The Threshold for the Balancedness of Colored Graphs

Konstantinos Panagiotou
panagiok@inf.ethz.ch

Angelika Steger
steger@inf.ethz.ch

Institute of Theoretical Computer Science, ETH Zürich
Universitätstrasse 6, CH - 8092 Zürich

Abstract
Let \( \text{Col}_n^\ell \) denote the set of all \( \ell \)-colored graphs on \( n \) vertices and let \( G \) be a graph drawn uniformly at random from this set. Provided that \( \ell \to \infty \), we prove that all color classes of \( G \) deviate roughly at most \( \sqrt{2 \log \ell} \) from \( \lfloor \frac{n}{\ell} \rfloor \), and there is at least one color class with such deviation. Furthermore, we prove similar results for the class \( \text{Col}_n^{\ell,M} \), which contains all \( \ell \)-colored graphs with precisely \( M = cn^2 \) edges, for \( 0 < c < \frac{1}{2} \).

1 Introduction & Results

We denote by \( G_{n,p} \) the binomial model of random graphs, where each edge between two out of \( n \) labeled vertices is included in the graph independently of the others with probability \( p \). This model was introduced in the beginning of the sixties by Erdős and Rényi [ER60]. One of the most studied parameters within the theory of random graphs is the chromatic number \( \chi(G_{n,p}) \). Plenty of deep and insightful work was devoted to this topic; the determination of this number had been for a long time an open problem, until Bollobás [Bol88] managed to determine the exact asymptotic behavior for every constant density.

In the following years, the result of Bollobás was refined by McDiarmid [McD89] and generalized by Łuczak [Luc91a] for (almost) every density. Parallel to this line of research, Shamir and Spencer [SS87] showed that the chromatic number is sharply concentrated around its expected value. These results were strengthened by Łuczak [Luc91b] and by Alon and Krivelevich [AK97], who proved the remarkable fact that as long as \( p \leq n^{-\frac{1}{2} - \varepsilon} \), the chromatic number is 2-point concentrated. More recently, Achlioptas and Naor [AN04] managed to determine these two values for very sparse random graphs (\( p = \frac{c}{n} \)).

Despite of the immense progress which was made in the determination of the chromatic number, many questions still remain open. The values of the chromatic number for \( p \gg \frac{1}{n} \) are still unknown; in case of constant density, the size of the gap between the best known bounds is of order \( \frac{n}{\log n} \), which can easily be verified by examining the statement of the main Theorem in e.g. [McD89]. This is somehow disappointing, as for constant density it is known [SS87] that the chromatic number is \( \sqrt{n} \)-concentrated. In a recent paper, Bollobás asked whether the size of this interval can be further improved, see [Bol04].

Very few is also known about the number of optimal colorings a random graph can have; for \( p = n^{-\alpha} \), where \( \frac{1}{2} < \alpha < 1 \), Krivelevich [Kri02] showed that most of the graphs have at least \( e^{n^{f(\alpha)}} \) colorings, where \( 0 < f(\alpha) < \frac{1}{2} \). Unfortunately, these lower bound seems not to be tight; furthermore, the dense case remains completely open.
In this paper, we study a closely related model. Instead of looking at the random graph $G_{n,p}$, we investigate properties of the class of $\ell$-colorable graphs $Col_n^\ell$. We believe that knowledge of structural properties of those graphs can be a powerful tool, which may eventually also help us to better understand coloring problems for the random graph $G_{n,p}$. The main result of our paper is that colored graphs have with high probability a strong structural property: the size of all color classes deviates from $n/\ell$ at most roughly $\sqrt{2 \log \ell}$. This means that almost all colored graphs are very balanced, in the sense that the overall deviation from a perfectly balanced coloring is small. Furthermore, we prove that this property exhibits a threshold, i.e., there is at least one color class with such deviation. Similar results are shown if we set restrictions on the number of edges.

1.1 Our Results

In this article, our main focus is to investigate properties of the class of colored graphs. Every $\ell$-colorable graph $G$ (i.e. it holds $\chi(G) \leq \ell$), which has $t$ different colorings with $\ell$ colors, counts exactly $t$ times as a colored graph (the permutations of the colors are counted only once). Therefore, a colored graph can be viewed as a tuple $(G, C)$, where $C$ is a valid partition for $G$, in the sense that all classes form independent sets in $G$. At this point, we note that in the remainder of the paper we use the terms ”partition” and ”coloring” interchangeably, as they are equivalent in our context.

In order to state our main results, some notation is required. In the following we denote by $Col_n^\ell$ the set of all $\ell$-colored graphs on $n$ labeled vertices. Intuitively, we can partition the graphs in $Col_n^\ell$ according to the deviation of each color class from a completely balanced coloring, where each color class has $\lfloor n/\ell \rfloor$ or $\lceil n/\ell \rceil$ vertices. For this, let

$$ D(f, \ell, n) = \left\{ (d_1, \ldots, d_\ell) \mid \forall i : d_i \in \mathbb{Z} \text{ and } \forall i : |d_i| \leq f \text{ and } \right. $$

$$ \left. \forall i : 0 \leq \left\lfloor \frac{n}{\ell} \right\rfloor + d_i \leq n \text{ and } \sum_{i=1}^\ell d_i = r \right\}, \quad (1.1) $$

where $r = r(n, \ell)$ is the remainder of $n$ and $\ell$. Additionally, let

$$ b(d) = \left\{ n \left\lceil \left\lfloor \frac{n}{\ell} \right\rfloor + d_1, \ldots, \left\lceil \left\lfloor \frac{n}{\ell} \right\rfloor + d_\ell \right\rfloor \right\} \quad (1.2) $$

and

$$ \log E(d) = \sum_{1 \leq i < j \leq \ell} \left( \left\lfloor \frac{n}{\ell} \right\rfloor + d_i \right) \left( \left\lfloor \frac{n}{\ell} \right\rfloor + d_j \right), \quad (1.3) $$

where the logarithms, if not stated otherwise, are to the base 2. We write

$$ col_n^\ell(f) = \frac{1}{\ell!} \sum_{d \in D(f, \ell, n)} b(d) E(d) \quad (1.4) $$

for the number of the $\ell$-colored graphs, which have the property that their color classes deviate at most $f$ from $\left\lfloor \frac{n}{\ell} \right\rfloor$; note that the number of all $\ell$-colored graphs is given by $|Col_n^\ell| = col_n^\ell(n - \left\lfloor \frac{n}{\ell} \right\rfloor)$, see [Wri61]. The following theorem states that the number of all colored graphs is asymptotically of the same order of magnitude as the number of ”very balanced” colored graphs; more precisely, if we draw a colored graph $G_C$ uniformly at random from the set $Col_n^\ell$, then with high probability every color class of $G_C$ will deviate roughly at most $\sqrt{2 \log \ell}$ from $\left\lfloor \frac{n}{\ell} \right\rfloor$, and there will be at least one such color class.
There exist constants \( c_1, c_2 > 0 \) such that for all integer functions \( \ell(n) \leq \frac{n}{\log n} \) and \( \ell(n) \to \infty \) for \( n \to \infty \), it holds

\[
\operatorname{col}_n^{\ell}(f) = \begin{cases} 
\lfloor \omega_n \log \frac{\ell}{\sqrt{n}} \rfloor, & f < \sqrt{2 \log \ell - c_1 \cdot (\log \ell)^{2/3}} \\
(1 - o(1)) \lfloor \log \frac{\ell}{\sqrt{n}} \rfloor, & f > \sqrt{2 \log \ell + c_2 \cdot (\log \ell)^{2/3}}.
\end{cases}
\] (1.5)

Theorem 1.1 gives a precise characterization of the balancedness of colored graphs and is therefore a strong generalization of the results in [Bol81], [Kuc89] and [PS95], where only the upper bound \( \operatorname{col}_n^{\ell}(f) = (1 - o(1)) \lfloor \log \frac{\ell}{\sqrt{n}} \rfloor \) for \( f = \varepsilon n \), \( f = \sqrt{n/\log n} \) and \( f = \varepsilon \log n \) was shown. Furthermore, note that Theorem 1.1 scales smoothly with the considered chromatic number: the smaller \( \ell \) is, the lesser is the deviation of each colors class from \( n/\ell \) with high probability. Observe also that if \( \ell \) is in the range of the chromatic number of the random graph \( G_n, \frac{1}{n} \) (i.e., \( \ell \sim \frac{n}{2\log n} \), see [Bol88]), the above theorem says almost surely the maximum deviation of each color class from \( n/\ell \sim 2 \log n \) is at most \( \approx \sqrt{2 \log n} \), which is asymptotically much smaller than the size of every color class.

