra to probability theory and random graphs.

2 Problems before the intermidiate finish

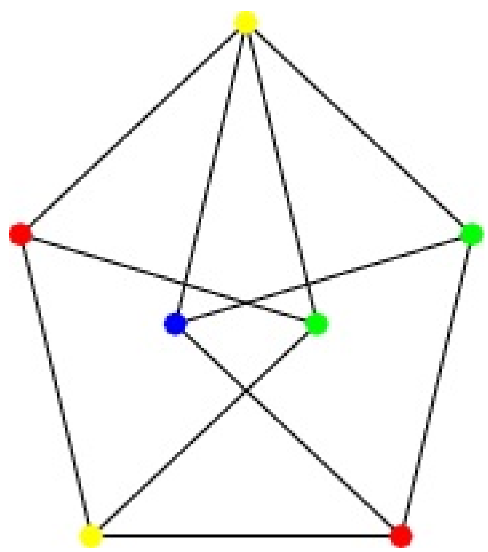
2.1 The simplest bounds for the chromatic number

Problem 1. Prove that $\chi(\mathbb{R}^1) = 2$.

Solution. Clearly, one color is not enough since points 1 and 2 should be of different colors. Two colors suffice. Indeed, let us split the line into semi-open intervals $[a; a+1)$, $a \in \mathbb{Z}$, and color them in turn into black and white.

Problem 2. Prove that $\chi(\mathbb{R}^2) \geqslant 4$.

Solution. Clearly, the figure below cannot be colored properly into three colors.



Problem 3. Prove that $\chi(\mathbb{R}^2) \leqslant 7$.

Solution. One can see that the following coloring is proper:

Problem 4. Prove that $\chi(\mathbb{R}^3) \leqslant 27$.

Solution. It is the particular case of Problem 7.

Problem 5. Prove that $\chi(\mathbb{R}^3) \geqslant 5$.

Solution. It is the particular case of Problem 9.

Problem 6. Prove that $\chi(\mathbb{R}^n)$ is finite for every n.

Solution. Problem 6 will follow from Problem 7.

Problem 7*. Prove that $\chi(\mathbb{R}^n) \leqslant (\lceil \sqrt{n} \rceil + 1)^n$.

Solution. Let us take some $k \in \mathbb{Z}$ and $p \in \mathbb{R}$, tile $\mathbb{R}^n$ with cubes of edge kp, split each edge of some cube into k parts, it will give a splitting of a cube into k^n small cubes of edge p, then color each small cube into its color and then shift this coloring to color the whole space. For the coloring to be proper, the following two inequalities should hold: $p\sqrt{n} \leqslant 1$, $(k-1)p \geqslant 1$. But for $k = \lceil \sqrt{n} \rceil + 1$ and $p = \frac{1}{\sqrt{n}}$ both inequalities are held.

Problem 8. Prove that in $\mathbb{R}^n$ there is a set of $n+1$ points, whose pairwise distances are equal to 1, and therefore, $\chi(\mathbb{R}^n) \geqslant n+1$.

Solution. Let us construct such a set of points explicitly. Let

$$A_1 = \left(\frac{\sqrt{2}}{2}; 0; 0; \ldots; 0; 0\right),$$
$$A_2 = \left(0; \frac{\sqrt{2}}{2}; 0; \ldots; 0; 0\right),$$
$$\ldots,$$
$$A_n = \left(0; 0; 0; \ldots; 0; \frac{\sqrt{2}}{2}\right).$$

Pairwise distances between these n points are all equal to 1. Now find the $(n+1)$th point of the form $S = (a; a; a; \ldots; a; a)$. For this we need

$$SA_i = \sqrt{(n-1)a^2 + \left(a - \frac{\sqrt{2}}{2}\right)^2} = 1,$$
$$na^2 - \sqrt{2}a - \frac{1}{2} = 0.$$

This system gives us a quadratic equation with positive discriminant. Thus, we will find two points S_1 and S_2, such that for each of them and for every $j = 1, 2, \ldots, n$, the distance between S_i and A_j equals 1.

Problem 9*. Prove that $\chi(\mathbb{R}^n) \geqslant n+2$.

Solution. Let us take two equal figures as in Problem 8, namely, $S_1S_2A_1\ldots A_n$ and $T_1T_2B_1\ldots B_n$. We can place them in $\mathbb{R}^n$ in such a way that S_1 coincides with T_1, and the distance S_2T_2 equals 1. Now, if we color $\mathbb{R}^n$ into $n+1$ color, then the points S_1 and S_2 will be of one color, and the points T_1 and T_2 will be of one color. Then S_2 and T_2 will be of one color, which is impossible.

2.2 Distance graphs of special type, their simplest properties, and the connection with the chromatic number of a space

Recall that the scalar product of vectors $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$ is the expression

$$(\mathbf{x}, \mathbf{y}) = x_1y_1 + \ldots + x_ny_n.$$

One can easily check that always

$$|\mathbf{x}-\mathbf{y}|^2 = (\mathbf{x},\mathbf{x}) + (\mathbf{y},\mathbf{y}) - 2(\mathbf{x},\mathbf{y}). \tag{1}$$

Let r, s be some natural numbers. For each $n \in \mathbb{N}$ denote by $G(n,r,s)$ the graph, whose set of vertices is

$$V(n,r) = \{\mathbf{x} = (x_1,\ldots,x_n):\ x_i \in \{0,1\},\ x_1+\ldots+x_n = r\}$$

and whose set of edges is

$$E(n,r,s) = \{\{\mathbf{x},\mathbf{y}\}:\ (\mathbf{x},\mathbf{y}) = s\}.$$

In other words, vertices are all possible vectors consisting of 0s and 1s, such that in every such vector one has exactly r 1s and $n-r$ 0s. In turn, those and only those vertices are joined by edges whose scalar product equals s. Due to formula (1) one can say that edges are those and only those pairs of vertices whose distance equals $\sqrt{2r-2s}$. This is why the graphs $G(n,r,s)$ are called *distance graphs*.

It is also convenient to have the following interpretation of a graph $G(n,r,s)$. Its vertices can be considered as all possible r-element subsets of the set $\mathcal{R}_n = \{1,2,\ldots,n\}$. Its edges can be considered as pairs of subsets whose intersections have cardinalities equal to s. Please make sure that you understand it!

Recall that an *independent set* of vertices of a graph is a set, in which every two vertices are not joined by an edge. The *independence number* $\alpha(G)$ of a graph G is the number of vertices in any maximal (by cardinality) independent set. The *chromatic number* $\chi(G)$ of a graph G is the minimum number of colors needed to color all the vertices of the graph in such a way that between any two vertices of the same color, there are no edges.

Problem 10. Prove that for any n, r, s, one has $\chi(\mathbb{R}^n) \geqslant \chi(G(n,r,s))$.

Solution. The graph $G(n,r,s)$ can be embedded into $\mathbb{R}^n$: first, we map every vector consisting of 0s and 1s to the corresponding point, then the distances between all points are equal to $\sqrt{2r-2s}$, and second, we make a homothety with coefficient $\frac{1}{\sqrt{2r-2s}}$, thus making all pairwise distances equal to 1. Hence, if we can properly color the space into k colors, then we can also properly color the vertices of the graph $G(n,r,s)$.

Problem 11. Prove that for every graph $G = (V,E)$, one has $\chi(G) \geqslant \frac{|V|}{\alpha(G)}$.

Solution. Every color should be an independent set, hence, its cardinality will not exceed $\alpha(G)$. Let us increase the quantity of vertices of every color and make it equal to $\alpha(G)$, then the total number of vertices will increase and become equal to $\chi(G)\cdot\alpha(G) \geqslant |V|$. It implies the desired inequality.

Problem 12. Find $\alpha(G(n,3,1))$. Derive from the obtained result a considerable improvement to the bound in Problem 9.

Solution. The answer depends on the residue of n modulo 4. We will prove by induction that

$$n \equiv 0 (\mathrm{mod}\ 4) \Rightarrow \alpha(G(n,3,1)) = n,$$
$$n \equiv 1 (\mathrm{mod}\ 4) \Rightarrow \alpha(G(n,3,1)) = n-1,$$
$$n \equiv 0 (\mathrm{mod}\ 4) \Rightarrow \alpha(G(n,3,1)) = n-2,$$
$$n \equiv 0 (\mathrm{mod}\ 4) \Rightarrow \alpha(G(n,3,1)) = n-2,$$

Every vector consists of zeroes and ones. Fix any independent set. It follows from the condition that every two distinct vectors in it either do not intersect, or intersect by 2 elements. First we give examples which show that $\alpha(G(n,3,1)$ is at least the required number. For $n=4$

$$\begin{aligned}&(0,1,1,1),\\&(1,0,1,1),\\&(1,1,0,1),\\&(1,1,1,0).\end{aligned}\tag{1}$$

For n divisible by 4 we split the coordinates into groups of 4 and in every group perform the same construction as above.

