

ARTICLE

Spanning F -cycles in random graphs

Alberto Espuny Díaz  and Yury Person 

Institut für Mathematik, Technische Universität Ilmenau, Ilmenau, Germany

Corresponding author: Alberto Espuny Díaz; Email: alberto.espuny-diaz@tu-ilmenau.de

(Received 17 September 2021; revised 16 December 2022; accepted 13 May 2023; first published online 9 June 2023)

Abstract

We extend a recent argument of Kahn, Narayanan and Park ((2021) *Proceedings of the AMS* 149 3201–3208) about the threshold for the appearance of the square of a Hamilton cycle to other spanning structures. In particular, for any spanning graph, we give a sufficient condition under which we may determine its threshold. As an application, we find the threshold for a set of cyclically ordered copies of C_4 that span the entire vertex set, so that any two consecutive copies overlap in exactly one edge and all overlapping edges are disjoint. This answers a question of Frieze. We also determine the threshold for edge-overlapping spanning K_r -cycles.

Keywords: Random graphs; threshold; spanning subgraph

2020 MSC Codes: Primary: 05C80

1. Introduction

The study of threshold functions for the appearance of spanning structures plays an important role in the theory of random graphs. Unlike in the case of small subgraphs, which was resolved by Erdős and Rényi [6] (for balanced graphs) and by Bollobás [4] (for general graphs), in the case of general spanning structures only sufficient conditions are known. These sufficient conditions lead to upper bounds for the threshold of a general spanning graph, although the expectation threshold conjecture of Kahn and Kalai [14], recently confirmed by Park and Pham [21], predicts the threshold for *any* graph up to a logarithmic factor.

Apart from particular structures where the thresholds are known, such as perfect matchings [7], F -factors [13], Hamilton cycles [16, 22] or spanning trees [18] (to name a few), the most general result providing upper bounds was, until recently, due to Riordan [23], giving in some cases asymptotically optimal upper bounds (lattices, hypercubes, k -th powers of Hamilton cycles for $k \geq 3$ [17]). An excellent survey by Böttcher [5] provides references to many other results, in particular algorithmic ones.

The recent breakthrough work by Frankston, Kahn, Narayanan and Park [8] (following ideas from a breakthrough of Alweiss, Lovett, Wu and Zhang [1] on the sunflower conjecture) established the fractional expectation threshold conjecture of Talagrand [26], leading in many cases to optimal thresholds, or providing an upper bound for the threshold which is off by at most a logarithmic factor. The subsequent work by Kahn, Narayanan and Park [15] exploited the proof approach of [8] in a more efficient way, allowing to erase the logarithmic factor in the case of the square of a Hamilton cycle, and thus proving the threshold for its appearance to be $n^{-1/2}$. Here, we generalise the approach from [15] and make it applicable to any structure satisfying a certain property, also erasing the logarithmic factor from the upper bound on their thresholds, and thus establishing the thresholds for several new structures. To showcase our result, we consider two specific families of spanning subgraphs.

In a recent paper, Frieze [10] studied thresholds for the containment of spanning K_r -cycles, i.e., cyclically ordered edge-disjoint copies of K_r with two consecutive copies sharing a vertex (assuming the right divisibility conditions on n). He proved the optimal threshold of the form $n^{-2/r} \log^{1/\binom{r}{2}} n$ by reducing this problem to another result of Riordan about coupling the random graph with the random r -uniform hypergraph [24] (see also the work of Heckel [11] for the triangle case). Frieze also raised the question about the threshold for the containment of a spanning C_4 -cycle, where the copies of C_4 are ordered cyclically and two consecutive cycles overlap in exactly one edge, whereby each cycle C_4 overlaps with two copies of C_4 in opposite edges (there are some possible variations, but this would be a canonically defined structure). Such C_4 -cycles are referred to in [10] as a C_4 -cycle with overlap 2, where it is also observed that the threshold for its appearance is at most $n^{-2/3} \log n$, which follows from [8].

The purpose of this paper is to contribute to the large body of work on thresholds for spanning structures by establishing thresholds for spanning 2-overlapping C_4 -cycles (which we denote by $C_{4,n}^e$), thus answering the question of Frieze [10], and also for 2-overlapping K_r -cycles (defined below) for $r \geq 4$. Both structures cannot be handled directly by the results in [8, 23]. We determine the thresholds for both of these structures.

Our first theorem answers the question of Frieze [10].

Theorem 1.1. *The threshold for the appearance of $C_{4,n}^e$ in $G(2n, p)$ is $\Theta(n^{-2/3})$.*

Our second result generalises the recent work of Kahn, Narayanan and Park [15] on the threshold for the square of a Hamilton cycle. The square of a Hamilton cycle can be seen as the particular case $r = 3$ of a structure which we call a 2-overlapping (or edge-overlapping) spanning K_r -cycle and denote by $K_{r,2,n}$, for $r \geq 3$. This consists of a set of cyclically ordered copies of K_r , where consecutive cliques share exactly one edge and, if $r \geq 4$, all other cliques are pairwise vertex-disjoint.

Theorem 1.2. *Let $r \geq 3$ and $n \in \mathbb{N}$ with $(r - 2) \mid n$. Then, the threshold for the appearance of $K_{r,2,n}$ in $G(n, p)$ is $\Theta(n^{-2/(r+1)})$.*

To prove Theorems 1.1 and 1.2, we state and prove a general lemma (the fragmentation lemma, Lemma 2.3), which has potential to handle more spanning structures. This lemma is a generalisation of the work of Kahn, Narayanan and Park on the square of a Hamilton cycle [15, Lemma 3.1] to handle structures for which constantly many rounds of exposure may be necessary, in contrast to [15], where only two rounds are used, and to [8], where logarithmically many rounds are necessary.

The organisation of the paper is as follows. In the next section, Section 2, we provide the main definitions, state a general lemma (the fragmentation lemma, Lemma 2.3) and use it to establish a general theorem (Theorem 2.2) about thresholds for certain spanning graphs. Theorem 2.2 is actually the main general result of the paper, and Theorems 1.1 and 1.2 are two of its applications. We prove these two applications in Section 3. Finally, in Section 4 we collect a few remarks, and in the appendices we provide the proof of Lemma 2.3 as well as several technical results for the applications.

2. A general theorem for thresholds

Given any real numbers a and b , we write $[a, b]$ to refer to the set $\{n \in \mathbb{Z} : a \leq n \leq b\}$. For an integer n , we often abbreviate $[n] := [1, n]$. We use standard O notation for asymptotic statements.

A hypergraph \mathcal{H} on the vertex set $V := V(\mathcal{H})$ is a subset of the power set 2^V . The elements of \mathcal{H} are referred to as edges. The hypergraph \mathcal{H} is said to be r -bounded if all its edges have cardinality at most r , and r -uniform if all the edges have exactly r vertices. Oftentimes, we will consider multihypergraphs \mathcal{H} on V , where we view \mathcal{H} as a multiset with elements from 2^V . We usually write

$|\mathcal{H}|$ for the number of edges of \mathcal{H} . To ease readability, we will often refer to multihypergraphs as hypergraphs. We also omit floor and ceiling signs whenever they do not affect our asymptotic computations.

Following [15], we say that a (multi-)hypergraph \mathcal{H} is q -spread if, for every $I \subseteq V(\mathcal{H})$, we have

$$|\mathcal{H} \cap \langle I \rangle| \leq q^{|I|} |\mathcal{H}|,$$

where $\langle I \rangle := \{J \subseteq V(\mathcal{H}) : I \subseteq J\}$ and $\mathcal{H} \cap \langle I \rangle$ is the set of edges of \mathcal{H} in $\langle I \rangle$ (with multiplicities if \mathcal{H} is a multihypergraph). The spreadness of \mathcal{H} is the minimum q such that \mathcal{H} is q -spread.

While the general framework developed in [8, 15] works for arbitrary hypergraph thresholds, here we focus on graphs. Let F be some (possibly spanning) subgraph of the complete graph K_n , and let \mathcal{F} denote the set of all copies of F in K_n . We will identify copies of F from \mathcal{F} with their edge sets, and we thus view \mathcal{F} as a k -uniform hypergraph, where $k = |E(F)|$, on the vertex set $M := \binom{[n]}{2}$.

With the notion of spreadness, (a consequence of) the main result of Frankston, Kahn, Narayanan and Park [8] can be stated as follows. Let F be a subgraph of K_n , and suppose that the hypergraph \mathcal{F} of all copies of F in K_n is q -spread. Then, there is an absolute constant C such that, if $p \geq Cq \log n$, then $\mathbb{P}[F \subseteq G(n, p)] \geq 1 - \varepsilon$. As mentioned in the introduction, our main contribution is removing the logarithmic factor in the bound on p above, under certain conditions on the hypergraph \mathcal{F} . We now define a strengthening of the notion of spreadness of hypergraphs which is key for our results.

Definition 2.1. For $q, \alpha, \delta \in (0, 1)$, we say that a k -bounded hypergraph \mathcal{F} on vertex set M is (q, α, δ) -superspread if it is q -spread and, for any $I \subseteq M$ with $|I| \leq \delta k$, we have

$$|\mathcal{F} \cap \langle I \rangle| \leq q^{|I|} k^{-\alpha c_I} |\mathcal{F}|,$$

where c_I is the number of components of I (when I is viewed as a subgraph of K_n).