For constant \( \ell \) the first statement of the above theorem is not true. Instead it follows easily from previous work of Wright [Wri61], see Section 5, that

**Theorem 1.2.** If \( \ell \) is constant, then for any function \( \omega_n = \omega(1) \) we have

\[
\operatorname{col}_n^{\ell}(f) = \begin{cases} 
(1 + o(1)) \cdot g(c) \cdot \lfloor \log \frac{\ell}{\sqrt{n}} \rfloor, & f = c \\
(1 - o(1)) \lfloor \log \frac{\ell}{\sqrt{n}} \rfloor, & f > \omega_n,
\end{cases}
\] (1.6)

where the function \( g \) satisfies \( 0 < g(c) < 1 \) for all \( c \geq 1 \) and \( g(0) = 0 \) if \( n \mod \ell > 0 \), otherwise \( 0 < g(0) < 1 \).

Furthermore, we prove a similar result as Theorem 1.1 for colored graphs with a fixed number of edges. For this, let \( \operatorname{Col}_{n,M}^{\ell} \) denote the set of all \( \ell \)-colored graphs with precisely \( M \) edges and

\[
\operatorname{col}_{n,M}^{\ell}(f) = \frac{1}{\ell!} \sum_{d \in D(f,\ell,n)} b(d) \bar{E}(d), \text{ where } \bar{E}(d) = \left( \sum_{1 \leq i < j \leq \ell} \left\lfloor \frac{n}{\ell} \right\rfloor + d_i \right) \left( \left\lfloor \frac{n}{\ell} \right\rfloor + d_j \right) - M.
\] (1.7)

The threshold reads for this case

**Theorem 1.3.** Let \( M = cn^2 \), where \( 0 < c < \frac{1}{2} \) and \( \varepsilon \) be a positive constant. Furthermore, let \( f(c) = \frac{2}{\log(1-2\varepsilon)} \). For all integer functions \( \ell(n) \leq \frac{n \log \frac{1}{\log n}}{\log n} \) with \( \ell(n) \to \infty \) for \( n \to \infty \), it holds

\[
\operatorname{col}_{n,M}^{\ell}(f) = \begin{cases} 
\lfloor \log \frac{\ell}{\sqrt{n}} \rfloor, & f < (1 - \varepsilon) \sqrt{\log \ell \cdot f(c)} \\
(1 - o(1)) \lfloor \log \frac{\ell}{\sqrt{n}} \rfloor, & f > (1 + \varepsilon) \sqrt{\log \ell \cdot f(c)}.
\end{cases}
\] (1.8)

In the above theorem, we can replace the \( \varepsilon \) with an appropriate function which goes to zero when \( \ell \to \infty \), as we did in Theorem 1.1. For the sake of a better presentation, we choose the above (slightly weaker) formulation.

Before we proceed with the proofs in the following sections, we need two auxiliary definitions, as they will be used extensively in the sequel:

\[
e(n, \ell) := \frac{\ell - 1}{2\ell} n^2 + \frac{\ell^2}{2\ell}
\] (1.9)

and \( p(n, \ell) := \left( n + \frac{\ell}{2} \right) \log \ell - \frac{\ell - 1}{2} \log n - \frac{\ell - 1}{2} \log 2\pi. \) (1.10)
The remainder of the paper is organized as follows. In Section 2 we introduce a few tools, which will help us proving the main results. In Sections 3 and 4 we prove Theorem 1.1. In Section 5 we deal with the case that $\ell$ is constant. Finally, in Section 6 we show how the ideas of the preceding sections can be used to prove Theorem 1.3.

2 Preliminaries

Before we prove Theorem 1.1 in Sections 3 and 4, we need a few technical tools, which will be used extensively in the sequel. Without further reference, we will use the following version of Stirling’s formula for the approximation of the factorial function and the estimates for binomial coefficients:

**Formula 2.1.** For $n > 0$

$$1 \leq \frac{n!}{\sqrt{2\pi n} \cdot n^n \cdot e^{-n}} \leq e^\frac{1}{12n}.$$

**Formula 2.2.** Let $a \geq b > 0$. Then it holds that

$$\left(\frac{a}{b}\right)^b \leq \left(\frac{a}{b}\right) \leq \left(\frac{ea}{b}\right)^b.$$

Furthermore, let $H(x) = -x \log x - (1 - x) \log(1 - x)$ denote the binary entropy function and $0 < \xi < 1$. Then it holds that

$$\left(\frac{a}{\xi a}\right) \leq 2^{aH(\xi)}.$$

**Proposition 2.3.** For every $n$, every integer function $\ell(n) \leq n$ and every integer sequence $d \in D\left(n - \left\lfloor \frac{n}{\ell} \right\rfloor, \ell, n\right)$ it holds that

$$\log E(d) = e(n, \ell) - \frac{1}{2} d^T d,$$

where $\log E(d)$ is defined in (1.3) and $e(n, \ell)$ in (1.9).

**Proof.** Note that $\left\lfloor \frac{n}{\ell} \right\rfloor = \frac{n - \ell}{\ell}$. By plugging in the definition of $E$ we obtain

$$\log E(d) = \left(\frac{\ell}{2}\right) \left(\frac{n - \ell}{\ell}\right)^2 + \frac{n - \ell}{\ell} \sum_{1 \leq i < j \leq \ell} (d_i + d_j) + \sum_{1 \leq i < j \leq \ell} d_i d_j.$$

Denote the first sum by $S_1$ and the second one by $S_2$. By double counting we obtain

$$2S_1 + 2 \sum_{i=1}^\ell d_i = \sum_{i,j} (d_i + d_j) = 2\ell r,$$

which shows $S_1 = (\ell - 1)r$. Furthermore, we obtain

$$2S_2 + \sum_{i=1}^\ell d_i^2 = \sum_{i,j} d_i d_j = r^2,$$

i.e., $S_2 = \frac{r^2}{2} - \frac{1}{2} d^T d$. Putting these together yields

$$\log E(d) = \left(\frac{\ell}{2}\right) \left(\frac{n - \ell}{\ell}\right)^2 + \frac{n - \ell}{\ell} (\ell - 1)r + \frac{r^2}{2} - \frac{1}{2} d^T d = \frac{\ell - 1}{2\ell} n^2 + \frac{r^2}{2\ell} - \frac{1}{2} d^T d,$$

which proves together with (1.9) the proposition.

In the following $0_r$ stands for a vector which has the property that all its entries are zero, except of $r$, which are equal to one; note that there are $\left(\frac{\ell}{r}\right)$ such vectors.
Proposition 2.4. For every sufficiently large \( n \), every integer function \( \ell(n) = o(n) \) and every \( d \in D \left( n - \left\lceil \frac{n}{\ell} \right\rceil, \ell, n \right) \) it holds that
\[
2^{p(n,\ell) + o(\ell)} \geq b(0_r) \geq b(d),
\]
where \( p(n,\ell) \) is defined in (1.10).

Proof. We first prove the second inequality; let \( \alpha = \left\lceil \frac{n}{\ell} \right\rceil \). Using the definition of \( b \) we obtain
\[
\frac{b(0_r)}{b(d)} = \frac{\prod_{i=1}^{\ell} (\alpha + d_i)!}{(\alpha + 1)^{\ell} \cdot (\alpha!)^{\ell-r}} \geq \frac{1}{(\alpha + 1)^r} \cdot \frac{\prod_{i:d_i>0} (\alpha + 1)^{d_i}}{\prod_{i:d_i<0} \alpha^{-d_i}} \geq \frac{1}{(\alpha + 1)^r} \cdot (\alpha + 1)^r = 1,
\]
as \( \sum_{i:d_i<0} d_i = \sum_{i:d_i>0} d_i = r \). The inequality \( 2^{p(n,\ell) + o(\ell)} \geq b(0_r) \) can be proven by a straightforward application of Stirling’s formula; we skip the calculations.