For $n \equiv 1(\mathrm{mod}\ 4)$ we forget one coordinate and repeat the same construction as above.

For other values of n, we give another construction:

$$\begin{aligned}(1,1,1,0,\ldots,0),\\(1,1,0,1,0,\ldots,0),\\(1,1,0,0,1,0,\ldots,0),\\\ldots,\\(1,1,0,0,\ldots,0,1,0),\\(1,1,0,0,\ldots,0,0,1),\end{aligned}\tag{2}$$

Here two first coordinates are always equal to 1, hence, every two vectors intersect in exactly two coordinates. This example contains $n-2$ vectors.

Now we prove that it is the maximal possible value. Take two intersecting vectors, without loss of generality they are

$$(1,1,1,0,\ldots,1) \text{ and } (1,1,0,1,0,\ldots,0).$$

Now take any other vector. Either it has a nonzero coordinate on the position 5–n, or it is as in (1). In the first case the vector is of form $(1,1,0,0,\ldots,0,1,0,\ldots)$, and the vectors of the second type are no more possible, in the second case we have no more than 4 vectors which have nonzero coordinates on the first 4 positions, and we can apply the induction step. In the first case we have no more than $n-2$ vectors, as in (2). It is not more than in the answer, and we are done.

We get the following bound for the chromatic number:

$$\chi(\mathbb{R}^n) \geqslant \chi(G(n,3,1)) \geqslant \frac{C_n^3}{\alpha(G(n,3,1))} \geqslant \frac{(n-1)(n-2)}{6}.$$

Problem 13*. Find $\chi(G(n,3,1))$ for $n=2^k$. **Hint.** Use Problems 11 and 12 as well as the following lemma and induction by k.

Solution. The proof of this theorem can be found in [6].

Lemma 1. *Let n be an even number and P_n be the set of all unordered pairs $\{a,b\}$ of natural numbers both of which do not exceed n. Then there exist such sets of pairs $B_1,\ldots,B_{n-1}$ that*

$$P_n = B_1 \sqcup \ldots \sqcup B_{n-1}.$$

Moreover, for any $i=1,\ldots,n-1$, no two pairs from B_i contain a common element. For odd n, we have a partition

$$P_n = B_1 \sqcup \ldots \sqcup B_n,$$

and, again for any $i=1,\ldots,n$, no two pairs from B_i contain a common element.

Problem 14.** Find sharpest possible bounds (ideally — a formula) for $\chi(G(n,3,1))$ for every n.

Solution. This problem has two stars, hence, if you have solved it, you could obtain a new result.

Problem 15. Prove that $\alpha(G(n,r,s)) \geqslant C_{n-s-1}^{r-s-1}$.

Solution. Consider only those vertices which have ones at the first $s+1$ positions. There are C_{n-s-1}^{r-s-1} such vertices. The scalar product of any two of them is at least $s+1$, hence, no two of them are joined by an edge.

Problem 16*. Prove that $\alpha(G(n,r,0)) = C_{n-1}^{r-1}$, if $2r \leqslant n$.

Solution. Let us denote by B any (fixed throughout the proof) subset in the set of vertices $V(n,r)$, which is independent in $G(n,r,0)$ and has cardinality $\alpha(G(n,r,0))$.

Lemma. For every $s, 1 \leqslant s \leqslant n$, let us consider the set $A_s = \{s, s+1, \ldots, s+r-1\}$, where the sum is modulo n. Then B cannot contain more than r subsets of form A_s.

Proof. Fix any A_s from $\alpha(G(n,r,0))$. Take all the other sets A_k intersecting A_s and form $r-1$ pairs $\{A_{s-i}, A_{s+r-i}\}$. Now the statement of Lemma follows from the fact that B contains at most one element from each pair.

Let us deduce the Problem statement from the Lemma. Take an arbitrary permutation σ of elements $\{1, 2, \ldots, n\}$ and fix $i \in \{1, 2, \ldots, n\}$. Let $A = \{\sigma(i), \sigma(i+1), \ldots, \sigma(i+r-1)\}$, where the sum is taken modulo n, as above. Since σ was an arbitrary permutation, it follows from Lemma that $P(A \in B) \leqslant \frac{r}{n}$. But A is taken equiprobably from all the r-element subsets. Hence,

$$\frac{r}{n} \geqslant P(A \in B) = \frac{|B|}{C_n^r}$$

and

$$|B| \leqslant \frac{r}{n} C_n^r = C_{n-1}^{r-1}.$$

Problem 17. Prove that $\chi(G(n,r,0)) \leqslant n - 2r + 2$, if $2r \leqslant n$.

Solution. Let us color all the vertices having 1st coordinate equal to one into the first color. Then let us color all the yet uncolored vertices, which have second coordinate equal to 1, into the second color, and so on till the $(n-2r+1)$th color. Clearly, there will be no edges between the vertices of one color. It remains to color some vertices, and we have one more color. But what vertices are yet uncolored? Only those who have only ones on the last $2r-1$ coordinates. Clearly, there are no edges between these vertices, hence, we can color them all in the remaining color.

Problem 18. Prove that $\chi(G(n,r,s)) \leqslant C_r^s C_{n-r}^{r-s} + 1$.

Solution. What is $C_r^s C_{n-r}^{r-s}$? It is the degree of every vertex of the graph $G(n,r,s)$. It is easy to see that if the degree of every vertex of the graph does not exceed d, then the graph can be properly colored into $d+1$ colors. Indeed, let us color the vertices one by one. For every yet uncolored vertex, there exists at least one color such that our vertex is not connected with vertices of this color. Hence, we can color it in this color.

Problem 19. Prove that $\chi(G(n,r,s)) \leqslant C_n^{s+1}$.

Solution. Let us enumerate all the $(s+1)$-element subsets of $\mathcal{R}_n$ with numbers $1,\ldots, C_n^{s+1}$. Now let us color all the vertices containing the first subset into the first color, all the yet uncolored vertices containing the second subset into the second color, and so on. Finally, all the vertices will be colored, and there will be no edges between the vertices of one color.

Problem 20*. Let $k = \left[\frac{r-1}{s}\right]$. Prove that $\chi(G(n,r,s)) \leqslant k \cdot C^{s+1}_{\lceil \frac{n}{k} \rceil}$.

Solution. The proof of this fact can be found in [6].

Problem 21*. Prove that $n-r+1 \leqslant \chi(G(n,r,r-1)) \leqslant n$ for $n = 2^k$. **Hint.** Use Lemma 1 (see above) and induction by r and k.

Solution. To prove the lower bound, we show that the graph contains a clique of size $n-r+1$. Indeed, let us take the vertices corresponding to the vectors having ones on the first $r-1$ coordinates. There are exactly $n-r+1$ such vertices, and they are all mutually joined by edges.

Consider the particular case $r=2$. Let us show that Lemma 1 implies the inequality $\chi(G(n,2,1)) \leqslant n$. Indeed, let us identify the vertices of $G(n,2,1)$ with the set of pairs of integers $1,\ldots,n$, and two vertices are joined by an edge if and only if the corresponding pairs have a common element. Hence, if we take colors $\chi_i = B_i$ from Lemma 1, then the graph will be colored into $n-1$ colors for $n = 2m$ and into n colors for $n = 2m+1$.

Now take arbitrary r and prove the upper bound by induction on r. Its base $r=2$ is proved above. Suppose that the upper bound is true for all $r < k$. Let us prove the step, for this we need to color the graph $G(n,k,k-1)$.

Let $n = 2^l$. To color the graph $G(n,k,k-1)$, we use induction on l. The base will be for the maximal l_0 such that $2^{l_0} < k$. For this l_0 we have $\chi(G(2^{l_0},k,k-1)) = 0$, and the upper bound is satisfied. Now we prove the induction step. We assume that $\chi(G(2^{l-1},k,k-1)) \leqslant 2^{l-1}$. Let us split the set of elements into two parts:

$$A_1 = \{1,\ldots,n/2\},$$
$$A_2 = \{n/2+1,\ldots,n\}.$$

Split the vertices of our graph into $k+1$ pairwise disjoint sets

$$V_j = \{v \in V(G(n,k,k-1))\colon |v \cap A_1| = j\}, \quad j = 0,\ldots,k.$$

We note that for $l = l_0+1$ some V_js are empty. It will only improve the bound, hence, below we assume that all V_j are nonempty.

Consider the graph $G(n/2,j,j-1)$ for $j = 1,\ldots,k-1$. By the induction assumption, $\chi(G(n/2,j,j-1)) \leqslant n/2$, and all the j-element subsets of the set A_1 can be colored into $n/2$ colors $\varphi_1,\ldots,\varphi_{n/2}$ in such a way that every two such subsets of one color intersect by a number of elements not equal to $j-1$. Let us denote by $W^{\nu}_{j,1}$ the set of all such subsets which are colored with color φ_ν. We also introduce the graph $G(n/2,k-j,k-j-1)$ constructed on the set of elements of A_2. Then all the $(k-j)$-element subsets of the set A_2 can be colored with colors $\psi_1,\ldots,\psi_{n/2}$ in such a way that subsets, colored with one color, intersect by a number of elements not equal to $k-j-1$. We denote by $W^{\mu}_{k-j,2}$ the set of all $(k-j)$-element subsets of A_2, colored with the color ψ_μ.

