Discussiones Mathematicae
Graph Theory xx (xxxx) 1–17
https://doi.org/10.7151/dmgt.2431

COVERING THE EDGES OF A RANDOM HYPERGRAPH BY CLIQUES

Vojtěch Rödl[1]

Emory University
Atlanta, GA

e-mail: rodl@mathcs.emory.edu

and

Andrzej Ruciński[2]

A. Mickiewicz University
Poznań, Poland

e-mail: rucinski@amu.edu.pl

Abstract

We determine the order of magnitude of the minimum clique cover of the edges of a binomial, r-uniform, random hypergraph $G^{(r)}(n,p)$, p fixed. In doing so, we combine the ideas from the proofs of the graph case ($r=2$) in Frieze and Reed [*Covering the edges of a random graph by cliques*, Combinatorica 15 (1995) 489–497] and Guo, Patten, Warnke [*Prague dimension of random graphs*, manuscript submitted for publication].

Keywords: r-uniform random hypergraph, clique covering.

2010 Mathematics Subject Classification: 05C65, 05C80,05C62, 05C69, 05C70, 05C75.

1. Introduction

For an r-uniform hypergraph (briefly, an r-graph) $H=(V,E)$ and a set S, a *representation of* H *on* S is an assignment of subsets $S_v \subset S$, $v \in V$, in such a way that for each $R \in \binom{V}{r}$ we have $\bigcap_{v\in R} S_v \neq \emptyset$ if and only if $R \in E$. To observe that any r-graph admits such a representation, assign to each vertex v the set $\{e:\ v \in e \in E\}$ of all edges e containing v. Then, $\{v_1,\ldots,v_r\} \in E$ if and only if $\bigcap_{j+1}^{r} S_{v_j} \neq \emptyset$.

[1]Research supported by NSF grant DMS 1764385.

[2]Research supported by Narodowe Centrum Nauki, grant 2018/29/B/ST1/00426 Part of research performed at Emory University, Atlanta.

Definition. The *representation number* $\theta_1(H)$ of H is the smallest cardinality of a set S which admits a representation of H. Equivalently, $\theta_1(H)$ is also the smallest number of cliques needed to cover all edges of H (see Appendix for a proof of the equivalence).

It is perhaps interesting to note (cf. [6]) that the maximum of $\theta_1(H)$ over all r-graphs H on n vertices equals the Turán number for the r-uniform clique $K_{r+1}^{(r)}$ on $r+1$ vertices which is unknown even for $r=3$.

We determine a typical order of magnitude of the parameter θ_1 for a class of large *random* r-graphs. Given integers $n, r \geq 2$, and a real $0 < p < 1$, let $G^{(r)}(n,p)$ denote the random r-graph obtained by independent inclusion of each r-set with probability p. In particular, the number of edges of $G^{(r)}(n,p)$ is binomially distributed with expectation $\binom{n}{r}p$. We say that a property of r-sets $\mathcal{P}$ holds *asymptotically almost surely*, abbreviated to *a.a.s.*, if the probability $\mathrm{Prob}(G^{(r)}(n,p) \in \mathcal{P}) \to 1$ as $n \to \infty$. Throughout the paper p remains independent of n, while all logarithms are natural and denoted by log.

Theorem 1. *For every integer $r \geq 2$ and a constant $0 < p < 1$, there exist positive constants c_1 and c_2 such that a.a.s.*

$$c_1 \frac{n^r}{(\log n)^{r/(r-1)}} \leq \theta_1(G^{(r)}(n,p)) \leq c_2 \frac{n^r}{(\log n)^{r/(r-1)}}.$$

The case $r=2$ was proved by Frieze and Reed [3], and, in a stronger form, by Guo, Patten, and Warnke [4]. Here we follow the ideas from there, some of which have already originated in a paper by Alon, Kim, and Spencer [1]. The lower bound follows immediately from the upper bound on the order of the largest clique in $G^{(r)}(n,p)$ (see below). Hence, in the remaining sections we focus on the upper bound only.

***Proof of the lower bound in Theorem* 1.** Recall that p and r are constants independent of n. The expected number of cliques of order

$$t := \left\lceil \left(\frac{r!}{\log(1/p)} \log n \right)^{1/(r-1)} \right\rceil + r$$

in $G^{(r)}(n,p)$ is

$$\binom{n}{t} p^{\binom{t}{r}} \leq \left(\frac{en}{t} p^{(t-r)^{r-1}/r!} \right)^t \leq \left(\frac{en}{t} e^{-\log n} \right)^t = o(1)$$

as $n \to \infty$. Hence, a.a.s., there are no cliques of order t (or higher). On the other hand, by Chebyshev's inequality, there are, a.a.s., at least $\frac{1}{2}p\binom{n}{r}$ edges in

$G^{(r)}(n,p)$. Therefore, a.a.s., one needs at least

$$\frac{\frac{1}{2}p\binom{n}{r}}{\binom{t}{r}} \geq \frac{1}{2}p\left(\frac{n}{t}\right)^r = \Omega\left(\frac{n^r}{(\log n)^{r/(r-1)}}\right)$$

cliques to cover all edges of $G^{(r)}(n,p)$. ■

For the proof of the upper bound we now define a crucial notion. Let an r-graph G on a vertex set V, two integers $0 \leq s < j \leq |V|$, and a set $S \in \binom{V}{s}$ be given. A subset $J \subset V$ is called an *(S,j)-clique in G* if

- $J \supset S$, $|J| = j$, and
- $E_J := \{f \in \binom{J}{r} : f \cap (J \setminus S) \neq \emptyset\} \subset E(G)$, that is, $E(G)$ contains all r-element subsets of J except those which are subsets of S.

Note that $|E_J| = \binom{j}{r} - \binom{s}{r}$. Moreover, for $s = 0$, an $(\emptyset, j)$-clique is just any copy of the clique $K_j^{(r)}$ in G, while for $j < r$ every $J \in \binom{V}{j}$, $J \supseteq S$, is a (trivial) (S,j)-clique (with $E_J = \emptyset$).