The role of the term $k^{-\alpha c_I}$ will become clear later, but, roughly speaking, it will be responsible for bounding the threshold by $O(q/\alpha)$. The value of the constant δ actually plays no role in the result, but we do need it to be bounded away from 0 for our approach to work.

With the definition above, we may already state our main result.

Theorem 2.2. Let $d, \alpha, \delta, \varepsilon > 0$ with $\alpha, \delta < 1$. Then, there exists some C_0 such that, for all $C \geq C_0$ and $n \in \mathbb{N}$, the following holds. If F is a subgraph of K_n with $\Delta(F) \leq d$ and $k_0 := |E(F)| = \omega(1)$ and the hypergraph \mathcal{F} of all copies of F is (q, α, δ) -superspread with $q \geq 4k_0/(Cn^2)$, then, for $p \geq Cq$,

$$\mathbb{P}[F \subseteq G(n, p)] \geq 1 - \varepsilon.$$

This result immediately provides an upper bound of Cq for the threshold for the appearance of F as a subgraph of $G(n, p)$. If a matching lower bound can be found (say, by the standard first moment method), then this establishes the threshold for the appearance of any graph F which satisfies the conditions in the statement.

It is worth mentioning here that, shortly after announcing our results, Spiro [25] independently introduced a different notion of spreadness which further generalises the notion of superspreadness given in Definition 2.1. As such, the main result in [25] also implies Theorem 2.2. One advantage of our approach is that the definition of superspreadness seems to be more readily applicable than the more general definition of Spiro [25].

Before proving Theorem 2.2, let us introduce some definitions. Given a hypergraph \mathcal{H} , let $S \in \mathcal{H}$ and $X \subseteq V(\mathcal{H})$. For any $J \in \mathcal{H}$ such that $J \subseteq S \cup X$, we call the set $J \setminus X$ an (S, X) -fragment. Given some $k \in \mathbb{N}$, we say that the pair (S, X) is k -good if some (S, X) -fragment has size at most k , and we say it is k -bad otherwise.

More generally, let \mathcal{H}_0 be some k_0 -bounded (multi-)hypergraph. Let $k_0 \geq k_1 \geq \dots \geq k_t$ be a sequence of integers and X_1, \dots, X_t be a sequence of subsets of $V(\mathcal{H}_0)$. Then, we define a

sequence of k_i -bounded multihypergraphs $\mathcal{H}_1, \dots, \mathcal{H}_t$ inductively as follows. Let $i \in [t]$, and assume the hypergraph \mathcal{H}_{i-1} is already defined. Then, consider each $S \in \mathcal{H}_{i-1}$ such that (S, X_i) is a k_i -good pair, and let \mathcal{H}_i be the multihypergraph which arises by including, for each k_i -good pair (S, X_i) , an (arbitrary) (S, X_i) -fragment of size at most k_i . That is, we define $\mathcal{G}_i := \{S \in \mathcal{H}_{i-1} : (S, X_i) \text{ is } k_i\text{-good}\}$ and, for each $S \in \mathcal{G}_i$, $\mathcal{J}_i(S) := \{J \setminus X_i : J \in \mathcal{H}_{i-1}, J \subseteq S \cup X_i, |J \setminus X_i| \leq k_i\}$. We then fix an arbitrary function $f_i : \mathcal{G}_i \rightarrow \bigcup_{S \in \mathcal{G}_i} \mathcal{J}_i(S)$ such that $f_i(S) \in \mathcal{J}_i(S)$ for every $S \in \mathcal{G}_i$ (for instance, we may simply pick the lexicographically smallest element in the set) and define

$$\mathcal{H}_i := \{f_i(S) : S \in \mathcal{G}_i\}.$$

We will refer to the sequence $(\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_t)$ as a *fragmentation process* with respect to (k_1, \dots, k_t) and (X_1, \dots, X_t) . In our applications, we will let X_1, \dots, X_t be random subsets of $V(\mathcal{H}_0)$ and choose a suitable sequence k_0, \dots, k_t which will guarantee that the hypergraphs in the sequence do not become very small (with high probability). Observe that the fragments at the i -th step of this process (i.e., the edges of \mathcal{H}_i) correspond to subsets of the edges of \mathcal{H}_0 which have not been covered by the sets X_1, \dots, X_i . In particular, for all $i \in [t]$ and all $I \subseteq V(\mathcal{H}_0)$ we have that

$$|\mathcal{H}_i \cap \langle I \rangle| \leq |\mathcal{H}_0 \cap \langle I \rangle|. \tag{2.1}$$

The proof of Theorem 2.2 follows along similar lines as the proofs in [8, 15]: one proceeds in rounds of sprinkling random edges by showing that, after each round of exposure (which corresponds to a step of the fragmentation process), the random graph contains larger pieces of the desired structure (or, conversely, the missing fragments become smaller). In the general proof in [8], the authors show that the progress in each round shrinks the percentage of the edges from the desired structure by a factor of 0.9, which results in logarithmically many steps and, thus, a $\log n$ factor with respect to the fractional expectation threshold of the structure (this result is quite general, though, and oftentimes a logarithmic factor is indeed needed, as in the case of spanning trees, Hamilton cycles and K_r -factors in random graphs, or of perfect matchings and loose Hamilton cycles in random hypergraphs). The threshold for the square of a Hamilton cycle $K_{3,2,n}$ happens to be $n^{-1/2}$, and in this case, as shown in [15], two rounds suffice: the shrinkage factor there is $n^{-1/2}$, so that after the first round a second moment computation suffices in the second round of exposure/sprinkling. We show that the threshold for the appearance of F is at most $O(q)$ and for this we will need $1/\alpha$ rounds (exposing each time edges with probability Cq , for some constant C): the shrinkage factor in all but the last round will be $n^{-\alpha}$, so that we can apply the second moment method in the last round.

The following result is the main lemma of the paper. It will be used to iteratively build a copy of F in $G(n, p)$ through a fragmentation process.

Lemma 2.3. *Let $d, \alpha, \delta > 0$ with $\alpha, \delta < 1$. Then, there exists some C_0 such that, for all $C \geq C_0$ and $n \in \mathbb{N}$, the following holds. Let F be some subgraph of K_n with $\Delta(F) \leq d$ and $k_0 := |E(F)| = \omega(1)$, and let \mathcal{F} be the set of all copies of F in K_n . Assume that \mathcal{F} is (q, α, δ) -superspread with $q \geq 4k_0/(Cn^2)$ and that $(\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_i)$ is some fragmentation process with $\mathcal{H}_0 := \mathcal{F}$ such that, for each $j \in [i]$, \mathcal{H}_j is k_j -bounded and $|\mathcal{H}_j| \geq |\mathcal{H}_{j-1}|/2$, and $k_i = \omega(k_0^\alpha)$. Then, for $w := Cq \binom{n}{2}$, $k := k_i k_0^{-\alpha}$ and X chosen uniformly at random from $\binom{M}{w}$, we have*

$$\mathbb{E} \left[\left| \{(S, X) : S \in \mathcal{H}_i, (S, X) \text{ is } k\text{-bad}\} \right| \right] \leq 2C^{-k/3} |\mathcal{H}_i|. \tag{2.2}$$

The proof of Lemma 2.3 closely follows the proofs of [15, Lemma 3.1] and [8, Lemma 3.1]. Therefore, for the sake of completeness, we give its proof in Appendix A, for the convenience of the interested reader.

In the proof of Theorem 2.2 we also make use of the following auxiliary lemma. Again, its proof follows similarly as the proof of [15, Proposition 2.2], and we thus defer it to Appendix A.

Lemma 2.4. *Let F be a graph with f edges and maximum degree d . Then, the number of subgraphs I of F with ℓ edges and c components is at most*

$$(4ed)^\ell \binom{f}{c}.$$

With these auxiliary lemmas in hand, we can already prove our main result.

Proof of Theorem 2.2. We first note that, by adjusting the value of C_0 , we may assume that n is sufficiently large, and therefore k_0 is sufficiently large too. We may also assume that $q < C_0^{-1}$. In the beginning, we will switch and work with the $G(n, m)$ model instead of $G(n, p)$. (A random graph $G(n, m)$ is simply a uniformly chosen graph from the set of all labelled graphs on n vertices with exactly m edges.) This can be done easily since these models are essentially equivalent for $m = p\binom{n}{2}$ (see, e.g., [12, Proposition 1.12]).

We proceed as follows. We consider $G(n, m_1) \cup G(n, m_2) \cup \dots \cup G(n, m_t)$ with $t = \lceil 1/\alpha \rceil - 1$ and $m_i = Kq\binom{n}{2}$ for each $i \in [t]$, where K is assumed to be sufficiently large throughout (and C_0 will be defined as $2(t + 1)K$). We then define a fragmentation process on \mathcal{F} with respect to (k_1, \dots, k_t) and $(G(n, m_1), \dots, G(n, m_t))$, where the integers k_1, \dots, k_t will be defined shortly. We prove that a.s. each step of this fragmentation process satisfies the conditions of Lemma 2.3, so that we may iteratively apply it and conclude that each of the subsequent hypergraphs is not ‘too small’. At the end of this process, we will be sufficiently ‘close’ to a copy of F that a second moment argument will yield the result.