From Proposition 2.3 follows that the larger the value of the dot product of the deviation vector \( d \) is, the less corresponding colored graphs exist. In the following proposition we make this more precise; the number of colored graphs with “sufficiently large” \( \frac{1}{2} d^T d \) is small compared to the number of all colored graphs. Let
\[
\mathcal{L}(n,\ell) := \left\{ d \in D \left( n - \left\lceil \frac{n}{\ell} \right\rceil, \ell, n \right) : \frac{1}{2} d^T d \geq 9\ell \right\} \quad \text{and} \quad S_\mathcal{L} := \frac{1}{\ell!} \sum_{d \in \mathcal{L}(n,\ell)} b(d) E(d). \tag{2.1}
\]

Proposition 2.5. Let \( S_\mathcal{L} \) be defined as in (2.1). For all integer functions \( \ell(n) = o(n) \) with \( \ell(n) \to \infty \) for \( n \to \infty \), it holds
\[
S_\mathcal{L} = o \left( \left\lceil \text{Col}_n^\ell \right\rceil \right).
\]

Proof. With Propositions 2.3 and 2.4 and (1.9) we estimate
\[
S_\mathcal{L} \leq \frac{1}{\ell!} b(0_r) 2^{p(n,\ell)} \cdot \sum_{d^T d \geq 18\ell} 2^{-\frac{1}{2} d^T d}. \tag{2.2}
\]
We now give an upper bound for the number of vectors \( d \) with \( d^T d = x \), for every \( x \geq 18\ell \). For this we determine all natural number solutions (including 0’es) of the equation \( \sum_{i=1}^{\ell} q_i = x \). It is known e.g. from [Juk01], that the number of such solutions is at most \( \left( \frac{x + \ell - 1}{\ell} \right) \leq \left( \frac{x}{\ell} \right)^\ell \). From these solutions we pick the ones with the property that all \( q_i \)’s are square numbers. Then we choose for every \( 1 \leq i \leq \ell \) the sign of the corresponding \( d_i \); there are at most \( 2^\ell \) ways of doing this, as not all possible choices of the signs result in valid integer sequences. We obtain
\[
\frac{\ell! \cdot S_\mathcal{L}}{b(0_r) 2^{p(n,\ell)}} \leq 2^\ell \sum_{x \geq 18\ell} \left( \frac{x + \ell}{\ell} \right)^\ell 2^{-\frac{x}{2}} \leq 2^\ell \sum_{x \geq 18\ell} \left( \frac{\ell(x + \ell)}{\ell} \right)^\ell 2^{-\frac{x}{2}} \leq (2e)^\ell \sum_{x \geq 18\ell} 2^{\ell \log \frac{x + \ell}{2} - \frac{x}{2}},
\]
Write now \( x = f_x \cdot \ell \); the exponent of 2 in the sum becomes \( \ell \left( \log(f_x + 1) - f_x/2 \right) \). It can be easily seen that this function is always smaller than \( -\ell f_x/4 \) whenever \( f_x \geq 18 \). Putting all together yields
\[
\frac{\ell! \cdot S_\mathcal{L}}{b(0_r) 2^{p(n,\ell)}} \leq (2e)^\ell \sum_{x \geq 18\ell} 2^{-x/4} \leq (2e)^\ell \cdot 2^{-\frac{18\ell}{4}} \sum_{x \geq 0} 2^{-x/4} \leq 2^{-2\ell},
\]

5
where the last step is true for sufficiently large \( \ell \) due to \( \log(2e) - \frac{18}{4} < -2 \).

The above calculation yields \( S_L \leq \frac{1}{\ell^2} b(0, r) e^{(n, \ell)} \cdot 2^{-2\ell} \); the proof completes with the fact \( |\text{Col}_n| > \frac{1}{r^2} b(0, r) E(0, r) > \frac{1}{\ell^2} b(0, r) e^{2(n, \ell) - 2} \), where the last step is due to Proposition 2.3.

With the above proposition, we have excluded a large class of possible choices for the vector \( d \) – it suffices to look at colored graphs with the property \( \frac{1}{2} d^T d < 9\ell \). The following proposition states that these colored graphs are "almost" balanced, i.e., the number of "large" \( d_i \)'s is "small". For this, let

\[
T_g(d) := \{1 \leq i \leq \ell : |d_i| \geq g\}, \tag{2.3}
\]

where \( g \geq 1 \) is any function which may depend on \( \ell \) and \( n \).

**Proposition 2.6.** Let \( d \in D(n - \lfloor n/\ell \rfloor, \ell, n) \) and \( d \notin L(n, \ell) \). For every \( g \geq 1 \), \( \ell(n) = \omega(1) \) and \( \ell(n) = o(n) \) it holds

\[
|T_g(d)| \leq \frac{18\ell}{g^2}.
\]

**Proof.** Consider the dot product \( \frac{1}{2} d^T d \). An upper bound follows easily from \( d \notin L(n, \ell) \); a lower bound is obtained from the assumption on the magnitude of the \( |d_i| \)'s. We obtain

\[
9\ell > \frac{1}{2} d^T d = \frac{1}{2} \sum_{i \in T_g(d)} d_i^2 \geq \frac{1}{2} |T_g(d)| \cdot g^2,
\]

which implies the statement.

Assume that \( \ell(n) \to \infty \) for \( n \to \infty \). Due to Proposition 2.6, for every function \( g \) with the property \( g \to \infty \), the number of positions in a given vector \( d \), which are larger than \( g \) in absolute value, is at most \( o(\ell) \). Therefore, provided that we restrict our attention to colored graphs with a "small" value of \( \frac{1}{2} d^T d \), we find \((1-o(1))\ell \) color classes, which deviate less than \( g \) from \( \lfloor n/\ell \rfloor \); this observation will be one of the main ingredients in the following proofs.

### 3 Proof of Theorem 1.1, first statement

In this section we will prove the first statement of Theorem 1.1, where we first will deal with the case \( \ell(n) = \omega(1) \). Let \( c_1 = 16 \sqrt{60} \) and \( f_0 = \sqrt{2\log \ell - c_1 \cdot (\log \ell)^2} \); with this notation, the theorem reads

\[
\text{col}_n^\ell(f_0) = o\left(|\text{Col}_n^\ell|\right).
\]

In the sequel we will assume that the function \( \ell(n) \) satisfies \( \ell(n) \leq n/\log n \) and \( \ell(n) \to \infty \) for \( n \to \infty \) throughout.

One main ingredient in the proof below is the observation that we can partition the set of all colored graphs according to the set of integer sequences \( D(n - \lfloor n/\ell \rfloor, \ell, n) \), which describe the deviation of the sizes of the color classes from \( \lfloor n/\ell \rfloor \), in disjoint classes. This will allow us to compare directly the number of colored graphs with maximum deviation smaller than \( f_0 \) in absolute value with the graphs which have deviations larger than \( f_0 \). In the proof, we will exploit the structure of the integer sequences and relate the number of colored graphs corresponding to "similar" pairs of such sequences, where the term similar will be made precise later.
Recall the definitions of $b$ and $E$ from (1.2) and (1.3). In order to formalize the above idea, we define a mapping $m$ which maps a given vector $d \in D (f_0, \ell, n)$ with $d^T d < 18 \ell$ onto a set of vectors $m(d) \subset D (n - \lceil n/\ell \rceil, \ell, n)$, such that the following three properties are fulfilled for $\ell$ sufficiently large:

(i) $m$ is injective, i.e., for all $d, d' \in D (f_0, \ell, n)$ with $d \neq d'$ it holds that $m(d) \cap m(d') = \emptyset$,

(ii) $|m(d)| \geq \frac{\ell}{2}$, i.e., $m(d)$ is large and

(iii) for all $d \in m(d)$ it holds that $b(d)E(d) \geq b(d)E(d) \cdot 2^{-\frac{1}{2}f_0^2 (1 + \frac{2}{\omega_0})}$, where $\omega_0 = \frac{16f_0^2/3}{c_1}$ denotes an auxiliary function.

Let $S(d) = \sum_{d \in m(d)} b(d)E(d)$. With (ii) and (iii) we obtain for all $d \in D (f_0, \ell, n)$

$$\frac{b(d)E(d)}{S(d)} \leq \frac{2 \cdot 2^{-\frac{1}{2}f_0^2 (1 + \frac{2}{\omega_0})}}{\ell} = o(1),$$

where the last step is due to the definitions of $f_0$ and $\omega_0$. Now we show how the theorem can be proved with (1.4) and (i):

$$col^\ell_n (f_0) = \frac{1}{\ell} \cdot \sum_{d \in D (f_0, \ell, n), d^T d \geq 18 \ell} b(d)E(d) + \frac{1}{\ell} \cdot \sum_{d \in D (f_0, \ell, n), d^T d < 18 \ell} b(d)E(d)$$

Prop. 2.5 $= 2 \frac{1}{\ell} \sum_{d \in D (f_0, \ell, n), d^T d \geq 18 \ell} b(d)E(d) \cdot \frac{S(d)}{S(d)} - S(d)$

$\leq (o(1) + \max_{d \in D (f_0, \ell, n), d^T d < 18 \ell} \frac{b(d)E(d)}{S(d)}) \cdot |col^\ell_n| = o \left( \left| Col^\ell_n \right| \right)$.