2. AN EXPANDING PROPERTY OF RANDOM r-GRAPHS

Throughout the paper V is an n-vertex set of the random r-graph $G^{(r)}(n,p)$. Given $s \geq r-1$, and a set $S \in \binom{V}{s}$, let $X(S)$ be the number of $(S, s+1)$-cliques in $G^{(r)}(n,p)$. In other words, $X(S)$ counts the common neighbors of all $(r-1)$-element subsets of S. Clearly,

$$\mathbf{E}(X(S)) = (n-s)p^{\binom{s}{r-1}} := \mu(s). \tag{1}$$

The next lemma asserts that, for a wide range of s and for every s-element set S of vertices there are roughly the same number of $(S, s+1)$-cliques in $G^{(r)}(n,p)$. Let

$$k = \left\lfloor (\alpha \log n)^{1/(r-1)} \right\rfloor$$

for sufficiently small $\alpha > 0$.

Claim 2. *Let $\mathcal{A}$ be the event that for all $r-1 \leq s \leq k-1$ and all $S \in \binom{V}{s}$*

$$|X(S) - \mathbf{E}(X(S))| \leq n^{-1/3}\mathbf{E}(X(S)).$$

Then, $\text{Prob}(\mathcal{A}) = 1 - o(1)$.

Proof. Note that $\mu(s)$ is a decreasing function of s and, by the definition of k above, $\binom{k-1}{r-1} \leq (k-1)^{r-1} \leq \alpha \log n$. Thus, for large n,

$$\mu(s) \geq \mu(k-1) \geq (n/2)p^{\binom{k-1}{r-1}} \geq (n/2)p^{\alpha \log n} \geq \frac{1}{2}n^{0.99}, \tag{2}$$

if only $\alpha \log(1/p) \le 0.01$.

Recall that, for every $S \in \binom{V}{s}$, the random variable $X(S)$ counts the number of $(S, s+1)$-cliques and so, it is binomially distributed with expectation given by (1). Thus, by Chernoff's bound (see, e.g., [5], Corollary 2.3, Ineq. (2.9)), assuming n is sufficiently large,

$$\mathrm{Prob}(|X(S) - \mu(s)| > n^{-1/3}\mu(s)) \le 2\exp\{-n^{-2/3}\mu(s)/3\} \le \exp\{-n^{1/4}\},$$

where the last inequality follows from (2). Finally, by the union bound, summing over all choices of s and S,

$$\mathrm{Prob}(\neg\mathcal{A}) \le kn^k \exp\{-n^{1/4}\} = o(1). \qquad \blacksquare$$

3. Proof of Theorem 1

The upper bound in Theorem 1 will be a consequence of Claim 2 given in Section 2 and Lemma 4 to be stated below.

3.1. Notation

Before stating Lemma 4, we introduce a few parameters used therein. Very roughly, the lemma will claim the existence of a sequence of r-graphs $G_1, \dots, G_{i_0}$,

$$i_0 = \left\lceil \frac{r+1}{r-1} \log\log n \right\rceil,$$

which begins with the random r-graph $G_1 := G^{(r)}(n,p)$ and maintains throughout certain properties. As the proof of Lemma 4 will reveal, each next graph G_{i+1} will be derived from G_i by a random deletion of cliques of order k_i, where, for sufficiently small $\alpha > 0$,

$$k_i = \left\lfloor \left(\frac{\alpha \log n}{i} \right)^{1/(r-1)} \right\rfloor.$$

(Note that $k_1 = k$ defined earlier.)

In addition, some random edges of G_i will be deleted as well. The random procedure will be designed in such a way that the graphs G_i will shrink at the rate of $1/e$, thus resembling random r-graphs $G^{(r)}(n, p_i)$, where

$$p_i = pe^{1-i}.$$

The resemblance will be manifested by the behavior of the number of (S, j)-cliques in G_i, which, for all $0 \le s < j \le k_i$, will be close to the quantity

$$\mu_i(s,j) = \binom{n-s}{j-s} p_i^{\binom{j}{r} - \binom{s}{r}}. \tag{3}$$

Note that $\mu_i(s,j)$ is the expected number of (S,j)-cliques in a random r-graph $G^{(r)}(n,p_i)$.

In particular, for $s \ge r-1$, the quantity $\mu_1(s,s+1) = \mu(s)$ has been defined in (1). Note also that

$$\frac{\mu_{i+1}(s,j)}{\mu_i(s,j)} = (1/e)^{\binom{j}{r}-\binom{s}{r}}. \tag{4}$$

Let us now prove some bounds on $\mu_i(s,j)$.

Claim 3. *For all $1 \le i \le i_0$, all $0 \le s < j \le k_i$, sufficiently small α and sufficiently large n,*

(a) $\frac{\mu_i(s+1,j)}{\mu_i(s,j)} \le n^{-0.99}$;

(b) $\mu_i(s,j) \ge n^{0.99}$.

Proof. (a) Since $s < k_i \le \left(\frac{\alpha \log n}{i}\right)^{1/(r-1)}$, we have $is^{r-1} < \alpha \log n$. Hence, for sufficiently small α and large n,

$$\frac{\mu_i(s+1,j)}{\mu_i(s,j)} = \frac{j-s}{n-s} p_i^{-\binom{s}{r-1}} < \frac{j}{n}(e^i/p)^{s^{r-1}} \le \frac{j}{n}(e/p)^{is^{r-1}} < \frac{j}{n}(e/p)^{\alpha \log n} \le n^{-0.99}.$$

(b) By (a), $\mu_i(s,j)$ decreases with growing s. Thus, similarly to (a), since

$$ij^{r-1} \le ik_i^{r-1} \le \alpha \log n,$$

we have

$$\mu_i(s,j) \ge \mu_i(j-1,j) = (n-j+1)p_i^{\binom{j-1}{r-1}} \ge \frac{n}{2}\left(p/e^{i-1}\right)^{j^{r-1}} \ge \frac{n}{2}(p/e)^{\alpha \log n} \ge n^{0.99},$$

for sufficiently small α. ∎

Finally, as an important part of the forthcoming lemma is a sequence of r-graphs $G_1, \ldots, G_{i_0}$, for given $0 \le s < j \le k_i$ and $S \in \binom{V}{s}$, we denote by $N_i(S,j)$ the number of (S,j)-cliques in G_i. Note that $N_1(S,s+1)$ is the deterministic counterpart of the random variable $X(S)$ appearing in Claim 2. In particular, if $G_1 \in \mathcal{A}$, then

$$|N_1(S,s+1) - \mu_1(s,s+1)| \le n^{-1/3}\mu_1(s,s+1). \tag{5}$$

Note also that, for $s=0$, $N_i(\emptyset,j)$ is just the number of cliques of order j in G_i, in particular, $N_i(\emptyset,r) = |G_i|$, the number of edges of G_i. (From now on, we will denote the number of edges of an r-graph G by $|G|$.) Finally, notice that by a comment at the end of Section 1 and by (3), for $j < r$,

$$N_i(S,j) = \binom{n-s}{j-s} = \mu_i(s,j). \tag{6}$$

3.2. Statement of Lemma 4 and proof of Theorem 1

Here we state a crucial, technical lemma from which Theorem 1 will follow. Out of the three properties listed therein, the second one, $\mathcal{Q}_i$, is there just to facilitate the proof. All parameters appearing in the statement have been defined in the previous subsection.