To be precise, we first consider the hypergraph $\mathcal{H}_0 := \mathcal{F}$ and take $X_1 := G(n, m_1)$ and $k_1 := k_0^{1-\alpha}$. We consider a first step in the fragmentation process. We obtain a multihypergraph \mathcal{H}_1 of (S, X_1) -fragments which is k_1 -bounded, where each S is an edge of \mathcal{H}_0 . In particular, by the assertion (2.2) of Lemma 2.3 and Markov’s inequality, we have that

$$\mathbb{P}[|\mathcal{H}_1| \geq |\mathcal{H}_0|/2] \geq 1 - 4K^{-k_1/3}. \tag{2.3}$$

Suppose now that we have already run the fragmentation process $(\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_i)$, for some $i \in [t - 1]$, and that $|\mathcal{H}_j| \geq |\mathcal{H}_{j-1}|/2$ for all $j \in [i]$. We run one further step of the fragmentation process with $X_{i+1} := G(n, m_{i+1})$ and $k_{i+1} := k_i k_0^{-\alpha}$ to obtain a k_{i+1} -bounded hypergraph \mathcal{H}_{i+1} of $(S, X_1 \cup \dots \cup X_{i+1})$ -fragments (where, again, each S is an edge of \mathcal{H}_0). By another application of Lemma 2.3 and Markov’s inequality, we obtain that

$$\mathbb{P}[|\mathcal{H}_{i+1}| \geq |\mathcal{H}_i|/2] \geq 1 - 4K^{-k_{i+1}/3}. \tag{2.4}$$

We say that the fragmentation process $(\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_t)$ is *successful* if $|\mathcal{H}_j| \geq |\mathcal{H}_{j-1}|/2$ for all $j \in [t]$. Let $\beta := 1 - t\alpha$, and note that, by the definition of $t = \lceil 1/\alpha \rceil - 1$, we have $0 < \beta \leq \alpha$. By (2.3) and (2.4), we conclude that the probability that the fragmentation process $(\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_{t-1})$ which we run is successful is at least

$$1 - 4 \sum_{i=1}^t K^{-k_i/3} \geq 1 - 4 \sum_{i=1}^{\lceil 1/\alpha \rceil - 1} K^{-k_0^{1-i\alpha}/3} = 1 - O\left(K^{-k_0^\beta/3}\right).$$

To summarise, a.s. the fragmentation process is successful and, thus, yields a k_t -bounded multihypergraph \mathcal{H}_t of $(S, X_1 \cup \dots \cup X_t)$ -fragments, where $k_t = k_0^\beta$, $|\mathcal{H}_t| \geq 2^{-t}|\mathcal{H}_0|$ and each S is an element of \mathcal{H}_0 .

We now apply one more round of sprinkling. In this final round we switch and work with the random set $X := G(n, p)$ with $p = Kq$. We may also assume that \mathcal{H}_t is k_t -uniform, since every set $S \in \mathcal{H}_t$ is contained in some $S' \in \mathcal{F}$ and thus we can add some arbitrary $k_t - |S|$ vertices from $S' \setminus S$ to S . The proof now will proceed along the same lines as the proof in [15, Theorem 1.2].

Define the random variable $Y := |\{S \in \mathcal{H}_t : S \subseteq G(n, p)\}|$. Our aim is to estimate the variance of Y and to show that $\mathbb{P}[Y = 0] \leq \varepsilon$. This would mean that the random graph $G(n, p) \cup$

$\bigcup_{i=1}^t G(n, m_i)$ contains a copy of F with probability at least $1 - \varepsilon$ (by [12, Proposition 1.12], this also applies to $G(n, C_0q)$).

We estimate the variance of Y as follows (recall that we work in $G(n, p)$ now). Let $R \in \mathcal{H}_t$, so $|R| = k_t = k_0^\beta$. Since $\beta \leq \alpha < 1$ and $k_0 = \omega(1)$, for sufficiently large n we have that $k_t \leq \delta k_0$. Then, using the fact that \mathcal{F} is (q, α, δ) -superspread, for each $\ell \in [k_t]$ we have that

$$\begin{aligned} |\{S \in \mathcal{H}_t : |S \cap R| = \ell\}| &\leq \sum_{L \subseteq R, |L| = \ell} |\mathcal{H}_t \cap \langle L \rangle| \stackrel{(2.1)}{\leq} \sum_{L \subseteq R, |L| = \ell} |\mathcal{F} \cap \langle L \rangle| \leq \sum_{L \subseteq R, |L| = \ell} q^{|L|} k_0^{-\alpha c_L} |\mathcal{F}| \\ &= \sum_{c=1}^{\ell} \sum_{L \subseteq R, |L| = \ell, c_L = c} q^\ell k_0^{-\alpha c} |\mathcal{F}| \stackrel{\text{Lemma 2.4}}{\leq} \sum_{c=1}^{\ell} (4ed)^\ell \binom{k_t}{c} q^\ell k_0^{-\alpha c} |\mathcal{F}| \\ &= q^\ell |\mathcal{F}| (4ed)^\ell \sum_{c=1}^{\ell} \binom{k_t}{c} k_0^{-\alpha c} \leq q^\ell |\mathcal{F}| (4ed)^\ell \sum_{c=1}^{\ell} \left(\frac{ek_t k_0^{-\alpha}}{c}\right)^c = q^\ell |\mathcal{F}| e^{O(\ell)}, \end{aligned}$$

where the implicit constants in the O notation are independent of K . We therefore get the following bound on the variance:

$$\text{Var}[Y] \leq p^{2k_t} \sum_{R, S \in \mathcal{H}_t, R \cap S \neq \emptyset} p^{-|R \cap S|} \binom{|\mathcal{H}_t|}{|R \cap S|} |\mathcal{F}|^2 p^{2k_t} \sum_{\ell=1}^{k_t} e^{O(\ell)} p^{-\ell} q^\ell = O(\mathbb{E}[Y]^2 / K),$$

where we use the facts that $p = Kq$, $\mathbb{E}[Y] = p^{k_t} |\mathcal{H}_t|$ and $|\mathcal{H}_t| \geq 2^{-t} |\mathcal{F}| = \Theta(|\mathcal{F}|)$. By letting K be sufficiently large, the result follows by Chebyshev’s inequality. \square

3. Applications of Theorem 2.2

We use this section to prove Theorems 1.1 and 1.2 as applications of Theorem 2.2.

3.1. Spanning C_4 -cycles

Throughout this section, we assume that n is even. Observe that the graph $C_{4,n}^e$ has $3n/2$ edges. Let \mathcal{C} be the $(3n/2)$ -uniform hypergraph on the vertex set $M = \binom{[n]}{2}$ where we see (the set of edges of) each copy of $C_{4,n}^e$ as an edge of \mathcal{C} . We write $|\mathcal{C}|$ for the number of its edges and notice that $|\mathcal{C}| = (n - 1)!/2$. Indeed, consider an arbitrary labelling v_1, \dots, v_n of the vertices. We define a copy of $C_{4,n}^e$ uniquely based on this ordering: we consider a matching between the set of even vertices and odd vertices (where we add the edge $v_{2i-1}v_{2i}$ for each $i \in [n/2]$), and then we define a cycle of length $n/2$ on the set of even vertices and another cycle in the set of odd vertices (where the edges of these cycles join the vertices which are closest in the labelling, seen cyclically). In this way, each of the C_4 ’s which conform the copy of $C_{4,n}^e$ is given by four consecutive vertices in the labelling, starting with an odd vertex v_{2i-1} , so that its edges are $\{v_{2i-1}v_{2i}, v_{2i+1}v_{2i+2}, v_{2i-1}v_{2i+1}, v_{2i}v_{2i+2}\}$. Now one can easily verify that there are $2n$ different labellings which yield the same copy of $C_{4,n}^e$ (there are $n/2$ possible starting points while maintaining the same cyclic ordering; if the ordering is reversed, the resulting graph is the same; and if all pairs of vertices $\{v_{2i-1}, v_{2i}\}$ are swapped, the resulting graph is also the same).

Recall that the hypergraph \mathcal{C} is q -spread, for some $q \in (0, 1)$, if $|\mathcal{C} \cap \langle I \rangle| \leq q^{|I|} |\mathcal{C}|$ for all $I \subseteq M$, where $\langle I \rangle$ denotes the set of all supersets of I . Moreover, for $\alpha, \delta \in (0, 1)$, we say \mathcal{C} is (q, α, δ) -superspread if it is q -spread and, for every $I \subseteq M$ with $|I| \leq 3\delta n/2$, we have

$$|\mathcal{C} \cap \langle I \rangle| \leq q^{|I|} (3n/2)^{-\alpha c_I} |\mathcal{C}|,$$

where c_I is the number of components of I . Our main goal now is to establish that the hypergraph \mathcal{C} is $(250n^{-2/3}, 1/3, 1/15)$ -superspread.

