On the Chromatic Number of a Random 3-Uniform Hypergraph

Yury A. Demidovich[1(✉)] and Dmitry A. Shabanov[1,2,3]

[1] Moscow Institute of Physics and Technology (National Research University), Dolgoprudnyi, Moscow Region, Russia
{demidovich.yua,dmitry.shabanov}@phystech.edu
[2] Lomonosov Moscow State University, Moscow, Russian Federation
[3] HSE University, Moscow, Russian Federation

Abstract. This paper is devoted to the problem concerning the chromatic number of a random 3-uniform hypergraph. We consider the binomial model $H(n,3,p)$ and show that if $p = p(n)$ decreases fast enough then the chromatic number of $H(n,3,p)$ is concentrated in 2 or 3 consecutive values which can be found explicitly as functions of n and p. This result is derived as an application of the solution of an extremal problem for doubly stochastic matrices.

Keywords: Random hypergraphs · Colorings · Second moment method · Doubly stochastic matrices

1 Introduction

The theory of random graphs and hypergraphs was always in the focus of study in probabilistic combinatorics. Recall that a *hypergraph* H is a pair of sets $H = (V, E)$, where V is a finite set whose elements are called *vertices*, and E is a family of subsets of V that are called edges of the hypergraph. If every edge consists of k vertices then a hypergraph is called *k-uniform*. An *r-coloring* of a vertex set is an arbitrary mapping $f : V \to \{1, \ldots, r\}$. It is said to be *proper* if no edge is monochromatic. The *chromatic number* $\chi(H)$ of a hypergraph H is the minimum number of colors required for a proper coloring of H.

One of main stochastic models of random hypergraphs is the well-known binomial model of a random k-uniform hypergraph $H(n,k,p)$, which can be viewed as the Bernoulli scheme on k-subsets of an n-element set: every subset is included into $H(n,k,p)$ as an edge independently with probability p. We study the asymptotic behaviour of the chromatic number of $H(n,k,p)$ for large n, when k is fixed, and $p = p(n)$ is a function of n.

1.1 Related Work

The chromatic numbers of random graphs and hypergraphs have been intensively studied since the 1970s. For known results concerning $\chi(H(n,k,p))$ in the graph

A. N. Shiryaev et al. (Eds.): ICSM-5 2020, PROMS 371, pp. 190–203, 2021.
https://doi.org/10.1007/978-3-030-83266-7_14

case, $k = 2$, the reader is referred to the recent paper [6]. In the current paper we concentrate only on the case $k \geq 3$. The asymptotics of the chromatic number of $H(n, k, p)$ in the dense case, when the expected number of edges is much larger than the number of vertices, i.e. when $pn^{k-1} \to +\infty$, was obtained by Shamir with coauthors [8,10] and by Krivelevich, Sudakov [7]. But we know much more about the limit distribution of $\chi(H(n, k, p))$ in the sparse case, when the expected number of edges is a linear function of n, i.e. $p = cn/\binom{n}{k}$ and $c > 0$ does not depend on n. Dyer, Frieze and Greenhill [5] proved that in this case $\chi(H(n, k, p))$ is concentrated in two consecutive numbers, moreover, for some values, there is a concentration in exactly one number. For given $c > 0$, let us denote $r_c = \min\{r \in \mathbb{N} : c < r^{k-1} \ln r\}$. Clearly, $c \in [(r_c - 1)^{k-1} \ln(r_c - 1), r_c^{k-1} \ln r_c)$. The authors of [5] established that

- if $c > r_c^{k-1} \ln r_c - \frac{1}{2} \ln r_c$ then
$$\mathsf{P}(\chi(H(n, k, p)) = r_c + 1) \to 1 \text{ as } n \to \infty;$$
- if $c < r_c^{k-1} \ln r_c - \frac{r_c - 1}{r_c} (1 + \ln r_c) - O\left(k^2 r_c^{1-k} \ln r_c\right)$ then
$$\mathsf{P}(\chi(H(n, k, p)) = r_c) \to 1 \text{ as } n \to \infty;$$
- if $c \in [r_c^{k-1} \ln r_c - \frac{r_c - 1}{r_c} (1 + \ln r_c) - O\left(k^2 r_c^{1-k} \ln r_c\right), r_c^{k-1} \ln r_c - \frac{1}{2} \ln r_c]$ then
$$\mathsf{P}(\chi(H(n, k, p)) \in \{r_c, r_c + 1\}) \to 1 \text{ as } n \to \infty. \tag{1}$$

So, in many cases we obtain the exact limit distribution of the chromatic number. Later, the bounds in the third ambiguous case (1) were improved by Ayre, Coja–Oghlan and Greenhill [2] and by Shabanov [9]. They proved that up to the value $r_c^{k-1} \ln r_c - \frac{1}{2} \ln r_c - O(1)$ we still have the chromatic number equal to r_c.

The non-sparse case when $pn^{k-1} \to +\infty$ is not studied so well. Krivelevich and Sudakov showed that if additionally $p \to 0$ then

$$\chi(G(n, p)) \cdot \left(\frac{(k-1)d}{k \ln d}\right)^{-\frac{1}{k-1}} \xrightarrow{\mathsf{P}} 1 \text{ as } n \to +\infty, \tag{2}$$

where $d = p\binom{n-1}{k-1}$. But they did not investigate the concentration effect. The authors of the current paper study the chromatic number of the random hypergraph $H(n, k, p)$ for $k \geq 4$ [4] and proved the following theorem.

