

On the chromatic index of random uniform hypergraphs

Valentas Kurauskas

Faculty of Mathematics and Informatics
Vilnius University, Vilnius, Lithuania

Katarzyna Rybarczyk

Faculty of Mathematics and Computer Science,
Adam Mickiewicz University, Poznań, Poland

Abstract

Let $\mathbb{H}^{(k)}(n, N)$, where $k \geq 2$, be a random hypergraph on the vertex set $[n] = \{1, 2, \dots, n\}$ with N edges drawn independently with replacement from all subsets of $[n]$ of size k . For $\bar{d} = kN/n$ and any $\varepsilon > 0$ we show that if $k = o(\ln(\bar{d}/\ln n))$ and $k = o(\ln(n/\ln \bar{d}))$, then with probability $1 - o(1)$ a random greedy algorithm produces a proper edge-colouring of $\mathbb{H}^{(k)}(n, N)$ with at most $\bar{d}(1 + \varepsilon)$ colours. This yields the asymptotic chromatic number of the corresponding uniform random intersection graph.

keywords: random uniform hypergraph, chromatic index, random intersection graph, greedy algorithm.

1 Introduction

A hypergraph is a pair $H = (V, E)$, where $V = V(H)$ is a set of vertices and $E = E(H)$ is a family of subsets of V , called *edges* (we will allow multiple edges). H is *k-uniform* if all of its edges are of size k . The *chromatic index* of H , denoted $\chi'(H)$, is the smallest number of colours needed to colour its edges so that no two intersecting edges share the same colour (edge-colourings that have the last property are called *proper*). Equivalently, $\chi'(H)$ is the smallest such number t such that the edges of H can be partitioned into t matchings. We consider the chromatic index of the random hypergraph $\mathbb{H}^{(k)}(n, N)$.

This problem is related to the generalisation of Vizing's theorem for hypergraphs. Some related results and conjectures may be found in [1] and [5]. Given a hypergraph H , let $\deg_H(x)$ denote the *degree* of a vertex x , that is $\deg_H(x) = |\{e : e \in E(H), x \in e\}|$ (in the case H contains repeated edges, we count the degree with multiplicity). Let

$$D(H) = \max_{x \in V(H)} \deg_H(x) \quad \text{and} \quad d(H) = \min_{x \in V(H)} \deg_H(x).$$

Also let $C(H) = \max_{x \neq y} |\{e \in E(G) : x, y \in e\}|$. Vizing's theorem [24] states that for any 2-uniform hypergraph without loops and multiple edges, $\chi'(H)$ is either $D(H)$ or $D(H) + 1$. Obviously, for any hypergraph

$$d(H) \leq D(H) \leq \chi'(H).$$

In 1989 Pippenger and Spencer [22] proved that Vizing's theorem may be extended in a certain sense to uniform hypergraphs in which degrees are concentrated around one value and $C(H)$ is small compared to $D(H)$.

Theorem 1.1 (Pippenger and Spencer [22]) *For every k and every $\varepsilon > 0$ there exist $\delta > 0$ and n_0 such that if H is a k -uniform hypergraph on $n \geq n_0$ vertices satisfying $d(H) \geq (1 - \delta)D(H)$ and $C(H) \leq \delta D(H)$ then*

$$\chi'(H) \leq (1 + \varepsilon)D(H).$$

In this article we are interested in asymptotic results concerning an algorithmic version of the theorem for random hypergraphs. By $\binom{[n]}{k}$ we will denote the family of all k -element subsets of $[n]$. It will also be convenient for us to write $a = b \pm c$ for $a \in [b - c, b + c]$.

Theorem 1.1 implies that $\chi'(\mathbb{H}^{(k)}(n, N)) = \bar{d}(1 \pm \varepsilon)$ when the degrees of $\mathbb{H}^{(k)}(n, N)$ are large and close to their mean $\bar{d} = \frac{kN}{n}$. This happens w.h.p. (with probability tending to 1 as $n \rightarrow \infty$) when $\ln n = o(\bar{d})$ for fixed k . Motivated by the problem of determining the chromatic number of random intersection graphs (see below), we ask whether a similar result holds for random uniform hypergraphs even when the set size k increases with n . The main result of the paper is the following theorem.

Theorem 1.2 *For any $\varepsilon > 0$ there is a constant $c_\varepsilon > 0$, such that the following holds. Let $\mathbb{H}^{(k)}(n, N)$, where $k, n, N \geq 2$, be a random hypergraph on the vertex set $[n]$ with N edges of size k drawn independently with replacement from the set $\binom{[n]}{k}$. Write $\bar{d} = \frac{kN}{n}$. Suppose*

$$k \leq c_\varepsilon \ln \left(\frac{n}{\ln \bar{d}} \right) \quad \text{and} \quad k \leq c_\varepsilon \ln \left(\frac{\bar{d}}{\ln n} \right). \quad (1)$$

Then a greedy algorithm properly colours all edges of $\mathbb{H}^{(k)}(n, N)$ with at most $\lceil \bar{d}(1 + \varepsilon) \rceil$ colours with probability at least $1 - \frac{2}{n} - \frac{2}{a}$.

The algorithm mentioned in Theorem 1.2 is a simple polynomial-time algorithm which for each edge just selects a random available colour. Its description is given in Section 2. Our proof is an application of the differential equations method [11, 25], and it differs from the method used in [22]. It should be pointed out that by a simple coupling argument analogous theorems follow for random hypergraphs with independent edges and N edges chosen without replacement.

As mentioned above, the motivation for our research was studying the chromatic number of uniform random intersection graphs. The random intersection graph model was introduced by Karoński, Scheinerman and Singer-Cohen [17] and further generalised by Godehardt and Jaworski [14]. For a survey of known results concerning properties of the model we refer the reader to two upcoming papers [6] and [7].

Let N , n and k be positive integers. Moreover, let $V = \{v_1, \dots, v_N\}$ and $\mathcal{W} = [n]$ be disjoint sets. By a *uniform random intersection graph* $\mathcal{G}(N, n, k)$ we mean a graph on the vertex set V in which each vertex $v \in V$ chooses a set S_v independently and uniformly at random from all k -element subsets of the set \mathcal{W} . Two vertices $v, v' \in V$ are connected by an edge in $\mathcal{G}(N, n, k)$ if and only if $S_v \cap S_{v'} \neq \emptyset$. Naturally, $\mathcal{G}(N, n, k)$ is a line graph of the hypergraph $\mathbb{H}^{(k)}(n, N)$. Therefore $\chi'(\mathbb{H}^{(k)}(n, N)) = \chi(\mathcal{G}(N, n, k))$, and Theorem 1.2 is immediately applicable. Results concerning the chromatic number of other models of random intersection graphs might be found in [4] and [21].

The chromatic number of any graph G on N vertices is related to its independence (stability) number $\alpha(G)$ and to the size of the largest clique $\omega(G)$ by the simple inequalities

$$\chi(G) \geq \frac{N}{\alpha(G)} \quad \text{and} \quad \chi(G) \geq \omega(G). \quad (2)$$

In a series of papers [9, 10, 12, 18, 19] it was shown that for $G = G(N, p)$, the Erdős–Rényi random graph with independent edges, the first inequality of (2) is nearly an equality w.h.p. when $Np \rightarrow \infty$. The independence number of $\mathcal{G}(N, n, k)$ was studied in [23], where it was shown that w.h.p. the greedy algorithm constructs an independent set of the optimal size $\frac{n}{k}(1 \pm \varepsilon)$ whenever the second inequality of (1) holds. Also, the main result of [2] implies that w.h.p. $\omega(\mathcal{G}(N, n, k)) = \frac{kN}{n}(1 \pm \varepsilon)$ whenever $k = o(n^{1/3})$ and $\frac{kN}{n} \rightarrow \infty$ not too slowly. Recently, the results of [2] have been extended to an even

wider range of parameters, see [3, 15]. Our result shows that subject to the assumptions of Theorem 1.2, both inequalities of (2) are nearly equalities w.h.p. when $G = \mathcal{G}(N, n, k)$.