For $\nu \in \{1, \ldots, n\}$, let $U_j^{1,\nu} = \{w_1 \sqcup w_2 \colon w_1 \in W_{j,1}^{\nu}, w_2 \in W_{j,2}^{\nu}\}$, and let $U_j^1 = \bigsqcup_{\nu=1}^{n/2} U_j^{1,\nu}$. It is clear that every two sets from the collection U_j^1 intersect by a number of elements not equal to $k-1$. Indeed, if this is not true, then there exist vertices $v, w \in U_j^1$ such that $|v \cap w| = k-1$. Then either $v \cap A_1 = w \cap A_1$, or $v \cap A_2 = w \cap A_2$, hence, the non-coinciding "halves" of these vertices are colored with the same color. Without loss of generality, we assume that $v \cap A_1 = w \cap A_1$, then $v \cap A_2$ and $w \cap A_2$ have at most $k-j-2$ common elements. Hence, the vertices v and w cannot intersect by $k-1$ elements, and we get a contradiction. The case $v \cap A_2 = w \cap A_2$ can be treated analogously.

In the same way we can obtain the collection U_j^2, taking the unions of sets from $W_{j,1}^1$ with sets from $W_{k-j,2}^2$, of sets from $W_{j,1}^2$ with sets from $W_{k-j,2}^3$, $\ldots$, of sets from $W_{1,j}^{n/2}$ with sets from $W_{k-j,2}^1$. Analogously, taking cyclical shifts, we define $U_j^3, \ldots, U_j^{n/2}$. Here $j = 1, \ldots, k-1$. By the assumption of the induction on n, there exist collections U_0^i U_k^i, which contain subsets of A_2 and A_1, respectively, and intersecting by a number of elements not equal to $k-1$ for every $i = 1, \ldots, n/2$.

Now we finally define the color χ_i for $i = 1, \ldots, n/2$ as the union of collections with even indices:

$$\chi_i = U_0^i \cup U_2^i \cup U_4^i \cup \ldots$$

For $i = n/2+1, \ldots, n$ we take the union of collections with odd indices:

$$\chi_i = U_1^{i-n/2} \cup U_3^{i-n/2} \cup U_5^{i-n/2} \ldots$$

Clearly, we colored our graph with not more than n colors, which ends the induction on the parameter n, and, hence, it end also the induction on r. The upper bound is proved.

Problem 22.** Find $\chi(G(n, r, r-1))$ or at least refine the bounds from Problem 21.

Solution. This problem has two stars, hence, if you have solved it, you could obtain a new result.

Please find out that none of the results that you have obtained allows you to improve the lower bounds of the value $\chi(\mathbb{R}^n)$ found in Problems 9 and 12. In view of Problem 11, it will be good to study upper bounds for the independence numbers of the graphs $G(n, r, s)$. It turns out that many of such bounds can be obtained with the help of the linear algebra method. Thus, in the next section, we will recall some basic notions of linear algebra.

2.3 Basics of linear algebra and its applications

We say that vectors $\mathbf{x}_1, \ldots, \mathbf{x}_t$ in $\mathbb{R}^n$ are *linearly independent*, if the equality $c_1\mathbf{x}_1 + \ldots + c_t\mathbf{x}_t = 0$ is possible only in the case when $c_1 = \ldots = c_t = 0$.

Problem 23. Prove that the maximum number of linearly independent vectors in $\mathbb{R}^n$ equals n.

Solution. The basis vectors $\mathbf{e}_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$, where one is at the ith position, $i = 1, 2, \ldots, n$, are linearly independent, since $c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \ldots + c_n\mathbf{e}_n = (c_1, c_2, \ldots, c_n)$. Now we prove by induction on n that every $n+1$ vectors in $\mathbb{R}^n$ are linearly dependent.

Base $n = 1$ is obvious.

Step from n to $n+1$. If the $(n+1)$th coordinate of all the $n+2$ vectors is zero, then we can forget it, and, using the induction assumption, find a nontrivial linear combination of the given $n+2$ vectors such that the resulting vector is zero. It will remain zero when we add back the last coordinate. Let a_k be the $(n+1)$th coordinate of the vector $\mathbf{x}_k$. Now, if the $(n+1)$th coordinate of some vector $\mathbf{x}_i$ is nonzero, then replace our set of vectors with $\mathbf{x}'_1, \mathbf{x}'_2, \ldots, \mathbf{x}'_{n+1}, \mathbf{x}'_{n+2}$, where $\mathbf{x}'_j = \mathbf{x}_j - (a_j/a_i)\mathbf{x}_i$, $\mathbf{x}'_i = 0$. Then by the induction assumption for the vectors $\mathbf{x}'_1, \ldots, \mathbf{x}'_{i-1}, \mathbf{x}'_{i+1}, \ldots, \mathbf{x}'_{n+2}$, we can find coefficients $c_1, \ldots, c_{n+1}$ not all vanishing such that

$$0 = c_1\mathbf{x}'_1 + \ldots, c_{i-1}\mathbf{x}'_{i-1} + c_i\mathbf{x}'_{i+1} + \ldots + c_{n+1}\mathbf{x}'_{n+2} = c_1(\mathbf{x}_1 - (a_1/a_i)\mathbf{x}_i) + \ldots + c_{n+1}(\mathbf{x}_{n+2} - (a_{n+2}/a_i)\mathbf{x}_i) =$$
$$= c_1\mathbf{x}_1 + \ldots + c_{i-1}\mathbf{x}_{i-1} + (-c_1(a_1/a_i) - c_2(a_2/a_i) - \ldots - c_{i-1}(a_{i-1}/a_i) - c_i(a_{i+1}/a_i) - \ldots - c_{n+1}(a_{n+2}/a_i))\mathbf{x}_i +$$
$$+ c_i\mathbf{x}_{i+1} + \ldots + c_{n+1}\mathbf{x}_{n+2}.$$

Hence, we presented a nontrivial linear combination of the initial vectors resulting into a zero vector. We get a contradiction.

Problem 24. Prove that if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ form an arbitrary system of linearly independent vectors in $\mathbb{R}^n$, then any vector $\mathbf{x} \in \mathbb{R}^n$ can be represented as $\mathbf{x} = c_1\mathbf{x}_1 + \ldots + c_n\mathbf{x}_n$, where $c_1, \ldots, c_n$ are real numbers. (The system $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is called a *basis* of the space and the expression $c_1\mathbf{x}_1 + \ldots + c_n\mathbf{x}_n$ is called *linear combination* of the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with *coefficients* $c_1, \ldots, c_n$. In these terms, any vector $\mathbf{x} \in \mathbb{R}^n$ can be represented as a linear combination of the vectors of the basis.)

Solution. On the contrary, suppose that there exists a vector $\mathbf{x}$ which cannot be represented as a linear combination of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Then the system of vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{x}\}$ will be linearly independent: indeed, let $c_1\mathbf{x}_1 + \ldots + c_n\mathbf{x}_n + c\mathbf{x} = 0$, then either $c = 0$ and all c_i vanish since $\mathbf{x}_i$, $i = 1, 2, \ldots, n$, are linearly independent, or $c \neq 0$, but then $\mathbf{x} = (c_1/c)\mathbf{x}_1 + \ldots (c_n/c)\mathbf{x}_n$, which contradicts our assumption. But the set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{x}\}$ contains $n+1$ vectors, and we get a contradiction with the previous problem.

Let p be a prime number. Let $\mathbb{Z}_p$ be the set of congruences modulo p. The space $\mathbb{Z}_p^n$, similarly to $\mathbb{R}^n$, is just the set of all the sequences of numbers from $\mathbb{Z}_p$. The operations of the sum of "vectors" $\mathbf{x}, \mathbf{y} \in \mathbb{Z}_p^n$ and of their product with elements of $\mathbb{Z}_p$ is done, as usual, coordinate by coordinate, but, this time, every coordinate is taken modulo p.

The notions of linear independence and of a basis for $\mathbb{Z}_p^n$ are defined in the same way as for $\mathbb{R}^n$. However, here all the numbers c_i are elements of $\mathbb{Z}_p$ — not $\mathbb{R}$, — and the equality to zero is understood as the equality to zero modulo p.

Problem 25. Prove that the maximum number of linearly independent vectors in $\mathbb{Z}_p^n$ equals n and that any maximal system forms a basis.

Solution. The proof is the same as the proof of two previous problems. Indeed, every element can be seen as a sequence of residues modulo p, the coefficients are not real numbers but residues modulo p, and the only difference is that by $1/a$ we now mean the inverse element modulo p.