Theorem 1 **([4]).** *Let $k \geq 4$ and $\varepsilon > 0$ be fixed. Denote $r_p = r_p(n) = \min\{r \in \mathbb{N} : c < r^{k-1} \ln r\}$ and $c = c(n) = p\binom{n}{k}\frac{1}{n}$. Suppose also that $c \leq n^{\frac{k-1}{2k+4} - \gamma}$ for some positive fixed γ, but $c \to +\infty$ as $n \to \infty$. Then we have the following concentration values for the chromatic number of $H(n, k, p)$:*

1. if $c \leq r_p^{k-1} \ln r_p - \frac{1}{2} \ln r_p - \frac{r_p - 1}{r_p} - O\left(\frac{k^2 \ln r_p}{r_p^{k/3-1}}\right)$ then
$$\mathsf{P}\left(\chi(H(n, k, p)) \in \{r_p, r_p + 1\}\right) \longrightarrow 1 \text{ as } n \to \infty;$$

2. *if* $c > r_p^{k-1} \ln r_p - \frac{1}{2} \ln r_p + \varepsilon$ *for some fixed positive* $\varepsilon > 0$ *then*

$$\mathsf{P}\left(\chi(H(n,k,p)) \in \{r_p+1, r_p+2\}\right) \longrightarrow 1 \text{ as } n \to \infty.$$

3. *finally, if*

$$c \in \left(r_p^{k-1} \ln r_p - \frac{1}{2} \ln r_p - \frac{r_p - 1}{r_p} - O\left(\frac{k^2 \ln r_p}{r_p^{k/3-1}}\right), r_p^{k-1} \ln r_p - \frac{1}{2} \ln r_p + \varepsilon\right]$$

then

$$\mathsf{P}\left(\chi(H(n,k,p)) \in \{r_p, r_p+1, r_p+2\}\right) \longrightarrow 1 \text{ as } n \to \infty.$$

So, we see almost the same picture as in the sparse case but every time we have one more value.

1.2 Extremal Problem for Doubly Stochastic Matrices

The key ingredient of the proof of Theorem 1 is some result concerning the doubly stochastic matrices. Suppose that $r \geq 3$ is an integer. Let $\mathcal{M}_r$ denote the set of $r \times r$ real-valued matrices $M = (m_{ij}, i, j = 1, \ldots, r)$ with nonnegative elements satisfying the following conditions:

$$\sum_{i=1}^{r} m_{ij} = \frac{1}{r}, \text{ for any } j = 1, \ldots, r; \quad \sum_{j=1}^{r} m_{ij} = \frac{1}{r}, \text{ for any } i = 1, \ldots, r. \quad (3)$$

So, for any $M \in \mathcal{M}_r$, the matrix $r \cdot M$ is doubly stochastic. Now, denote the following functions

$$\mathcal{H}_r(M) = -\sum_{i,j=1}^{r} m_{ij} \ln(r \cdot m_{ij}); \quad \mathcal{E}_{r,k}(M) = \ln\left(1 - \frac{2}{r^{k-1}} + \sum_{i,j=1}^{r} m_{ij}^k\right). \quad (4)$$

Denote for $c > 0$, $\mathcal{G}_{c,r,k}(M) = \mathcal{H}_r(M) + c \cdot \mathcal{E}_{r,k}(M)$. It is known that if $c = c(r,k)$ is not too large then $\mathcal{G}_{c,r,k}(M)$ reaches its maximal value at the matrix J_r which has all entries equal to $1/r^2$. The first result of this type was obtained by Achlioptas and Naor in the breakthrough paper [1] for the graph case $k = 2$. Recently, it was improved by Kargaltsev, Shabanov and Shaikheeva [6]. For $k \geq 4$, Shabanov [9] proved the following.

Theorem 2 **([9]).** *There exists an absolute constant* d *such that if* $k \geq 4$, $\max(r,k) > d$ *and*

$$c < r^{k-1} \ln r - \frac{1}{2} \ln r - \frac{r-1}{r} - O(k^2 r^{1-k/3} \ln r) \quad (5)$$

then for any $M \in \mathcal{M}_r$, $\mathcal{G}_{c,r,k}(M) \leq \mathcal{G}_{c,r,k}(J_r)$.

The aim of our work was to generalize Theorems 1 and 2 to the missed case $k = 3$.

1.3 New Results

The first new result of the paper provides the solution for the extremal problem concerning $\mathcal{G}_{c,r,k}$ in the case $k = 3$.

Theorem 3. *There exists an absolute constant r_0 such that if $r \geq r_0$ and*

$$c < r^2 \ln r - \frac{1}{2} \ln r - 1 - r^{-1/6} \tag{6}$$

then for any $M \in \mathcal{M}_r$, $\mathcal{G}_{c,r,3}(M) \leq \mathcal{G}_{c,r,3}(J_r)$.

Note that the obtained result is best possible in the following sense: if for some fixed $\varepsilon > 0$, it holds that $c > r^2 \ln r - \frac{1}{2} \ln r - 1 + \varepsilon$ then for any large enough r, there is $M \in \mathcal{M}_r$ such that $\mathcal{G}_{c,r,3}(M) > \mathcal{G}_{c,r,3}(J_r)$.

Theorem 3 and the second moment method allow us to estimate the chromatic number of the random 3-uniform hypergraph from above when $p = p(n)$ does not decrease too slowly.