It remains to define the mapping $m$ and to prove properties (i), (ii) and (iii). For this, consider for a given vector $d = (d_1, \ldots, d_i) \in D (f_0, \ell, n)$ with $d^T d < 18 \ell$ the subset of indexes

$$P(d) = \left\{ 1 \leq i \leq \ell - f_0 - 1 : \left\{ i \leq j \leq i + f_0 + 1 : |d_j| \geq \frac{f_0}{\omega_0} \right\} \leq \frac{f_0}{\omega_0} \right\}.$$

$m$ maps a given vector $d$ onto the set of vectors

$$\left\{ \ldots, d_{k-1}, d_k + g_k, d_{k+1} + 1_k, d_{k+\lceil g_k \rceil} + 1_k, d_{k+\lceil g_k \rceil + 1}, \ldots \right\} : k \in P(d) \text{ and } \left| d_k \right| \leq \frac{f_0}{\omega_0} \right\}.$$

where $g_k = \text{sign}(d_k) \cdot (f_0 + 1)$ and $1_k = \text{sign}(d_k)$; here we assume $\text{sign}(0) = +1$. It is easy to verify that for all pairs $d, d' \in D (f_0, \ell, n)$ of distinct vectors it holds $m(d) \cap m(d') = \emptyset$, i.e., we construct no integer sequence more than once by $m$; this shows that property (i) is fulfilled by $m$. Furthermore, note that every element in $m(d)$ is an admissible integer sequence, in the sense that the invariant $\sum_{i=1}^\ell d_i = r(n, \ell)$ is preserved, where $r(n, \ell)$ stands for the remainder of $n$ and $\ell$.

Next we show that property (ii) is fulfilled for the above defined mapping, i.e., $\left| \left\{ k \in P(d) \text{ and } \left| d_k \right| \leq \frac{f_0}{\omega_0} \right\} \right| \geq \ell/2$. Note that it suffices to show that $|P(d)| \geq \frac{\ell}{2} - f_0$; we then can conclude with Proposition 2.6 that the number of entries of $d$, which are larger than $\frac{f_0}{\omega_0}$ in absolute value, fulfills for large enough $\ell$ the claimed bound.
Proposition 3.1. Let \( d \in D(f_0, \ell, n) \) with \( d^T d < 18\ell \). Then \( |P(d)| \geq \frac{20\ell}{3} - f_0 \).

Proof. Let \( B(d) := \{1, \ldots, \ell - f_0 - 1\} \setminus P(d) \) and assume that \( |B(d)| > \frac{\ell}{3} = 20\ell\omega^2/f_0^2 \). Let \( x \in B(d) \); then there are at least \( f_0/\omega \) indexes \( i \) in the interval \([x, x + f_0 + 1]\) with \( |d_i| \geq f_0/\omega\). With the notation of (2.3), we obtain with the above assumptions for sufficiently large \( \ell \) the following lower bound for the number of "large" entries in \( d \):

\[
|T_{\ell}^d(d)| \geq \sum_{x \in B(d)} \frac{f_0}{\omega} \geq \frac{19}{20} \frac{|B(d)|}{\omega\ell} > \frac{19\ell\omega^2}{f_0},
\]

where the first inequality follows from the fact that we count each "large" \( d_i \) at most \( f_0 + 2 \) times. But this is clearly a contradiction; with Proposition 2.6 we obtain that the cardinality of the set of indexes, which are larger than \( f_0/\omega\ell \) in absolute value, is at most \( 18\ell\omega^2/f_0^3 \).

We conclude that \( |B(d)| \leq 40\ell\omega^2/f_0^3 = \ell/3 \), which proves the proposition.

Finally we show that (iii) holds for the mapping \( m \) defined above. Let for the moment \( d \in m(d) \) be a vector which differs from the "very" balanced vector \( d \) at the position \( k \) by \( f_0 + 1 \) and at the positions \( k + 1, \ldots k + f_0 + 1 \) by \( -1 \) (the symmetric case can be handled in the same way, where we negate all signs). With Proposition 2.3 and the definition of the multinomial coefficient we obtain

\[
b(d)E(d) = b(d)E(d) \cdot \prod_{i=1}^{f_0+1} \left( \left\lfloor \frac{n}{\ell} \right\rfloor + d_{k+i} \right) \left( \left\lceil \frac{n}{\ell} \right\rceil + d_k + f_0 + 1 \right)^{f_0+1} \cdot 2^{-\frac{1}{2}(f_0+1)^2 - d_k(f_0+1) + \sum_{x=k+1}^{k+f_0+1} d_x - \frac{f_0+1}{2}}. \quad (3.4)
\]

Recall the definition (3.2) of \( P(d) \). As \( k \in P(d) \), there are at most \( f_0/\omega \) indexes \( i_k \) in the interval \([k + 1, k + f_0 + 1]\) with \( |d_{i_k}| \geq f_0/\omega \). Therefore

\[
\sum_{x=k+1}^{k+f_0+1} d_x \geq \left( f_0 + 2 \right) - \frac{f_0}{\omega\ell}, \quad \left( f_0 + 2 \right) - \frac{f_0}{\omega\ell} \cdot (-f_0) \geq -\frac{2f_0^2}{\omega\ell},
\]

where the last inequality holds for sufficiently large \( \ell \). With this, the exponent of the power-of-2 term in (3.4) can be (crudely) bounded from below with \( -\frac{1}{2}f_0^2(1 + \frac{2}{\omega\ell}) \), as due to (3.3) we can assume \( |d_k| \leq f_0/\omega \). To give a lower bound for the term with the product in (3.4) observe that \( f_0 = o(\lfloor n/\ell \rfloor) \) due to the assumptions on \( \ell \); therefore, \( |d_i| = o(\lfloor n/\ell \rfloor) \) holds for all \( 1 \leq i \leq \ell \). We obtain

\[
\prod_{i=1}^{f_0+1} \left( \left\lfloor \frac{n}{\ell} \right\rfloor + d_{k+i} \right) \left( \left\lceil \frac{n}{\ell} \right\rceil + d_k + f_0 + 1 \right)^{f_0+1} \geq 2^{-f_0+1}.
\]

Putting all together yields \( b(d)E(d) \geq b(d)E(d) \cdot 2^{-\frac{1}{2}f_0^2(1 + \frac{2}{\omega\ell})} \), i.e., \( m \) fulfills (iii).

4 Proof of Theorem 1.1, second statement

In this section we will prove the second statement of Theorem 1.1, namely

\[
\frac{1}{\ell} \sqrt{\frac{2}{\ell} \log \ell + 274(\log \ell)^{2/3}} = (1 - o(1)) \left| \text{Col}_n^\ell \right|.
\]

8
where again we first consider the case \( \ell(n) = \omega(1) \). Furthermore, as in Section 3, we assume throughout that \( \ell(n) \leq \frac{n}{\alpha(n)} \), where \( \alpha(n) \geq \log n \). For brevity, we will abbreviate \( f_0 = (2 \log \ell + 2^{14}(\log \ell)^{2/3})^{1/2} \).

Before we proceed with the proof of Theorem 1.1, we exclude a large class of admissible vectors \( \mathbf{d} \), i.e., we show that the number of corresponding colored graphs is negligible. Recall definition (2.3) and let

\[
Q(n, \ell) = \left\{ \mathbf{d} \in \mathcal{D} \left( n - \left\lfloor \frac{n}{\ell} \right\rfloor, \ell, n \right) : \left| \sum_{i \in T_{f_0}(\mathbf{d})} d_i \right| \geq \frac{19\ell}{f_0} \right\}. \tag{4.1}
\]

Proposition 4.1. Let \( Q(\ell, n) \) be defined as above. Then

\[
\mathbf{d} \in S_Q \Rightarrow \frac{1}{2} \mathbf{d}^T \mathbf{d} > 9\ell, \tag{4.2}
\]

or equivalently

\[
S_Q := \frac{1}{\ell!} \sum_{\mathbf{d} \in Q(n, \ell)} b(\mathbf{d})E(\mathbf{d}) = o \left( |\text{Col}_n^\ell| \right).
\]

Proof. We show that for every integer sequence \( \mathbf{d} \) the statement in (4.2) holds, which implies with Proposition 2.5 the second statement. Recall definition (2.3) and consider the dot product \( \frac{1}{2} \mathbf{d}^T \mathbf{d} \); a lower bound can easily be obtained with the assumption on the absolute value of the "large" \( d_i \)'s:

\[
\mathbf{d}^T \mathbf{d} \geq \sum_{i \in T_{f_0}(\mathbf{d})} d_i^2 \geq \sum_{i \in T_{f_0}(\mathbf{d})} d_i \cdot \min_{i \in T_{f_0}(\mathbf{d})} |d_i| \geq \frac{19\ell}{f_0} \cdot f_0 = 19\ell.
\]

We conclude \( \frac{1}{2} \mathbf{d}^T \mathbf{d} > 9\ell \). \hfill \( \blacksquare \)

Now we prove the second part of Theorem 1.1. Let

\[
\tilde{\mathcal{D}}(\ell, n) = \left\{ \mathbf{d} \in \mathcal{D} \left( n - \left\lfloor \frac{n}{\ell} \right\rfloor, \ell, n \right) : \mathbf{d}^T \mathbf{d} < 18\ell \right\}.
\]

Now we partition the set of all integer sequences, which contain at least one value larger than \( f_0 \) according to the positions of that values and according to their magnitude. More precisely, recall (2.3), fix a non-empty index set \( T \subseteq \{1, \ldots, \ell\} \) and an integer sequence \( \mathbf{x}_T = (x_1, \ldots, x_{|T|}) \) with \( |x_j| \geq f_0 \) for all \( 1 \leq j \leq |T| \); now consider the set of integer sequences

\[
\mathcal{D}_{T, \mathbf{x}_T} = \left\{ \mathbf{d} \in \tilde{\mathcal{D}}(\ell, n) : T_{f_0}(\mathbf{d}) = T = \{i_1, \ldots, i_{|T|}\} \right\}\tag{4.3}
\]

and \( \forall 1 \leq j \leq |T| : d_{i_j} = x_j \) and \( \forall j \notin T : |d_j| < f_0 \).