Lemma 4. *For every n-vertex r-graph $G_1 \in \mathcal{A}$, where the event $\mathcal{A}$ is defined in Claim 2, there exist a descending sequence of r-graphs*

$$G_1 \supset G_2 \supset \cdots \supset G_{i_0}$$

and an ascending sequence of families of cliques

$$\emptyset = \mathcal{C}_1 \subset \mathcal{C}_2 \subset \cdots \subset \mathcal{C}_{i_0}$$

such that the three properties below hold.

($\mathcal{P}_i$) *For all $2 \le i \le i_0$, $\mathcal{C}_i$ is a clique cover of $G_1 - G_i$ and*

$$|\mathcal{C}_i \setminus \mathcal{C}_{i-1}| \le \frac{2p_{i-1}n^r}{k_{i-1}^r}.$$

($\mathcal{Q}_i$) *For all $1 \le i \le i_0$, all $0 \le s \le k_i - 1$, and all $S \in \binom{V}{s}$,*

$$|N_i(S, s+1) - \mu_i(s, s+1)| \le in^{-1/3}\mu_i(s, s+1). \tag{7}$$

($\mathcal{R}_i$) *For all $1 \le i \le i_0$, all $0 \le s < j \le k_i$, and all $S \in \binom{V}{s}$,*

$$|N_i(S, j) - \mu_i(s, j)| \le n^{-1/4}\mu_i(s, j). \tag{8}$$

Note that for $j = s + 1$, property $\mathcal{R}_i$ is overwritten by $\mathcal{Q}_i$. This is because, $i \le i_0 \le \frac{r+1}{r-1} \log\log n$ and thus, for all n, the right-hand side of (7) is smaller than the right-hand side of (8). Also, by (6), for $j < r$ the left-hand side of (8) equals 0. For the same reason, whenever $s \le r - 2$, the left-hand side of (7) equals 0. This means that in these cases properties $\mathcal{Q}_i$ and $\mathcal{R}_i$ hold trivially.

We defer the proof of Lemma 4 for later. Now, we give a short proof of Theorem 1 based on Claim 2 and Lemma 4.

***Proof of Theorem* 1.** By Claim 2, the random r-graph $G_1 = G^{(r)}(n, p)$ a.a.s. satisfies event $\mathcal{A}$ and so, we are in position to fix $G_1 \in \mathcal{A}$ and apply Lemma 4. Obviously, the union $\mathcal{C}$ of the clique cover $\mathcal{C}_{i_0}$ and the edge set of the graph G_{i_0} form a clique cover of G_1. Further, recalling that $|G_{i_0}| = N_{i_0}(\emptyset, r)$, we have, by

$\mathcal{P}_i$, $i = 2, \ldots, i_0$, and by $\mathcal{R}_{i_0}$ applied only in one special case of $S = \emptyset$, $j = r$, and $i = i_0$,

$$|\mathcal{C}| = |\mathcal{C}_{i_0}| + |G_{i_0}| = \sum_{i=1}^{i_0-1} |\mathcal{C}_{i+1} \setminus \mathcal{C}_i| + N_{i_0}(\emptyset, r) \leq \sum_{i=1}^{i_0-1} \frac{2p_i n^r}{k_i^r} + (1 + n^{-1/4})\mu_{i_0}(0, r)$$
$$\leq \sum_{i=1}^{i_0-1} \frac{2pn^r i^{\frac{r}{r-1}}}{e^{i-1}(\alpha \log n)^{r/(r-1)}} + 2\binom{n}{r} p_{i_0} \leq \frac{2epn^r}{(\alpha \log n)^{r/(r-1)}} \sum_{i=1}^{\infty} \frac{i^{\frac{r}{r-1}}}{e^i} + \frac{2epn^r}{e^{i_0}}$$
$$\leq \frac{2epCn^r}{(\alpha \log n)^{r/(r-1)}} + \frac{2epn^r}{(\alpha \log n)^{(r+1)/(r-1)}} = \frac{2epC(1 + o(1))n^r}{(\alpha \log n)^{r/(r-1)}},$$

where $C = \sum_{i=1}^{\infty} \frac{i^{\frac{r}{r-1}}}{e^i}$. ∎

4. PREPARATIONS FOR THE PROOF OF LEMMA 4

First, we are going to show that property $\mathcal{R}_i$ is, in some sense, redundant. Nevertheless, we found it convenient to state it explicitly in Lemma 4.

4.1. $\mathcal{Q}_i$ implies $\mathcal{R}_i$

This short subsection is devoted to proving the following implication.