Theorem 4. *Let $0 < \gamma < 1/5$ be fixed. Denote $c = c(n) = p\binom{n}{k}\frac{1}{n}$ and $r_p = r_p(n) = \min\{r \in \mathbb{N} : c < r^{k-1} \ln r\}$. Suppose that $c \leq n^{\frac{1}{5}-\gamma}$ and $c \to \infty$ as $n \to \infty$. If*

$$c < r_p^2 \ln r_p - \frac{1}{2} \ln r_p - 1 - r_p^{-1/6}, \tag{7}$$

then

$$\mathsf{P}\left(\chi\left(H\left(n, 3, p\right)\right) \leq r_p + 1\}\right) \longrightarrow 1 \text{ as } n \to \infty.$$

Together with a theorem from [4] (see Theorem 1 in [4]) our second theorem extends Theorem 1 to the missed case $k = 3$. For $c \leq n^{\frac{1}{5}-\gamma}$, we obtain the following values of the chromatic number of a random 3-uniform hypergraph:

1. if $c \leq r_p^2 \ln r_p - \frac{1}{2} \ln r_p - 1 - r_p^{-1/6}$ then
$$\mathsf{P}\left(\chi(H(n, 3, p)) \in \{r_p, r_p + 1\}\right) \longrightarrow 1 \text{ as } n \to \infty;$$
2. if $c > r_p^2 \ln r_p - \frac{1}{2} \ln r_p + \varepsilon$ for some fixed positive $\varepsilon > 0$ then
$$\mathsf{P}\left(\chi(H(n, 3, p)) \in \{r_p + 1, r_p + 2\}\right) \longrightarrow 1 \text{ as } n \to \infty.$$
3. if $c \in \left(r_p^2 \ln r_p - \frac{1}{2} \ln r_p - 1 - r_p^{-1/6}, r_p^2 \ln r_p - \frac{1}{2} \ln r_p + \varepsilon\right]$ then
$$\mathsf{P}\left(\chi(H(n, 3, p)) \in \{r_p, r_p + 1, r_p + 2\}\right) \longrightarrow 1 \text{ as } n \to \infty.$$

In the next section we will prove Theorem 3.

2 Proof of Theorem 3

Note that (3) implies that the total sum of m_{ij} is equal to 1. Since $\mathcal{G}_{c,r,3}(J_r) = \ln r + c \cdot \ln\left(1 - \frac{1}{r^2}\right)^2$, we have

$$\begin{aligned}
&\mathcal{G}_{c,r,3}(J_r) - \mathcal{G}_{c,r,3}(M) = \mathcal{H}_r(J_r) - \mathcal{H}_r(M) - c\left(\mathcal{E}_{r,3}(M) - \mathcal{E}_{r,3}(J_r)\right) \\
&= \ln r + \sum_{i,j=1}^{r} m_{ij}\ln(r \cdot m_{ij}) + c\left(\ln\left(1 - \frac{2}{r^2} + \sum_{i,j=1}^{r} m_{ij}^3\right) - \ln\left(1 - \frac{1}{r^2}\right)^2\right) \\
&= \sum_{i,j=1}^{r} m_{ij}\ln(r^2 \cdot m_{ij}) - c \cdot \ln\left(1 + \frac{\sum_{i,j=1}^{r} m_{ij}^3 - r^{-4}}{\left(1 - \frac{1}{r^2}\right)^2}\right). \qquad (8)
\end{aligned}$$

We need to show that this value is nonnegative for any $M \in \mathcal{M}_r$. In fact, we prove a more precise statement and show that there exist some function $a = a(r) > 0$ such that given the condition (6) the following inequality holds for any $M \in \mathcal{M}_r$,

$$\mathcal{G}_{c,r,3}(J_r) - \mathcal{G}_{c,r,3}(M) \geq a(r) \cdot \sum_{i,j=1}^{r}\left(m_{ij} - \frac{1}{r^2}\right)^2. \qquad (9)$$

Our proof strategy follows the proof of Theorem 2 from [9], however we need to make some changes that allow to extend the result to the case $k = 3$.

2.1 Row Functions

Let us denote $\varepsilon_{ij} = m_{ij} - 1/r^2$. Due to (3), for any $i, j = 1, \ldots, r$, we have

$$\varepsilon_{ij} \in \left[-\frac{1}{r^2}, \frac{1}{r} - \frac{1}{r^2}\right], \quad \sum_{j'=1}^{r} \varepsilon_{ij'} = 0, \quad \sum_{i'=1}^{r} \varepsilon_{i'j} = 0. \qquad (10)$$

Let us also define the following "row" functions: for any $i = 1, \ldots, r$,

$$\begin{aligned}
H_i(M) &= \sum_{j=1}^{r} m_{ij}\ln(r^2 \cdot m_{ij}) = \sum_{j=1}^{r}\left(\frac{1}{r^2} + r^2\varepsilon_{ij}\right)\ln(1 + r^2\varepsilon_{ij}), \\
E_i(M) &= \frac{\sum_{j=1}^{r} m_{ij}^3 - r^{-5}}{\left(1 - \frac{1}{r^2}\right)^2} = \left(1 - \frac{1}{r^2}\right)^{-2}\left(\frac{3}{r^2}\sum_{j=1}^{r}\varepsilon_{ij}^2 + \sum_{j=1}^{r}\varepsilon_{ij}^3\right). \qquad (11)
\end{aligned}$$

Clearly,

$$\mathcal{H}_r(J_r) - \mathcal{H}_r(M) = \sum_{i=1}^{r} H_i(M), \quad \mathcal{E}_{r,3}(M) - \mathcal{E}_{r,3}(J_r) \leq \sum_{i=1}^{r} E_i(M). \qquad (12)$$

Now we are going to estimate the differences $H_i(M) - c \cdot E_i(M)$, $i = 1, \ldots, r$, in various cases. The value $H_i(M) - c \cdot E_i(M)$ depends only on the i-th row of the matrix M. The classification of rows is the following. The row $M_i = (m_{ij}; j = 1, \ldots, r)$ is said to be

1. *central* if

$$\max_{j=1,\ldots,r} m_{ij} < \frac{1}{r} - \frac{1}{r\sqrt{\ln r}};$$

2. *good* if

$$\max_{j=1,\ldots,r} m_{ij} \in \left[\frac{1}{r} - \frac{1}{r\sqrt{\ln r}}, \frac{1}{r} - r^{-11/4}\right];$$

3. *bad* if

$$\max_{j=1,\ldots,r} m_{ij} > \frac{1}{r} - r^{-11/4}.$$

Now let us consider these three types of rows separately. Throughout the paper we use the estimates from [9] whenever it is possible. We also assume that r is large enough.