Lemma 3.1. *Let $I \subseteq C_{4,n}^e$ be a graph with $\ell \leq n/10$ edges and c components. Then, we have*

$$|V(I)| - c \geq \frac{2}{3}\ell + \frac{c}{3}.$$

Proof. Let I_1, \dots, I_c be the components of I with at least one edge, and let v_1, \dots, v_c be the number of vertices spanned by I_1, \dots, I_c , respectively. Since for all $j \in [c]$ we have $|I_j| \leq |I| \leq n/10$, we conclude the following easy bound on any component:

$$|I_j| \leq \frac{1}{2} (4 \cdot 2 + (v_j - 4) \cdot 3) = \frac{3}{2}v_j - 2. \tag{3.1}$$

Indeed, this holds since the maximum degree of I is at most 3 and in every component I_j with $4 \leq v_j \leq n/10$ there are four vertices whose sum of degrees is at most 8. For $v_j \in [2, 3]$ we have $|I_j| = v_j - 1$, hence the bound given in (3.1) holds in these cases as well. Summing over all $j \in [c]$, we obtain that

$$\ell = |I| \leq \frac{3}{2}|V(I)| - 2c,$$

which yields the desired result by rearranging the terms. □

Lemma 3.2. *Let $I \subseteq C_{4,n}^e$ be a graph with ℓ edges and c components. Then, we have*

$$|V(I)| - c \geq \frac{2}{3}\ell - 1.$$

Proof. Since every vertex of $C_{4,n}^e$ has degree 3, the bound in the statement holds trivially if I has only one component. We may thus assume that I contains at least two components with at least one edge each. But then, one can directly check that (3.1) must hold, and we can argue exactly as in Lemma 3.1, which leads to a better bound than claimed in the statement. □

Lemma 3.3. *The hypergraph \mathcal{C} is $(250n^{-2/3}, 1/3, 1/15)$ -superspread.*

Proof. Let $I \subseteq M$. We need to obtain upper bounds for $|\mathcal{C} \cap \langle I \rangle|$. If I is not contained in any copy of $C_{4,n}^e$, then $|\mathcal{C} \cap \langle I \rangle| = 0$, so we may assume I is a subgraph of some copy of $C_{4,n}^e$. Recall that a copy of $C_{4,n}^e$ can be defined by an ordering of $[n]$ and that exactly $2n$ such orderings define the same copy of $C_{4,n}^e$. Thus, it suffices to bound the number of orderings of $[n]$ which define a copy of $C_{4,n}^e$ containing I .

Let I_1, \dots, I_c be the components of I which contain at least one edge, and let v_1, \dots, v_c be the number of vertices spanned by I_1, \dots, I_c , respectively. For each $j \in [c]$, choose a vertex $x_j \in V(I_j)$ (note that there are v_j possible choices for this, which leads to a total of

$$\prod_{j=1}^c v_j \leq 2^{|I|} \tag{3.2}$$

choices for $\{x_1, \dots, x_c\}$, where the inequality holds since $2^{|I_j|} \geq 2^{v_j-1} \geq v_j$). Now, each ordering σ of $[n]$ (recall this defines a copy of $C_{4,n}^e$) induces an ordering on the set consisting of the vertices x_1, \dots, x_c as well as all isolated vertices in I . Let us denote this induced ordering as $\tau = \tau(\sigma)$. We now want to bound the total number of possible orderings σ by first bounding the number of orderings τ (which depend on the choice of x_1, \dots, x_c) and then the number of orderings σ with $\tau = \tau(\sigma)$.

After the choice of x_1, \dots, x_c , the number of possible orderings τ is

$$(n - |V(I)| + c)!. \tag{3.3}$$

Now, in order to obtain some σ such that $\tau = \tau(\sigma)$, it suffices to ‘insert’ the vertices which are missing into the ordering, and this must be done in a way which is consistent with the structure of the components I_j . For each $j \in [c]$, consider a labelling of the vertices of I_j starting with x_j and such that each subsequent vertex has at least one neighbour with a smaller label. Then, we insert the vertices of I_j into the ordering following this labelling, and note that, for each vertex, there are at most three choices, as $\Delta(I_j) \leq 3$. This implies there are at most $3^{|V(I_j)|-1} \leq 3^{|I_j|}$ possible ways to fix the ordering of the vertices of I_j . By considering all $j \in [c]$, we conclude that there are at most

$$\prod_{j=1}^c 3^{|I_j|} \leq 3^{|I|} \tag{3.4}$$

possible orderings σ which result in the same τ .

Combining (3.2), (3.3) and (3.4) with the fact that there are $2n$ distinct orderings σ which result in the same copy of $C_{4,n}^e$, we conclude that

$$|\mathcal{C} \cap \langle I \rangle| \leq \frac{6^{|I|}}{2n} (n - |V(I)| + c)! \leq 6^{|I|} (n - |V(I)| + c - 1)! \tag{3.5}$$

We can now estimate the spreadness of \mathcal{C} . Consider first any $I \subseteq M$ with $|I| \leq n/10 = |C_{4,n}^e|/15$, and let c be its number of components. Then, by substituting the bound given by Lemma 3.1 into (3.5), we conclude that

$$|\mathcal{C} \cap \langle I \rangle| \leq 6^{|I|} \left(n - \frac{2}{3}|I| - \frac{c}{3} - 1 \right)!$$

By using the bound on $|I|$ and taking into account that $|\mathcal{C}| = (n - 1)!/2$ and $|C_{4,n}^e| = 3n/2$, we conclude that $|\mathcal{C} \cap \langle I \rangle| \leq q^{|I|} |C_{4,n}^e|^{-c/3} |\mathcal{C}|$ for $q \geq 250n^{-2/3}$.

Similarly, assume $I \subseteq M$ has $|I| > n/10$ edges and c components. By substituting the bound given by Lemma 3.2 into (3.5), we now have that

$$|\mathcal{C} \cap \langle I \rangle| \leq 6^{|I|} \left(n - \frac{2}{3}|I| \right)!$$

Now, as above, we conclude that $|\mathcal{C} \cap \langle I \rangle| \leq q^{|I|} |\mathcal{C}|$ for $q \geq 12n^{-2/3}$.

Combining the two statements above, it follows by definition that \mathcal{C} is $(250n^{-2/3}, 1/3, 1/15)$ -superspread, as we wanted to see. □

Proof of Theorem 1.1. Lemma 3.3 establishes that \mathcal{C} is $(250n^{-2/3}, 1/3, 1/15)$ -superspread. By Theorem 2.2 we have that, if $p \geq Cn^{-2/3}$, where C is a sufficiently large constant, then

$$\mathbb{P} [C_{4,n}^e \subseteq G(n, p)] \geq 1/2.$$

To finish the argument, one can employ a general result of Friedgut [9] (see, e.g., a recent paper of Narayanan and Schacht [19]) which allows to establish that

$$\mathbb{P} [C_{4,n}^e \subseteq G(n, (1 + o(1))p)] = 1 - o(1). \tag{3.6} \quad \square$$

3.2. Spanning K_r -cycles

In the following we will study copies of K_r arranged in a cyclic way. Since there are several ways how two consecutive copies of K_r can overlap, we provide a precise definition of what will be called an s -overlapping K_r -cycle.

Definition 3.4. Let $r > s \geq 0$ and $n \in \mathbb{N}$ with $(r - s) \mid n$ be integers. A $K_{r,s,n}$ -cycle is a graph on vertex set $\mathbb{Z}_n = [0, n - 1]$ whose edge set is the union of the edge sets of $n/(r - s)$ copies of K_r , where for each $i \in [0, n/(r - s) - 1]$ there is a copy of K_r on the vertices $[i(r - s), i(r - s) + r - 1]$ (modulo n).

In other words, the $n/(r - s)$ copies of K_r are arranged cyclically on the vertex set \mathbb{Z}_n , so that two consecutive copies of K_r intersect in exactly s vertices and two non-consecutive cliques intersect in as few vertices as possible.

The case $s = 0$ corresponds to a K_r -factor. The threshold for the property of containing a K_r -factor was famously determined by Johansson, Kahn and Vu [13]. For the case $s = 1$, the copies of K_r in $K_{r,1,n}$ are edge-disjoint. As mentioned in the introduction, the threshold for the appearance of $K_{r,1,n}$ in $G(n, p)$ was recently determined by Frieze [10]. When $s = r - 1$, the s -overlapping K_r -cycles are usually referred to as the $(r - 1)$ -th power of a Hamilton cycle C_n , where the k -th power of some arbitrary graph G is obtained by connecting any two vertices of G which are at distance at most k with an edge. The threshold for the appearance of $K_{r,r-1,n}$ is known to be $n^{-1/r}$. This was observed by Kühn and Osthus [17] for $r \geq 4$, while the case $r = 3$ was solved recently by Kahn, Narayanan and Park [15].

We determine the threshold for the appearance of $K_{r,s,n}$ for all the remaining values of r and s . Whenever $s \geq 3$, the result follows from a general result of Riordan [23]; see Section 4. Our main focus here is on the cases when $s = 2$. The overall strategy follows the same structure as in Section 3.1.

We denote the set of all unlabelled copies of $K_{r,s,n}$ on $[n]$ by $\mathcal{C}_{r,s,n}$. First, we observe the following simple facts about $K_{r,s,n}$.

Fact 3.5. *Let $r > s \geq 0$ and $n \in \mathbb{N}$ with $(r - s) \mid n$. Then, the number of edges in $K_{r,s,n}$ is exactly*

$$\left(\binom{r}{2} - \binom{s}{2} \right) \frac{n}{r - s} = \frac{1}{2}(r + s - 1)n.$$

In particular, for $s \leq r/2$, the number of vertices of degree $2r - s - 1$ is exactly $sn/(r - s)$ (such vertices belong to two copies of K_r), whereas the remaining $(r - 2s)n/(r - s)$ vertices have degree $r - 1$ (these vertices belong to exactly one copy of K_r). \square

Fact 3.6. *Let $r > s \geq 1$ and $n \in \mathbb{N}$ with $(r - s) \mid n$ and $s \leq r/2$. We have*

$$|\mathcal{C}_{r,s,n}| = \frac{(n - 1)!(r - s)}{2((r - 2s)!)^{n/(r-s)}(s!)^{n/(r-s)}} = \frac{r - s}{2} d_{r,s}^n (n - 1)!,$$

where $d_{r,s} \in (0, 1]$ is some absolute constant that depends on s and r only. \square

The main technical part of the proof of Theorem 1.2 is to obtain estimates which are crucial for studying the spreadness of $\mathcal{C}_{r,2,n}$. Here we only include the statements of the two estimates which we require; we postpone their proofs to Appendix B.