Claim 5. *For all $1 \leq i \leq i_0$, if $G_i \in \mathcal{Q}_i$, then $G_i \in \mathcal{R}_i$*

Proof. The key idea is to view (S, j)-cliques as a result of an iterative process of vertex by vertex "extensions" of the set S (with all required edges present). Fix $0 \leq s < j \leq k$, $S = \{v_1, \ldots, v_s\} \in \binom{V}{s}$ and let $\mathcal{N}_i(S, j)$ be the set of all (S, j)-cliques in G_i. Recall that $N_i(S, j) = |\mathcal{N}_i(S, j)|$, that is, $N_i(S, j)$ counts the number of sets $\{v_{s+1}, \ldots, v_j\} \subset V \setminus S$ such that $J = S \cup \{v_{s+1}, \ldots, v_j\}$ is an (S, j)-clique in G_i. Similarly, we define $N'_i(S, j)$ as the number of *sequences* $(v_{s+1}, \ldots, v_j)$ of $j - s$ *distinct* vertices in $V \setminus S$ such that, again, $J = S \cup \{v_{s+1}, \ldots, v_j\}$ is an (S, j)-clique in G_i. Equivalently,

$$N'_i(S, j) = |\{(v_{s+1}, \ldots, v_j) : S \cup \{v_{s+1}, \ldots, v_j\} \in \mathcal{N}_i(S, j)\}|.$$

We have, obviously,

$$N'_i(S, j) := (j - s)! N(S, j). \tag{9}$$

For all $s + 1 \leq \ell \leq j$ and $v_{s+1}, \ldots, v_\ell$, by property $\mathcal{Q}_i$ applied to the set $S_\ell := S \cup \{v_{s+1}, \ldots, v_\ell\}$, setting $V_\ell = \{v_{\ell+1} : S_\ell \cup \{v_{\ell+1}\} \in \mathcal{N}_i(S_\ell, \ell + 1)\}$,

$$\Big| |V_\ell| - \mu_i(s + \ell, s + \ell + 1) \Big| \leq in^{-1/3} \mu_i(s + \ell, s + \ell + 1).$$

Note that, by definition (3) and the standard combinatorial identity $\sum_{h=0}^{t-1}\binom{h}{r-1} = \binom{t}{r}$,

$$\prod_{h=s}^{j-1}\mu_i(h,h+1) = (n-s)_{(j-s)}p_i^{\sum_{h=s}^{j-1}\binom{h}{r-1}} = (j-s)!\mu_i(s,j).$$

Thus, observing that $N_i'(S,j) = \prod_{\ell=s}^{j-1}|V_\ell|$, we arrive at

$$\left(1-in^{-1/3}\right)^{j-s}(j-s)!\mu_i(s,j) \le N_i'(S,j) \le \left(1+in^{-1/3}\right)^{j-s}(j-s)!\mu_i(s,j).$$

Comparing with (9) and canceling $(j-s)!$ sidewise, this yields

$$\left(1-in^{-1/3}\right)^{j-s}\mu_i(s,j) \le N_i(S,j) \le \left(1+in^{-1/3}\right)^{j-s}\mu_i(s,j).$$

Finally, as $(j-s)in^{-1/3} \le k_i(i_0+1)n^{-1/3} = o(n^{-1/4})$, we conclude that

$$|N_i(S,j) - \mu_i(s,j)| \le n^{-1/4}\mu_i(s,j)$$

which means that G_i, indeed, satisfies property $\mathcal{R}_i$. ■

4.2. A random procedure

We intend to prove Lemma 4 by induction on i. Suppose that for some $1 \le i \le i_0 - 1$, a graph G_i and a clique cover $\mathcal{C}_i$ of $G_1 - G_i$ satisfy properties $\mathcal{P}_i$, $\mathcal{Q}_i$, and $\mathcal{R}_i$. To obtain $(G_{i+1}, \mathcal{C}_{i+1})$, we apply a random procedure during which we simultaneously select

- $\mathcal{K}_i$ — a random collection of cliques in G_i of order k_i, each chosen independently with probability

$$q_i := \frac{1}{(1+n^{-1/4})\mu_i(r,k_i)} \tag{10}$$

  and

- $\mathcal{E}_i$ — a random collection of edges $f \in G_i$, viewed as r-vertex cliques, each f chosen independently with probability

$$q_{i,f} = 1 - (1-q_i)^{(1+n^{-1/4})\mu_i(r,k_i) - N_i(f,k_i)}. \tag{11}$$

  Then, we set

- $\mathcal{C}_{i+1} := \mathcal{C}_i \cup \mathcal{K}_i \cup \mathcal{E}_i$, and
- $G_{i+1} = G_i - (\bigcup\mathcal{K}_i \cup \mathcal{E}_i)$, where $\bigcup\mathcal{K}_i$ is the set of edges covered by the union of cliques in $\mathcal{K}_i$.

(The idea of using such a random procedure has appeared in a similar context already in [1].)

The selections of $\mathcal{K}_i$ and $\mathcal{E}_i$ are performed simultaneously, that is, independently of each other. Note also that the exponent in (11) is, due to property $\mathcal{R}_i$, nonnegative. Finally, observe that for an edge $f \in G_i$, the probability that $f \in G_{i+1}$ equals

$$(1-q_i)^{N_i(f,k_i)}(1-q_{i,f}) = (1-q_i)^{(1+n^{-1/4})\mu_i(r,k_i)} \sim \frac{1}{e},$$

which explains the definition of p_i given earlier.

4.3. $\mathcal{R}_i$ implies $\mathcal{P}_{i+1}$

The following result is the first ingredient of the forthcoming probabilistic proof of Lemma 4.

Claim 6. *For all* $1 \le i \le i_0 - 1$, *if* $G_i \in \mathcal{R}_i$, *then, with probability at least* 0.49, *the pair* $(G_{i+1}, \mathcal{C}_{i+1})$ *satisfies property* $\mathcal{P}_{i+1}$.

Proof. Recall that, by the random procedure described in Subsection 4.2,

$$\mathcal{C}_{i+1} \setminus \mathcal{C}_i = \mathcal{K}_i \cup \mathcal{E}_i, \tag{12}$$

where $\mathcal{K}_i$ is a collection of k_i-cliques and $\mathcal{E}_i$ is a collection of edges selected randomly and independently from G_i.