Note that (1) can be rewritten as

$$\frac{n}{k} e^{k/c_\varepsilon} \ln n \leq N \leq \frac{n}{k} e^{ne^{-k/c_\varepsilon}}, \quad (3)$$

which is possible for all n large enough, as long as, e.g., $k \leq c_\varepsilon(\ln n - 3 \ln \ln n)$. For such k the upper bound for N , equivalently the left inequality of (1), is very generous: it allows $N \geq \binom{n}{k}$. Also, the results of [18] indicate that the second constraint should be not far from the best possible even for larger k . Naturally, it would be interesting to determine $\chi(\mathcal{G}(N, n, k))$ or, equivalently, $\chi'(\mathbb{H}^{(k)}(n, N))$ for other ranges of parameters, including those covered in [2, 8, 23].

Finally, let us make the following observation. The greedy colouring algorithms for $G(N, p)$ usually use about twice the chromatic number of colours (see, e.g., [13] or Section 7.2 of [16]). Our randomised greedy algorithm (see also [4] and [21]) produces an asymptotically optimal colouring of $\mathcal{G}(N, n, k)$ w.h.p. One could ask whether such an algorithm also exists for other ranges of k and N , perhaps in all cases where the clique number of $\mathcal{G}(N, n, k)$ is w.h.p. approximately equal to its chromatic number.

2 The algorithm

The edges of $\mathbb{H}^{(k)}(n, N)$ can be represented as a sequence e_1, \dots, e_N of independent identically distributed sets, where each set is selected uniformly at random from $\binom{[n]}{k}$. We imagine e_1, \dots, e_N being added to the hypergraph one by one and coloured by a random valid colour. More precisely, we fix a positive integer q , the number of possible colours. The set e_1 is coloured with a uniformly random colour from $[q]$.

The random hypergraph obtained by adding and colouring the first i edges is denoted by $H(i)$. We always consider $H(i)$ together with the (random) colouring of its edges $C(i)$. For any set $S \subseteq [n]$ let $\mathcal{M}_S(i)$ denote the set of all colours not used on edges of $H(i)$ incident to the vertices in S .

Let $M_S(i) = |\mathcal{M}_S(i)|$. Thus, $M_S(i)$ is the number of ‘‘available’’ colours for the set S after the step i .

For $i \geq 1$ the edge $e = e_{i+1}$ is coloured with a uniformly random colour $c(e)$ from the subset $\mathcal{M}_e(i)$ (given $\mathcal{M}_e(i)$, the colour $c(e)$ is conditionally independent of $H(i)$ and $C(i)$). If the set $\mathcal{M}_e(i)$ is empty for the random edge $e = e_{i+1}$, then the colour of e remains unassigned.

For a colour $c \in [q]$, let $L_c(i)$ be the number of vertices from $[n]$ which do not belong to an edge coloured c in the hypergraph $H(i)$. In the beginning we have $M_e(0) = q$ for each k -element subset of $[n]$ and $L_c(0) = n$ for each $c \in [q]$. We will prove Theorem 1.2 by showing that for any $\varepsilon > 0$ the above algorithm with a large enough probability succeeds to colour every edge e_i for $i \leq N(1 - \varepsilon)$ when $q = \lceil \frac{kN}{n} \rceil$. From this, it follows that the same algorithm succeeds with the claimed probability to colour all N edges when we start with $q = \lceil (1 + 2\varepsilon)\bar{d} \rceil$.

3 Proofs

In the following sections we assume that the integers $N \geq 1, k \geq 2$ and $n \geq 3$ are fixed and satisfy $k^2 < n/2, \frac{kN}{n} \geq 1$.

3.1 One-step differences

We are going to use the differential equations method, see [11, 25]. We will analyse the randomised colouring algorithm with $q = \lceil \frac{kN}{n} \rceil$ colours. The final result of one run is a random object in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega = \binom{[n]}{k}^N \times \{0, \dots, q\}^N$, \mathcal{F} is the σ -field generated by the outcomes of all N edges and their colours¹, and \mathbb{P} is as described in Section 2. The associated natural filtration is

$$\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_N = \mathcal{F},$$

where \mathcal{F}_i is the σ -field generated by the random edges e_1, \dots, e_i and their colours $c(e_1), \dots, c(e_i)$. Corresponding to each \mathcal{F}_i is a natural partition of Ω into blocks that generate \mathcal{F}_i , that is, a *block* (or an atom) of \mathcal{F}_i is the set of all $\omega \in \Omega$ corresponding to a particular sequence of the first i edges and their colours.

Given $i \in \{0, \dots, N - 1\}$, some non-negative $\mu_M(i), \mu_L(i)$ and small non-negative $k_1(i), k_2(i)$, we will consider the events that for each k -set $e \in \binom{[n]}{k}$

$$M_e(i) = \mu_M(i) \pm k_1(i), \tag{4}$$

and for each colour $c \in [q]$

$$L_c(i) = \mu_L(i) \pm k_2(i). \tag{5}$$

¹A special value 0 is assigned to the edges which the algorithm fails to colour.

The numbers $\mu_M(i)$ and $\mu_L(i)$ will be defined later, but one can think for now that they are the expected values of $M_e(i)$ and $L_c(i)$ respectively, while $k_1(i)$ and $k_2(i)$ are the corresponding “errors”. So we will assume that $\mu_L(i) \leq n$.

Observe that the difference $M_e(i+1) - M_e(i)$ is either -1 or 0 ; and $L_c(i+1) - L_c(i)$ is either $-k$ or 0 . If $M_e(i+1) - M_e(i) < 0$ or $L_c(i+1) - L_c(i) < 0$ we will say that M_e or, respectively, L_c decreases (at step $i+1$).

Lemma 3.1 *Suppose $i \in \{0, 1, \dots, N-1\}$ and positive numbers $\mu_M(i)$, $\mu_L(i)$, $k_1(i)$, $k_2(i)$ satisfy*

$$k_1(i) < \mu_M(i)/2 \quad \text{and} \quad k^2 + kk_2(i) < \mu_L(i) \leq n. \quad (6)$$

Let B_i be a block of the partition generating \mathcal{F}_i . Suppose (4) and (5) are satisfied on B_i . Then for each $e \in \binom{[n]}{k}$

$$\begin{aligned} & \mathbb{P}(M_e(i+1) - M_e(i) = -1 | B_i) \\ &= \frac{k^2 \mu_L(i)^{k-1}}{n^k} \left(1 \pm 30 \left(\frac{k_1(i)}{\mu_M(i)} + \frac{k^2 + kk_2(i)}{\mu_L(i)} \right) \right) \end{aligned}$$

and for each $c \in [q]$

$$\begin{aligned} & \mathbb{P}(L_c(i+1) - L_c(i) = -k | B_i) \\ &= \frac{\mu_L(i)^k}{\mu_M(i)n^k} \left(1 \pm 8 \left(\frac{k_1(i)}{\mu_M(i)} + \frac{k^2 + kk_2(i)}{\mu_L(i)} \right) \right). \end{aligned}$$

Proof Since i is fixed, we write $\mu_L = \mu_L(i)$, $\mu_M = \mu_M(i)$, $k_j = k_j(i)$, $M_S = M_S(i)$, $L_c = L_c(i)$ and $\mathcal{M}_S = \mathcal{M}_S(i)$. Conditionally on B_i , all random variables that are \mathcal{F}_i -measurable are constant. This includes M_S for any $S \subseteq [n]$ and L_c for any $c \in [q]$, etc. Note that (4–6) imply that $M_e > 0$ for any $e \in \binom{[n]}{k}$, so each edge of $H(i)$ is assigned some colour on B_i .