2.2 Central Rows

Proposition 1. *For any central row M_i,*

$$H_i(M) - c \cdot E_i(M) \geq \frac{r^2}{4} \sum_{j:\varepsilon_{ij}<0} \varepsilon_{ij}^2 + \left(2r\sqrt{\ln r} + O\left(r \ln\ln r\right)\right) \sum_{j:\varepsilon_{ij}\geq 0} \varepsilon_{ij}^2. \quad (13)$$

Proof. First, let us estimate the value $c \cdot E_i(M)$. Since every $\varepsilon_{ij} < \frac{1}{r} - \frac{1}{r^2} - \frac{1}{r\sqrt{\ln r}}$ and $c < r^2 \ln r$, we have

$$c \cdot E_i(M) = c \cdot \left(1 - \frac{1}{r^2}\right)^{-2} \left(\frac{3}{r^2} \sum_{j=1}^{r} \varepsilon_{ij}^2 + \sum_{j=1}^{r} \varepsilon_{ij}^3\right)$$

$$\leq r^2 \ln r \left(1 - \frac{1}{r^2}\right)^{-2} \left(\frac{3}{r^2} \sum_{j:\varepsilon_{ij}<0} \varepsilon_{ij}^2 + \left(\frac{3}{r^2} + \frac{1}{r} - \frac{1}{r^2} - \frac{1}{r\sqrt{\ln r}}\right) \sum_{j:\varepsilon_{ij}>0} \varepsilon_{ij}^2\right)$$

$$\leq 4 \ln r \sum_{j:\varepsilon_{ij}<0} \varepsilon_{ij}^2 + \left(r \ln r - r\sqrt{\ln r} + O(\ln r)\right) \sum_{j:\varepsilon_{ij}>0} \varepsilon_{ij}^2. \quad (14)$$

Now proceed to H_i. We need to estimate the value of $\left(\frac{1}{r^2} + \varepsilon_{ij}\right) \ln(1 + r^2\varepsilon_{ij})$ from below. In [9] it was proved that

1. if $\varepsilon_{ij} < 0$ then (see (34) in [9])

$$\left(\frac{1}{r^2} + \varepsilon_{ij}\right) \ln(1 + r^2\varepsilon_{ij}) \geq \varepsilon_{ij} + \frac{r^2}{2}\varepsilon_{ij}^2; \quad (15)$$

2. if $\varepsilon_{ij} \geq 0$ and $\varepsilon_{ij} \leq \frac{1}{r \ln r} - \frac{1}{r^2}$ then (see (34) in [9])

$$\left(\frac{1}{r^2} + \varepsilon_{ij}\right) \ln(1 + r^2\varepsilon_{ij}) \geq \varepsilon_{ij} + \frac{3r \ln r}{2(1 + 2\ln r/r)}\varepsilon_{ij}^2. \quad (16)$$

Assume that $\varepsilon_{ij} > \frac{1}{r\ln r} - \frac{1}{r^2}$. Then

$$\begin{aligned}\left(\frac{1}{r^2}+\varepsilon_{ij}\right)\ln(1+r^2\varepsilon_{ij}) &\geq \varepsilon_{ij}\ln\left(\frac{r}{\ln r}\right) = \varepsilon_{ij}+\varepsilon_{ij}(\ln r-\ln\ln r-1)\\ &\geq \varepsilon_{ij}+r\varepsilon_{ij}^2\left(1-\frac{1}{r}-\frac{1}{\sqrt{\ln r}}\right)^{-1}(\ln r-\ln\ln r-1)\\ &= \varepsilon_{ij}+r\ln r\cdot\varepsilon_{ij}^2\left(1+\frac{1}{\sqrt{\ln r}}+O\left(\frac{\ln\ln r}{\ln r}\right)\right). \end{aligned} \tag{17}$$

The obtained bounds (15), (16), (17) imply that for all large enough r,

$$H_i(M) \geq \frac{r^2}{2}\sum_{j:\varepsilon_{ij}<0}\varepsilon_{ij}^2 + \left(r\ln r + r\sqrt{\ln r} + O\left(r\ln\ln r\right)\right)\sum_{j:\varepsilon_{ij}\geq 0}\varepsilon_{ij}^2. \tag{18}$$

Together (14) and (18) provide the required estimate:

$$H_i(M) - c\cdot E_i(M) \geq \frac{r^2}{4}\sum_{j:\varepsilon_{ij}<0}\varepsilon_{ij}^2 + \left(2r\sqrt{\ln r} + O\left(r\ln\ln r\right)\right)\sum_{j:\varepsilon_{ij}\geq 0}\varepsilon_{ij}^2.$$

2.3 Good Rows

For good or bad row M_i, its maximal element is very close to $\frac{1}{r}$. So, it is convenient to define the value

$$m_i = \frac{1}{r} - \max_{j=1,\dots,r} m_{ij}. \tag{19}$$

The inequality (28) from [9] estimates the value $H_i(M)$ in terms of the value m_i as follows:

$$H_i(M) \geq \frac{\ln r}{r} + m_i\ln m_i + m_i\ln\left(\frac{r}{r-1}\right) - m_i. \tag{20}$$

Note that these bounds hold for any row. We will use it very often in the remaining proof.