Lemma 3.7. *Let $r \geq 4$ and $n \in \mathbb{N}$ with $(r - 2) \mid n$. Let $I \subseteq K_{r,2,n}$ be a subgraph with $\ell \leq n/(2r)$ edges and c components. Then,*

$$|V(I)| - c \geq \frac{2}{r + 1} \ell + \frac{c}{r + 1}.$$

Lemma 3.8. *Let $r \geq 4$ and $n \in \mathbb{N}$ with $(r - 2) \mid n$. Let $I \subseteq K_{r,2,n}$ be a subgraph with ℓ edges and c components. Then,*

$$|V(I)| \geq \frac{2}{r + 1} \ell.$$

Combining the previous two lemmas, we show that $\mathcal{C}_{r,2,n}$ is a $(O(n^{-2/(r+1)}), 1/(r + 1), 1/(r(r + 1)))$ -superspread hypergraph.

Lemma 3.9. *Let $r \geq 4$ and $n \in \mathbb{N}$ with $(r - 2) \mid n$. Then, the hypergraph $\mathcal{C}_{r,2,n}$ of all copies of $K_{r,2,n}$ in $M = \binom{[n]}{2}$ is $(O(n^{-2/(r+1)}), 1/(r + 1), 1/(r(r + 1)))$ -superspread.*

Proof. Let $I \subseteq M$. Our first aim is to obtain a general upper bound on $|\mathcal{C}_{r,2,n} \cap \langle I \rangle|$. If I is not contained in any copy of $K_{r,2,n}$, we automatically have $|\mathcal{C}_{r,2,n} \cap \langle I \rangle| = 0$, so we may assume I is a subgraph of some copy of $K_{r,2,n}$. Recall that each copy of $K_{r,2,n}$ can be defined by an ordering of the n vertices, so it suffices to bound the number of orderings which yield a copy of $K_{r,2,n}$ containing I .

Let I_1, \dots, I_c be the components of I with at least one edge, and let v_1, \dots, v_c be the number of vertices spanned by I_1, \dots, I_c , respectively. For each $j \in [c]$, let $x_j \in V(I_j)$ (there are v_j such possible choices, which leads to a total of

$$\prod_{j=1}^c v_j \leq 2^{|I|} \tag{3.6}$$

choices for $\{x_1, \dots, x_j\}$). Then, each ordering σ of $[n]$ which defines a copy of $K_{r,2,n}$ containing I induces a unique ordering $\tau = \tau(\sigma)$ on the set consisting of x_1, \dots, x_j and all other isolated vertices. The total number of such orderings τ is

$$(n - |V(I)| + c)! \tag{3.7}$$

so now it suffices to bound, for each such τ , the number of orderings σ with $\tau = \tau(\sigma)$.

Given an ordering τ , in order to obtain an ordering σ with $\tau = \tau(\sigma)$, it suffices to ‘insert’ the missing vertices into the ordering. That is, for each $j \in [c]$, we need to ‘insert’ the other vertices of $V(I_j)$ into the ordering. By considering a labelling of the vertices of I_j in such a way that each subsequent vertex is a neighbour of at least one previously included vertex (and taking into account that x_j is already included), we note that there are at most $2r$ choices for each vertex (recall that $\Delta(K_{r,2,n}) < 2r$). This leads to a total of at most $(2r)^{|V(I_j)|-1} \leq (2r)^{|I_j|}$ possible ways to include the component I_j . By considering all $j \in [c]$, we conclude that there are at most

$$\prod_{j=1}^c (2r)^{|I_j|} = (2r)^{|I|} \tag{3.8}$$

orderings σ with $\tau = \tau(\sigma)$.

Combining (3.6), (3.7) and (3.8) with the fact that each copy of $K_{r,2,n}$ is given by $d_{r,2}^{-n} 2n/(r - 2)$ distinct orderings (see Fact 3.6), we conclude that

$$|\mathcal{C}_{r,2,n} \cap \langle I \rangle| \leq \frac{r - 2}{2} d_{r,2}^n (4r)^{|I|} (n - |V(I)| + c - 1)!. \tag{3.9}$$

We can now estimate the spreadness of $\mathcal{C}_{r,2,n}$. Consider first any $I \subseteq M$ with $|I| > n/(2r) = |K_{r,2,n}|/(r(r + 1))$ (see Fact 3.5). Note that I has at most $r(r + 1)$ components I_j of size larger than $n/(2r)$. For each of these components, we use Lemma 3.8 to bound $|V(I_j)|$. For the remaining components I_j , we simply use the bound $|V(I_j)| - 1 \geq 2|I_j|/(r + 1)$, which follows by Lemma 3.7. By substituting these bounds into (3.9), we conclude that

$$|\mathcal{C}_{r,2,n} \cap \langle I \rangle| \leq \frac{r - 2}{2} d_{r,2}^n (4r)^{|I|} \left(n - \frac{2}{r + 1} |I| + r(r + 1) \right)!$$

By comparing this with the expression given in Fact 3.6 (and taking into account the bound on $|I|$), we conclude that $|\mathcal{C}_{r,2,n} \cap \langle I \rangle| \leq q^{|I|} |\mathcal{C}_{r,2,n}|$ whenever $q \geq c_1 n^{-2/(r+1)}$, where c_1 is a constant that depends only on r .

Consider now some $I \subseteq M$ with $|I| \leq n/(2r) = |K_{r,2,n}|/(r(r + 1))$, and let c be its number of components. By making use of Lemma 3.7 and (3.9), we have that

$$|\mathcal{C}_{r,2,n} \cap \langle I \rangle| \leq \frac{r - 2}{2} d_{r,2}^n (4r)^{|I|} \left(n - \frac{2}{r + 1} |I| - \frac{c}{r + 1} - 1 \right)!$$

As above, by comparing this with the expression given in Fact 3.6, and taking into account also Fact 3.5, for $q \geq c_2 n^{-2/(r+1)}$, where c_2 depends only on r , we have that $|C_{r,2,n} \cap I| \leq q^{|I|} |K_{r,2,n}|^{-c/(r+1)} |C_{r,2,n}|$. \square

Proof of Theorem 1.2. By Lemma 3.9, we have that $C_{r,2,n}$ is $(O(n^{-2/(r+1)}), 1/(r+1), 1/(r(r+1)))$ -superspread. By Theorem 2.2 it follows that, if C is sufficiently large, for $p \geq Cn^{-2/(r+1)}$ we have

$$\mathbb{P} [K_{r,2,n} \subseteq G(n, p)] \geq 1/2.$$

To finish the argument, one can employ a general result of Friedgut [9] (see also [19]) which allows to establish that $\mathbb{P} [K_{r,2,n} \subseteq G(n, (1 + o(1))p)] = 1 - o(1)$. \square

4. Concluding remarks

4.1. Dense overlapping K_r -cycles

As mentioned in the introduction, a general result of Riordan [23] provides a sufficient condition for a spanning graph to be contained in $G(n, p)$. For a graph $H = (V, E)$, let $v(H) := |V|$ and $e(H) := |E|$. For each integer v , let $e_H(v) := \max\{e(F) : F \subseteq H, v(F) = v\}$. Then, the following parameter will be responsible for the upper bound on the threshold for the property that $H \subseteq G(n, p)$:

$$\gamma(H) := \max_{3 \leq v \leq n} \left\{ \frac{e_H(v)}{v - 2} \right\}.$$

Riordan proved the following (see also [20] for its generalisation to hypergraphs).

Theorem 4.1. *Let $H = H^{(i)}$ be a sequence of graphs with $n = n(i)$ vertices (where n tends to infinity with i), $e(H) = \alpha \binom{n}{2} = \alpha(n) \binom{n}{2}$ edges and $\Delta = \Delta(H)$. Let $p = p(n) : \mathbb{N} \rightarrow [0, 1)$. If H has a vertex of degree at least 2 and $np^{\gamma(H)} \Delta^{-4} = \omega(1)$, then a.a.s. the random graph $G(n, p)$ contains a copy of H .*

From Fact 3.5 it follows that $\gamma(K_{r,s,n}) \geq \frac{r+s-1}{2}$ and, since $\gamma(K_r) > \frac{r+s-1}{2}$ for $s \in [2]$, Riordan’s theorem does not provide optimal bounds on the threshold in the case of our Theorem 1.2 (nor for $K_{r,1,n}$, for which the threshold was determined by Frieze [10]). However, in the cases for $s \geq 3$, Riordan’s theorem suffices and yields the correct threshold $n^{-2/(r+s-1)}$ for the property that $G(n, p)$ contains a copy of $K_{r,s,n}$.