Problem 26. Let $W = \{\mathbf{x}_1, \ldots, \mathbf{x}_t\}$ be an arbitrary independent set of vertices of the graph $G(n, 3, 1)$. Prove that the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_t$ are linearly independent in $\mathbb{Z}_2^n$ and thus $\alpha(G(n, 3, 1)) \leqslant n$, which is only by an "epsilon" weaker than the result of Problem 12!

Solution. On the contrary, suppose that there exist coefficients $c_1, \ldots, c_t \in \mathbb{Z}_2$, not all vanishing, such that

$$c_1\mathbf{x}_1 + \ldots + c_n\mathbf{x}_n = 0.$$

Without loss of generality we may assume that $c_1 \neq 0$. Let us take the scalar product of the both hands of this equality by $\mathbf{x}_1$. Note that if we take the scalar product of two vectors, each has three nonzero coordinates equal to 1, then it equals 3, if the vectors are equal, and 0, 1, or 2 in the other cases. Since we take vectors from an independent subset of the graph $G(n, 3, 1)$, the scalar product cannot equal 1. If we take the obtained equality modulo two, then the first summand of the left hand side will be equal to 1 (since $3 \equiv 1 (\mathrm{mod}\ 2)$), and all the others will be equal to 0, but on the right hand side we have 0. We get a contradiction.

Let $F \in \{\mathbb{R}, \mathbb{Z}_p\}$. Let $x_1, \ldots, x_n$ be "variables". By a *monomial* depending on these n variables we mean an expression of the form $x_1^{a_1} \cdot \ldots \cdot x_n^{a_n}$, where $a_1, \ldots, a_n$ are some non-negative integers. A *polynomial* is an arbitrary linear combination of monomials. More precisely, a polynomial P belongs to $F[x_1, \ldots, x_n]$, if its coefficients are from F. Polynomials are added and multiplied according to the usual rules. Also if $P \in F[x_1, \ldots, x_n]$, then it can be multiplied by any element of F. In any case, the rules of summation and multiplication of the coefficients of polynomials are defined by the rules of summation and multiplication in the set F. The *degree* of a monomial is the sum of the degrees of its variables. The degree of a polynomial is the maximum of the degrees of its monomials. A polynomial $P \in F[x_1, \ldots, x_n]$ *equals zero*, if all its coefficients are equal to zero in F. Polynomials $P_1 \in F[x_1, \ldots, x_n], \ldots, P_t \in F[x_1, \ldots, x_n]$ are *linearly independent* over F, if $c_1P_1 + \ldots + c_tP_t = 0$ only in the case when all the numbers $c_1 \in F, \ldots, c_t \in F$ are equal to zero in F. It is obvious that any polynomial is generated by the *basis* consisting of its monomials.

Problem 27. Prove that if some polynomials are linearly independent over their F, then their number does not exceed the number of monomials in a basis, which generates all these polynomials.

Solution. Every polynomial is a linear combination of basis monomials, and suppose that there exist n basis monomials. Then every polynomial corresponds to a unique sequence of n coefficients from F. In other words, any polynomial can be viewed as a vector in F^n. Then by Problems 24 and 25, the maximal possible number of linearly independent vectors in this space is n, which gives the required bound.

Problem 28. Let $W = \{\mathbf{x}_1, \ldots, \mathbf{x}_t\}$ be an arbitrary independent set of vertices of the graph $G(n, 5, 2)$. Let polynomials $P_1 \in \mathbb{Z}_3[y_1, \ldots, y_n]$, $P_2 \in \mathbb{Z}_3[y_1, \ldots, y_n]$, $\ldots$, $P_t \in \mathbb{Z}_3[y_1, \ldots, y_n]$ be given by formulae

$$P_i(\mathbf{y}) = P_i(y_1, \ldots, y_n) = (\mathbf{x}_i, \mathbf{y})((\mathbf{x}_i, \mathbf{y}) - 1), \quad i = 1, \ldots, t.$$

For example, if $\mathbf{x}_1 = (1, 1, 1, 1, 1, 0, \ldots, 0)$, $\mathbf{x}_2 = (0, \ldots, 0, 1, 1, 1, 1, 1)$, then

$$P_1(y_1, \ldots, y_n) = (y_1+y_2+y_3+y_4+y_5)(y_1+y_2+y_3+y_4+y_5-1) = y_1^2+\ldots+y_5^2+2y_1y_2+\ldots+2y_4y_5-y_1-\ldots-y_5,$$

$$P_2(y_1, \ldots, y_n) = (y_{n-4} + y_{n-3} + y_{n-2} + y_{n-1} + y_n)(y_{n-4} + y_{n-3} + y_{n-2} + y_{n-1} + y_n - 1) =$$
$$= y_{n-4}^2 + \ldots + y_n^2 + 2y_{n-4}y_{n-3} + \ldots + 2y_{n-1}y_n - y_{n-4} - \ldots - y_n.$$

Prove that the polynomials $P_1, \ldots, P_t$ are linearly independent over $\mathbb{Z}_3$ and therefore $\alpha(G(n, 5, 2)) \leqslant C_n^2 + 2C_n^1$.

Solution. As we did in Problem 26, assume the contrary: suppose that there exists a linear dependency, which has a nonzero coefficient at some P_i. Let us substitute $\mathbf{y}$ by $\mathbf{x}_i$. Then the right hand side will remain zero, and all P_j, $j \neq i$, will vanish in the left hand side. Indeed, the scalar product $(\mathbf{x}_i, \mathbf{y})$ can equal 0, 1, 3, or 4, and then the polynomial vanishes modulo 3. The scalar product cannot be equal to two since we started from an independent subset of $G(n, 5, 2)$. It can equal 5 in the only case when two vectors coincides, which gives us the only nonzero summand, and we get a contradiction.

Now let us count the number of possible monomials. We have products of form $y_i y_j$, squares y_i^2, and linear terms y_i. This gives us $C_n^2 + 2C_n^1$ monomials, which implies the required bound for $\alpha(G(n, 5, 2))$.

Problem 29. Assume that in the conditions of the previous problem the polynomials P_i are substituted by P_i' according to the following rule: every monomial of the form y_i^2 is changed by y_i, and after that monomials of the same form are added. Prove that the polynomials $P_1', \dots, P_t'$ corresponding to the vectors from an independent set of vertices W of the graph $G(n, 5, 2)$ are also linearly independent over $\mathbb{Z}_3$, similarly to the initial polynomials $P_1, \dots, P_t$. Derive from this fact the bound $\alpha(G(n, 5, 2)) \leqslant C_n^2$ and compare it with the bound from the problem 15.

Solution. Note that in the previous solution, when we used linear dependencies, we used only the values of the polynomials at the points having coordinates equal either to 0 or to 1. Clearly, the values of the new polynomials will coincide with the values of the old polynomials in these points, hence, the same proof can be applied.

Now note that, if we open the brackets and perform the required substitutions, every polynomial will contain only pairwise products of variables, hence, the basis of the corresponding space contains at most C_n^2 monomials, which gives the required bound.

Now let us compare the bound with the bound of Problem 15. We get that $\alpha(G(n, 5, 2)) \sim \frac{1}{2}n^2$.

Problem 30. Derive from the result of the previous problem a lower bound for $\chi(\mathbb{R}^n)$, which considerably improves the bound from Problem 12. Ensure yourselves, however, that, in view of Problem 19, substantial further advances based on the graph $G(n, 5, 2)$ cannot be done.

Solution. Let us apply the homothety with the coefficient $1/\sqrt{6}$ to the graph $G(n, 5, 2)$ and then apply Problem 11. We obtain that

$$\chi(\mathbb{R}^n) \geqslant \frac{|V|}{\alpha(G(n, 5, 2))} \geqslant C_n^5 / C_n^2 = 60n^3 + O(n^2).$$

On the other hand, using this graph, we cannot bound the chromatic number of order greater than C_n^3.

Problem 31. Let $W = \{\mathbf{x}_1, \dots, \mathbf{x}_t\}$ be an arbitrary independent set of vertices of the graph $G(n, 9, 4)$. Let polynomials $P_1 \in \mathbb{Z}_5[y_1, \dots, y_n]$, $P_2 \in \mathbb{Z}_5[y_1, \dots, y_n]$, $\dots$, $P_t \in \mathbb{Z}_5[y_1, \dots, y_n]$ be given by the formulae

$$P_i(\mathbf{y}) = P_i(y_1, \dots, y_n) = (\mathbf{x}_i, \mathbf{y})((\mathbf{x}_i, \mathbf{y}) - 1)((\mathbf{x}_i, \mathbf{y}) - 2)((\mathbf{x}_i, \mathbf{y}) - 3), \quad i = 1, \dots, t.$$

Prove that the polynomials $P_1, \dots, P_t$ are linearly independent over $\mathbb{Z}_5$.