As each k_i-clique is drawn with the same probability q_i, the quantity $|\mathcal{K}_i|$ is binomially distributed with expectation $\mathbf{E}|\mathcal{K}_i| = N_i(0,k_i) \times q_i$. This, for large n can be estimated, using property $\mathcal{R}_i$, the definition (3) of $\mu_i(s,j)$, and the divergence $k_i \to \infty$ as $n \to \infty$ (cf. definitions of k_i and i_0 in Subsection 3.1), as follows:

$$\begin{aligned}\mathbf{E}|\mathcal{K}_i| = N_i(0,k_i) \times q_i = \frac{N_i(0,k_i)}{(1+n^{-1/4})\mu_i(r,k_i)} &\le \frac{\mu_i(0,k_i)}{\mu_i(r,k_i)} = \frac{(n)_r}{(k_i)_r} p_i \\ \le \left(\frac{n}{k_i - r + 1}\right)^r p_i = (1+o(1))(n/k_i)^r p_i &\le 1.01 (n/k_i)^r p_i.\end{aligned}$$

Similarly, quantity $|\mathcal{E}_i|$ has a general binomial distribution with

$$\mathbf{E}|\mathcal{E}_i| = \sum_{f \in G_i} q_{i,f}. \tag{13}$$

For $f \in G_i$, by property $\mathcal{R}_i$ and Bernoulli's inequality, we have

$$q_{i,f} \le 1 - (1-q_i)^{2n^{-1/4}\mu_i(r,k_i)} \le 2n^{-1/4}\mu_i(r,k_i) q_i \le 2n^{-1/4}. \tag{14}$$

Consequently, bounding crudely $|G_i| \le n^r$, by (13) and (14),

$$\mathbf{E}|\mathcal{E}_i| \le n^r \times 2n^{-1/4} = O(n^{r-1/4}).$$

Further, observe that by the definitions of p_i, k_i, and i_0,

$$\frac{k_i^r}{p_i} \le \frac{e^i k_i^r}{pe} \le \frac{e^{i_0} k_i^r}{pe} = O\left((\log n)^{\frac{r+1}{r-1}+\frac{r}{r-1}}\right) = O\left((\log n)^{\frac{2r+1}{r-1}}\right).$$

Thus, $\mathbf{E}|\mathcal{E}_i| = O(n^{r-1/4}) = o\left((n/k_i)^r p_i\right)$, and, by (12), we have

$$\mathbf{E}|\mathcal{C}_{i+1} \setminus \mathcal{C}_i| = \mathbf{E}|\mathcal{K}_i| + \mathbf{E}|\mathcal{E}_i| \le (1.01 + o(1))(n/k_i)^r p_i.$$

Finally, by Markov's inequality,

$$\mathrm{Prob}(|\mathcal{C}_{i+1} \setminus \mathcal{C}_i| > 2(n/k_i)^r p_i) \le 0.51.$$

It means that property $\mathcal{P}_{i+1}$ holds for $(G_{i+1}, \mathcal{C}_{i+1})$ with probability at least 0.49. ∎

5. Proof of Lemma 4

We are going to show the existence of sequences $G_1 \supset G_2 \supset \cdots \supset G_{i_0}$ and $\emptyset = \mathcal{C}_1 \subset \mathcal{C}_2 \subset \cdots \subset \mathcal{C}_{i_0}$ satisfying properties $\mathcal{P}_i$, $\mathcal{Q}_i$, and $\mathcal{R}_i$, by induction on $i = 1, \ldots, i_0$.

Let us begin with the base case $i = 1$. For a fixed $G_1 \in \mathcal{A}$, property $\mathcal{Q}_1$ follows from Claim 2 (cf. (5)), while property $\mathcal{R}_1$ is implied by $\mathcal{Q}_1$, as shown in Claim 5.

Assuming now that for some $i \ge 1$ a pair $(G_i, \mathcal{C}_i)$ satisfies properties $\mathcal{P}_i$ (only for $i \ge 2$), $\mathcal{Q}_i$, and $\mathcal{R}_i$. We are going to show that with positive probability the pair $(G_{i+1}, \mathcal{C}_{i+1})$, chosen randomly according to the procedure described in Subsection 4.2, satisfies $\mathcal{P}_{i+1}$, $\mathcal{Q}_{i+1}$, and $\mathcal{R}_{i+1}$, and thus, such a pair exists.

By Claim 5, property $\mathcal{Q}_{i+1}$ implies property $\mathcal{R}_{i+1}$. Moreover, by Claim 6, property $\mathcal{P}_{i+1}$ holds for $(G_{i+1}, \mathcal{C}_{i+1})$ with probability at least 0.49. Thus, it suffices to prove that G_{i+1} satisfies property $\mathcal{Q}_{i+1}$ with probability strictly greater than 0.51. In fact, the latter probability will turn out to be $1 - o(1)$.

We begin with estimating the expectation of

$$X := N_{i+1}(S, s+1)$$

the (random) number of $(S, s+1)$-cliques in G_{i+1}.

Claim 7. *For all $1 \le i \le i_0 - 1$, if $G_i \in \mathcal{Q}_i$, then, for all $r - 1 \le s < k_i$, and all $S \in \binom{V}{s}$,*

$$|\mathbf{E}X - \mu_{i+1}(s, s+1)| \le (i + 0.5)n^{-1/3}\mu_{i+1}(s, s+1).$$

Proof. Fix $r-1 \le s < k_i$ and $S \in \binom{V}{s}$ and recall our notation E_J for the set of all edges of an (S,j)-clique J and $\mathcal{N}_i(S,j)$ for the family of all (S,j)-cliques in G_i. By linearity of expectation,

$$\mathbf{E}X = \sum_{J \in \mathcal{N}_i(S,s+1)} \mathrm{Prob}(E_J \subset G_{i+1}). \tag{15}$$

To estimate $\mathrm{Prob}(E_J \subset G_{i+1})$, observe that an $(S,s+1)$-clique J of G_i "survives" into G_{i+1} if none of its edges was selected to $\mathcal{E}_i$ or belonged to some k_i-clique selected to $\mathcal{K}_i$. The probability of the former event is $\prod_{f\in E_J}(1-q_{i,f})$, while the probability of the latter event is $(1-q_i)^{|\mathcal{U}|}$, where $\mathcal{U} := \bigcup_{f\in E_J}\mathcal{N}_i(f,k_i)$. Set

$$m_1 = |\mathcal{U}| - \sum_{f\in E_J} N_i(f,k_i) \quad \text{and} \quad m_2 = (1+n^{-1/4})\mu_i(r,k_i)|E_J|.$$

Then, using (11), we infer that

$$\mathrm{Prob}(E_J \subset G_{i+1}) = (1-q_i)^{|\mathcal{U}|}\prod_{f\in E_J}(1-q_{i,f}) = (1-q_i)^{m_1+m_2}. \tag{16}$$