4.2. Spanning $C_{2\ell}$ -cycles

As a generalisation of Theorem 1.1, we may consider copies of $C_{2\ell}$ arranged in a cyclic manner, in such a way that they form a spanning graph. More precisely, for $\ell \geq 2$ and $n \in \mathbb{N}$ with $(2\ell - 2) \mid n$, we define a $C_{2\ell,n}^e$ -cycle as $n/(2\ell - 2)$ copies of $C_{2\ell}$ arranged cyclically on n vertices so that only consecutive copies intersect, any two consecutive copies intersect in exactly one edge and the two edges that a cycle C shares with its neighbouring cycles are as far apart in C as possible.

We may also use the ideas from this paper to determine the threshold for the appearance of $C_{2\ell,n}^e$ for all $\ell \geq 3$. Indeed, we have the following result.

Theorem 4.2. *Let $\ell \geq 2$ and $n \in \mathbb{N}$ with $(2\ell - 2) \mid n$. Then, the threshold for the appearance of $C_{2\ell,n}^e$ in $G(n, p)$ is $\Theta \left(n^{-\frac{2\ell-2}{2\ell-1}} \right)$.*

The proof follows the ideas from Section 3; in particular, it follows the steps in Section 3.2 very closely. We thus omit the details here.

4.3. Extensions: hypergraphs and rainbow thresholds

Throughout this paper, for simplicity, we have focused on properties of random graphs. However, we believe that Theorem 2.2 extends to random hypergraphs without much issue. In fact, this extension also follows from the work of Spiro [25].

Very recently, Bell, Frieze and Marbach [3] and Bell and Frieze [2] partially extended the results from [8, 15] to rainbow versions, where the vertices of some r -uniform hypergraph are coloured randomly with r colours. It is shown in [2, 3] that, under certain conditions, the upper bounds on the thresholds as proved in [8, 15] remain asymptotically the same to yield a rainbow hyperedge or a rainbow copy of a spanning structure (e.g., bounded degree spanning tree, square of a Hamilton cycle), and the result is also extended to a rainbow version of the containment of the k -th power of a Hamilton cycle. We believe that it should be possible to adapt our techniques to prove a rainbow version of Theorem 2.2, but we leave working out the details of this extension as an open problem.

References

- [1] Alweiss, R., Lovett, S., Wu, K. and Zhang, J. (2020) Improved bounds for the sunflower lemma, pp. 624–630, *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020*, Association for Computing Machinery, New York, NY, USA. doi: [10.1145/3357713.3384234](https://doi.org/10.1145/3357713.3384234), ISBN 9781450369794.
- [2] Bell, T. and Frieze, A. (2022) Rainbow powers of a Hamilton cycle in $G_{n,p}$. *arXiv e-prints*. arXiv: [2210.08534](https://arxiv.org/abs/2210.08534).
- [3] Bell, T., Frieze, A. and Marbach, T. G. (2021) Rainbow thresholds. *arXiv e-prints*. arXiv: [2104.05629](https://arxiv.org/abs/2104.05629).
- [4] Bollobás, B. (1981) Threshold functions for small subgraphs. *Math. Proc. Camb. Philos. Soc.* **90**(2) 197–206. doi: [10.1017/S0305004100058655](https://doi.org/10.1017/S0305004100058655).
- [5] Böttcher, J. (2017) Large-scale structures in random graphs. *Surveys in combinatorics 2017*, University of Strathclyde, Glasgow, UK, Cambridge: Cambridge University Press, pp. 87–140, *Papers Based on the 26th British Combinatorial Conference*, ISBN 978-1-108-41313-8/pbk; 978-1-108-33103-6/ebook.
- [6] Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci. Ser. A* **5** 17–61. doi: [10.1515/9781400841356.38](https://doi.org/10.1515/9781400841356.38).
- [7] Erdős, P. and Rényi, A. (1966) On the existence of a factor of degree one of a connected random graph. *Acta Math. Acad. Sci. Hungar.* **17**(3-4) 359–368. doi: [10.1007/BF01894879](https://doi.org/10.1007/BF01894879).
- [8] Frankston, K., Kahn, J., Narayanan, B. and Park, J. (2021) Thresholds versus fractional expectation-thresholds. *Ann. Math.* **194**(2) 475–495. doi: [10.4007/annals.2021.194.2.2](https://doi.org/10.4007/annals.2021.194.2.2).
- [9] Friedgut, E. (2005) Hunting for sharp thresholds. *Random Struct. Algorithms* **26**(1-2) 37–51. doi: [10.1002/rsa.20042](https://doi.org/10.1002/rsa.20042).
- [10] Frieze, A. (2020) A note on spanning K_r -cycles in random graphs. *AIMS Math.* **5**(5) 4849–4852. doi: [10.3934/math.2020309](https://doi.org/10.3934/math.2020309).
- [11] Heckel, A. (2021) Random triangles in random graphs. *Random Struct. Algorithms* **59**(4) 616–621. doi: [10.1002/rsa.21013](https://doi.org/10.1002/rsa.21013).
- [12] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization. New York, Wiley-Interscience. doi: [10.1002/9781118032718](https://doi.org/10.1002/9781118032718), ISBN 0-471-17541-2.
- [13] Johansson, A., Kahn, J. and Vu, V. (2008) Factors in random graphs. *Random Struct. Algorithms* **33**(1) 1–28. doi: [10.1002/rsa.20224](https://doi.org/10.1002/rsa.20224).
- [14] Kahn, J. and Kalai, G. (2007) Thresholds and expectation thresholds. *Combin. Probab. Comput.* **16**(3) 495–502. doi: [10.1017/S0963548307008474](https://doi.org/10.1017/S0963548307008474).
- [15] Kahn, J., Narayanan, B. and Park, J. (2021) The threshold for the square of a Hamilton cycle. *Proc. Am. Math. Soc.* **149**(8) 3201–3208. doi: [10.1090/proc/15419](https://doi.org/10.1090/proc/15419).
- [16] Koršunov, A. D. (1977) Solution of a problem of P. Erdős and A. Rényi on Hamiltonian cycles in undirected graphs. *Metody Diskretn. Anal.* **31** 17–56.
- [17] Kühn, D. and Osthus, D. (2012) On Pósa’s conjecture for random graphs. *SIAM J. Discrete Math.* **26**(3) 1440–1457. doi: [10.1137/120871729](https://doi.org/10.1137/120871729).
- [18] Montgomery, R. (2019) Spanning trees in random graphs. *Adv. Math.* **356** 92pp. doi: [10.1016/j.aim.2019.106793](https://doi.org/10.1016/j.aim.2019.106793). Id/No 106 793.
- [19] Narayanan, B. and Schacht, M. (2020) Sharp thresholds for nonlinear Hamiltonian cycles in hypergraphs. *Random Struct. Algorithms* **57**(1) 244–255. doi: [10.1002/rsa.20919](https://doi.org/10.1002/rsa.20919).
- [20] Parczyk, O. and Person, Y. (2016) Spanning structures and universality in sparse hypergraphs. *Random Struct. Algorithms* **49**(4) 819–844. doi: [10.1002/rsa.20690](https://doi.org/10.1002/rsa.20690).
- [21] Park, J. and Pham, H. T. (2022) A proof of the Kahn-Kalai conjecture, *Journal of the American Mathematical Society*, to appear
- [22] Pósa, L. (1976) Hamiltonian circuits in random graphs. *Discrete Math.* **14**(4) 359–364. doi: [10.1016/0012-365X\(76\)90068-6](https://doi.org/10.1016/0012-365X(76)90068-6).

[23] Riordan, O. (2000) Spanning subgraphs of random graphs. *Combin. Probab. Comput.* **9**(2) 125–148. doi: [10.1017/S0963548399004150](https://doi.org/10.1017/S0963548399004150).
 [24] Riordan, O. (2022) Random cliques in random graphs and sharp thresholds for F -factors. *Random Struct. Algorithms* **61**(4) 619–637. doi [10.1002/rsa.21111](https://doi.org/10.1002/rsa.21111).
 [25] Spiro, S. (2021) A smoother notion of spread hypergraphs. *arXiv e-prints*. arXiv: [2106.11882](https://arxiv.org/abs/2106.11882).
 [26] Talagrand, M. (2010) Are many small sets explicitly small? Association for Computing Machinery (ACM), New York, NY, pp. 13–36, *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing, STOC'10, June 5-8, 2010, Cambridge, MA, USA*, ISBN 978-1-60558-817-9.

Appendix A. Proof of Lemma 2.3

We begin with the proof of the auxiliary Lemma 2.4.

Proof of Lemma 2.4. The proof follows along the same lines as the proof from [15, Proposition 2.2]: one uses the fact that the number of connected h -edge subgraphs of a graph G containing a given vertex is less than $(e\Delta(G))^h$.

We now bound the number of ways in which we may construct subgraphs with ℓ edges and c components. First, we specify the roots of the c components of the ℓ -edge subgraph of F , which can be done in at most $2^c \binom{\ell}{c}$ ways. Next, we choose the sizes of the components; this can be done in at most $\binom{\ell-1}{c-1}$ ways (this is the number of c -compositions of ℓ). Say that, to each of the components $j \in [c]$, we have assigned size ℓ_j . We finally choose the subgraphs which conform each component along F , which, by the fact mentioned in the first paragraph, can be done in at most $\prod_{j=1}^c (e\ell_j)^{\ell_j} = (e\ell)^\ell$ ways. This yields a total of at most $(e\ell)^\ell \binom{\ell-1}{c-1} 2^c \binom{\ell}{c} \leq (4e\ell)^\ell \binom{\ell}{c}$ subgraphs. \square

The proof of Lemma 2.3 is essentially from [15, Lemma 3.1]: the only changes we need to make are replacing $2n$ in [15] by k_i and slightly adapting the computations for the verification of Claim A.1 in the pathological case below.