Solution. As we did in Problem 28, we take a nontrivial linear combination resulting into zero vector. Then we find a nonzero term in this linear combination, substitute the corresponding vector x_i instead of y into the whole linear combination and obtain that on one hand we have identically zero, while on the other hand we have exactly one nonzero term corresponding to x_i. We get a contradiction.

Problem 32. Which upper bound for $\alpha(G(n,9,4))$ follows from the previous problem?

Solution. We need to count the number of monomials necessary to generate all the vectors $P_1, P_2, \dots, P_t$. It suffices to take all the monomials of the form

$$y_iy_jy_ky_l,\ y_i^2y_jy_k,\ y_iy_jy_k,\ y_i^3y_j,\ y_i^2y_j^2,\ y_i^2y_j,\ y_iy_j,\ y_i^4,\ y_i^3,\ y_i^2,\ y_i,\ const.$$

The number of such monomials equals $C_n^4 + 4C_n^3 + 6C_n^2 + 4C_n^1 + 1$, hence, $\alpha(G(n,9,4))$ is bounded by the same value.

Problem 33. Let in the conditions of Problem 31 the polynomials P_i be replaced by some P_i' according to the following rule: in them, every monomial, which appears after opening the brackets and summing similar monomials, has of course the form $y_1^{a_1} \cdot \ldots \cdot y_n^{a_n}$; if among the numbers a_i, one has some numbers greater than or equal to 2, then we replace all of them by 1s and sum up similar monomials. For example, the monomial $y_1^2y_2^2$ is transformed to y_1y_2 and the same is true for the monomials $y_1^2y_2, y_1y_2$, etc. Prove that the polynomials $P_1', \dots, P_t'$ corresponding to the vectors from an independent set W of the graph $G(n,9,4)$ are also linearly independent over $\mathbb{Z}_5$, just as it was with the initial polynomials $P_1, \dots, P_t$. Derive from this fact the bound $\alpha(G(n,9,4)) \leqslant C_n^4 + C_n^3 + C_n^2 + C_n^1 + C_n^0$ and compare it with the bound from Problem 15.

Solution. To prove linear independency in Problem 31, we substituted 0 and 1 instead of y_i into some inequalities. In the both cases, the value of the polynomial will not change during this operation. Hence, the proof remains valid, and we get a new estimate on the number of such polynomials: it is bounded by the number of generating monomials, which is equal to $C_n^4 + C_n^3 + C_n^2 + C_n^1 + C_n^0$. Asymptotically this coincides with the bound from Problem 15.

Problem 34. Derive from the result of the previous problem a lower bound for $\chi(\mathbb{R}^n)$, which considerably refines the bound from Problem 30. Ensure yourselves, however, that, in view of Problem 19, substantial further advances based on the graph $G(n,9,4)$ cannot be done.

Solution. In the same way as we did in Problem 30, we get $\chi(\mathbb{R}^n) \geqslant C_n^9/C_n^4 = \frac{4!}{9!}n^5 + O(n^4)$. On the other hand, if we apply Problem 19, we get a bound not better than C_n^5, which asymptotically gives the same order of magnitude.

Problem 35. Let r and s be such that $r - s = p$, where p is a prime number and $r - 2p < 0$. Prove that $\alpha(G(n,r,s)) \leqslant \sum\limits_{k=0}^{p-1} C_n^k$. Compare this bound with the bound from Problem 15.

Solution. Fix any independent subset $W = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ and consider polynomials $P_i \in \mathbb{Z}_p[y_1, \dots, y_n]$ given by formulae

$$P_i(\mathbf{y}) = (\mathbf{x}_i, \mathbf{y})((\mathbf{x}_i, \mathbf{y}) - 1) \ldots ((\mathbf{x}_i, \mathbf{y}) - (p-1)).$$

As we did above, we prove that they are linearly independent and that their number is bounded from above by the required value. On the other hand, in Problem 15 the same value was bounded from below, which gives asymptotically the same order of magnitude: $\alpha(G(n,r,s)) \sim \frac{1}{(p-1)!}n^{p-1}$.

Problem 36*. Explore lower bounds for the quantity $\chi(\mathbb{R}^n)$, which follow from the results of the previous problem. How do these bounds correlate with the bounds from Problem 19?

Solution. We use the same notation as in the previous problem. We have

$$\chi(\mathbb{R}^n) \geqslant \chi(G(n,r,s)) \geqslant \frac{|V(n,r)|}{\alpha(G(n,r,s))} \geqslant \frac{C_n^r}{\sum\limits_{k=0}^{p-1} C_n^k} = (1+o(1))(\frac{C_n^r}{C_n^{r-s-1}}),$$

i.e., there exists a constant c_1 such that $\chi(G(n,r,s)) \geqslant c_1 n^{s+1}$, but, on the other hand, we know by Problem 19 that $\chi(G(n,r,s)) \leqslant c_2 n^{s+1}$ for some other constant c_2. This means that we know the order of growth of the chromatic number for the graphs $G(n,r,s)$ with $r-s=p$, p prime, $r-2p<0$.

3 Problems after the intermediate finish

Before the intermediate finish we have seen how important are the independence numbers of graphs for obtaining lower bounds on the chromatic number of a space. Moreover, we considered different sequences of distance graphs — the sequences $\{G(n,r,s)\}_{n=1}^{\infty}$ with given r and s. It is interesting to understand how will be changed the independence numbers, if, instead of the graphs $G(n,r,s)$, we take their spanning subgraphs, i.e., if we take the same sets of vertices and partially remove the edges. It seems to be evident that the independence numbers must become substantially larger, provided we remove many edges. However, surprisingly, sometimes this is true, but sometimes this is completely wrong! In order to obtain the corresponding results, let us learn some random graphs and some probability theory.

3.1 The random Erdős–Rényi graph and some of its probabilistic characteristics

Let $V_n = \{1,\ldots,n\}$ be a set of vertices. In principle, one can construct on this set C_n^2 edges, if we do not allow multiple edges, loops and orientation. Let us each of these C_n^2 potential edges draw with *probability* $p \in [0,1]$, which is the same for every edge. Different edges appear independently. Let $G = (V_n, E)$ be a graph which can appear as a result of the just-described probabilistic procedure. Denote the probability of its appearance by $\mathbb{P}(G)$. Clearly it is equal to $p^{|E|}(1-p)^{C_n^2-|E|}$. If A is a property of a graph, then its probability — $\mathbb{P}(A)$ — is the sum over all the graphs G, which have property A, of the probabilities of these graphs.

Denote by Ω_n the set of all graphs on the vertices V_n. Any function X defined on Ω_n and taking real values is called *random variable*. For example, the number of triangles in a graph or its number of connected components, or its independence number are random variables. Here you must understand that the variables are random only due to the fact that apriori we do not know which graph will (randomly) come to us. When a graph is already born the value of X is determined uniquely!

Any random variable has some "mean value" — the so-called *(mathematical) expectation*. The expectation of a variable X is the number $\mathbb{M}X$ defined by the formula $\mathbb{M}X = \sum\limits_{G\in\Omega_n} X(G)\mathbb{P}(G)$. We just sum up the values of the function X on all the graphs multiplied by the probabilities of these graphs. Of course this is some weighted average (weights are probabilities) — a kind of the mass centre). Let us learn to calculate expectations and apply the obtained results to the study of the properties of random graphs.

Problem 37. Prove that if a random variable is just a constant c, then $\mathbb{M}c = c$.

Solution. What is the average value of the constant? Obviously, it is constant itself.

Problem 38. Let X_1, X_2 be random variables and c_1, c_2 be some fixed numbers. Certainly $c_1X_1 + c_2X_2$ is also a random variable. Prove that its expectation equals $c_1\mathbb{M}X_1 + c_2\mathbb{M}X_2$. This property is called *linearity of expectation*.

Solution.

$$\mathbb{M}(c_1X_1 + c_2X_2) = \sum_G (c_1X_1(G)P(G) + c_2X_2(G)P(G)) = c_1 \sum_G X_1(G)P(G) + c_2 \sum_G X_2(G)P(G) =$$

$$= c_1\mathbb{M}X_1 + c_2\mathbb{M}X_2.$$

Problem 39. Using linearity of expectation find expectations of a) the number of triangles in the random graph; b) the number of connected components of the random graph, each of which is a cycle on k vertices (k is a fixed given number); c) the number of independent sets in the random graph, each of which has cardinality k (k is a fixed given number).