In order for M_e to decrease at step $i+1$, e_{i+1} has to be incident to e . So

$$\begin{aligned} \mathbb{P}(M_e \text{ decreases} | B_i) &= \sum_{f: f \sim e} \mathbb{P}(e_{i+1} = f) \mathbb{P}(M_e \text{ decreases} | e_{i+1} = f, B_i) \\ &= \binom{n}{k}^{-1} \sum_{f: f \sim e} \mathbb{P}(M_e \text{ decreases} | e_{i+1} = f, B_i) \end{aligned}$$

Here f ranges over all k -element subsets of $[n]$ intersecting e (we write $e \sim f$ if edges f and e share at least one vertex). Since the algorithm picks

a colour $c = c(e_{i+1})$ uniformly at random from all available ones, and since M_e decreases only in the case that in the hypergraph $H(i)$ both e and e_{i+1} could be coloured with colour c , we get

$$\mathbb{P}(M_e \text{ decreases} \mid e_{i+1} = f, B_i) = \frac{M_{f \cup e}}{M_f}.$$

Using (4) and (6) we can approximate

$$M_f^{-1} = (\mu_M \pm k_1)^{-1} = \mu_M^{-1} \left(1 \pm \frac{2k_1}{\mu_M}\right) \quad (7)$$

to get

$$\mathbb{P}(M_e \text{ decreases} \mid B_i) = \binom{n}{k}^{-1} \mu_M^{-1} \left(1 \pm \frac{2k_1}{\mu_M}\right) \sum_{f: f \sim e} M_{f \cup e}. \quad (8)$$

For any set $S \subseteq [n]$ and any $c \in [q]$ let $A_{c,S} = A_{c,S}(i)$ be the indicator of the event that the colour c is not used on edges of H_i incident to vertices in S . Then

$$M_S = \sum_{c \in [q]} A_{c,S}.$$

Let $\mathcal{L}_c = \mathcal{L}_c(i)$ be the set of vertices which are not contained in an edge coloured c in $H(i)$. Recall that $|\mathcal{L}_c| = L_c$. Note that $A_{c,e \cup f} = A_{c,e} A_{c,f} = 1$ if both e and f are subsets of \mathcal{L}_c . If $e \subseteq \mathcal{L}_c$ then the number of k -subsets f of \mathcal{L}_c intersecting e is at least $k \binom{L_c - k}{k-1}$ and at most $k \binom{L_c - 1}{k-1}$. Writing $r = L_c - \mu_L$ and using (5) and (6) we have

$$\begin{aligned} \binom{L_c - k}{k-1} &= \frac{\mu_L^{k-1}}{(k-1)!} \left(1 - \frac{k-r}{\mu_L}\right) \times \cdots \times \left(1 - \frac{2k-2-r}{\mu_L}\right) \\ &\geq \frac{\mu_L^{k-1}}{(k-1)!} \left(1 - \frac{(k-r) + \cdots + (2k-2-r)}{\mu_L}\right) \\ &\geq \frac{\mu_L^{k-1}}{(k-1)!} \left(1 - \frac{kk_2 + 1.5k^2}{\mu_L}\right). \end{aligned}$$

Also, by (6) we have $x = k_2/\mu_L \leq 1/k$, so using the simple inequalities $(1+x)^{k-1} \leq 1 + kx(1+x)^{k-1}$ and $(1+x)^{k-1} \leq e$

$$\binom{L_c - 1}{k-1} \leq \frac{\mu_L^{k-1}}{(k-1)!} \left(1 + \frac{ek_2}{\mu_L}\right).$$

Thus if $A_{c,e} = 1$ then

$$\sum_{f:f \sim e} A_{c,f} = \frac{k\mu_L^{k-1}}{(k-1)!} \left(1 \pm \frac{3(k^2 + kk_2)}{\mu_L} \right).$$

Therefore

$$\begin{aligned} \sum_{f:f \sim e} M_{f \cup e} &= \sum_{f:f \sim e} \sum_{c \in [q]} A_{c,f \cup e} \\ &= \sum_{c \in [q]} \sum_{f:f \sim e} A_{c,f \cup e} = \sum_{c \in [q]} A_{c,e} \frac{k\mu_L^{k-1}}{(k-1)!} \left(1 \pm \frac{3(k^2 + kk_2)}{\mu_L} \right) \\ &= \frac{k\mu_L^{k-1}\mu_M}{(k-1)!} \left(1 \pm \frac{3(k^2 + kk_2)}{\mu_L} \right) \left(1 \pm \frac{k_1}{\mu_M} \right). \end{aligned}$$

In the last step we used $\sum_{c \in [q]} A_{c,e} = M_e = \mu_M \pm k_1$. Putting the last estimate into (8) we obtain

$$\begin{aligned} &\mathbb{P}(M_e \text{ decreases} \mid B_i) \\ &= \frac{k^2\mu_L^{k-1}}{n^k} \left(1 \pm \frac{2k_1}{\mu_M} \right) \left(1 \pm \frac{k^2}{n} \right) \left(1 \pm \frac{3(k^2 + kk_2)}{\mu_L} \right) \left(1 \pm \frac{k_1}{\mu_M} \right) \\ &= \frac{k^2\mu_L^{k-1}}{n^k} \left(1 \pm 30 \left(\frac{k_1}{\mu_M} + \frac{k^2 + kk_2}{\mu_L} \right) \right). \end{aligned}$$

Now denote by $\mathcal{P}_c = \mathcal{P}_c(i)$ the set of k -element subsets of vertices in $H(i)$ which do not touch an edge of colour c . We have

$$\begin{aligned} \mathbb{P}(L_c \text{ decreases} \mid B_i) &= \sum_{e \in \mathcal{P}_c} \mathbb{P}(L_c \text{ decreases} \mid e_{i+1} = e, B_i) \mathbb{P}(e_{i+1} = e) \\ &= \sum_{e \in \mathcal{P}_c} \frac{1}{M_e} \frac{1}{\binom{n}{k}} = \mu_M^{-1} \left(1 \pm \frac{2k_1}{\mu_M} \right) \binom{L_c}{k} \binom{n}{k}^{-1} \\ &= \frac{\mu_L^k}{\mu_M n^k} \left(1 \pm \frac{2k_1}{\mu_M} \right) \left(1 \pm \frac{4(k^2 + kk_2)}{\mu_L} \right) \\ &= \frac{\mu_L^k}{\mu_M n^k} \left(1 \pm 8 \left(\frac{k_1}{\mu_M} + \frac{k^2 + kk_2}{\mu_L} \right) \right). \end{aligned}$$

Here we used (5), (6), (7) and

$$\frac{\mu_L^k}{n^k} \left(1 - \frac{kk_2 + k^2}{\mu_L} \right) \leq \frac{(L_c)_k}{(n)_k} \leq \frac{L_c^k}{n^k} \leq \frac{\mu_L^k}{n^k} \left(1 + \frac{4kk_2}{\mu_L} \right).$$

□

3.2 Differential equations

If M_e and L_c were always concentrated around the functions μ_M and μ_L , respectively, then by Lemma 3.1 we would have, informally,

$$\mu_M(i+1) - \mu_M(i) \approx -\frac{k^2 \mu_L(i)^{k-1}}{n^k}$$

and

$$\mu_L(i+1) - \mu_L(i) \approx -\frac{k \mu_L(i)^k}{n^k \mu_M(i)}.$$

We can rescale the “time” i and the random variables. Define new functions f and g by $t = i/N$, $\mu_M(i) = qf(t)$ and $\mu_L(i) = ng(t)$, so that

$$\begin{aligned} q(f(t+1/N) - f(t)) &\approx -\frac{k^2 g(t)^{k-1}}{n} \\ n(g(t+1/N) - g(t)) &\approx -\frac{k g(t)^k}{q f(t)} \end{aligned}$$

or

$$\begin{aligned} N(f(t+1/N) - f(t)) &\approx -k g(t)^{k-1} \\ N(g(t+1/N) - g(t)) &\approx -\frac{g(t)^k}{f(t)}. \end{aligned}$$

The left-hand sides above can be approximated by $f'(t)$ and $g'(t)$ respectively (assuming that f and g are differentiable).

This suggests a system of differential equations:

$$\begin{cases} f'(t) = -k g(t)^{k-1} \\ g'(t) = -\frac{g(t)^k}{f(t)} \end{cases}$$

with the initial condition $f(0) = 1$ and $g(0) = 1$.

The solution is $f(t) = (1-t)^k$ and $g(t) = (1-t)$. This argument is not formal, but indicates the choice for μ_M and μ_L . It will be formalised below.