Since we will follow the proof from [15, Lemma 3.1] closely, we use the same notation wherever possible, so that the reader familiar with the argument can quickly verify the validity of Lemma 2.3.

Proof of Lemma 2.3. Without loss of generality, we may assume that n is sufficiently large, and thus k_i and k are also sufficiently large. We may also assume that \mathcal{H}_i is k_i -uniform (indeed, every set $S \in \mathcal{H}_i$ is contained in some $S' \in \mathcal{F}$, so we can add some arbitrary $k_i - |S|$ vertices from $S' \setminus S$ to S). Let $M := \binom{[n]}{2}$ and $m := |M|$.

Now, consider pairs (S, X) with $S \in \mathcal{H}_i$ and $X \subseteq M$ such that $|S \cap X| = t$, for some $t \in [0, k_i]$. Our aim is to prove that, for each t ,

$$|\{(S, X) : |S \cap X| = t, (S, X) \text{ is } k\text{-bad}\}| \leq 2C^{-k/3} |\mathcal{H}_i| \binom{k_i}{t} \binom{m - k_i}{w - t}. \tag{A.1}$$

Summing over all $t \in [0, k_i]$ yields the counting version of the bound (2.2).

Let us bound the number of bad pairs (S, X) with $|S \cap X| = t$, for some fixed $t \in [0, k_i]$. Set $w' := w - t$, so $|X \setminus S| = w'$ and $|X \cup S| = w' + k_i$. A set $Z \in \binom{M}{w'+k_i}$ will be called *pathological* if

$$|\{S \in \mathcal{H}_i : S \subseteq Z, (S, Z \setminus S) \text{ is } k\text{-bad}\}| > C^{-k/3} |\mathcal{H}_i| \binom{m - k_i}{w'} \bigg/ \binom{m}{w' + k_i}.$$

A pair (S, X) with $|S \cup X| = w' + k_i$ will be called *pathological* if $S \cup X$ is pathological. We estimate the contributions of pathological and nonpathological pairs to (A.1) separately.

Nonpathological pairs. To specify a nonpathological pair (S, X) with $|S \cap X| = t$, one first specifies $Z = S \cup X$, then S , and finally X . We use the crude bound $\binom{m}{w'+k_i}$ for the number of sets Z .

Since Z is nonpathological, there are at most $C^{-k/3} |\mathcal{H}_i| \binom{m-k_i}{w'} / \binom{m}{w'+k_i}$ choices for S , since $(S, Z \setminus S)$ is k -bad if (S, X) is k -bad. Finally, one can specify $X \subseteq Z$ in at most $\binom{k_i}{t}$ ways, since $|X \cap S| = t$ and $|S| = k_i$. Altogether, this gives a total of at most

$$\binom{m}{w'+k_i} \cdot \binom{k_i}{t} \cdot C^{-k/3} |\mathcal{H}_i| \binom{m-k_i}{w'} / \binom{m}{w'+k_i} \stackrel{w'=w-t}{=} C^{-k/3} |\mathcal{H}_i| \binom{m-k_i}{w-t} \binom{k_i}{t} \tag{A.2}$$

k -bad pairs.

Pathological pairs. The following claim will assist in estimating pathological contributions.

Claim A.1. *For a given $S \in \mathcal{H}_i$ and Y chosen uniformly at random from $\binom{M \setminus S}{w'}$, we have that*

$$\mathbb{E} [|\{J \in \mathcal{H}_i : J \subseteq S \cup Y, |J \cap S| \geq k\}|] \leq C^{-2k/3} |\mathcal{H}_i| \binom{w'+k_i}{k_i} / \binom{m}{k_i}. \tag{A.3}$$

The number of pathological pairs (S, X) can now be estimated as follows. First, we choose S and $S \cap X$, which can be done in at most $|\mathcal{H}_i| \binom{k_i}{t}$ ways. We then need to choose $X \setminus S$. Since we are counting pairs (S, X) which are k -bad, for every $J \in \mathcal{H}_i$ with $J \subseteq S \cup X$ we have that $|J \cap S| \geq |J \setminus X| \geq k$. Since we are only considering pathological pairs (S, X) , by definition this yields the following lower bound on the number of such $J \in \mathcal{H}_i$:

$$\begin{aligned} |\{J \in \mathcal{H}_i : J \subseteq S \cup (X \setminus S), |J \cap S| \geq k\}| &\geq C^{-k/3} |\mathcal{H}_i| \binom{m-k_i}{w'} / \binom{m}{w'+k_i} \\ &= C^{-k/3} |\mathcal{H}_i| \binom{w'+k_i}{k_i} / \binom{m}{k_i}. \end{aligned}$$

When choosing a uniformly random set $X \setminus S$ of size w' , Claim A.1 and Markov's inequality yield

$$\mathbb{P} \left[|\{J \in \mathcal{H}_i : J \subseteq S \cup (X \setminus S), |J \cap S| \geq k\}| \geq C^{-k/3} |\mathcal{H}_i| \binom{m-k_i}{w'} / \binom{m}{w'+k_i} \right] \leq C^{-k/3},$$

which gives us at most $C^{-k/3} \binom{m-k_i}{w'} = C^{-k/3} \binom{m-k_i}{w-t}$ choices for $X \setminus S$. Altogether this yields at most

$$C^{-k/3} \binom{m-k_i}{w-t} |\mathcal{H}_i| \binom{k_i}{t} \tag{A.4}$$

choices for pathological fragments.

From the bounds in each case, (A.2) and (A.4), we obtain (A.1). □

Finally, we turn to the proof of Claim A.1.

Proof of Claim A.1. For each $j \in [0, k_i]$, let f_j denote the fraction of $J \in \mathcal{H}_i$ with $|J \cap S| = j$. Then, the left-hand side of (A.3) is

$$\sum_{j=k}^{k_i} f_j |\mathcal{H}_i| \frac{\binom{m-k_i-(k_i-j)}{w'-(k_i-j)}}{\binom{m-k_i}{w'}} = \sum_{j=k}^{k_i} f_j |\mathcal{H}_i| \frac{(w')_{k_i-j}}{(m-k_i)_{k_i-j}}. \tag{A.5}$$

Hence, it suffices to show that, for each $j \in [k, k_i]$,

$$f_j \frac{(w')_{k_i-j}}{(m-k_i)_{k_i-j}} \frac{\binom{m}{k_i}}{\binom{m}{w'+k_i}} = e^{O(j)} C^{-j}, \tag{A.6}$$

where the implied constants in the O notation are independent of C .

We rewrite the left-hand side of (A.6) as follows:

$$f_j \frac{(w')_{k_i-j}}{(m-k_i)_{k_i-j}} \cdot \frac{(m)_{k_i}}{(w'+k_i)_{k_i}} = f_j \frac{(w')_{k_i-j}}{(w'+k_i)_{k_i-j}} \cdot \frac{(m)_{k_i-j}}{(m-k_i)_{k_i-j}} \cdot \frac{(m-k_i+j)_j}{(w'+j)_j}.$$

We next use the bounds $\frac{(w')_{k_i-j}}{(w'+k_i)_{k_i-j}} \leq 1$, $\frac{(m)_{k_i-j}}{(m-k_i)_{k_i-j}} \leq \exp(k_i(k_i-j)/(m-k_i)) \leq e^{d^2} = O(1)$ and

$$\frac{(m-k_i+j)_j}{(w'+j)_j} \leq \left(\frac{m}{Cq \binom{n}{2} - k_i} \right)^j \stackrel{k_i=O(k_0), q \geq 4k_0/(Cn^2)}{\leq} e^{O(j)} C^{-j} q^{-j}$$

to bound the left-hand side of (A.6) from above by

$$f_j e^{O(j)} C^{-j} q^{-j}. \tag{A.7}$$

Finally we observe that, since \mathcal{F} is q -spread and $(\mathcal{F}, \mathcal{H}_1, \dots, \mathcal{H}_i)$ is a fragmentation process with $|\mathcal{H}_\ell| \geq |\mathcal{H}_{\ell-1}|/2$ for all $\ell \in [i]$, by (2.1), for every $I \subseteq M$ we have

$$|\mathcal{H}_i \cap \langle I \rangle| \leq |\mathcal{F} \cap \langle I \rangle| \leq q^{|\mathcal{F}|} |\mathcal{F}| \leq 2^i q^{|\mathcal{F}|} |\mathcal{H}_i|.$$

Consider now sets $I \subseteq M$ with $|I| = j$. For $j \in [\delta e(F), k_i]$, since $\delta e(F) \geq \delta k_i$, we have that

$$f_j \leq \binom{k_i}{j} 2^i q^{|\mathcal{F}|} = e^{O_\delta(j)} q^j, \tag{A.8}$$

where we have used the fact that $2^i = O(1)$. Now recall that $k_0 = e(F)$. For $j \in [k, \min\{k_i, \delta k_0\}]$, we use the fact that \mathcal{F} is (q, α, δ) -superspread and, thus, by (2.1), for each $I \subseteq M$ with $|I| = j$ and c_I components we have

$$|\mathcal{H}_i \cap \langle I \rangle| \leq |\mathcal{F} \cap \langle I \rangle| \leq q^{|\mathcal{F}|} k_0^{-\alpha c_I} |\mathcal{F}| \leq 2^i k_0^{-\alpha c_I} q^{|\mathcal{F}|} |\mathcal{H}_i|.$$