Our strategy is now to show that for every such set of integer sequences \( \mathcal{D}_{T, \mathbf{x}_T} \), the set of corresponding colored graphs is "small", namely

\[
\sum_{\mathbf{d} \in \mathcal{D}_{T, \mathbf{x}_T}} b(\mathbf{d})E(\mathbf{d}) \leq \ell! \cdot 2^{-\frac{1-\ell(\ell)}{2}} \sum_{i=1}^{\frac{|T|}{2}} x_i^2 \cdot |\text{Col}_n^\ell|, \tag{4.4}
\]
where \( d(\ell) = \frac{212}{f_0^{1/2}} \). Observe that due to Proposition 2.6 it is suffient to consider sets \( \mathcal{T} \) with \(|\mathcal{T}| \leq 18\ell/f_0^2\). With the above inequality, the theorem can be proved as follows: we count the number of the \( \ell \)-colored graphs, which contain at least one color class, that deviates more than \( f_0 \) in absolute value from \([n/\ell]\) by first choosing the number \( t \) of entries which have a large deviation. Second, we choose their positions, i.e., the set \( \mathcal{T} \) of size \( t \). Third, we fix their values, i.e., the vector \( x_{\mathcal{T}} \). Finally, we sum up for all integer sequences in \( \mathcal{D}_{\mathcal{T}, x_{\mathcal{T}}} \) the number of corresponding colored graphs, i.e., we obtain with (1.4)

\[
|\text{Col}_{n}^\ell| - \text{col}_{n}^\ell(f_0) = \frac{1}{\ell!} \cdot \sum_{d \in \mathcal{D}(f_0, \ell, n)} b(d)E(d) + \frac{1}{\ell!} \cdot \sum_{d \in \mathcal{D}(f_0, \ell, n)} b(d)E(d) \\
\leq o \left( |\text{Col}_{n}^\ell| \right) + |\text{Col}_{n}^\ell| \cdot \sum_{t=1}^{18\ell/f_0^2} \cdot \sum_{x = \{x_1, \ldots, x_t\}} \sum_{|x_i| \geq f_0} 2^{-1 - d(\ell) t} x_i^t \\
\leq o \left( |\text{Col}_{n}^\ell| \right) + |\text{Col}_{n}^\ell| \cdot \sum_{t=1}^{18\ell/f_0^2} t \left( \sum_{|x| \geq f_0} 2^{-1 - d(\ell) x^2} \right)^t.
\]

Let \( z = \frac{1 - d(\ell)}{2} \). The inner sum above can be estimated as follows:

\[
\sum_{|x| \geq f_0} 2^{-zx^2} = 2 \cdot 2^{-z f_0^2} \sum_{x \geq 0} 2^{-z(x^2 + 2xf_0)} \leq 2 \cdot 2^{-z f_0^2} \sum_{x \geq 0} 2^{-zx^2}.
\]

However, it can be easily seen that the last sum converges rapidly:

\[
\sum_{x \geq 0} 2^{-zx^2} \leq 1 + \sum_{x \geq 1} 2^{-x^2} \leq 1 + \sum_{x \geq 1} \frac{4}{x^2} = 1 + \frac{4}{6} \pi^2 \leq 8.
\]

Recall the definition of \( f_0 \) at the beginning of this chapter. By putting all together, we obtain from (4.5)

\[
|\text{Col}_{n}^\ell| - \text{col}_{n}^\ell(f_0) \leq o \left( |\text{Col}_{n}^\ell| \right) + |\text{Col}_{n}^\ell| \cdot \sum_{t=1}^{18\ell/f_0^2} (8\ell)^t \cdot 2^{-1 - d(\ell) f_0^2 + t} \\
\leq o \left( |\text{Col}_{n}^\ell| \right) + |\text{Col}_{n}^\ell| \cdot \sum_{t \geq 1} 2^{-t \Theta((\log \ell)^2/3)} = o \left( |\text{Col}_{n}^\ell| \right),
\]

which concludes the proof of the theorem.

What remains is to prove is (4.4). The core idea is as in Section 3 to define a mapping \( m \) from the set of the "unbalanced" integer sequences \( d \) to the set of "mainly balanced" integer sequences. By applying this mapping, the sum counting the unbalanced colored graphs will be transformed into a weighted sum over all colored graphs, where the weights correspond to the "profit" we make by switching from unbalanced to balanced partitions. We then will show that this weights are "small enough", i.e., the inequality (4.4).
The mapping works in this case as follows. Let $1 \leq t \leq 18\ell/f_0^2$ and fix for the remainder of this section a non-empty index set $T = \{i_1, \ldots, i_t\} \subseteq \{1, \ldots, \ell\}$ and an integer sequence $x_T = (x_1, \ldots, x_t)$ with $|x_j| \geq f_0$ for all $1 \leq j \leq |T|$. For every sequence $d \in D_{T,x_T}$ we look at the "excess" $E(x) = \sum_{i=1}^{t} x_i$ generated by the entries with "large" deviation, and split it up into small parts, which are distributed among the entries of $d$ with indexes in $\{1, \ldots, \ell\} \setminus T$. Intuitively, in this way we "transform" colored graphs with unbalanced partitions into graphs with (almost) balanced partitions, if we can guarantee that among the color classes, which are modified, there are many with "sufficiently small" deviation.

In order to formalize this idea, we define for every set $D_{T,x_T}$ given in (4.3) an injective mapping to the set of all admissible integer sequences, i.e.,

$$m_{T,x_T} : D_{T,x_T} \rightarrow D \left(n - \left\lfloor \frac{n}{\ell} \right\rfloor, \ell, n\right).$$

Due to symmetry, we assume in the following that $T = \{1, \ldots, t\}$ for some $t \geq 1$, as this will ease the calculations and shorten notation. Furthermore, for brevity we write $m_{t,x}$ for $m_{T,x_T}$ and similarly, we write $D_{t,x}$ instead of $D_{T,x}$. Let $s(x) = |E(x)|/|E(x)|$, if $E(x) \neq 0$ and $s(x) = 1$ otherwise. With these assumptions, the mapping is given by

$$m_{t,x}(d) \rightarrow (0, \ldots, 0, d_{t+1} + s(x), \ldots, d_{t+|E(x)|+1} + s(x), d_{t+|E(x)|+2}, \ldots, d_{\ell}). \quad (4.6)$$

The main idea behind the above mapping is to "split" up $E(d)$ in parts of size 1 or -1, depending on $s(d)$; then it replaces all $d_j$’s for $1 \leq j \leq t$ by zero; finally, it modifies the first $|E(d)|$ entries $d_{t+1}, \ldots$, which are smaller than $f_0$ in absolute value, by $s(d)$. The mapping is always well-defined, as due to Proposition 4.1 it suffices to consider only integer sequences $d \in D_{t,x}$ with $|E(x)| < 19\ell/f_0$ (otherwise $D_{t,x}$ would be empty, and (4.4) would hold trivially). Note also that $m_{t,x}(d) \in D \left(n - \left\lfloor \frac{n}{\ell} \right\rfloor, \ell, n\right)$, as the invariant $\sum_{i=1}^{\ell} (m(d))_i = r(n, \ell)$ is preserved through the mapping (recall that $r(n, \ell)$ denotes the remainder of $n \in \ell$).

Observe that every integer sequence $d \in D \left(n - \left\lfloor \frac{n}{\ell} \right\rfloor, \ell, n\right)$ is constructed at most once by $m_{t,x}$, i.e., for distinct integer sequences $d, d' \in D_{t,x}$ we have $m_{t,x}(d) \neq m_{t,x}(d')$. Hence $m_{t,x}$ is injective.

Intuitively, it can happen that the above defined mapping is sometimes in a specific sense "bad". This case occurs, if there are many large entries among the first $|E(d)|$ entries, which are modified; in such a case, the image $m_{t,x}(d)$ of the considered integer sequence can be even more "unbalanced" than $d$ itself, which would imply that the "profit" made by switching from $d$ to $m_{t,x}(d)$ is not large enough in order to yield the inequality (4.4).