Next, we separately find lower and upper bounds on $(1-q_i)^{m_1}$ and $(1-q_i)^{m_2}$. By Bonferroni's inequality, property $\mathcal{R}_i$ (which follows from $\mathcal{Q}_i$, see Claim 5), and the monotonicity of $\mu_i(t,k_i)$ as a function of t (see Claim 3(a)), the quantity $-m_1$ can be bounded as follows:

$$\begin{aligned} 0 \le -m_1 &\le \sum_{g,h\in E_J, g\ne h} N_i(g\cup h,k_i) \le \sum_{g,h\in E_J,g\ne h}(1+n^{-1/4})\mu_i(|g\cup h|,k_i) \\ &\le \sum_{g,h\in E_J,g\ne h}(1+n^{-1/4})\mu_i(r+1,k_i) \le |E_J|^2(1+n^{-1/4})\mu_i(r+1,k_i). \end{aligned} \tag{17}$$

(Above, we maximized $\mu_i(|g\cup h|,k_i)$ by minimizing $|g\cup h|$ which achieves minimum at $r+1$.) Note that

$$|E_J| = \binom{s}{r-1} \le k^{r-1} = \Theta(\log n). \tag{18}$$

Consequently, by Claim 3(a) and the definition (10) of q_i,

$$\begin{aligned} 1 &\le (1-q_i)^{m_1} \le \exp\left\{q_i|E_J|^2\left(1+n^{-1/4}\right)\mu_i(r+1,k_i)\right\} \\ &= \exp\left\{\frac{|E_J|^2\mu_i(r+1,k_i)}{\mu_i(r,k_i)}\right\} \overset{Cl.3(\mathrm{a})}{\le} \exp\left\{|E_J|^2 n^{-0.99}\right\} = 1+o\left(n^{-0.98}\right). \end{aligned} \tag{19}$$

Further, by Claim 3(b), (10), and (18),

$$\tfrac{q_i|E_J|}{1-q_i} \leq 2q_i|E_J| = \frac{O(\log n)}{\left(1+n^{-1/4}\right)\mu_i(r,k_i)} \leq 2n^{-0.99}\Theta(\log n) = o\left(n^{-0.98}\right).$$

This implies that

$$\begin{aligned} e^{-|E_J|} &\geq (1-q_i)^{m_2} \geq \exp\left\{-\frac{|E_J|}{1-q_i}\right\} \geq e^{-|E_J|}\left(1-\frac{q_i|E_J|}{1-q_i}\right) \\ &\geq e^{-|E_J|}\left(1-o(n^{-0.98})\right). \end{aligned} \tag{20}$$

Thus, by (16), (19), and (20),

$$\left(1-o(n^{0.98})\right)e^{-|E_J|} \leq \mathrm{Prob}(E_J \subset G_{i+1}) \leq \left(1+o(n^{-0.98})\right)e^{-|E_J|}. \tag{21}$$

Recall that, by property $\mathcal{Q}_i$, $N_i(S,s+1) \leq (1+in^{-1/3})\mu_i(s,s+1)$, while, by (4), $\mu_i(s,s+1)e^{-\binom{s}{r-1}} = \mu_{i+1}(s,s+1)$. Thus, using also (15) and (21), and recalling that $|E_J| = \binom{s}{r-1}$, we finally have

$$\begin{aligned} \mathbf{E}X &\leq N_i(S,s+1)\left(1+o(n^{-0.98})\right)e^{-\binom{s}{r-1}} \\ &\leq \left(1+o(n^{-0.98})\right)(1+in^{-1/3})\mu_i(s,s+1)e^{-\binom{s}{r-1}} \\ &\overset{(4)}{=} \left(1+o(n^{-0.98})\right)\left(1+in^{-1/3}\right)\mu_{i+1}(s,s+1) \\ &\leq \left(1+(i+0.5)n^{-1/3}\right)\mu_{i+1}(s,s+1) \end{aligned}$$

and, similarly, $\mathbf{E}X \geq \left((1-(i+0.5)n^{-1/3}\right)\mu_{i+1}(s,s+1)$. ■

In view of the above claim, to establish property $\mathcal{Q}_{i+1}$ of G_{i+1}, it remains to show that X is concentrated around its expectation with probability very close to 1. In doing so, similarly to [4], we will utilize the following Azuma-type concentration inequality which can be deduced from [7], Theorem 3.8 (see also [8], Corollary 1.4).

Lemma 8. *Let $X_1,\dots,X_M$ be 0–1 independent random variables and let $f : \{0,1\}^{[M]} \to \mathbb{R}$ satisfy Lipschitz condition (L) with constants $c_1,\dots,c_M$:*

(L) for all $(z_1,\dots,z_M) \in \{0,1\}^{[M]}$ and $(z'_1,\dots,z'_M) \in \{0,1\}^{[M]}$, and all $1 \leq m \leq M$,

$$|f(z_1,\dots,z_M) - f(z'_1,\dots,z'_M)| \leq c_m, \quad \textit{whenever } z_h = z'_h \textit{ for all } h \neq m.$$

Set

$$X = f(X_1, \dots, X_M), \quad W = \sum_{m=1}^{M} c_m^2 \operatorname{Prob}(X_m = 1), \quad \textit{and} \quad C = \max_{1 \le m \le M} c_m.$$

Then, for every $t \geq 0$,

$$\operatorname{Prob}(|X - \mathbf{E}X| \geq t) \leq 2 \exp\left\{ -\frac{t^2}{2(W + Ct)} \right\}. \qquad \blacksquare$$

Now, we are ready to provide the last ingredient of the proof of Lemma 4.

Claim 9. *For all* $1 \le i \le i_0 - 1$, *if* $G_i \in \mathcal{Q}_i$, *then, a.a.s.,* $G_{i+1} \in \mathcal{Q}_{i+1}$.

Proof. Fix $r - 1 \le s < k_i$ and $S \in \binom{V}{s}$, and notice that if J is an $(S, s+1)$-clique in G_{i+1}, then it must have also been an $(S, s+1)$-clique in G_i, whose all edges "survived" the random procedure described in Subsection 4.2.