Solution. a) There are C_n^3 choices of three vertices, and each of them will give a triangle with probability p^3. From linearity of expectation, it follows that $\mathbb{M}X = C_n^3 p^3$, where X is the required random variable.

b) There are C_n^k choices of k vertices in the given graph, and each of them can form $(k-1)!/2$ cycles (here we count the order of the vertices in the cycle). The probability of such an ordered set to become a connected component being a cycle equals $p^k(1-p)^{(C_k^2-k)+k(n-k)}$, since there should be exactly k edges between the vertices of this cycle and since there should be no edges between this subgraph and the other vertices of the graph. It follows from linearity that

$$\mathbb{M}X = C_n^k (k-1)! p^k (1-p)^{(C_k^2-k)+k(n-k)},$$

where X is the required random variable.

c) There are C_n^k choices of k vertices, and each of them is independent with probability $(1-p)^{C_k^2}$. Hence, $\mathbb{M}X = C_n^k(1-p)^{C_k^2}$, where X is the required random variable.

Problem 40. Prove *Markov's inequality*: if X is a random variable taking non-negative values and a is a positive number, then $\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{M}X}{a}$.

Solution.

$$\mathbb{M}X = \sum_z zP(X=z) = \sum_{z \geqslant a} zP(X=z) + \sum_{z<a} zP(X=z) \geqslant \sum_{z \geqslant a} aP(X=z) = aP(x \geqslant a).$$

This is equivalent to Markov's inequality.

Problem 41. Using Markov's inequality prove that if $p = \frac{1}{2}$, then $\mathbb{P}(\alpha(G) \leqslant 2\log_2 n) \to 1$ as $n \to \infty$ (it is used to say that "almost surely $\alpha(G) \leqslant 2\log_2 n$").

Solution. Introduce the random variable X_k which is equal to the number of independent sets of vertices of size k. Then from problems 39 and 40 can be seen that

$$P(X_k \geqslant 1) \leqslant \mathbb{M}X_k = C_n^k \left(\frac{1}{2}\right)^{\frac{k(k-1)}{2}} \leqslant \frac{n^k}{k!}\left(\frac{1}{2}\right)^{\frac{k(k-1)}{2}} \leqslant \frac{n^k e^k}{k^k}\left(\frac{1}{2}\right)^{\frac{k(k-1)}{2}} = \left(\frac{ne2^{\frac{k-1}{2}}}{k}\right)^k$$

.

It is easy to see that for $k = 2\log_2 n$ this value tends to 0. So, with probability tending to 1, an independent set of vertices of size k in the graph does not exist.

Note that actually for $p = \frac{1}{2}$ almost surely $\alpha(G) \geqslant (2-\varepsilon)\log_2 n$ for any given arbirarily small $\varepsilon > 0$. More precisely, — and we shall use it later, — the following theorem holds.

Theorem 1. *For any $\varepsilon > 0$ and large enough n the inequality*

$$\mathbb{P}(\alpha(G) \geqslant (2-\varepsilon)\log_2 n) > 1 - 2^{-n}$$

holds.

Thus, for $p = \frac{1}{2}$, almost surely the independence number is approximately $2\log_2 n$. In other words, what's happening? We take the complete graph on n vertices and we delete some of its edges, each with probability $\frac{1}{2}$. A typical graph, which appears as a result of this procedure, has about $\frac{C_n^2}{2}$ edges — two times less than the complete graph has. And the independence number of a typical graph is in a logarithm times greater than the independence number of the initial graph (the independence number of the complete graph is of course 1). OK, the number of edges is two times smaller and the independence number increases: quite natural. It turns out that for some $G(n,r,s)$ when deleting their edges randomly, we get the same "expected" result (the independence number becomes approximately log times bigger). But the miracle is that this happens not everytime! For many $G(n,r,s)$ the independence number does not change at all! Below, we will study together examples of both situations.

3.2 Random subgraphs of the graph $G(n,3,0)$

Let $G_{1/2}(n,3,0)$ be a random subgraph of the graph $G(n,3,0)$ obtained by deleting mutually independently some edges of the graph $G(n,3,0)$, each with probability 1/2.

Problem 42. If you solved Problem 16, then just derive from it that $\alpha(G(n,3,0)) = C_{n-1}^2$. Otherwise, try to solve this particular case of that problem.

Solution. See the solution of Problem 16.

In a series of problems below, we will prove that almost surely $\alpha(G_{1/2}(n,3,0)) \leqslant C_{n-1}^2\left(1 + \frac{1}{\ln n}\right)$. This is exactly what we mentioned above speaking about a surprising phenomenon: there is no growth in log times; if there is some growth, then it is only in such a number of times that itself tends to 1 as $n \to \infty$! One can prove even stronger facts, but this is very difficult, and we do not want to do it here: we just want to see the essence!

In what follows, we will for conciseness omit integer parts around quantities which must be integer. For example, writing $C_{\log_2 n}^k$ means, depending on context, that actually we take an upper or a lower integer part of the number $\log_2 n$. None of the computations will become false after this roughening.

Problem 43. Put $k = C_{n-1}^2\left(1 + \frac{1}{\ln n}\right)$. Assume that k is integer (cf. a remark before the problem). Let $A \subset V(n,3)$ be an arbitrary set of vertices of the graph $G(n,3,0)$ having cardinality k. Denote by $r(A)$ the number of edges of the graph $G(n,3,0)$, whose both ends are in A. Since $|A| = k > \alpha(G(n,3,0))$, it is clear that $r(A) > 0$. Let X_k be the random variable equal to the number of independent sets of size k in a graph $G_{1/2}(n,3,0)$. Prove that

$$\mathbb{M}X_k = \sum_{A \subset V(n,3):\ |A|=k} \left(\frac{1}{2}\right)^{r(A)}. \tag{2}$$

Solution. Let $t = C_{|V|}^k$. Let each subgraph on k vertices correspond to the random variable equal to 1 if the corresponding subgraph on k vertices became independent, and 0 otherwise. Let these random variables be $Y_1, Y_2, \ldots, Y_t$. Then $X_k = Y_1 + Y_2 + \cdots + Y_t$. But the expectation of Y_i is obviously equal to $\frac{1}{2}^{|Y_i|}$, where $|Y_i|$ is the number of edges in corresponding to the i-th random variable subgraph. It immediately follows from the linearity of the expectation that

$$\mathbb{M}X_k = \sum_{A \subset V(n,3):\ |A|=k} \left(\frac{1}{2}\right)^{r(A)}.$$

Problem 44. Prove that our aim will be attained, if we will prove that $\mathbb{M}X_k \to 0$ as $n \to \infty$.

Solution. We want to prove that with high probability there are no independent sets of size k, i.e, $P(X_k = 0) \to 1$ as $n \to \infty$. But it is clear that $P(X_k = 0) + P(X_k \geqslant 1) = 1$. Then we need to prove that $P(X_k \geqslant 1) \to 0$. But Markov's inequality implies that $P(X_k \geqslant 1) \leqslant MX_k$. It remains to show that $\mathbb{M}X_k \to 0$ as $n \to \infty$.

Clearly we have to learn bounding *from below* the quantity $r(A)$. For each A, denote by $B = B(A)$ any (chosen for everytime) subset of A, which is independent in $G(n,3,0)$ and has the maximum cardinality among all analogous subsets of the set A.

Problem 45. Let $A \subset V(n,3)$, $|A| = k$. Let $B = B(A)$. Note that $k \approx \frac{n^2}{2}$. Assume that $|B|$ is substantially smaller than k: for example, let $|B| < n^{1.9}$ (here the strange number 1.9 is taken almost arbitrarily; important is that it is strictly smaller than 2). Prove that for large enough n the inequality $r(A) \geqslant \frac{k^2}{3|B|}$ holds (being more careful one can replace 3 in the denominator by "almost" 2, but this does not matter).

Solution. First, let us prove the following lemma:

Lemma. Any subgraph $\widetilde{\mathcal{G}}$ of graph $\mathcal{G}$, with mn vertices, and with $\alpha(G) \leqslant n$, contains not less than $\frac{m^2n}{2} - \frac{mn}{2}$ edges ($m \in N$).

Proof.
Let us prove this by induction on m.
Base. $m = 1$.
$\frac{m^2n}{2} - \frac{mn}{2} = 0$. It is obvious.
The induction step. $m \to m + 1$.

Consider a maximal independent set in the subgraph $\widetilde{\mathcal{G}}$. If there is more than one such set, consider any of them. It is obvious that the independence number of the subgraph does not exceed the independence number of the initial graph. Therefore, $\alpha(\widetilde{\mathcal{G}}) \le n$. So, there is at least 1 edge from each of the remaining $(m+1)nk$ vertices to this set, otherwise we could increase the independent set, but by the assumption it is maximum. So, we have not less than $(m+1)n-k$ edges. And no one from this edges is lying in the subgraph on remaining $(m+1)n-k$ vertices. But $(m+1)n-k \ge mn$, and so in the subgraph on remaining $(m+1)n-k$ vertices there are not less than $\frac{m^2n}{2} - \frac{mn}{2}$ edges (by the induction hypothesis). Then, there are not less then

$$\frac{m^2n}{2} - \frac{mn}{2} + (m+1)n - k \ge \frac{m^2n}{2} - \frac{mn}{2} + mn = \frac{m^2n + 2mn + n}{2} - \frac{n}{2} - \frac{mn}{2} = \frac{(m+1)^2n}{2} - \frac{(m+1)n}{2}$$

edges. The induction step is proved.