3.3 Martingales

In the rest of this paper we write $t = t(i) = i/N$ and

$$\mu_M(i) = q(1-t)^k \quad \mu_L(i) = n(1-t) \quad \text{for } i = 0, 1, \dots, N. \quad (9)$$

Lemma 3.2 Suppose $i \in \{0, 1, \dots, N-1\}$, (6) holds and let B_i be as in Lemma 3.1.

$$\begin{aligned} & \mathbb{P}(M_e(i+1) - M_e(i) = -1 | B_i) \\ &= \frac{k^2(1-t)^{k-1}}{n} \pm 30 \left(\frac{(k^4 + k^3 k_2(i))(1-t)^{k-2}}{n^2} + \frac{k k_1(i)}{N(1-t)} \right) \end{aligned} \quad (10)$$

and

$$\mathbb{P}(L_c(i+1) - L_c(i) = -k | B_i) = \frac{1}{q} \pm 30 \left(\frac{k_1(i)}{q^2(1-t)^k} + \frac{k + k_2(i)}{N(1-t)} \right).$$

Proof Simply use (9) in Lemma 3.1. \square

Let $\omega, C, K, \varepsilon$ be positive reals, $0 < \varepsilon < 1$. Write $N_1 = \lfloor (1-\varepsilon)N \rfloor$ and define for $i \in \{0, \dots, N_1\}$

$$E_1(i) = \frac{kq}{\omega} e^{Ct} \quad \text{and} \quad E_2(i) = \frac{n}{\omega} \frac{e^{Ct}}{(1-t)^k}.$$

We want to show that for each edge e the process $M_e(i)$ remains at distance at most $E_1(i)$ from the deterministic function $\mu_M(i)$, and similarly $L_c(i)$ remains at distance at most $E_2(i)$ from $\mu_L(i)$. The values of E_1 and E_2 were chosen so that the difference $E_j(i+1) - E_j(i)$ cancels out the error resulting from our approximation, see e.g., inequality (20) below.

For any $i \in \{0, \dots, N_1\}$ and $e \in \binom{[n]}{k}$ set

$$M_e^-(i) = M_e(i) - \mu_M(i) - E_1(i); \quad (11)$$

$$M_e^+(i) = M_e(i) - \mu_M(i) + E_1(i); \quad (12)$$

and for each colour $c \in [q]$:

$$L_c^-(i) = L_c(i) - \mu_L(i) - E_2(i); \quad (13)$$

$$L_c^+(i) = L_c(i) - \mu_L(i) + E_2(i). \quad (14)$$

We will consider a stopping time T_S given by

$$\begin{aligned} T_S = 1 + \sup_i \left\{ M_e^-(i) \leq Kk\sqrt{q \ln n}, M_e^+(i) \geq -Kk\sqrt{q \ln n}, \right. \\ \left. L_c^-(i) \leq K\sqrt{kn \ln q}, L_c^+(i) \geq -K\sqrt{kn \ln q}, \right. \\ \left. \text{for all } e \in \binom{[n]}{k}, c \in [q] \right\}. \end{aligned}$$

Note that T_S is well defined since $i = 0$ is always in the set.

Now for each $e \in \binom{[n]}{k}$ define the stopped processes $\{\tilde{M}_e^-\} = \{\tilde{M}_e^-(i), i = 0, \dots, N_1\}$ and $\{\tilde{M}_e^+\} = \{\tilde{M}_e^+(i), i = 0, \dots, N_1\}$ where

$$\tilde{M}_e^-(i) = M_e^-(i \wedge T_S), \quad \tilde{M}_e^+(i) = M_e^+(i \wedge T_S)$$

and $x \wedge y = \min(x, y)$. Also, for each $c \in [q]$ define stopped processes $\{\tilde{L}_c^-\} = \{\tilde{L}_c^-(i), i = 0, \dots, N_1\}$ and $\{\tilde{L}_c^+\} = \{\tilde{L}_c^+(i), i = 0, \dots, N_1\}$ where

$$\tilde{L}_c^-(i) = L_c^-(i \wedge T_S) \quad \text{and} \quad \tilde{L}_c^+(i) = L_c^+(i \wedge T_S).$$

Lemma 3.3 *Suppose that $2k^2/n \leq \varepsilon < 1$, $K \geq 1$, $(kN)/n \geq 3$, $C = 240k\varepsilon^{-2}$ and ω is such that*

$$4ke^C \varepsilon^{-(k+1)} \leq \omega \leq K^{-1} \min \left(\frac{q}{\ln n}, \frac{n}{k \ln q} \right)^{1/2}.$$

Then for any $e \in \binom{[n]}{k}$ and any $c \in [q]$ the processes $\{\tilde{M}_e^-\}, \{\tilde{L}_c^-\}$ are supermartingales and $\{\tilde{M}_e^+\}, \{\tilde{L}_c^+\}$ are submartingales.

Note that the existence of ω satisfying the condition of this lemma (and the next one) will be proved only in the proof of Theorem 1.2, given in Section 3.4.

Proof Suppose $i \leq \min(T_S, N_1)$. Then

$$\begin{aligned} M_e(i) &= \mu_M(i) \pm k_1(i) \quad \forall e \in \binom{[n]}{2}, \\ L_c(i) &= \mu_L(i) \pm k_2(i) \quad \forall c \in [q], \end{aligned}$$

where

$$\begin{aligned} k_1(i) &= E_1(i) + Kk\sqrt{q \ln n}, \\ k_2(i) &= E_2(i) + K\sqrt{kn \ln q}. \end{aligned}$$

Write $\phi(\tau) = e^{C\tau}$ and $\gamma(\tau) = e^{C\tau}/(1-\tau)^k$. Both ϕ and γ are increasing for $\tau \in [0, 1)$. Since $\omega \leq K^{-1}\sqrt{q/\ln n}$ and $\omega \leq K^{-1}\sqrt{n/(k \ln q)}$ we have

$$k_1(i) \leq 2E_1(i); \tag{15}$$

$$k_2(i) \leq 2E_2(i). \tag{16}$$

We claim that (6) is satisfied for each $i \in [0, N_1]$. Indeed, since $\varepsilon \geq 2k^2/n$, $\omega \geq 4ke^C\varepsilon^{-(k+1)}$ and ϕ, γ are increasing

$$\begin{aligned} \frac{k_1(i)}{\mu_M(i)} &\leq \frac{2E_1(i)}{q(1-t)^k} = \frac{2k\gamma(t)}{\omega} \leq \frac{2ke^C}{\omega\varepsilon^k} \leq \frac{1}{2}; \\ \frac{kk_2(i)}{\mu_L(i)} &\leq \frac{2kE_2(i)}{n(1-t)} \leq \frac{2k\gamma(t)}{\omega(1-t)} \leq \frac{2ke^C}{\omega\varepsilon^{k+1}} \leq \frac{1}{2}; \\ \frac{k^2}{\mu_L(i)} &= \frac{k^2}{n(1-t)} \leq \frac{k^2}{\varepsilon n} \leq \frac{1}{2}. \end{aligned} \tag{17}$$

So $k_1(i) \leq \mu_M(i)/2$ and $k^2 + kk_2(i) \leq \mu_L(i)$ as required.

We will first show that $\{\tilde{M}_e^+\}$ is a submartingale and $\{\tilde{M}_e^-\}$ is a supermartingale. Since the increments are zero for $i \geq T_S$, it suffices to prove that

$$\mathbb{E} \left(\tilde{M}_e^-(i+1) - \tilde{M}_e^-(i) | i < T_S \right) \leq 0, \quad \mathbb{E} \left(\tilde{M}_e^+(i+1) - \tilde{M}_e^+(i) | i < T_S \right) \geq 0.$$

On the event $i < T_S$ (4) and (5) hold by the definition of T_S .

Write

$$R_1(i) = \frac{30k^4(1-t)^{k-2}}{n^2} + \frac{30k^3k_2(i)(1-t)^{k-2}}{n^2} + \frac{30kk_1(i)}{N(1-t)}.$$

By Lemma 3.2

$$\begin{aligned} &\mathbb{E} \left(\tilde{M}_e^+(i+1) - \tilde{M}_e^+(i) | i < T_S \right) \\ &= \left(-\frac{k^2(1-t)^{k-1}}{n} \pm R_1(i) \right) + (\mu_M(i) - \mu_M(i+1)) + (E_1(i+1) - E_1(i)). \end{aligned}$$

Some of the terms cancel out:

$$\begin{aligned} &-\frac{k^2(1-t)^{k-1}}{n} + \mu_M(i) - \mu_M(i+1) \\ &= -\frac{k^2}{n} \left(1 - \frac{i}{N}\right)^{k-1} + q \left(\left(1 - \frac{i}{N}\right)^k - \left(1 - \frac{i+1}{N}\right)^k \right) \\ &\geq -\frac{k^2}{n} \left(1 - \frac{i}{N}\right)^{k-1} + \frac{kN}{n} \left(1 - \frac{i}{N}\right)^k \left(1 - \left(1 - \frac{1}{N-i}\right)^k\right) \\ &\geq -\frac{k^2(1 - \frac{i}{N})^{k-1}}{n} + \frac{kN}{n} \left(1 - \frac{i}{N}\right)^k \left(\frac{k}{N-i} - \frac{k^2}{2(N-i)^2} \right) \\ &= -\frac{k^3}{2Nn} \left(1 - \frac{i}{N}\right)^{k-2} \geq -\frac{k}{N}. \end{aligned} \tag{18}$$