On the other hand, given an integer sequence $d$ with $d^T d < 18\ell$, Proposition 2.6 states that the number of entries greater than some given $g$ in absolute value is "small", namely at most $18\ell/g^2$. In the following, we show that this is in almost all cases also a local property of those integer sequences, namely, that within a subsequence of a given length we have with high probability only proportionally many "large" entries.

In order to make this more precise, observe that for any fixed $x$ of size $t$, we can group the integer sequences in $D_{t,x}$ according to the values of their remaining $\ell - t$ entries. This grouping has the advantage that for any pair $d$ and $d'$ of integer sequences in such a group, the number of corresponding colored graphs is equal, i.e., $b(d)E(d) = b(d')E(d')$, and therefore allows us to compare directly the number of graphs for which the mapping (4.6) is "good" with the number of graphs for which it is "bad". More formally, consider the definition below.
\textbf{Definition 4.2.} Let $1 \leq t \leq 18\ell/f_0^2$. Furthermore, let $\mathbf{x} = (x_1, \ldots, x_t)$ be an integer sequence and $\mathbf{y} = (y_{-\lfloor n/\ell \rfloor}, \ldots, y_{n-\lfloor n/\ell \rfloor})$ be a sequence of non-negative integers. We call the tuple $\mathbf{T} = (\mathbf{x}, \mathbf{y})$ a descriptor, if the following properties are fulfilled:

\begin{equation}
\forall 1 \leq j \leq t : |x_j| \geq f_0 \quad \text{and} \quad \left| \sum_{j=1}^{t} x_j \right| < \frac{19\ell}{f_0} \quad \text{and} \quad \left(4.7\right)
\end{equation}

\begin{equation}
\sum_{i=-\lfloor n/\ell \rfloor}^{n-\lfloor n/\ell \rfloor} y_i = \ell - t \quad \text{and} \quad \sum_{i=1}^{t} x_i + \sum_{i=-\lfloor n/\ell \rfloor}^{n-\lfloor n/\ell \rfloor} y_i \cdot i^2 < 18\ell \quad \text{and} \quad \sum_{i=1}^{t} x_i + \sum_{i=-\lfloor n/\ell \rfloor}^{n-\lfloor n/\ell \rfloor} y_i \cdot i = r(n, \ell). \quad \left(4.8\right)
\end{equation}

Every descriptor $\mathbf{T}$ defines as follows a set of integer sequences:

\begin{equation}
\mathbf{q}(\mathbf{T}) = \{ \mathbf{d} : \forall 1 \leq j \leq t : d_j = x_j \quad \text{and} \quad \forall j \in \left[-\frac{n}{\ell}, \frac{n}{\ell} \right] : |\{ x : d_x = j \}| = y_j \}. \quad \left(4.9\right)
\end{equation}

Let $1 \leq g \leq \ell/2$ be a function, which may depend on $\ell$ and $\omega_\ell$ a function with $\omega_\ell \to \infty$ when $\ell \to \infty$. We call an integer sequence $\mathbf{d} \in \mathbf{q}(\mathbf{T})$ $(g, \omega_\ell)$-bad, if among the values $d_{t+1}, \ldots, d_{t+g}$ there are more than $\left\lfloor g \cdot \frac{256}{\omega_\ell^2} \right\rfloor$ values larger than $\omega_\ell$ in absolute value. More precisely, $\mathbf{d}$ is $(g, \omega_\ell)$-bad, if

\begin{equation}
|\{ j \in \{ t+1, \ldots, t+g \} : |d_j| > \omega_\ell \}| \geq \left\lfloor g \cdot \frac{256}{\omega_\ell^2} \right\rfloor. \quad \left(4.10\right)
\end{equation}

The following corollary is an immediate consequence of the above definition.

\textbf{Corollary 4.3.} Let $\mathbf{T}$ be a descriptor and $\mathbf{d}, \mathbf{d}' \in \mathbf{q}(\mathbf{T})$. Then $b(\mathbf{d})E(\mathbf{d}) = b(\mathbf{d}')E(\mathbf{d}')$, where $b$ and $E$ are defined in (1.2) and (1.3).

The lemma below states that almost all integer sequences in $\mathbf{q}(\mathbf{T})$ will be ”well-behaved”, in the sense that they do not have big ”clusters” of large values.

\textbf{Lemma 4.4.} Let $\mathbf{T} = (\mathbf{x}, \mathbf{y})$ with $\mathbf{x} = (x_1, \ldots, x_t)$ be a descriptor and $g \geq 1$ a function which may depend on $\ell$ with $g < 19\ell/f_0$. Furthermore, define $B_{(g, \omega_\ell)}(\mathbf{T}) = \{ \mathbf{d} \in \mathbf{q}(\mathbf{T}) : \mathbf{d}$ is $(g, \omega_\ell)$-bad $\}$. Then $|B_{(g, \omega_\ell)}(\mathbf{T})| \leq \frac{|\mathbf{q}(\mathbf{T})|}{2g}$.

\textbf{Proof.} Firstly observe that $|\mathbf{q}(\mathbf{T})| = \frac{\ell f_0^2}{\text{aut}(\mathbf{T})}$, where we set $\text{aut}(\mathbf{T}) = \prod_{j=-\lfloor n/\ell \rfloor}^{\lfloor n/\ell \rfloor} y_j!$. Furthermore, due to (4.8), every $\mathbf{d} \in \mathbf{q}(\mathbf{T})$ satisfies $\mathbf{d}^T \mathbf{d} < 18\ell$; it follows with Proposition 2.6, that the number of indexes $1 \leq x \leq \ell$ with $|d_x| > \omega_\ell$ is at most $18\ell/\omega_\ell^2$ and therefore $\sum_{|j| > \omega_\ell} y_j \leq 18\ell/\omega_\ell^2$.

An upper bound for the size of $B_{(g, \omega_\ell)}(\mathbf{T})$ can be given as follows. We count the number of integer sequences $\mathbf{d} \in \mathbf{q}(\mathbf{T})$ which are $(g, \omega_\ell)$-bad by choosing $z = \left\lfloor g \cdot \frac{256}{\omega_\ell^2} \right\rfloor > 0$ positions $p_1, \ldots, p_z$ out of the positions $t+1, \ldots, t+g$. Then we choose the values $v_1, \ldots, v_z$ for these positions, such that $|v_i| \geq \omega_\ell$ for all $1 \leq i \leq z$; with the above observation, there are at most $(18\ell/\omega_\ell^2)^z$ ways to do that. Finally, we permute all the remaining $(\ell - t - z)$ values. It is clear that in this way we have constructed each vector in $B_{(g, \omega_\ell)}(\mathbf{T})$ exactly $\text{aut}(\mathbf{T})$ times. Hence,

\begin{equation}
\frac{|B_{(g, \omega_\ell)}(\mathbf{T})|}{|\mathbf{q}(\mathbf{T})|} \leq \frac{\left(\frac{\ell^z}{(\ell - t)!}\right)^z (\ell - t - z)!}{\ell^z} \leq \frac{\ell^x \cdot \left(18e \frac{\omega_\ell^2}{256}\right)^z}{(\ell - t)!} \leq \frac{\ell^x \cdot \omega_\ell^{2z}}{(\ell - t)!} \cdot 2^{(x-8+\log 18e)}.
\end{equation}
Observe that due to the assumptions on \( T \), for sufficiently large \( \ell \) it holds \( \ell - t - z \geq \frac{\ell}{2} \). Therefore
\[
\frac{\ell^z}{(\ell - t)^z} \leq \left( \frac{\ell}{\ell - t - z} \right)^z \leq 2^z.
\]
Putting all together yields \(|B_{(g,\omega_\ell)}(T)| \leq 2^{-z/2} \cdot |q(T)|\), which completes the proof.

\[\text{Proof.}\]

Lemma 4.5. For this, we estimate the ratios of the functions \( \rho \) for every descriptor \( T \)

\[
\sum_{d \in q(T)} b(d)E(d) \leq 2 \cdot \sum_{d \not\in (g,\omega_\ell)-bad} b(d)E(d).
\]

(4.11)

Furthermore, given any \( x \) of size \( t \), we can write the set of integer sequences \( D_{t,x} \) as a disjoint union over all admissible descriptors \( (x,y) \), i.e., it suffices to consider integer sequences which are not \( (g,\omega_\ell) \)-bad. With this observation, we can now prove (4.4), where in the remainder of the section we assume that \( \omega_\ell = \frac{1}{\ell} \) :

\[
\sum_{d \in D_{t,x}} b(d)E(d) \leq 2 \cdot \sum_y \sum_{d \not\in (\sum_{i=1}^t x_i, \omega_\ell)-bad} b(d)E(d)
\]

\[
\leq 2 \cdot \sum_y \sum_{d \not\in (\sum_{i=1}^t x_i, \omega_\ell)-bad} \frac{b(d)E(d)}{b(m(d))}b(m(d))E(m(d))
\]

(4.12)

where

\[
M(t;x) = \max_{d \in \mathcal{X}(t;x)} \frac{E(d)}{E(m(d))} \quad \text{and} \quad \mathcal{X}(t;x) = \{d \in D_{t,x} : d \not\in (\sum_{i=1}^t x_i, \omega_\ell)-bad\}.
\]

(4.13)

Note that the \( \ell! \) in the last inequality in (4.12) is due to the fact that we construct every admissible integer sequence \( d \in \mathcal{D} (n - \lfloor n/\ell \rfloor, \ell, n) \) at most once through \( m_{t,x} \) – therefore we count all colored graphs at most \( \ell! \) times on the left-hand side.