So it follows from lemma that

$$r(A) \geqslant \frac{\left[\frac{|A|}{|B|}\right]^2 |B|}{2} - \frac{\left[\frac{|A|}{|B|}\right] |B|}{2} = \frac{\left[\frac{|A|}{|B|}\right] |B|}{2}\left(\left[\frac{|A|}{|B|}\right] - 1\right) \geqslant \frac{k-|B|}{2}\left(\frac{k}{|B|} - 2\right) = (1+o(1))\frac{k^2}{2|B|}.$$

Problem 46. Let $A \subset V(n,3)$, $|A| = k$. Let $B = B(A)$. Let $|B| > 9n$ or, even better (in addition to the previous problem), $|B| \geqslant n^{1.9}$. Prove that $r(A) \geqslant (|B| - 9n)(|A| - |B|)$.

Solution. Let us estimate the number of edges of the graph $G(n,3,0)$ in A. Recall that B is the *maximum* independent set of vertices of graph $G(n,3,0)$ in A. Because of this, for any vertex $\mathbf{x} \in A \setminus B$ there is a vertex $\mathbf{y} \in B$ such that $\{\mathbf{x},\mathbf{y}\} \in E(n,3,0)$. Let us show that the vertex $\mathbf{y}$ is not unique with this property. Indeed, the vertices $\mathbf{x}$ and $\mathbf{y}$ are connected by an edge, and so they do not intersect as 3-element sets. Let us estimate how many vertices $\mathbf{z} \in B$ which **aren't connected** with $\mathbf{x}$ can exist. On the one hand, they must intersect with $\mathbf{x}$ by at least one element. On the other hand, they must intersect with $\mathbf{y}$ too, cause B is an independent set. But $\mathbf{x}$ and $\mathbf{y}$ haven't any intersection. So, there are not more than $3^2 n^{3-2}$ such vertices $\mathbf{z}$ in the current situation $|B| > n^{3-1.1.}$. Consequently, the number of vertices of B, which is connected to the given vertex $\mathbf{x} \in A \setminus B$, is not less than $|B| - 9n$. Thus,

$$|\{\{\mathbf{x},\mathbf{y}\} \in E(n,3,0):\ \mathbf{x},\mathbf{y} \in A\}| \geqslant (|A| - |B|)(|B| - 9n) \geqslant \frac{C_{n-1}^2}{\ln n} n^{1.9} = (1+o(1))\frac{n^{3.9}}{2\ln n},$$

(last transitions are possible, provided that $|B| \geqslant n^{1.9}$. But for inequality $r(A) \geqslant (|B| - 9n)(|A| - |B|)$ it suffices that $|B| > 9n$).

Problem 47. Decompose sum (2) into two parts: in the first part, only those A will be taken, for which $|B| < n^{1.9}$; in the second part, all the other sets A will be considered. To the summands in both parts apply the bounds from the corresponding problems and make it sure that the whole sum (2) does really tend to zero, which means that we are done!

How can we change the threshold $n^{1.9}$ in order to still get the same result?

Solution. It is sufficient to verify that in the random graph $\mathcal{G}(G(n,3,0),\ 1/2)$ there is an independent set of vertices of size k with probability tending to zero. It is known that this probability is not greater than

$$\sum_{A \subset V(n,r), |A|=k} \mathbb{P}(A \text{ is independent in } \mathcal{G}(G(n,3,0), 1/2)) = \sum_{A \subset V(n,r), |A|=k} 2^{-|\{\{\mathbf{x},\mathbf{y}\} \in E(n,3,0):\ \mathbf{x},\mathbf{y} \in A\}|}.$$

Let us show that the last sum tends to zero too.

There are two options: either for the A holds $|B| \leqslant n^{1.9} = o\left(C_{n-1}^{2}\right)$ (A is *set of the first type*), or for the A holds $|B| > n^{1.9}$ (A is *set of the second type*).

Let us select terms corresponding to the sets A of the first type in the sum above. This sum does not exceed

$$C_{C_n^3}^{k} 2^{-(1+o(1))\frac{k^2}{2n^{1.9}}} < 3^k n^{3k} 2^{-(1+o(1))\frac{k^2}{2n^{1.9}}} = 3^k 2^{(1+o(1))3\frac{n^2}{2}\log_2 n} 2^{-(1+o(1))\frac{n^4}{2(2)^2 n^{1.9}}} \to 0.$$

Here we use the fact that r is constant and

$$n^2 \log_2 n = o\left(n^{2.1}\right).$$

Now let A be a set of the second type. Using Problem 45,

$$|\{\{\mathbf{x}, \mathbf{y}\} \in E(n,3,0) : \ \mathbf{x}, \mathbf{y} \in A\}| \geqslant \frac{n^{3.9}}{2\ln n},$$

So, the sum of terms corresponding to the sets A of the second type (in the sum above) does not exceed

$$C_{C_n^3}^{k} 2^{-(1+o(1))\frac{n^{3.9}}{2\ln n}} < 3^k n^{3k} 2^{-(1+o(1))\frac{n^{3.9}}{2\ln n}} = 3^k 2^{(1+o(1))3\frac{n^2}{2}\log_2 n} 2^{-(1+o(1))\frac{n^{3.9}}{2\ln n}} \to 0.$$

The result follows.

3.3 Random subgraphs of the graph $G(n,3,1)$

Let $G_{1/2}(n,3,1)$ be a random subgraph of the graph $G(n,3,1)$ obtained by deleting mutually independently some edges of the graph $G(n,3,1)$, each with probability $1/2$. It seems that everything is the same as with the graphs $G(n,3,0)$. This is completely wrong!

Recall that $\alpha(G(n,3,1)) \approx n$ (see Problem 12).

Problem 48*. Write down an analog of inequality (2) and prove an analog of the bound from Problem 45. Ensure yourselves eventually that there exists a $c > 0$, with which almost asurely $\alpha(G_{1/2}(n,3,1)) \leqslant cn\log_2 n$.

Solution. We explained in detail how to prove the bound in Problem 45, so we will not re-do it. We assume that the quantity of edges in the set A of cardinality k is not less than $(1+o(1))\frac{k^2}{2n}$. If somebody does not believe this, he should repeat inference from Problem 45 by himself.

Let $X_k = X_k(\mathcal{G}(G(n,3,1),1/2))$ be a function of random graph which is equal to quantity of independent sets with k vertices in the graph (i. e., of sets such that their elements are not mutually connected by edges). Let us estimate its expectation and use Markov's inequality:

$$\mathbb{E}X_k = \sum_{A \subset V(n,3),\, |A|=k} \mathbb{P}(A \text{ is an independent set in } \mathcal{G}(G(n,3,1),1/2)) =$$

$$= \sum_{A \subset V(n,3),\, |A|=k} 2^{-|\{\{\mathbf{x},\mathbf{y}\} \in E(n,3,1):\, \mathbf{x},\mathbf{y} \in A\}|},$$

so in exponential factor there is a quantity of edges of subgraph of graph $G(n,3,1)$, which is induced by the concrete set of vertices A. As we remember

$$|\{\{\mathbf{x}, \mathbf{y}\} \in E(n,3,1) : \ \mathbf{x}, \mathbf{y} \in A\}| > (1+o(1))\frac{k^2}{2\alpha(G(n,3,1))} = (1+o(1))\frac{k^2}{2n}.$$

So, cause of this bound, we obtain that

$$\mathbb{E}X_k < \sum_{A\subset V(n,3),\,|A|=k} 2^{-(1+o(1))\frac{k^2}{2n}} = C^k_{C^3_n} 2^{-(1+o(1))\frac{k^2}{2n}}.$$

It is well known that $C_a^b \leqslant \left(\frac{ea}{b}\right)^b$, where e is a base of the natural logarithm. So,

$$\mathbb{E}X_k < \left(\frac{n^3}{k}\right)^k 2^{-(1+o(1))\frac{k^2}{2n}} = 2^{3k\log_2 n - k\log_2 k - (1+o(1))\frac{k^2}{2n}}.$$

I.e., there exists a function $k = k(n)$, which is asymptotically equal to $4n\log_2 n$ and such that for this subsequence, $\mathbb{E}X_k \to 0$ where $n \to \infty$. The theorem follows from this statement and Markov's inequality:

$$\mathbb{P}(\alpha(\mathcal{G}(G(n,3,1),1/2)) \leqslant 4(1+o(1))n\log_2 n) = \mathbb{P}(X_k = 0) \geqslant 1 - \mathbb{E}X_k \to 1, \;\; n \to \infty.$$

QED.

Problem 49. Prove that in the graph $G(n,3,1)$, there exist approximately $\frac{n}{2}$ complete subgraphs, each having about $\frac{n}{4}$ vertices, such that every two of them are not connected by an edge.