Proposition 2. *For any good row M_i,*

$$H_i(M) - c\cdot E_i(M) \geq \frac{1}{4}r^{-11/4}\ln r. \tag{21}$$

Proof. Let us estimate $c\cdot E_i(M)$. For a good row, we have $m_i \in [r^{-11/4}, 1/r\sqrt{\ln r}]$, so $m_i = o(r^{-1})$ and $m_i = \omega(r^{-3})$. Suppose that $m_{ij_0} = 1/r - m_i$ is the maximal element of M_i. Then (3) implies that $\sum_{j\neq j_0} m_{ij} = 1/r - m_{ij_0} = m_i$. Thus,

$$c \cdot E_i(M) = c \cdot \left(1 - \frac{1}{r^2}\right)^{-2} \left(\sum_{j=1}^{r} m_{ij}^3 - r^{-5}\right) \le c \cdot \left(1 - \frac{1}{r^2}\right)^{-2} \sum_{j=1}^{r} m_{ij}^3$$
$$= c \cdot \left(1 - \frac{1}{r^2}\right)^{-2} \left(\left(\frac{1}{r} - m_i\right)^3 + \sum_{j \ne j_0}^{r} m_{ij}^3\right)$$
$$\le c \cdot \left(1 - \frac{1}{r^2}\right)^{-2} \left(\frac{1}{r^3} - \frac{3m_i}{r^2} + \frac{3m_i^2}{r}\right).$$

Here we use the fact that $\sum_{j \ne j_0}^{r} m_{ij}^3 \le m_i^3$. Since $m_i = o(r)$ and $c < r^2 \ln r$ we obtain that

$$c \cdot E_i(M) \le c \cdot \left(\frac{1}{r^3} - \frac{3m_i}{r^2} + \frac{3m_i^2}{r} + O\left(\frac{1}{r^5}\right)\right)$$
$$\le \frac{\ln r}{r} - 3m_i \ln r(1 + o(1)) + O\left(\frac{\ln r}{r^3}\right). \quad (22)$$

The general estimate (20) and the condition $m_i \ge r^{-11/4}$ imply that

$$H_i(M) \ge \frac{\ln r}{r} + m_i \ln m_i(1 + o(1)) \ge \frac{\ln r}{r} - \frac{11}{4} m_i \ln r(1 + o(1)). \quad (23)$$

The bounds (22) and (23) provide the required inequality:

$$H_i(M) - c \cdot E_i(M) \ge \frac{1}{4} m_i \ln r(1 + o(1)) + O\left(\frac{\ln r}{r^3}\right)$$
$$\ge \frac{1}{8} m_i \ln r \ge \frac{1}{4} r^{-11/4} \ln r.$$

2.4 Bad Rows

Now it is time to deal with bad rows. Recall that in every bad row M_i there is an index j_0 such that $m_{ij_0} = \max_{j=1,\dots,r} > \frac{1}{r} - r^{-11/4}$. The main problem here is that in this case the difference $H_i(M) - c \cdot E_i(M)$ can be negative. For $k \ge 4$, this negative value can be compensated by the bounds (13), (21) for central and good rows, if there is at least one non-bad row (see [9]). So, it remains to consider the case when all the rows are bad. Unfortunately, this is not the way for $k = 3$. Here we had to consider all the bad rows simultaneously.

Let $D \subset \{1, \dots, r\}$ denote the set of indices of the bad rows in M. Introduce the following values:

$$H_D(M) = \sum_{i \in D} H_i(M), \quad E_D(M) = \ln\left(1 + \sum_{i \in D} E_i(M)\right). \quad (24)$$

The following statement estimates their difference.

Proposition 3. *Under the condition* (6) *the following inequality holds:*

$$H_D(M) - c \cdot E_D(M) \geq -\frac{|D| \ln r}{2r^3} + \frac{|D|^2 \ln r}{2r^4} + \frac{|D|}{r^{19/6}} + O\left(r^{-3}\right). \tag{25}$$

Proof. For $H_i(M)$, $i \in D$, we have the bound (20). So, it remains to estimate $c \cdot \mathcal{E}_D(M)$. Again, for any $i \in D$, we consider the maximal element of M_i,

$$m_{ij_0(i)} = \max_{j=1,\dots,r} m_{ij} = \frac{1}{r} - m_i,$$

where $m_i \in [0, r^{-11/4}]$. Using (11) we get

$$\begin{aligned} \sum_{i \in B} E_i(M) &= \left(1 - \frac{1}{r^2}\right)^{-2} \sum_{i \in D} \left(\sum_{j=1}^{r} m_{ij}^3 - \frac{1}{r^5} \right) \\ &= \left(1 - \frac{1}{r}\right)^{-2} \sum_{i \in D} \left(\left(\frac{1}{r} - m_i\right)^3 + \sum_{j \neq j_0(i)} m_{ij}^3 - \frac{1}{r^5} \right). \end{aligned}$$

Note that $\sum_{j \neq j_0(i)} m_{ij}^3 \leq m_i^3 = O(r^{-33/4})$. Therefore,

$$\begin{aligned} \sum_{i \in B} \mathcal{E}_i(M) &= \left(1 - \frac{1}{r^2}\right)^{-2} \sum_{i \in D} \left(\frac{1}{r^3} - \frac{3m_i}{r^2} + \frac{3m_i^2}{r} - \frac{1}{r^5} + O(r^{-33/4}) \right) \\ &= \sum_{i \in D} \left(\frac{1}{r^3} - \frac{3m_i}{r^2} - \frac{1}{r^5} + O(r^{-13/2}) \right) \left(1 + \frac{2}{r^2} + O(r^{-4}) \right). \end{aligned}$$

Now, we have

$$\left(\frac{1}{r^3} - \frac{3m_i}{r^2} - \frac{1}{r^5} + O(r^{-13/2}) \right) \left(\frac{2}{r^2} + O(r^{-4}) \right) = \frac{2}{r^5} + O(r^{-27/4}).$$

Consequently,

$$\begin{aligned} \sum_{i \in D} E_i(M) &= \sum_{i \in D} \left(\frac{1}{r^3} - \frac{3m_i}{r^2} + \frac{1}{r^5} + O(r^{-13/2}) \right) \\ &= \frac{|D|}{r^3} - \frac{3}{r^2} \sum_{i \in D} m_i + \frac{|D|}{r^5} + O(|D| r^{-13/2}). \end{aligned} \tag{26}$$