In the remainder of this section we show that \( M(t,x) \leq 2^{-1 - 2t/\ell} \sum_{i=1}^t s_i^2 \), which proves (4.4); for this, we estimate the ratios of the functions \( \rho \) and \( E \) seperately.

Lemma 4.5. Let \( \xi_\ell = 2^{1/2} \omega_\ell^{-2} \). With the above definitions, for every \( t \geq 1 \) and sufficiently large \( n \) it holds

\[
2 \left( \log E(d) - \log E(m(d)) \right) = \sum_{i=1}^\ell \left( (m(d))_i^2 - d_i^2 \right)
\]

(4.14)

Proof. Recall the Definition (1.3) of \( E \) and Proposition 2.3. By considering the logarithm of the left-hand side of the statement and using the definition of \( m \), we obtain

\[
2 \left( \log E(d) - \log E(m(d)) \right) = \sum_{i=1}^\ell \left( (m(d))_i^2 - d_i^2 \right)
\]

(4.14)
Now recall (4.10): as we consider only \((|\mathcal{E}(x)|, \omega_t)\)-good sequences \(d\), we can bound the last sum with

\[
2s(d) \sum_{j=1}^{\frac{|\mathcal{E}(x)|}{2}} d_{t+j} \leq 2 \left( \frac{256}{\omega_t^2} |\mathcal{E}(x)| \cdot f_0 + \left( 1 - \frac{256}{\omega_t^2} \right) |\mathcal{E}(x)| \cdot \omega_t \right)
\]

\[
\leq 2 |\mathcal{E}(x)| f_0 \cdot \frac{256}{\omega_t^2} \cdot \left( 1 + \frac{\omega_t^2}{f_0} \right) \leq 2^{10} \frac{|\mathcal{E}(x)| f_0}{\omega_t^2},
\]

(4.15)

where the last step is due to \(\omega_t = f_0^{\frac{1}{4}}\). By putting all together we obtain

\[
2 \cdot (\log E(d) - \log E(m(d))) \leq -(1 - \xi_t) \sum_{j=1}^{t} x_j^2 - \xi_t \sum_{j=1}^{t} x_j^2 + 2^{11} \cdot \frac{|\mathcal{E}(x)| f_0}{\omega_t^2}
\]

\[
\leq -(1 - \xi_t) \sum_{j=1}^{t} x_j^2 - \xi_t |\mathcal{E}(x)| \cdot \min_{1 \leq j \leq t} |x_j| + 2^{11} \cdot \frac{|\mathcal{E}(x)| f_0}{\omega_t^2},
\]

which completes the proof, as we assumed \(|x_j| \geq f_0\) for all \(1 \leq j \leq t\).

**Lemma 4.6.** With the above definitions, for every \(t \geq 1\) and sufficiently large \(n\) it holds

\[
\max_{d \in \mathcal{X}(t; x)} \frac{b(d)}{b(m(d))} \leq \exp \left\{ (1 + o(1)) \frac{2^{10} \cdot |\mathcal{E}(x)| \cdot f_0}{\omega_t^2 \cdot \left[ \frac{n}{t} \right]} \right\}. \tag{4.16}
\]

Proof. Abbreviate \(\alpha = \frac{n}{\ell}\) and recall the definition (1.2) of \(b\); we obtain

\[
\frac{b(d)}{b(m(d))} = \frac{\prod_{j=1}^{t} (\alpha + (m(d)))_{d_j}!}{\prod_{j=1}^{t} (\alpha + d_j)!} = \prod_{i=1}^{t} (\alpha + x_j)! \cdot \begin{cases} \prod_{j=1}^{t} (\alpha + d_{t+j} + 1) & \text{if } s(d) = 1 \\ \prod_{j=1}^{t} \frac{1}{\alpha + d_{t+j}} & \text{if } s(d) = -1. \end{cases}
\]

Let \(x^\sharp = x(x - 1) \ldots (x - y + 1)\). We treat each term above separately:

\[
\frac{(\alpha !)^t}{\prod_{i=1}^{t} (\alpha + x_j)!} = \frac{\prod_{x_i \leq f_0} \alpha^{-x_i}}{\prod_{x_i \geq f_0} (\alpha + x_i)_{\omega}} \leq \alpha^{-\sum_{x_i \leq f_0} x_i} = \alpha^{-\mathcal{E}(x)}.
\]

When \(s(d) = 1\), the second term can be estimated as follows:

\[
\prod_{j=1}^{t} (\alpha + d_{t+j} + 1) \leq \alpha^{\mathcal{E}(x)} \prod_{j=1}^{t} \left( 1 + \frac{d_{t+j} + 1}{\alpha} \right) \leq \alpha^{\mathcal{E}(x)} \cdot e^{\sum_{j=1}^{t} d_{t+j}}. \tag{4.16}
\]

If \(s(d) = -1\), observe that \(\mathcal{E}(x) < 0\); we obtain with the inequality \(1 - x \geq e^{-x - x^2}\), which is valid for small \(x\)

\[
\prod_{j=1}^{t} \frac{1}{\alpha + d_{t+j}} \leq \alpha^{-\mathcal{E}(x)} \prod_{i=1}^{t} \frac{1}{1 - \frac{d_{t+j}}{\alpha}} \leq \alpha^{-\mathcal{E}(x)} \cdot e^{\sum_{i=1}^{t} \frac{|d_{t+j}|}{\alpha} + \frac{d_{t+j}^2}{\alpha^2}}. \tag{4.17}
\]
As we consider only \(|E(x)|, \omega\)-good sequences, with (4.10) we can estimate the common sum in the exponent of \(e\) from (4.16) and (4.17)

\[
\sum_{j=1}^{\lfloor n/\ell \rfloor} dt_{t+j} \leq \sum_{j=1}^{\lfloor n/\ell \rfloor} |dt_{t+j}| \leq \frac{2^{10} \cdot |E(x)| f_0}{\omega \ell},
\]

where the derivation is essentially the same as in (4.15). Similarly, the additional error term in (4.17) can be crudely estimated from above with

\[
\frac{1}{\alpha^2} \sum_{i=1}^{\lfloor n/\ell \rfloor} d^2_{t+i} \leq \frac{|E(x)| f_0^3}{\alpha^2} \leq 3 \frac{|E(x)|}{\alpha},
\]

as \(|dt_{i+j}| \leq f_0\) and \(\alpha = \lfloor \frac{n}{\ell} \rfloor \geq \log n\). By putting all together, the lemma follows immediately with \(\omega = f_0^{1/3}\), as this implies \(|E(x)| f_0^2\).

Let \(\xi_\ell\) be defined as in Lemma 4.5. With Lemma 4.6 we obtain the following upper bound for \(M\) from (4.13):

\[
\log M(t; x) \leq \frac{1}{\alpha^2} \sum_{i=1}^{\lfloor n/\ell \rfloor} x_i^2 + \frac{1^{11} \cdot |E(x)| \cdot f_0}{\omega \ell \cdot \lfloor n/\ell \rfloor} \leq \frac{(1 + o(1)) \xi_\ell}{2} \sum_{i=1}^{\lfloor n/\ell \rfloor} x_i^2,
\]

(4.18)
as \(|x_i| \geq f_0\) and \(\sum_{i=1}^{\lfloor n/\ell \rfloor} x_i^2 \geq \sum_{i=1}^{\lfloor n/\ell \rfloor} x_i \cdot f_0 \geq |E(x)| f_0\). This concludes with (4.12) the proof of inequality (4.4) and therefore also the proof of Theorem 1.1.