Recall that $\mathcal{N}_i(\emptyset, k_i)$ denotes the set of all k_i-cliques in G_i and that $N_i(\emptyset, k_i) = |\mathcal{N}_i(\emptyset, k_i)|$. We set $M_1 = N_i(\emptyset, k_i)$ and $\mathcal{N}_i(\emptyset, k_i) = \{K_1, \dots, K_{M_1}\}$. Let X_m, $m = 1, \dots, M_1$, be the indicator random variable which equals 1 if $K_m \in \mathcal{K}_i$ and 0 otherwise. Similarly, set $M_2 = |G_i|$ and $G_i = \{f_1, \dots, f_{M_2}\}$, and denote by Y_m, $m = 1, \dots, M_2$, the indicator random variable equal to 1 if $f_m \in \mathcal{E}_i$ and 0 otherwise.

As the events $K_m \in \mathcal{K}_i$, $m = 1, \dots, M_1$, and $f_m \in \mathcal{E}_i$, $m = 1, \dots, M_2$, fully determine the number of (S, s)-cliques left in G_{i+1}, there exists a function $f : \{0,1\}^{[M_1 + M_2]} \to \mathbb{R}$, such that

$$X = N_{i+1}(S, s+1) = f(X_1, \dots, X_{M_1}, Y_1, \dots, Y_{M_2}).$$

The explicit form of function f is not important for us.

As we are aiming at applying Lemma 8 to X, we need to find constants for which the Lipschitz condition (L) holds and then estimate W. Set

$$c_m = \max |f(x_1, \dots, x_{M_1}, y_1, \dots, y_{M_2}) - f(x'_1, \dots, \dots, x'_{M_1}, y_1, \dots, y_{M_2})|,$$

where the maximum is taken over all $(x_1, \dots, x_{M_1})$, $(x'_1, \dots, x'_{M_1}) \in \{0,1\}^{[M_1]}$ and $(y_1, \dots, y_{M_2}) \in \{0,1\}^{[M_2]}$ such that $x_h = x'_h$ for all $h \neq m$. Similarly, we set

$$d_m = \max |f(x_1, \dots, x_{M_1}, y_1, \dots, y_{M_2}) - f(x_1, \dots, \dots, x_{M_1}, y'_1, \dots, y'_{M_2})|,$$

where the maximum is taken over all $(x_1, \dots, x_{M_1}) \in \{0,1\}^{[M_1]}$ and $(y_1, \dots, y_{M_2})$, $(y'_1, \dots, y'_{M_2}) \in \{0,1\}^{[M_2]}$ such that $y_h = y'_h$ for all $h \neq m$. In other words, c_m and d_m are, respectively, upper bounds on the change of X due to flipping the outcome of the event $K_m \in \mathcal{K}_i$, respectively, $f_m \in \mathcal{E}_i$.

Now we estimate the Lipschitz parameters c_m and d_m taking into account the position of K_m and f_m with respect to the given set S. We begin with d_m as this case is easier. As an edge of G_i may belong to at most one $(S, s+1)$-clique J, the values of $X = N_{i+1}(S, s+1)$ for G_{i+1} with or without the edge f_m may differ by at most one. Thus,

$$d_m \le 1 \tag{22}$$

for all $m = 1, \dots, M_2$.

On the other hand, since $|J| = s+1$, any such J contains exactly $\binom{s}{r-1}$ edges of G_i, so there are altogether

$$N_i(S, s+1) \times \binom{s}{r-1} \le N_i(S, s+1)k^{r-1} \tag{23}$$

edges $f_m \in G_i$ whose removal could affect $X = N_{i+1}(S, s+1)$. Thus, for that many edges we put $d_m = 1$, while for all other edges $d_m = 0$.

Turning to c_m, by the same token, very crudely, for all $m = 1, \dots, M_1$,

$$c_m \le |K_m| = \binom{k_i}{r} \le k^r, \tag{24}$$

as every edge of K_m may belong to at most one $(S, s+1)$-clique J.

Moreover, $c_m > 0$ only if K_m contains at least one edge of some $(S, s+1)$-clique of G_i. There are $N_i(S, s+1)$ such $(S, s+1)$-cliques and each contains $\binom{s}{r-1}$ edges. In turn, by property $\mathcal{R}_i$, each edge f is contained in $N_i(f, k_i) \le (1 + n^{-1/4})\mu_i(r, k_i)$ k_i-cliques of G_i. Hence, there are at most

$$N_i(S, s+1) \times \binom{s}{r-1} \times \max_{f \in G_i} N_i(f, k_i) \le N_i(S, s+1)k^{r-1}(1+n^{-1/4})\mu_i(r, k_i) \tag{25}$$

cliques K_m in G_i which share an edge with some $(S, s+1)$-clique. This implies that for at most that many indices $m \in [M_1]$ we have $c_m > 0$.

Putting (22)–(25) together, one can bound the parameter W appearing in Lemma 8, using again property $\mathcal{R}_i$, the definition (10) of q_i, and the estimate (14) of $q_{i,f}$, as follows.

(26)
$$\begin{aligned} W = \sum_{m=1}^{M_1} c_m^2 q_i + \sum_{m=1}^{M_2} d_m^2 q_{i,f} &\overset{(14)}{\le} N_i(S, s+1)k^{r-1}(1 + n^{-1/4})\mu_i(r, k_i) \times k^{2r} \times q_i \\ + N_i(S, s+1)k^{r-1} \times 1^2 \times 2n^{-1/4} &\overset{(10)}{\le} (1+o(1))N_i(S, s+1)k^{3r-1} \overset{\mathcal{R}_i}{\le} \mu_i(s, s+1)k^{3r}. \end{aligned}$$

Recall that, by definition (3), $\mu_i(s,s+1) = (n-s)p_i^{\binom{s}{r-1}} \le n$, while, by equality (4), $\mu_{i+1}(s,s+1) = e^{-\binom{s}{r-1}}\mu_i(s,s+1) \le \mu_i(s,s+1)$. Moreover, by Claim 3(b) applied with $i+1$, $\mu_{i+1}(s,s+1) \ge n^{0.99}$. Putting all these facts together, we have

$$n^{0.99} \le \mu_{i+1}(s,s+1) \le \mu_i(s,s+1) \le n. \tag{27}$$

Thus, in view of (26), by Lemma 8 with

$$t = \frac{1}{2}n^{-1/3}\mu_{i+1}(s,s+1) \quad \text{ and } \quad C = \max\left\{\max_{1\le m\le M_1} c_m, \max_{1\le m\le M_2} d_m\right\} \le k^r,$$

noting that $Ct = o\left(\mu_i(s,s+1)k^{3r}\right)$ and taking n sufficiently large,

$$\begin{aligned}
&\mathrm{Prob}(X - \mathbf{E}X \ge t) \le 2\exp\left\{-\frac{\frac{1}{4}n^{-2/3}\mu_{i+1}^2(s,s+1)}{2(\mu_i(s,s+1)k^{3r} + Ct)}\right\} \\
&\le 2\exp\left\{-\frac{\mu_{i+1}^2(s,s+1)}{9\mu_i(s,s+1)n^{2/3}k^{3r}}\right\} \overset{(27)}{\le} 2\exp\left\{-\frac{n^{1.98}}{n^{5/3}k^{3r}}\right\} \le \exp\left\{-n^{0.3}\right\}.
\end{aligned}$$