Similarly, since $q = \lceil \frac{kN}{n} \rceil \leq \frac{kN}{n} + 1$

$$-\frac{k^2(1-t)^{k-1}}{n} + \mu_M(i) - \mu_M(i+1) \leq \frac{k}{N}.$$

We will also need later that since $\frac{k}{N} \leq \frac{k^2}{3n}$ we have

$$\mu_M(i) - \mu_M(i+1) < \frac{2k^2}{n}. \quad (19)$$

Therefore

$$\mathbb{E} \left(\tilde{M}_e^+(i+1) - \tilde{M}_e^+(i) | i < T_S \right) \geq (E_1(i+1) - E_1(i)) - \frac{k}{N} - R_1(i).$$

Similarly,

$$\mathbb{E} \left(\tilde{M}_e^-(i+1) - \tilde{M}_e^-(i) | i < T_S \right) \leq -(E_1(i+1) - E_1(i)) + \frac{k}{N} + R_1(i).$$

It remains to verify that

$$\frac{k}{N} + R_1(i) \leq E_1(i+1) - E_1(i). \quad (20)$$

Since $\phi''(t) > 0$ we have

$$\begin{aligned} E_1(i+1) - E_1(i) &= \frac{kq}{\omega} \left(\phi \left(t + \frac{1}{N} \right) - \phi(t) \right) \\ &\geq \frac{qk\phi'(t)}{\omega N} = \frac{Cqk\phi(t)}{\omega N} \geq \frac{Ck^2\phi(t)}{\omega n}. \end{aligned} \quad (21)$$

Now, firstly,

$$\frac{k}{N} \leq \frac{1}{4}(E_1(i+1) - E_1(i)).$$

This is because by (21) the ratio of the left side and the right is at most

$$\frac{k}{N} \frac{4\omega N}{Cqk\phi(t)} \leq \frac{4\omega}{Cq\phi(0)} \leq \frac{4}{q^{1/2}C} \leq 1.$$

Here we used that ϕ is increasing, $\omega \leq q^{1/2}$ and $C > 8$ from the assumption of the lemma. Secondly, for the first term of $R_1(i)$ we have

$$\frac{30k^4(1-t)^{k-2}}{n^2} \leq \frac{1}{4}(E_1(i+1) - E_1(i)).$$

To see this, note that by (21), the ratio of the left and the right side expressions is at most

$$\frac{120k^2(1-t)^{k-2}\omega}{nCe^{Ct}} \leq \frac{120k^2\omega}{nC} \leq \frac{k\omega}{2n} \leq \frac{1}{2K} \sqrt{\frac{k}{n \ln q}} \leq 1.$$

Here we used the facts that $e^{C\tau}/(1-\tau)^{k-2}$ is increasing for $\tau \in [0, 1)$ and the earlier assumptions about n, q, k, C and ω .

Thirdly,

$$\frac{30k^3k_2(i)(1-t)^{k-2}}{n^2} \leq \frac{1}{4}(E_1(i+1) - E_1(i)).$$

Indeed, by (21), the ratio of the left and the right side is at most

$$\frac{120kk_2(i)\omega(1-t)^{k-2}}{nCe^{Ct}} \leq \frac{240kE_2(i)\omega(1-t)^{k-2}}{nCe^{Ct}} \leq \frac{240k}{C(1-t)^2} \leq 1.$$

Here we used (16) in the first inequality and $t \leq 1-\varepsilon$ in the last one. Finally, for the last term of $R_1(i)$

$$\frac{30kk_1(i)}{N(1-t)} \leq \frac{1}{4}(E_1(i+1) - E_1(i)).$$

This is because the ratio of the left side and the right by (21) and (15) is at most

$$\frac{120k_1(i)\omega}{Cqe^{Ct}(1-t)} \leq \frac{240kqe^{Ct}}{Cqe^{Ct}(1-t)} \leq \varepsilon \leq 1.$$

Now (20) follows by combining bounds for each of the four terms of $k/N + R_1(i)$.

Let us now show that $\{\tilde{L}_c^-\}$ is a supermartingale. Again it suffices to consider only the blocks of \mathcal{F}_i where $i < T_S$. By Lemma 3.2

$$\begin{aligned} & \mathbb{E} \left(\tilde{L}_c^-(i+1) - \tilde{L}_c^-(i) \mid i < T_S \right) \\ &= \mathbb{E} (L_c(i+1) - L_c(i) \mid i < T_S) + (\mu_L(i) - \mu_L(i+1)) + (E_2(i) - E_2(i+1)) \\ &= -\frac{k}{q} \pm R_2(i) + (\mu_L(i) - \mu_L(i+1)) + (E_2(i) - E_2(i+1)), \end{aligned}$$

where

$$R_2(i) = \frac{Dkk_1(i)}{q^2(1-t)^k} + \frac{30k^2}{N(1-t)} + \frac{30kk_2(i)}{N(1-t)}$$

Now

$$\begin{aligned} -\frac{k}{q} + \mu_L(i) - \mu_L(i+1) &= -\frac{k}{q} + n \left(\left(1 - \frac{i}{N}\right) - \left(1 - \frac{i+1}{N}\right) \right) \\ &= -\frac{k}{q} + \frac{n}{N} \in \left[0, \frac{2k}{q^2}\right], \end{aligned} \quad (22)$$

since $\frac{kN}{n} \geq 3$ and $\frac{n}{kN} \leq \frac{1}{q-1}$. Therefore

$$\mathbb{E} \left(\tilde{L}_c^-(i+1) - \tilde{L}_c^-(i) | i < T_S \right) = E_2(i) - E_2(i+1) \pm \left(R_2(i) + \frac{2k}{q^2} \right).$$

We need to show that the above quantity is non-positive. Since $\gamma''(t) > 0$ for $t \in [0, 1)$, we have similarly as in (21)

$$E_2(i+1) - E_2(i) \geq \frac{\gamma'(t)n}{N\omega} \geq \frac{\gamma'(t)k}{q\omega} \geq \frac{C\gamma(t)k}{q\omega}. \quad (23)$$

We have

$$\frac{30kk_1(i)}{q^2(1-t)^k} \leq \frac{1}{4}(E_2(i+1) - E_2(i)),$$

since by (15) and (23) the ratio of the left and the right side is at most

$$\frac{120k}{q^2(1-t)^k} \frac{2kq\phi(t)}{\omega} \frac{q\omega}{C\gamma(t)k} = \varepsilon^2 \leq 1.$$

Let us now check that

$$\frac{30k^2}{N(1-t)} \leq \frac{1}{4}(E_2(i+1) - E_2(i)).$$

Indeed, the ratio of the left and the right side is by (23) at most

$$\frac{120k^2}{N(1-t)} \frac{q\omega}{C\gamma(t)k} \leq \frac{q\omega}{2N} \leq \frac{1}{K} \sqrt{\frac{k}{n \ln q}} \leq 1.$$

The first inequality follows, among others, because $(1-\tau)\gamma(\tau)$ is increasing for $\tau \in [0, 1)$. Next,

$$\frac{30kk_2(i)}{N(1-t)} \leq \frac{1}{4}(E_2(i+1) - E_2(i)).$$

Indeed, by (16) and (23) the ratio of the left and the right side is at most

$$\frac{120k}{N(1-t)} \frac{2n\gamma(t)}{\omega} \frac{N\omega}{\gamma'(t)n} \leq \varepsilon \leq 1.$$

Finally,

$$\frac{2k}{q^2} \leq \frac{1}{4} (E_2(i+1) - E(i)).$$

Here, again, the ratio of the two sides is by (23) at most

$$\frac{2k}{q^2} \frac{4q\omega}{C\gamma(t)k} \leq \frac{\omega}{30qk} \leq 1$$

since $\omega \leq q^{1/2}$.

We have shown that

$$R_2(i) + \frac{2k}{q^2} \leq E_2(i+1) - E_2(i) \tag{24}$$

and hence

$$\mathbb{E} \left(\tilde{L}_c^-(i+1) - \tilde{L}_c^-(i) | i < T_S \right) \leq 0$$

as required, so $\{\tilde{L}_c^-\}$ is a supermartingale. The bounds above also show that $\{\tilde{L}_c^+\}$ is a submartingale. \square

3.4 Applying concentration results

In this section we prove the main lemma of this paper. We will use a concentration result that takes into account conditional variance of martingale differences (see McDiarmid [20]).