5 Proof of Theorem 1.2

Let \(\ell > 1\) and \(c \geq 0\) be constants, and let \(r(n, \ell) = r = n \mod \ell\). Furthermore, let us for the moment assume \(c > 0\). Wright [Wri61, Thm. 1] showed that for sufficiently large \(n\)

\[
\text{col}_n^\ell(c) = \frac{1}{\ell!} \sum_{d \in D(c, \ell, n)} b(d) E(d) = (1 + o(1)) \frac{1}{\ell!} 2^{e(n, \ell) + p(n, \ell)} \cdot L(r, c),
\]

(5.1)

where

\[
L(r, c) = \sum_{d \in D(c, \ell, n)} 2^{-\frac{1}{2} d^2 \cdot d}.
\]

(5.2)

For \(c \to \infty\), the above sum converges rapidly to a constant [Wri61]; therefore, \(\text{col}_n^\ell(c)\) is a constant fraction of all \(\ell\)-colored graphs. If \(c = 0\) and \(r > 0\), then the sum in (5.2) is empty, i.e., the number of corresponding colored graphs is zero. Finally, if \(c = 0\) and \(r = 0\), the same argument as above yields that we get a constant fraction of all colored graphs. This proves the lower bound stated in Theorem 1.1 for constant \(\ell\).

The upper bound follows immediately from equations (5.1) and (5.2), and the fact mentioned previously: \(L(r, c)\) converges to a constant for \(c \to \infty\) (see [Wri61]).
6 Proof of Theorem 1.3

In this section we are going to show how Theorem 1.3 can be proved. For this, observe that the number of \( \ell \)-colored graphs with a fixed number of edges is given by a very similar expression to the number of all \( \ell \)-colored graphs – the power-of-two term in (1.4) has to be replaced with the binomial coefficient defined in (1.7). But these two functions behave very similar – recall the approximation of the binomial coefficient

\[
\left( \frac{n}{\alpha n} \right) \approx 2^{nH(\alpha)},
\]

where \( H \) denotes the binary entropy function and \( 0 < \alpha < 1 \), see e.g. [Juk01]. Note that \( H(\alpha) \) is a constant, if \( \alpha \) is constant; it is therefore not surprising that the proof carries over if we add the additional restriction on the number of edges. In the sequel we will use the following estimates for fractions of binomial coefficients:

\[
\left( 1 - \frac{\beta + x}{\alpha} \right)^x \leq \left( \frac{\alpha - x}{\beta} \right)^x \leq \left( 1 - \frac{\beta - x}{\alpha - x} \right)^x.
\]

We now outline how the counterparts of the needed lemmas and propositions can be proved; we assume that \( M = cn^2 \) for \( 0 < c < \frac{1}{2} \) throughout.

Proposition 6.1 (Prop. 2.5). Let \( C(c) = \frac{30}{\log(1-2c)} + \frac{1}{2} \) and define

\[
\mathcal{L}'(n, \ell) := \left\{ d \in \mathcal{D} \left( n - \left\lfloor \frac{n}{\ell} \right\rfloor, \ell, n \right) : \frac{1}{2} d^T d \geq C(c) \ell \right\}.
\]

For all integer functions \( \ell(n) \) with \( \ell(n) \to \infty \) for \( n \to \infty \), it holds

\[
\sum_{d \in \mathcal{L}'(n, \ell)} b(d) \tilde{E}(d) = o \left( \left| \text{Col}_{n,M}^{\ell} \right| \right).
\]

Proof. Recall (1.9) and (1.10); Let \( r \) be the remainder of \( n \) and \( \ell \) and \( 0 \) be the vector which has only zero entries except for the first \( r \), which are one. We first derive a lower bound for the number of graphs in \( \text{Col}_{n,M}^{\ell} \) by

\[
\left| \text{Col}_{n,M}^{\ell} \right| \geq \frac{1}{\ell!} \cdot b(0_r) \cdot \left( e(n, \ell) - \frac{r}{2} \right).
\]

First consider the case \( d^T d \geq \frac{2}{\ell} n \log n \). As the sum over all multinomial coefficients is \( \ell^n \), we can estimate with the inequality \( \left( \frac{\alpha - x}{\beta} \right) \leq e^{-x\beta/\alpha} \cdot \left( \frac{\alpha}{\beta} \right) \)

\[
\frac{1}{\ell!} \sum_{d \in \mathcal{L}'(n, \ell)} b(d) \tilde{E}(d) \leq \frac{\ell^n}{\ell!} \left( e(n, \ell) - \frac{2}{\ell} n \log n \right)
\]

\[
\leq \frac{\ell^n}{\ell!} \left( e(n, \ell) - \frac{r}{2} \right) \cdot e^{-s n \log n} = o \left( \frac{1}{\ell!} \left( e(n, \ell) - \frac{r}{2} \right) \right).
\]
Next we consider the case $C(c)\ell \leq d^T d < \frac{\epsilon}{2} n \log n$. Observe that
\[
\sum_{d \in L^{(n, \ell)}} b(d) \tilde{E}(d) \leq \frac{1}{\ell!} b(0.) \left( \frac{e(n, \ell) - \frac{\epsilon}{2}}{M} \right) \cdot \sum_{d^T d \geq 2C(c)\ell} \left( \frac{e(n, \ell) - \frac{\epsilon}{2} \sum_{i=1}^{\ell} d_i^2}{(e(n, \ell) - \frac{\epsilon}{2} M)} \right).
\]
However, the term in the sum becomes with estimate (6.1) and the assumption $d^T d = O(n \log n)$
\[
(1 - (1 + \Theta(\ell^{-1})) 2\epsilon) \frac{1}{2} \sum_{i=1}^{\ell} d_i^2 - \frac{\epsilon}{2}
\]
and the remainder of the proof remains due to the choice of $C(c)$ the same (observe that the additional $\frac{1}{2}$ in the definition of $C(c)$ cancels the $\frac{\epsilon}{2}$, which can be at most $\frac{\epsilon}{2}$).

Now we can mimic the proof of Proposition 2.6 and we obtain for every integer sequence $d$ with the desired properties
\[
|T_g(d)| \leq C(c) \cdot \frac{\ell}{g^2},
\]
where $T_g$ is defined in (2.3) and $C$ is a constant which depends only on $c$.

Let $f_0 = (1 - \epsilon) \sqrt{\log \ell - \log(1 - 2c)}$, where $\epsilon > 0$ is an arbitrarily small constant. The first statement of Theorem 1.3 can be proved with the same mapping as in Section 3. For this, observe that as above, the proof of Proposition 3.1 can be rewritten with different constants for the current case. The remainder of the proof remains the same – in (3.4) we replace the power-of-2-term by
\[
\frac{\tilde{E}(d)}{E(d)} \leq (1 - (1 + \Theta(\ell^{-1})) 2c) \frac{1}{2} (f_0 + 1)^2 + d_k (f_0 + 1) - \sum_{x=0}^{k} d_x + f_0 + 1 - k + \frac{f_0 + 1}{2},
\]
where we used again (6.1). The proof of the first statement then completes with the assumption on the value of $f_0$ and (3.1). Observe that we used the correction factor $(- \log(1 - 2c))^{-\frac{1}{2}}$ in the threshold function in order to compensate for change in the basis function (from $\frac{1}{2}$ to roughly $(1 - 2c)$).

To complete the proof, we show how the calculations in Section 4 can be modified to prove the second part of Theorem 1.3. Here we set $f_0 = (1 + \epsilon) \sqrt{\log \ell - \log(1 - 2c)}$, where $\epsilon > 0$ is again an arbitrarily small constant. As a first step, we show that the absolute value of the "large" entries defined in (4.1) (where we substitute the constant 19 by $2C(c) + 1$) cannot become large, i.e. we can assume
\[
\left| \sum_{i \in T_{f_0}(d)} d_i \right| = O(\ell/f_0),
\]
where the constant in the $O$ depends only on $c$; the proof is essentially the same as in Proposition 4.1. The remainder of the proof and especially the definition of the mapping (4.6) remain the same, as we make the same assumptions. A slight modification has to be made in (4.11) and (4.12), where we simply substitute the $E$-term with the appropriate
binomial coefficient $\tilde{E}(d) = \binom{\binom{n}{d} - \frac{1}{2}d^T d}{M}$. Finally, in Lemma 4.5 we estimate with (6.1) the ratio

$$\frac{\tilde{E}(d)}{E(m(d))} = \frac{\binom{\binom{n}{d} - \frac{1}{2}d^T d}{M}}{\binom{\binom{n}{d} - \frac{1}{2}(m(d))^T m(d)}{M}} \leq (1 - (1 + \Theta(\ell^{-1}))2c)^{\frac{1}{2}}(d^T d - (m(d))^T m(d)),$$

where we set $x = \frac{1}{2}(d^T d - (m(d))^T m(d))$ in (6.1) and used $d^T d = O(\ell)$. The remainder of the proof of the lemma follows the lines of (4.14) and (4.15). The proof of Lemma 4.6 does not change at all, as it is independent of $E$. Finally, the theorem is proved by the same calculations as in (4.18).

References