In view of the above and using Claim 7 and the union bound, a.a.s., for all s and $S \in \binom{V}{s}$,

$$\begin{aligned}
X = N_{i+1}(S,s+1) &\le \mathbf{E}X + t \le \left(1 + (i+0.5)n^{-1/3} + 0.5n^{-1/3}\right)\mu_{i+1}(s,s+1) \\
&\le \left(1 + (i+1)n^{-1/3}\right)\mu_{i+1}(s,s+1)
\end{aligned}$$

and, similarly, $X \ge \left(1 - (i+1)n^{-1/3}\right)\mu_{i+1}(s,s+1)$, which completes the proof of Claim 9. ∎

Acknowledgments

We would like to thank Lutz Warnke for sharing with us an early draft of [4], as well as for enlightening correspondence on the topic of the paper. Thanks are also due to an anonymous referee for their remarks which led to an improvement of the manuscript.

References

[1] N. Alon, J.-H. Kim and J. Spencer, *Nearly perfect matchings in regular simple hypergraphs*, Israel J. Math. **100** (1997) 171–187. https://doi.org/10.1007/BF02773639

[2] B. Bollobás, P. Erdős, J. Spencer and D.B. West, *Clique coverings of the edges of a random graph*, Combinatorica **13** (1993) 1–5. https://doi.org/10.1007/BF01202786

[3] A. Frieze and B. Reed, *Covering the edges of a random graph by cliques*, Combinatorica **15** (1995) 489–497 (see arXiv:1103.4870v1 for corrected version). https://doi.org/10.1007/BF01192522

[4] H. Guo, K. Patten and L. Warnke, *Prague dimension of random graphs*, manuscript submitted for publication.

[5] S. Janson, T. Łuczak and A. Ruciński, Random Graphs (Wiley, 2000). https://doi.org/10.1002/9781118032718

[6] J. Lehel, *Covers in hypergraphs*, Combinatorica **2** (1982) 305–309. https://doi.org/10.1007/BF02579237

[7] C. McDiarmid, *Concentration*, in: Probabilistic Methods for Algorithmic Discrete Mathematics, (Springer, Berlin, 1998) 195–248. https://doi.org/10.1007/978-3-662-12788-9_6

[8] L. Warnke, *On the method of typical bounded differences*, Combin. Probab. Comput. **25** (2016) 269–299. https://doi.org/10.1017/S0963548315000103

Received 24 March 2021
Revised 6 September 2021
Accepted 6 September 2021

Appendix

The following folklore result was observed by many authors for graphs ($r = 2$) but there seems to be no published proof of the general case. Here we fill that gap.

Fact 1. *Let H be an r-graph and let $\theta_1(H)$ and $\tilde{\theta}_1(H)$ stand, respectively, for its representation number and minimum (edge) clique cover number. Then $\theta_1(H) = \tilde{\theta}_1(H)$.*

For the proof we need a simple observation.

Observation. *Let $H = (V, E)$ be an r-graph and let $\mathcal{S} = \{S_v \subset S : v \in V\}$ be a representation of H with the smallest set S. Then every element $s \in S$ belongs to at least r sets in $\mathcal{S}$.*

Proof. Suppose there is an $s \in S$ belonging to fewer than r sets in $\mathcal{S}$. Then $\mathcal{S}_s = \{S_v \setminus \{s\} : v \in V\}$ would also be a representation of H which contradicts the minimality of S. Indeed, for such an s and any R with $|R| = r$,

$$\bigcap_{v \in R} S_v \neq \emptyset \quad \text{if and only if} \quad \bigcap_{v \in R} (S_v \setminus \{s\}) \neq \emptyset.$$

■

***Proof of Fact* 1.** Let $\mathcal{S} = \{S_v \subset S : v \in V\}$ be a minimum representation of H, that is, a representation of size $|\mathcal{S}| = \theta_1(H)$. By the above Observation, for each $s \in S$ the set $C(s) = \{v : s \in S_v\}$ has size $|C(s)| \geq r$. What is more important, $C(s)$ is a clique in H. Indeed, if $\{v_1, \ldots, v_r\} \subset C(s)$, then $S_{v_1} \cap \cdots \cap S_{v_r} \ni s$, thus $\{v_1, \ldots, v_r\} \in H$. Moreover, each edge $\{v_1, \ldots, v_r\} \in H$ is covered by a clique $C(s)$, where $s \in S_{v_1} \cap \cdots \cap S_{v_r}$. Hence, $\tilde{\theta}_1(H) \leq \theta_1(H)$.

Conversely, let $\{C(s) : s \in S\}$ be a clique cover of H indexed by some (abstract) set S. For every vertex $v \in V$ consider the set

$$S_v = \{s \in S : v \in C(s)\}.$$

Next, observe that $\{v_1, \ldots, v_r\} \in E$ if and only if there is some $s \in S$ with $\{v_1, \ldots, v_r\} \subset C(s)$. We will draw two consequences of this equivalence. First, if $\{v_1, \ldots, v_r\} \in E$, then there exists $s \in S_{v_1} \cap \cdots \cap S_{v_r}$, implying that

$$S_{v_1} \cap \cdots \cap S_{v_r} \neq \emptyset.$$

However, if $\{v_1, \ldots, v_r\} \notin E$ then $\{v_1, \ldots, v_r\} \not\subset C(s)$ for all $s \in S$, which means that $S_{v_1} \cap \cdots \cap S_{v_r} = \emptyset$. Consequently, $\{S_v : v \in V\}$ is a representation of H, yielding $\theta_1(H) \leq \tilde{\theta}_1(H)$. ∎