Now, we want to estimate the square of this expression. Since $|D| \leq r$, the last three summands have the order $O(r^{-15/4})$. Therefore, (26) implies that

$$\left(\sum_{i \in D} E_i(M) \right)^2 = \frac{|D|^2}{r^6} + O(r^{-23/4}), \quad \left(\sum_{i \in D} E_i(M) \right)^3 = O(r^{-6}). \tag{27}$$

Now we are ready to estimate $c \cdot E_D(M)$. Using (26), (27) and applying Taylor expansion for the logarithm, we obtain

$$
\begin{aligned}
c \cdot E_D(M) &= c \cdot \ln\left(1 + \sum_{i \in D} E_i(M)\right) \\
&= c \cdot \left(\sum_{i \in D} E_i(M) - \frac{1}{2}\left(\sum_{i \in D} E_i(M)\right)^2 + O\left(\left(\sum_{i \in D} E_i(M)\right)^3\right)\right) \\
&= c \cdot \left(\frac{|D|}{r^3} - \frac{3}{r^2}\sum_{i \in D} m_i + \frac{|D|}{r^5} + O(|D| r^{-13/2}) - \frac{|D|^2}{2r^6} + O(r^{-23/4}) + O(r^{-6})\right) \\
&= c \cdot \left(\frac{|D|}{r^3} + \frac{|D|}{r^5} - \frac{|D|^2}{2r^6} - \frac{3}{r^2}\sum_{i \in D} m_i + O(r^{-11/2})\right).
\end{aligned}
$$

The condition (6) states that $c < r^2 \ln r - \frac{1}{2}\ln r - 1 - r^{-1/6}$. Thus,

$$
\begin{aligned}
c \cdot E_D(M) &< \left(r^2 \ln r - \frac{1}{2}\ln r - 1 - r^{-1/6}\right) \\
&\times \left(\frac{|D|}{r^3} + \frac{|D|}{r^5} - \frac{|D|^2}{2r^6} - \frac{3}{r^2}\sum_{i \in D} m_i + O(r^{-11/2})\right) \\
&= \frac{|D| \ln r}{r} + \frac{|D| \ln r}{r^3} - \frac{|D|^2 \ln r}{2r^4} - (3 \ln r)\sum_{i \in D} m_i + O\left(\frac{\ln r}{r^{7/2}}\right) \\
&- \frac{|D| \ln r}{2r^3} - \frac{|D|}{r^3} - \frac{|D|}{r^{19/6}} + O\left(\ln r \cdot r^{-15/4}\right) \\
&= \frac{|D| \ln r}{r} + \frac{|D| \ln r}{2r^3} - \frac{|D|^2 \ln r}{2r^4} - \frac{|D|}{r^3} - \frac{|D|}{r^{19/6}} - (3 \ln r)\sum_{i \in D} m_i + O\left(\frac{\ln r}{r^{7/2}}\right).
\end{aligned}
\tag{28}
$$

Let us complete the proof. Due to (20) we have the following lower bound for $H_D(M)$:

$$
H_D(M) \geq \frac{|D| \ln r}{r} + \sum_{i \in D}\left[m_i \ln m_i + m_i \ln\left(\frac{r}{r-1}\right) - m_i\right].
$$

Using (28), we obtain that

$$
\begin{aligned}
H_D(M) - c \cdot E_D(M) \geq & -\frac{|D| \ln r}{2r^3} + \frac{|D|^2 \ln r}{2r^4} + \frac{|D|}{r^3} + \frac{|D|}{r^{19/6}} + O\left(\frac{\ln r}{r^{7/2}}\right) \\
& + \sum_{i \in D}\left[m_i \ln m_i + m_i \ln\left(\frac{r}{r-1}\right) - m_i + 3m_i \ln r\right].
\end{aligned}
$$

The function $f(x) = x \ln x + x \ln\left(\frac{r}{r-1}\right) - x + 3x \ln r$ is minimized when $x = (r-1)/r^4 \in [0, r^{-11/4})$. So, the minimal value is attained when $m_i = (r-1)/r^4$ for any $i \in D$. Hence,

$$\sum_{i\in D}\left[m_i\ln m_i + m_i\ln\left(\frac{r}{r-1}\right) - m_i + 3m_i\ln r\right]$$
$$\geq \sum_{i\in D}\left(\frac{r-1}{r^4}\left(\ln\left(\frac{r-1}{r^4}\right)+\ln\left(\frac{r}{r-1}\right)-1+3\ln r\right)\right)$$
$$= -\sum_{i\in D}\frac{r-1}{r^4} = -\frac{(r-1)|D|}{r^4} = -\frac{|D|}{r^3}+O(r^{-3}).$$

This finally implies the required inequality

$$H_D(M) - c\cdot E_D(M) \geq -\frac{|D|\ln r}{2r^3} + \frac{|D|^2\ln r}{2r^4} + \frac{|D|}{r^{19/6}} + O\left(r^{-3}\right).$$

2.5 Completion of the Proof

It remains to summarize the obtained information. Now everything depends on the number of bad rows $|D|$ in the matrix M. Let $C, G \subset \{1,\ldots,r\}$ denote the set of indices of central and good rows, respectively. Recall (see (8), (11), (24)) that

$$\mathcal{G}_{c,r,3}(J_r) - \mathcal{G}_{c,r,3}(M) = \sum_{i=1}^{r} H_i(M) - c\cdot\ln\left(1+\sum_{i=1}^{r}E_i(M)\right)$$
$$\geq \sum_{i\in C\cup G}\left(H_i(M) - c\cdot E_i(M)\right) + H_D(M) - c\cdot E_D(M). \tag{29}$$

Note that it is sufficient to show that $\mathcal{G}_{c,r,3}(J_r) - \mathcal{G}_{c,r,3}(M) \geq b$ for some $b = b(r) > 0$. This also implies the required inequality (9), because $\sum_{i,j=1}^{r}(m_{ij} - r^{-2})^2 < 1$.