Solution. For convenience let us assume that n is divisible by 4. Let $m = \frac{n}{2}$. We divide $\mathcal{R}_n$ into parts $R_1 = \mathcal{R}_m$ and $R_2 = \mathcal{R}_n \setminus R_1$. First, we describe the construction of one clique (independent subset) Q_1. For this, we take disjoint pairs $\{1,2\}, \{3,4\}, \{5,6\}, \ldots, \{m-1, m\}$ (cause of m is even) in R_1. We add element $m+1 \in R_2$ to each of these pairs. This is the desired clique. It contains $\frac{m}{2} = \frac{n}{4}$ vertices. We similarly construct $n-m-1$ cliques $Q_2, \ldots, Q_{n-m}$ by adding to all the pairs in R_1 element $m+2 \in R_2$, element $m+3 \in R_2$, etc. Obviously, there are no edges between $\mathbf{x}, \mathbf{y}$ for any i, j, $i \neq j$, $\mathbf{x}$ from Q_i, $\mathbf{y}$ from Q_j: these triples either do not intersect at all or intersect by some pair of R_1.

Hence, we have constructed $\frac{n}{2}$ cliques, each containing $\frac{n}{4}$ elements.

Problem 50. With the help of the previous problem and Theorem 1 prove that for every $\varepsilon > 0$, almost surely $\alpha(G_{1/2}(n,3,1)) \geqslant (1-\varepsilon)n\log_2 n$.

Solution. As we know, a random graph $\mathcal{G}(G(n,3,1),1/2)$ is obtained from the graph $G(n,3,1)$ as a result of mutually independent choice of edges from $E(n,3,1)$ with the same probability $\frac{1}{2}$. Therefore, there are independent copies of a random graph of Erdős–Rényi $G(m/2,1/2)$ on cliques $Q_1, \ldots, Q_{n-m}$. Let us note that these copies are independent from the point of view of probability theory (as random elements), and from the point of view of the theory of graphs (there are no edges between them).

When $p = \frac{1}{2}$, Theorem 1 says that with the asymptotic probability 1, we have $\alpha(G(m/2,1/2)) \sim 2\log_2 m$ when $m \to \infty$, but m is only by a constant factor less than n, so $\alpha(G(m/2,1/2)) \sim 2\log_2 n$ when $n \to \infty$. Moreover, the rate of tending of probability to one is very high. For an appropriate choice of the infinitesimal parameter and large n, we get the estimate

$$\mathbb{P}(\alpha(G(m/2,1/2)) \geqslant 2(1+o(1))\log_2 n) \geqslant 1 - 2^{-n}.$$

This means that

$$\mathbb{P}(\forall\, i = 1, \ldots, n-m \quad \alpha(\mathcal{G}(Q_i,1/2)) \geqslant 2(1+o(1))\log_2 n) \geqslant \left(1 - 2^{-n}\right)^{n-m} \to 1, \;\; n \to \infty.$$

Therefore, with the asymptotic probability 1 there are $n-m$ independent sets of size $2(1+o(1))\log_2 n$ in a random graph $\mathcal{G}(G(n,3,1),1/2)$, and there are no edges between them. Together they make up one independent set of size $2(n-m)(1+o(1))\log_2 n \sim n\log_2 n$, QED.

Thus, for $G_{1/2}(n,3,1)$, just as it was for the Erdős–Rényi random graph, we again have the growth of the independence number in approximately logarithm of the number of vertices times!

Problem 51*. Try to improve into a constant number of times the bound in the problem 50.

Solution. To improve the bound, it suffices to repeat the proofs of Problems 49 and 50, replacing $m=\frac{n}{2}$ by $m=2\left[\frac{n}{2\log_2 n}\right]$, here $[x]$ is the usual integer part of x. Then we obtain an estimate of exactly two times better than in Problem 50.

3.4 Random subgraphs of the graph $G(n,2,1)$

Problem 52. Find $\alpha(G(n,2,1))$.

Solution. Obviously, this is just a cover of our set by disjoint pairs. Hence, $\alpha(G(n,2,1))=\frac{n}{2}$ for even n, and $\alpha(G(n,2,1))=\frac{n-1}{2}$ for odd n.

Problem 53. Let $r(A)$ be the same as in formula (2). Prove that always $r(A)\geqslant\frac{2|A|^2}{n}-|A|$.

Solution. The set A is a subset of $\mathcal{R}_n$. For every $i\in\mathcal{R}_n$, let k_i be the number of vertices of the graph $G(n,2,1)$ (“ twos”) which belong to A and contain i. It is clear that for fixed i, any two “ twos” corresponding to this i form an edge in $G(n,2,1)$. Therefore,

$$|\{\{\mathbf{x},\mathbf{y}\}\in E(n,2,1):\ \mathbf{x},\mathbf{y}\in A\}|\geqslant\sum_{i=1}^{n}C_{k_i}^2.$$

At the same time $k_1+\ldots+k_n=2|A|$. It is easy to show that, under these constraints, the minimum of $\sum\limits_{i=1}^{n}C_{k_i}^2$ is achieved at $k_i=\frac{2|A|}{n}$, $i=1,\ldots,n$, which implies the bound

$$|\{\{\mathbf{x},\mathbf{y}\}\in E(n,2,1):\ \mathbf{x},\mathbf{y}\in A\}|\geqslant\sum_{i=1}^{n}C_{k_i}^2\geqslant\frac{2|A|^2}{n}-|A|.$$

Problem 54. Derive from the previous problem the inequality $\alpha(G_{1/2}(n,2,1))\leqslant\left(\frac{1}{2}+\varepsilon\right)n\log_2 n$, which is true for every $\varepsilon>0$ almost surely.

Solution. We know the estimate from Problem 53, so we have

$$\sum_{A\subset V(n,2),|A|=k}2^{-|\{\{\mathbf{x},\mathbf{y}\}\in E(n,2,1):\ \mathbf{x},\mathbf{y}\in A\}|}<C_{C_n^2}^k 2^{-(1+o(1))\frac{2k^2}{n}}<\left(\frac{eC_n^2}{k}\right)^k 2^{-(1+o(1))\frac{2k^2}{n}}<2^k n^{2k}k^{-k}2^{-(1+o(1))\frac{2k^2}{n}}=$$

$$=2^{2k\log_2 n-(1+o(1))k\log_2 k-(1+o(1))\frac{2k^2}{n}}.$$

Therefore, for a suitable $k \sim \frac{1}{2}n\log_2 n$ we obtain

$$2\log_2 n - (1+o(1))\log_2 k - (1+o(1))\frac{2k}{n} < 0,$$

then

$$2k\log_2 n - (1+o(1))k\log_2 k - (1+o(1))\frac{2k^2}{n} = k\left(2\log_2 n - (1+o(1))\log_2 k - (1+o(1))\frac{2k}{n}\right) \to -\infty.$$

That is, we have shown that $P(X_k \geqslant 1) \leqslant MX_k \to 0$ and the estimate is proved.

Problem 55*. Prove a lower bound for $\alpha(G_{1/2}(n,2,1))$ which has order of magnitude $cn\log_2 n$ with some $c > 0$.

Solution. The proof of this fact is quite long. If you want you can find it in [5].

Problem 56**. Find a constant c in the assertion: for any $\varepsilon > 0$, almost surely

$$(c-\varepsilon)n\log_2 n \leqslant \alpha(G_{1/2}(n,2,1)) \leqslant (c+\varepsilon)n\log_2 n.$$

Solution. Unfortunately, by the beginning of the conference, this problem has not been solved. And the students did not change this fact.

References

[1] A.M. Raigorodskii, *The chromatic numbers*, Moscow Centre for Continuous Mathematical Education (MCCME), Moscow, Russia, 2003 (in Russian).

[2] A.M. Raigorodskii, *The linear algebra method in combinatorics*, Moscow Centre for Continuous Mathematical Education (MCCME), Moscow, Russia, 2007 (in Russian).

[3] A.M. Raigorodskii, *Probability and algebra in combinatorics*, Moscow Centre for Continuous Mathematical Education (MCCME), Moscow, Russia, 2008 (in Russian).

[4] A.M. Raigorodskii, *Models of random graphs*, Moscow Centre for Continuous Mathematical Education (MCCME), Moscow, Russia, 2011 (in Russian).

[5] L. I. Bogolyubskiy, A. S. Gusev, M. M. Pyaderkin, and A. M. Raigorodskii, The independence numbers and the chromatic numbers of random subgraphs of some distance graphs, Sbornik:Math, 2014, to appear (current Russian version will be on the disc, English journal version will be available at http://www.mathnet.ru/php/person.phtml?&personid=8601&option_lang=eng)

[6] J. Balogh, A.V. Kostochka, A.M. Raigorodskii, *Coloring some finite sets in* $\mathbb{R}^n$, Discussiones Mathematicae Graph Theory, 33 (2013), N1, 25 - 31.