Let X be a \mathcal{F} -measurable bounded random variable. Let $(\emptyset, \omega) = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_n$ be a filtration in \mathcal{F} . Let X_0, \dots, X_n be a martingale obtained by setting $X_i = \mathbb{E}(X | \mathcal{F}_i)$. For $1 \leq i \leq n$ set $Y_i = X_i - X_{i-1}$ and define \mathcal{F}_{i-1} -measurable functions $\text{dev}_i^+ = \sup(Y_i | \mathcal{F}_{i-1})$ and $\text{var}_i = \text{Var}(Y_i | \mathcal{F}_{i-1})$. The number $\hat{v} = \sup \sum_{i=1}^n \text{var}_i$ is called *the maximum sum of conditional variances*. We call the quantity $\max \text{dev}^+ = \sup_i \text{dev}_i^+$ *the maximum conditional positive deviation*.

Theorem 3.4 (Theorem 3.15 of [20]) *Let X be a random variable with $\mathbb{E} X = \mu$ and let $(\emptyset, \omega) = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_n$ be a filtration in \mathcal{F} . Assume the maximum conditional positive deviation $b = \max \text{dev}^+$ and the maximum sum of conditional variances \hat{v} are finite. Then for any $x \geq 0$*

$$\mathbb{P}(X - \mu \geq x) \leq \exp \left(-\frac{x^2}{2\hat{v}(1 + bx/(3\hat{v}))} \right).$$

Lemma 3.5 *Suppose that $\varepsilon \in (0, 1)$, $C = 240k\varepsilon^{-2}$, and*

$$2k^2 e^C \varepsilon^{-(k+2)} < \omega < \frac{1}{8\sqrt{30}} \min\left(\frac{q}{\ln n}, \frac{n}{k \ln q}\right)^{1/2}.$$

Then

$$\mathbb{P}(T_S < \lfloor (1 - \varepsilon)N \rfloor) \leq \frac{2}{q} + \frac{2}{n}. \quad (25)$$

Proof Set $K = 8\sqrt{30}$. The condition of the lemma implies that $2k^2/n < \varepsilon^{k+2}n^{-1/2} < \varepsilon$ and $q = \lceil kN/n \rceil > k^2 \geq 4$, so the assumptions of Lemma 3.3 are satisfied. By Lemma 3.3 $\{\tilde{M}_e^-\}$, $\{\tilde{L}_c^-\}$ are supermartingales and $\{\tilde{M}_e^+\}$, $\{\tilde{L}_c^+\}$ are submartingales for every $e \in \binom{[n]}{k}$ and $c \in [q]$. Observe that $\tilde{M}_e^\pm(i)$ and $\tilde{L}_c^\pm(i)$ are \mathcal{F}_i -measurable (where \mathcal{F}_i is as in Section 3.1) and T_S is a well-defined stopping time for these processes.

Recall that $N_1 = \lfloor (1 - \varepsilon)N \rfloor$. We shall apply Theorem 3.4 to show that for any $e \in \binom{[n]}{k}$ and $c \in [q]$:

$$\begin{aligned} \mathbb{P}(\tilde{M}_e^-(N_1) \geq Kk\sqrt{q \ln n}) &\leq \frac{1}{n^{k+1}}; \\ \mathbb{P}(\tilde{M}_e^+(N_1) \leq -Kk\sqrt{q \ln n}) &\leq \frac{1}{n^{k+1}}; \\ \mathbb{P}(\tilde{L}_c^-(N_1) \geq K\sqrt{kn \ln q}) &\leq \frac{1}{q^2}; \\ \mathbb{P}(\tilde{L}_c^+(N_1) \leq -K\sqrt{kn \ln q}) &\leq \frac{1}{q^2}. \end{aligned}$$

Notice that if $T_S < N_1$ then at least one of the above events occurs (for some e or for some c), since all the processes remain “frozen” after the time T_S . Since there are $\binom{n}{k} \leq n^k$ sets $e \in \binom{[n]}{k}$ and q colours c , the above inequalities and the union bound imply (25). We will modify the submartingales and supermartingales slightly to turn them into martingales.

Set $\beta = 241$. Then

$$\omega > 2k^2 e^C \varepsilon^{-(k+2)} > 2kC e^C \varepsilon^{-k} \beta^{-1}.$$

Notice that the bounds on ω in the assumption imply

$$\left(\frac{\ln n}{q}\right)^{1/2} \leq \frac{1}{k^2 K} \quad \text{and} \quad \left(\frac{k \ln q}{n}\right)^{1/2} \leq \frac{1}{k^2 K}. \quad (26)$$

For $i = 0, \dots, N_1 - 1$ define random variables

$$Z_M^-(i+1) = -\mathbb{E}(\tilde{M}_e^-(i+1) - \tilde{M}_e^-(i) | \mathcal{F}_i).$$

Since $\{\tilde{M}_e^-\}$ is a supermartingale, $Z_M^-(i) \geq 0$. Also, since $\{\tilde{M}_e^+\}$ is a submartingale by Lemma 3.3

$$\mathbb{E}(\tilde{M}_e^-(i+1) - \tilde{M}_e^-(i) | \mathcal{F}_i) + 2(E_1(i+1) - E_1(i)) = \mathbb{E}(\tilde{M}_e^+(i+1) - \tilde{M}_e^+(i) | \mathcal{F}_i) \geq 0.$$

Therefore

$$0 \leq Z_M^-(i+1) \leq 2(E_1(i+1) - E_1(i)). \quad (27)$$

The sequence $\{\hat{M}_e^-\} = \{\hat{M}_e^-(i), i = 0, \dots, N_1\}$ where

$$\hat{M}_e^-(i) = \tilde{M}_e^-(i) + \sum_{j=1}^i Z_M^-(j)$$

is a martingale (here and below we define the sum when j ranges from 1 to 0 to be equal to 0) and

$$\mathbb{P}\left(\tilde{M}_e^-(N_1) \geq Kk\sqrt{q \ln n}\right) \leq \mathbb{P}\left(\hat{M}_e^-(N_1) \geq Kk\sqrt{q \ln n}\right).$$

Let us estimate the maximum conditional positive deviation and the maximum conditional variance of $\{\hat{M}_e^-\}$.

Notice that since $(kN)/n \geq 3$, we have $n/(kN) \leq 3/(2q)$. Let $\phi(\tau), \gamma(\tau)$ be as in the proof of Lemma 3.3. Since $\phi'(\tau), \gamma'(\tau)$ are increasing, the following inequalities hold for $i = 0, \dots, N_1 - 1$:

$$E_1(i+1) - E_1(i) \leq \frac{\phi'(1-\varepsilon)kq}{\omega N} \leq \frac{k^2 C e^C}{n \omega} \leq \frac{\beta k}{n} \leq \frac{\beta k^2}{n}; \quad (28)$$

$$E_2(i+1) - E_2(i) \leq \frac{\gamma'(1-\varepsilon)n}{\omega N} \leq \frac{3k}{2q\omega} \left(\frac{C e^C}{\varepsilon^k} + \frac{k e^C}{\varepsilon^{k+1}} \right) \leq \frac{\beta}{q}. \quad (29)$$

We have $\hat{M}_e^-(i+1) - \hat{M}_e^-(i) = 0$ for $i \geq T_S$. If $i < T_S$, using (19), (27) and (28) the difference $\hat{M}_e^-(i+1) - \hat{M}_e^-(i)$ is

$$\begin{aligned} & M_e(i+1) - M_e(i) + \mu_M(i) - \mu_M(i+1) \\ & + E_1(i) - E_1(i+1) + Z_M^-(i+1) \\ & \leq \frac{2k^2}{n} + \frac{\beta k^2}{n} \leq \frac{k^2(2+\beta)}{n}. \end{aligned}$$

Let R_1, R_2 be as in the proof of Lemma 3.3. For $i \in \{0, \dots, N_1 - 1\}$, the conditional variance of the differences of $\{\hat{M}_e^-\}$ is

$$\begin{aligned} \text{Var}(\hat{M}_e^-(i+1) - \hat{M}_e^-(i) | \mathcal{F}_i) & \leq \max_{B_i} \text{Var}(M_e(i+1) - M_e(i) | B_i) \\ & \leq \max_{B_i} \mathbb{P}(M_e(i+1) - M_e(i) = -1 | B_i) \\ & \leq \frac{k^2(1-t)^{k-1}}{n} + R_1(i) \leq \frac{k^2(1+\beta)}{n}. \end{aligned}$$