Let us consider the following four cases.

1. If $|D| = 0$ then (9) follows from (13) and (21).
2. If $|D| \geq r - \frac{r^{5/6}}{\ln r}$ then (25) implies that the total contribution of bad rows is positive. Indeed, for large enough r,

$$H_D(M) - c\cdot E_D(M) \geq \frac{|D|\ln r}{2r^3}\left(\frac{|D|}{r} - 1\right) + \frac{|D|}{r^{19/6}} + O\left(r^{-3}\right)$$
$$\geq \frac{|D|\ln r}{2r^3}\left(-\frac{1}{r^{1/6}\ln r}\right) + \frac{|D|}{r^{19/6}} + O\left(r^{-3}\right) = \frac{|D|}{2r^{19/6}} + O\left(r^{-3}\right) > \frac{1}{3}r^{-13/6}.$$

   Hence, again (9) follows from (13) and (21).
3. Suppose $|D| < r - \frac{r^{5/6}}{\ln r}$, but $|G| \geq r^{4/5}$. Then

$$H_D(M) - c\cdot E_D(M) \geq -\frac{|D|\ln r}{2r^3} + O\left(r^{-3}\right) \geq -\frac{\ln r}{2r^2} + O\left(r^{-3}\right) > -\frac{\ln r}{r^2}.$$

The inequality (29) and the obtained bounds (13), (21) imply that for large enough r,

$$\mathcal{G}_{c,r,3}(J_r) - \mathcal{G}_{c,r,3}(M) \geq \sum_{i \in G} (H_i(M) - c \cdot E_i(M)) + H_D(M) - c \cdot E_D(M)$$
$$\geq r^{4/5} \frac{1}{4} r^{-11/4} \ln r - \frac{\ln r}{r^2} \geq \frac{1}{5} r^{-39/20} \ln r.$$

4. It remains to consider the situation when $|G| < r^{4/5}$ and $0 < |D| < r - \frac{r^{5/6}}{2 \ln r}$. In this case there is at least $\frac{r^{5/6}}{\ln r} - r^{4/5}$ central rows in M. Suppose that $i_1, i_2 \in D$ are two indices corresponding to bad rows. Recall that any bad row has an element greater than $r^{-1} - r^{-11/4}$. Suppose that $m_{i_1 j_1}$ and $m_{i_2 j_2}$ are both greater than $1/r - r^{-11/4}$. Then it is straightforward to verify that the double stochastic property (3) implies that $j_1 \neq j_2$. So, the maximal elements of bad rows should be in different columns of matrix M. Without loss of generality we may assume that these elements are diagonal, i.e. for any $i \in D$,

$$m_{ii} = \max_{j=1,\ldots,r} m_{ij} > \frac{1}{r} - r^{-11/4}.$$

Recall the notation $\varepsilon_{ij} = m_{ij} - \frac{1}{r^2}$. If $j \in D$ then due to (10) we obtain that

$$\sum_{i \in C} \varepsilon_{ij} = -\varepsilon_{jj} - \sum_{i \in G \cup D; i \neq j} \varepsilon_{ij}.$$

We know that $\varepsilon_{jj} = m_{jj} - \frac{1}{r^2} \geq \frac{1}{r} - \frac{1}{r^2} - r^{-11/4}$ and any other element is at least $-r^{-2}$. Hence,

$$\sum_{i \in C} \varepsilon_{ij} \leq -\frac{1}{r} + \frac{1}{r^2} + r^{-11/4} + \frac{1}{r^2} \left(|G| + |D| \right).$$

In our case $|G| < r^{4/5}$ and $|D| < r - \frac{r^{5/6}}{\ln r}$, so, we get

$$\sum_{i \in C} \varepsilon_{ij} \leq -\frac{1}{r} + \frac{1}{r^2} + r^{-11/4} + \frac{1}{r^2} \left(r^{4/5} + r - \frac{r^{5/6}}{\ln r} \right)$$
$$= -\frac{r^{-7/6}}{\ln r}(1 + o(1)) < -\frac{r^{-7/6}}{2 \ln r} < 0.$$

Hence, the sum over all negative summands is also less than $-\frac{r^{-7/6}}{2 \ln r}$:

$$\sum_{i \in C : \varepsilon_{ij} < 0} \varepsilon_{ij} \leq -\frac{r^{-7/6}}{2 \ln r}.$$

By Cauchy–Schwarz inequality

$$\sum_{i \in C : \varepsilon_{ij} < 0} \varepsilon_{ij}^2 \geq \frac{1}{r} \left(\frac{r^{-7/6}}{2 \ln r} \right)^2 = \frac{r^{-10/3}}{4 (\ln r)^2}. \tag{30}$$