Here B_i ranges over all blocks of the partition corresponding to \mathcal{F}_i ; to get the third inequality we used Lemma 3.2 and to get the fourth inequality we used (20) and (28). Thus the maximum sum of conditional variances \hat{v} for \hat{M}_e^- satisfies

$$\hat{v} \leq \frac{Nk^2(1+\beta)}{n} \leq kq(1+\beta)$$

and the maximum conditional positive deviation b satisfies

$$b \leq \frac{k^2(2+\beta)}{n} \leq 2+\beta.$$

By Theorem 3.4, since $\mathbb{E} \hat{M}_e^-(N_1) = \mathbb{E} \tilde{M}_e^-(0) = -E_1(0) < 0$,

$$\mathbb{P}(\hat{M}_e^-(N_1) > x) \leq \mathbb{P}(\hat{M}_e^-(N_1) - \mathbb{E} \hat{M}_e^-(N_1) > x) \leq \exp\left(-\frac{x^2}{2\hat{v}(1 + \frac{bx}{3\hat{v}})}\right).$$

Therefore

$$\begin{aligned} \mathbb{P}(\hat{M}_e^-(N_1) \geq Kk\sqrt{q \ln n}) &\leq \exp\left(-\frac{k^2 K^2 q \ln n}{2kq(1+\beta) + \frac{2}{3}(2+\beta)Kk\sqrt{q \ln n}}\right) \\ &\leq \exp\left(-\frac{kK^2 \ln n}{2(1+\beta) + \frac{2}{3}(2+\beta)k^{-2}}\right) \\ &\leq e^{-\frac{kK^2 \ln n}{3(1+\beta)}} \leq e^{-(k+1) \ln n} \end{aligned}$$

by (26) since $k \geq 2$, $D \geq 1$ and $K = 8\sqrt{30} > (12 + 48 \cdot 30)^{1/2}$.

Now consider the submartingale $\{\tilde{M}_e^+\}$ and define for $i = 0, \dots, N_1 - 1$

$$Z_M^+(i+1) = -\mathbb{E}(\tilde{M}_e^+(i+1) - \tilde{M}_e^+(i) | \mathcal{F}_i).$$

Since \tilde{M}_e^+ is a submartingale and \tilde{M}_e^- is a supermartingale we have

$$\mathbb{E}(\tilde{M}_e^+(i+1) - \tilde{M}_e^+(i) | \mathcal{F}_i) - 2(E_1(i+1) - E_1(i)) \leq 0$$

so

$$2(E_1(i) - E_1(i+1)) \leq Z_M^+(i+1) \leq 0. \quad (30)$$

The sequence

$$\hat{M}_e^+ = \{\tilde{M}_e^+(i) + \sum_{j=1}^i Z_M^+(j), \quad i = 0, 1, \dots, N_1\}$$

is a martingale and

$$\mathbb{P}\left(\tilde{M}_e^+(N_1) \leq -Kk\sqrt{q \ln n}\right) \leq \mathbb{P}\left(-\hat{M}_e^+(N_1) \geq Kk\sqrt{q \ln n}\right).$$

Using (28) and (30) the difference $-\hat{M}_e^+(i+1) - (-\hat{M}_e^+(i))$ is

$$\begin{aligned} & M_e(i) - M_e(i+1) + \mu_M(i+1) - \mu_M(i) + E_1(i) - E_1(i+1) - Z_M^+(i+1) \\ & \leq 1 + E_1(i+1) - E_1(i) \leq 1 + \frac{k^2\beta}{n} \leq 2 + \beta. \end{aligned}$$

Furthermore, the conditional variance of $-\hat{M}_e^+(i+1) + \hat{M}_e^+(i)$ is the same as the conditional variance of $M_e^-(i+1) - M_e^-(i)$ so $\hat{v} \leq kq(1 + \beta)$. Now Theorem 3.4 yields

$$\mathbb{P}\left(\hat{M}_e^+(N_1) \leq -Kk\sqrt{q \ln n}\right) \leq e^{-(k+1)\ln n}$$

by the same calculation as in the corresponding bound for \hat{M}_e^- .

Now consider the supermartingale $\{\tilde{L}_c^-\}$. Similarly as above, define a martingale $\{\hat{L}_c^-\} = \{\hat{L}_c^-(i), i = 0, \dots, N_1\}$, where

$$\hat{L}_c^-(i) = \tilde{L}_c^-(i) + \sum_{j=1}^i Z_L^-(j)$$

and $Z_L^-(i+1) = -E(\tilde{L}_c^-(i+1) - \tilde{L}_c^-(i) | \mathcal{F}_i)$.

Using the fact that $L_c^+(i) = L_c^-(i) + 2E_2(i)$ we obtain similarly as above

$$0 \leq Z_L^-(i) \leq 2(E_2(i+1) - E_2(i)).$$

Therefore using (22) and (29) the difference $\hat{L}_c^-(i+1) - \hat{L}_c^-(i)$ is

$$\begin{aligned} & L_c(i+1) - L_c(i) + \mu_L(i) - \mu_L(i+1) - E_2(i+1) + E_2(i) + Z_L^-(i+1) \\ & \leq \frac{k}{q} + \frac{2k}{q^2} + \frac{\beta}{q} \leq \frac{k(1+\beta)}{q}. \end{aligned}$$

Here we used $2k/q \leq (k-1)\beta$, since $k \geq 2, q \geq 3$ and $\beta > 8$. Now using Lemma 3.2 and the bounds (24) and (29) for $i \in \{0, \dots, N_1 - 1\}$ we get

$$\begin{aligned} & \text{Var}(\hat{L}_c^-(i+1) - \hat{L}_c^-(i) | \mathcal{F}_i) \leq \max_{B_i} \text{Var}(L_c(i+1) - L_c(i) | B_i) \\ & \leq k^2 \cdot \max_{B_i} \mathbb{P}(L_c \text{ decreases at step } i+1 | B_i) \\ & \leq \frac{k^2}{q} + kR_2(i) \leq \frac{k^2(1+\beta)}{q} \leq k(1+\beta). \end{aligned}$$

Hence the maximum sum of conditional variances \hat{v} satisfies

$$\hat{v} \leq \frac{k^2(1+\beta)N}{q} \leq (1+\beta)kn. \quad (31)$$

So by Theorem 3.4, (26) and our choice of K

$$\begin{aligned} \mathbb{P}\left(\tilde{L}_c^-(N_1) \geq K\sqrt{kn \ln q}\right) &\leq \mathbb{P}\left(\hat{L}_c^-(N_1) \geq K\sqrt{kn \ln q}\right) \\ &\leq \exp\left(-\frac{K^2kn \ln q}{2(1+\beta)kn + \frac{2}{3}k(1+\beta)K\sqrt{kn \ln q}}\right) \\ &\leq \exp\left(-\frac{K^2 \ln q}{2(1+\beta) + \frac{2}{3}(1+\beta)k^{-2}}\right) \leq \exp\left(-\frac{K^2 \ln q}{3(1+\beta)}\right) \leq e^{-2 \ln q}. \end{aligned}$$

Finally, let us bound in exactly the same way the probability that $\{\tilde{L}_c^+\}$ ever attains a large negative value. Define a martingale $\{\hat{L}_c^+\} = \{\hat{L}_c^+(i), i = 0, \dots, N_1\}$, where

$$\hat{L}_c^+(i) = \tilde{L}_c^+(i) + \sum_{j=1}^i Z_L^+(j)$$

and $Z_L^+(i+1) = -\mathbb{E}(\tilde{L}_c^+(i+1) - \tilde{L}_c^+(i) | \mathcal{F}_i)$.

As before, by Lemma 3.3:

$$2(E_2(i) - E_2(i+1)) \leq Z_L^+(i+1) \leq 0.$$

Using (29) we get that the difference $(-\hat{L}_c^+(i+1) - (-\hat{L}_c^+(i)))$ is at most

$$\begin{aligned} &L_c(i) - L_c(i+1) - \mu_L(i) + \mu_L(i+1) + E_2(i) - E_2(i+1) - Z_L^+(i+1) \\ &\leq k + \frac{\beta}{q} \leq k(1+\beta). \end{aligned}$$

The conditional variance of $-\hat{L}_c^+(i+1) + \hat{L}_c^+(i)$ is the same as the conditional variance of $\hat{L}_c^-(i+1) - \hat{L}_c^-(i)$, so the estimate (31) still holds.