Finally, by using (13), (21), (25), (29) and (30) we establish the required estimate

$$\begin{aligned}
\mathcal{G}_{c,r,3}(J_r) - \mathcal{G}_{c,r,3}(M) &\geq \sum_{i \in C} (H_i(M) - c \cdot E_i(M)) + H_B(M) - c \cdot E_B(M) \\
&\geq \sum_{i \in C} \frac{r^2}{4} \sum_{j : \varepsilon_{ij} < 0} \varepsilon_{ij}^2 - \frac{|D| \ln r}{2r^3} + \frac{|D|^2 \ln r}{2r^4} + \frac{|D|}{r^{19/6}} + O\left(r^{-3}\right) \\
&\geq \frac{r^2}{4} \sum_{i \in C} \sum_{j : \varepsilon_{ij} < 0} \varepsilon_{ij}^2 - \frac{|D| \ln r}{2r^3} + O\left(r^{-3}\right) \\
&\geq \frac{r^2}{4} \sum_{i \in C} \sum_{j \in D : \varepsilon_{ij} < 0} \varepsilon_{ij}^2 - \frac{|D| \ln r}{2r^3} + O\left(r^{-3}\right) \\
&= \frac{r^2}{4} \sum_{j \in D} \sum_{i \in C : \varepsilon_{ij} < 0} \varepsilon_{ij}^2 - \frac{|D| \ln r}{2r^3} + O\left(r^{-3}\right) \\
&\geq \frac{r^2}{4} |D| \cdot \frac{r^{-10/3}}{4(\ln r)^2} - \frac{|D| \ln r}{2r^3} + O\left(r^{-3}\right) \\
&= |D| \cdot \frac{r^{-4/3}}{16(\ln r)^2} (1 + o(1)) + O\left(r^{-3}\right).
\end{aligned}$$

Since $|D| \geq 1$, the obtained value is at least $\frac{r^{-4/3}}{16 \ln^2 r}(1 + o(1))$.

Theorem 3 is proved.

3 Sketch of the Proof of Theorem 4

In the last section we give a short sketch of the proof of Theorem 4. We just follow the proof of Theorem 1 which can be found in [4]. The general scheme was first developed by Coja-Oghlan, Panagiotou and Steger [3] in the case of graphs.

First of all, we have to estimate from below the probability that the chromatic number of the random hypergraph does not exceed r_p. By using the second moment method and Theorem 3 we prove the following lemma.

Lemma 1. *Suppose $pn^2 \to +\infty$ and $p \to 0$ as $n \to +\infty$. If the condition (7) holds then for all large enough n,*

$$\mathsf{P}\left(\chi(H(n, 3, p)) \leq r_p\right) \geq n^{-2r_p^2}.$$

Lemma 1 helps to estimate the proportion of vertices of our hypergraph that can be properly colored with r_p colors. Let V_n denote the set of vertices of $H(n, 3, p)$. The following statement is true.

Lemma 2. *Suppose that the conditions of Lemma 1 hold. Then with probability tending to 1, there exists a vertex subset U_0 with size at most $2r_p\sqrt{n \ln n}$ such that the chromatic number of the subhypergraph induced by $H(n, 3, p)$ on $V_n \setminus U_0$ does not exceed r_p.*

Finally, we need to estimate the number of edges in any small induced subhypergraph in $H(n,k,p)$. Here we prove the following.

Lemma 3. *Suppose the conditions of Theorem 4 hold. Suppose that fixed $\delta = \delta(\gamma) > 0$ satisfies the inequality*

$$\delta < \frac{25\gamma}{18 + 60\gamma}.$$

Then with probability tending to 1, any vertex subset U in $H(n,3,p)$ with size at most $r_p\sqrt{n}(\ln n)$ has at most $\left(\frac{2}{3} - \delta\right)|U|$ edges inside.

Theorem 4 is easily deduced from Lemmas 1–3. So, we know that with probability tending to 1, almost whole hypergraph can be properly colored with r_p colors. The remained small vertex subset U has size at most $2r_p\sqrt{n\ln n}$. Therefore, there is a small number of edges inside U.

Now we can increase this set U in such a way that there are no edges in the set of neighbors of the extended set U' and the size of U' is still less than $r_p\sqrt{n}(\ln n)$. By Lemma 3 the set U' can be properly colored with colors $\{1, 2\}$, its neighborhood W—with reserved color $r_p + 1$ and the remaining subset $V_n \setminus (U \cup W)$—with colors $\{1, 2, \ldots, r_p\}$. Clearly, this is a proper coloring of $H(n,3,p)$ with $r_p + 1$ colors. Theorem 4 is proved.

Acknowledgments. Research of Yu.A. Demidovich was supported by the RFBR, project number 19-31-90016. Research of D.A. Shabanov was supported by the grant of the President of Russian Federation no. MD-1562.2020.1

References

1. Achlioptas, D., Naor, A.: The two possible values of the chromatic number of a random graph. Ann. Math. **162**(3), 1335–1351 (2005)
2. Ayre, P., Coja-Oghlan, A., Greenhill, C.: Hypergraph coloring up to condensation. Random Struct. Algorithms **54**(4), 615–652 (2019)
3. Coja-Oghlan, A., Panagiotou, K., Steger, A.: On the chromatic number of random graphs. J. Combin. Theory Ser. B **98**, 980–993 (2008)
4. Demidovich, Y.A., Shabanov, D.A.: On the chromatic number of random hypergraphs. Dokl. Math. **102**(2), 380–383 (2020)
5. Dyer, M., Frieze, A., Greenhill, C.: On the chromatic number of a random hypergraph. J. Combin. Theory Ser. B **113**, 68–122 (2015)
6. Kargaltsev, S., Shabanov, D., Shaikheeva, T.: Two values of the chromatic number of a sparse random graph. Acta Math. Univ. Comenian. **88**(3), 849–854 (2019)
7. Krivelevich, M., Sudakov, B.: The chromatic numbers of random hypergraphs. Random Struct. Algorithms **12**(4), 381–403 (1998)
8. Schmidt-Pruzan, J., Shamir, E., Upfal, E.: Random hypergraph coloring algorithms and the weak chromatic number. J. Graph Theory **8**, 347–362 (1985)
9. Shabanov, D.A.: Estimating the r-colorability threshold for a random hypergraph. Discret. Appl. Math. **282**, 168–183 (2020)
10. Shamir, E.: Chromatic number of random hypergraphs and associated graphs. Adv. Comput. Res. **5**, 127–142 (1989)