Once again applying Theorem 3.4

$$\mathbb{P}(\tilde{L}_c^+(N_1) \leq -K\sqrt{kn \ln q}) \leq \mathbb{P}(-\hat{L}_c^+(N_1) \geq K\sqrt{kn \ln q}) \leq e^{-2 \ln q},$$

as in the corresponding bound for $\{\hat{L}_c^-\}$. \square

Proof of Theorem 1.2 We may assume that $\varepsilon < 1$. Let ε' be such that $\frac{1}{1-\varepsilon'} = 1 + \frac{\varepsilon}{2}$ and define

$$W(\varepsilon, k) = 16\sqrt{30}k^3 e^{240k/\varepsilon'^2} \varepsilon'^{-(k+2)}.$$

Let $c_\varepsilon > 0$ be a constant such that for all $k \geq 2$ we have $W(\varepsilon, k)^2 < \frac{1}{2}e^{k/c_\varepsilon}$ and observe that this implies $c_\varepsilon \leq \varepsilon'/8$. Suppose (1) holds. Then $n \geq 3$, $\bar{d} \geq 4$ and $\frac{k}{n} \leq \frac{\varepsilon'}{2} \leq \frac{\varepsilon}{2}$. Let

$$N' = \left\lceil \frac{N}{1 - \varepsilon'} \right\rceil \quad \text{and} \quad q = \left\lceil \frac{kN'}{n} \right\rceil.$$

Then

$$\bar{d} \leq q \leq \lceil \bar{d}(1 + \varepsilon/2) + k/n \rceil \leq \lceil (1 + \varepsilon)\bar{d} \rceil \leq 3\bar{d} \leq \bar{d}^2,$$

and using the definition of c_ε and (1)

$$W(\varepsilon, k)^2 < \frac{1}{2}e^{k/c_\varepsilon} \leq \frac{1}{2} \min \left(\frac{n}{\ln \bar{d}}, \frac{\bar{d}}{\ln n} \right) \leq \min \left(\frac{n}{\ln q}, \frac{q}{\ln n} \right).$$

Setting $\omega = 8^{-1}(30k)^{-1/2}W(\varepsilon, k)$ we see that

$$2k^2 e^{240k/\varepsilon'^2} \varepsilon^{-(k+2)} < \omega < \frac{1}{8\sqrt{30}} \min \left(\frac{n}{\ln q}, \frac{q}{k \ln n} \right)^{1/2}$$

and so Lemma 3.5 applies for the random colouring process described in Section 3.3 with N' random hyperedges, q colours and n vertices. By that lemma, the probability that the process hits the stopping time until step $i = N - 1 \leq \lfloor N'(1 - \varepsilon') \rfloor$ is at most $\frac{2}{q} + \frac{2}{n}$. If $T_S \geq N - 1$, then by (17) and the definition of the process

$$M_e(N - 1) \geq \mu_M(N - 1) - (E_1(N - 1) + Kk\sqrt{q \ln n}) \geq \frac{1}{2}\mu_M(N - 1) > 0,$$

for any edge $e \in \binom{[n]}{k}$. So the N -th edge can be coloured successfully (which means all of the previous edges have been coloured successfully as well). The claim follows by a trivial observation, that the random hypergraph obtained after adding the first N out of N' edges has distribution exactly $\mathbb{H}^{(k)}(n, N)$. \square

Acknowledgements

We would like to thank all the referees for thorough and helpful comments, especially for suggesting (3).

Valentas Kurauskas acknowledges partial support of the Lithuanian Research Council grant MIP-067/2013.

Katarzyna Rybarczyk acknowledges partial support of the National Science Center grant DEC-2011/01/B/ST1/03943.

References

- [1] N. Alon and J. H. Kim, *On the degree, size, and chromatic index of a uniform hypergraph*, J. Combin. Theory Ser. A, 77 (1997), pp. 165–170.
- [2] J. Balogh, T. Bohman and D. Mubayi, *Erdős-Ko-Rado in random hypergraphs*, Combin. Probab. Comput., 18 (2009), 629–646.
- [3] J. Balogh, S. Das, M. Delcourt, H. Liu, M. Sharifzadeh, *The typical structure of intersecting families of discrete structures*, arXiv:1408.2559 (2014).
- [4] M. Behrisch, A. Taraz and M. Ueckerdt, *Colouring random intersection graphs and complex networks*, SIAM J. Discrete Math. 23 (2009), 288–299.
- [5] C. Berge, *On two conjectures to generalize Vizing’s Theorem*, Le Matematiche 45(1) (1990), pp. 11-24.
- [6] M. Bloznelis, J. Jaworski, E. Godehardt, V. Kurauskas, and K. Rybarczyk, *Recent Progress in Complex Network Analysis Models of Random Intersection Graphs*, to appear in Studies in Classification, Data Analysis, and Knowledge Organization.
- [7] M. Bloznelis, J. Jaworski, E. Godehardt, V. Kurauskas, and K. Rybarczyk, *Recent Progress in Complex Network Analysis Properties of Random Intersection Graphs*, to appear in Studies in Classification, Data Analysis, and Knowledge Organization.
- [8] M. Bloznelis and V.Kurauskas, *Large cliques in sparse random intersection graphs*, arXiv:1302.4627 (2013).
- [9] B. Bollobás, *The chromatic number of random graphs*, Combinatorica 8 (1988), pp. 49–56.
- [10] B. Bollobás and P. Erdős, *Cliques in random graphs*, Math. Proc. Camb. Phil. Soc. 80 (1976), pp. 419–427.
- [11] T. Bohman, *The triangle-free process*, Adv. Math.,221 (2009), pp. 1653–1677.
- [12] A.M. Frieze, *On the independence number of random graphs*, Discrete Math. 81 (1990), pp. 171–175.

- [13] A.M. Frieze and C. McDiarmid, *Algorithmic theory of random graphs*, Random Structures Algorithms, 10 (1997), pp. 5–42.
- [14] E. Godehardt and J. Jaworski, *Two models of random intersection graphs for classification*. In: Studies in Classification, Data Analysis and Knowledge Organization 22, O. Optiz and M. Schwaiger, ed., Springer, Berlin, 2003, pp. 67–82.
- [15] M. Gauy, H. Hàn and I. Oliveira, *Erdős-Ko-Rado for random hypergraphs: asymptotic bounds and stability*, arXiv:1409.3634 (2014).
- [16] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, Wiley, 2001.
- [17] M. Karoński, E. R. Scheinerman and K. B. Singer-Cohen, *On Random Intersection Graphs: The Subgraph Problem*, Combin. Probab. Comput., 8 (1999), pp. 131–159.
- [18] T. Łuczak, *The chromatic number of random graphs*, Combinatorica, 11 (1991), pp. 45–54.
- [19] D. Matula, *The largest clique size in a random graph*, Tech. Rep., Dept. Comp. Sci., Southern Methodist University, Dallas, Texas (1976).
- [20] C. McDiarmid, *Concentration*, in Probabilistic Methods for Algorithmic Discrete Mathematics, Algorithms Combin. 16, Springer, Berlin, 1998, pp. 195–248.
- [21] S. E. Nikolettseas, C. Raptopoulos and P. G. Spirakis, *Colouring non-sparse random intersection graphs*, in Mathematical Foundations of Computer Science 2009, LNCS 5734, Springer Berlin Heidelberg, 2009, pp. 600–611.
- [22] N. Pippenger and J. Spencer, *Asymptotic behavior of the chromatic index for hypergraphs*, J. Combin. Theory Ser. A, 51 (1989), pp. 24–42.
- [23] K. Rybarczyk, *Constructions of independent sets in random intersection graphs*, Theoret. Comput. Sci., 524 (2014), pp. 103–125.
- [24] V. G. Vizing, *On an estimate of the chromatic class of a p -graph*, Diskret. Analiz., 3 (1964), pp. 25–30.
- [25] N. C. Wormald, *Differential equations for random processes and random graphs*, Ann. Appl. Probab. 5 (1995), pp. 1217–1235.