This yields

$$\begin{aligned} f_j &= |\mathcal{H}_i|^{-1} |\{J \in \mathcal{H}_i : |J \cap S| = j\}| \leq |\mathcal{H}_i|^{-1} \sum_{I \subseteq S, |I|=j} |\mathcal{H}_i \cap \langle I \rangle| \leq |\mathcal{H}_i|^{-1} \sum_{I \subseteq S, |I|=j} 2^i k_0^{-\alpha c_I} q^{|\mathcal{F}|} |\mathcal{H}_i| \\ &= O(q^j) \sum_{I \subseteq S, |I|=j} k_0^{-\alpha c_I} = O(q^j) \sum_{c=1}^j k_0^{-\alpha c} |\{I \subseteq S : |I| = j, c_I = c\}| \stackrel{\text{Lemma 2.4}}{\leq} e^{O(j)} q^j \sum_{c=1}^j k_0^{-\alpha c} \binom{k_i}{c} \\ &\leq e^{O(j)} q^j \sum_{c=1}^j \left(\frac{e \cdot k_i \cdot k_0^{-\alpha}}{c} \right)^c = e^{O(j)} q^j \sum_{c=1}^j \left(\frac{e \cdot k}{c} \right)^c \leq e^{O(j)} q^j \sum_{c=1}^j e^{k \cdot j \geq k} e^{O(j)} q^j. \end{aligned} \tag{A.9}$$

Combining (A.8) and (A.9), we simplify (A.7) as follows:

$$f_j e^{O(j)} C^{-j} q^{-j} \leq e^{O(j)} q^j e^{O(j)} C^{-j} q^{-j} = e^{O(j)} C^{-j}.$$

This confirms (A.6) and, thus, concludes the proof of the claim. □

Appendix B. Proofs of Lemmas 3.7 and 3.8

We present here all the details needed to prove Lemmas 3.7 and 3.8. When talking about subgraphs of $K_{r,s,n}$, we refer to sets of consecutive vertices as *segments*. The *length* of a segment is the number of vertices it contains. We will make use of the following simple fact.

Fact B.1. Let $r \geq 4$ and $n \in \mathbb{N}$ with $(r - 2) \mid n$. Let $V \subseteq V(K_{r,2,n})$ be a segment starting in the first vertex of some clique K_r with $|V| \leq n/(2r) + 1$. Then,

$$e(K_{r,2,n}[V]) = \binom{r}{2} - 1 \cdot a + \binom{b}{2} - \max\{2 - b, 0\} \cdot (r - 2), \tag{B.1}$$

where $|V| = (r - 2)a + b$ with $a, b \in \mathbb{N}_0$ and $0 \leq b < r - 2$. □

Instead of using (B.1), we will make use of the following estimate to streamline our calculations.

Proposition B.2. Let $r \geq 4$ and $v \in \mathbb{N}$ with $v = (r - 2)a + b$, where $a, b \in \mathbb{N}_0$ and $0 \leq b < r - 2$. Then,

$$\left(\binom{r}{2} - 1 \right) a + \binom{b}{2} - \max\{2 - b, 0\} \cdot (r - 2) \leq \frac{r + 1}{2} v - \frac{r + 2}{2}. \tag{B.2}$$

Proof. We can rewrite the LHS of (B.2) as

$$\frac{(r + 1)(r - 2)}{2} a + \binom{b}{2} - \max\{2 - b, 0\} \cdot (r - 2).$$

By substituting $v = (r - 2)a + b$ in the RHS, we see that (B.2) is equivalent to

$$\binom{b}{2} - \max\{2 - b, 0\} \cdot (r - 2) \leq \frac{r + 1}{2} b - \frac{r + 2}{2}.$$

To verify this, we consider

$$\begin{aligned} f(b) &:= 2 \left(\frac{r + 1}{2} b - \frac{r + 2}{2} - \left(\binom{b}{2} - \max\{2 - b, 0\} \cdot (r - 2) \right) \right) \\ &= (r + 2 - b)b - (r + 2) + 2 \max\{2 - b, 0\} \cdot (r - 2). \end{aligned}$$

For all $b \neq 2$ we have $f'(b) = (r + 2) - 2b - 2(r - 2) \cdot \mathbb{1}_{\{b < 2\}}$ and $f''(b) = -2$. Since f is concave in $(-\infty, 2)$ and $(2, \infty)$, in order to verify that $f(b) \geq 0$ for all $b \in [0, r - 3]$ (which is then equivalent to (B.2)) it suffices to check the value of $f(b)$ at 0, 1, 2 and $r - 3$ (assuming this is larger than 2). Indeed,

$$\begin{aligned} f(0) &= -(r + 2) + 4(r - 2) = 3r - 10 > 0, \\ f(1) &= (r + 1) - (r + 2) + 2(r - 2) = 2r - 5 > 0, \\ f(2) &= 2r - (r + 2) = r - 2 > 0, \end{aligned}$$

and, if $r - 3 > 2$,

$$f(r - 3) = 5(r - 3) - (r + 2) = 4r - 17 > 0.$$

□

Our goal now is to establish that the densest subgraphs of $K_{r,2,n}$ are precisely those described in Fact B.1. The next lemma establishes that, among all subgraphs of $K_{r,2,n}$ induced by segments of length at most $n/(2r) + 1$ (i.e., there is no ‘wrapping around the cycle’), the densest ones are those where the segment starts in a ‘new’ K_r or ends in a ‘full’ K_r .

Lemma B.3. Let $r \geq 4$ and $n \in \mathbb{N}$ with $(r - 2) \mid n$. Let $V \subseteq V(K_{r,2,n})$ be a segment with $r \leq |V| \leq n/(2r) + 1$. Then, the number of edges induced by V is maximised when V starts in the first vertex of some clique K_r or ends in the last vertex of some clique K_r .

Proof. Let V be a segment which induces the maximum possible number of edges from $K_{r,2,n}$. By the symmetries of $K_{r,2,n}$, we may assume that $V \cap ([0, r - 1] \cup [n - r, n - 1]) = \emptyset$. Assume that V is not of the form described in the claim (i.e., it neither begins in the first vertex nor ends in the last vertex of some clique K_r). Let i_1 and i_2 be the first and last vertices of V , and let j_1 and j_2 be the number of vertices which V contains in the (last) clique K_r which contains i_1 and in the (first)

clique which contains i_2 , respectively. We may assume, without loss of generality, that $j_1 \leq j_2$ (and recall that $j_1, j_2 < r$). However, then the set $(V \setminus \{i_1\}) \cup \{i_2 + 1\} = [i_1 + 1, i_2 + 1]$ induces more edges than V . But this contradicts our choice of V as a set of consecutive vertices which induces the maximum possible number of edges. \square

Now we prove that no subgraph of $K_{r,2,n}$ is denser than the subgraphs induced by segments. In the coming proofs, we will call vertices of $K_{r,2,n}$ with degree $2r - 3$ *heavy* and those with degree $r - 1$ *light*.

Lemma B.4. *Let $r \geq 4$ and $n \in \mathbb{N}$ with $(r - 2) \mid n$. Let $V \subseteq V(K_{r,2,n})$ with $r + 1 \leq v := |V| \leq n/(2r) + 1$. Then, the number of edges induced by V is at most the number induced by a segment of length v .*

Proof. Let $V \subseteq V(K_{r,2,n})$ be a set of cardinality v inducing the maximum possible number of edges. Let I be the graph induced by V . Of course, we may assume I has no isolated vertices. Let S be a smallest segment containing V . By the symmetries of $K_{r,2,n}$, we may assume that $S \cap ([0, r - 1] \cup [n - r, n - 1]) = \emptyset$. We use a compression-type argument to show that we can modify V into a segment which induces at least as many edges as V . We achieve this by consecutively creating new sets V' which are contained in shorter segments but induce at least as many edges as the previous set.

Let i_1 and i_2 be the first and last vertices of V (i.e., $S = [i_1, i_2]$) and notice that i_1 and i_2 do not form an edge (since $r + 1 \leq v \leq n/(2r) + 1$). Observe, then, that $\deg_I(i_1), \deg_I(i_2) \in [r - 1]$. By an argument as in Lemma B.3, we may assume, without loss of generality, that $\deg_I(i_1) \geq \deg_I(i_2)$ and that i_1 is the first vertex of some clique K_r completely contained in I : otherwise, we could replace the vertex i_2 with some missing vertex from such a clique and increase the number of edges.

Let $K^{(1)}$ be the copy of K_r contained in I with the smallest indices (in particular, it contains i_1). Assume that V is not a segment. Consider the vertex $i' \in S \setminus V$ with the smallest index. Observe that $i' \notin V(K^{(1)})$. We distinguish two cases, depending on whether i' is heavy or light.

If i' is heavy, let K' and K'' be the two copies of K_r from $K_{r,2,n}$ with $i' \in V(K') \cap V(K'')$ and K' containing smaller indices than K'' . If $E(K'') \cap E(I) = \emptyset$, then we can shift all edges of I induced by $V \cap [i' + 1, i_2]$ to the left $r - 2$ positions, yielding a graph contained in a segment of length $i_2 - i_1 + 1 - r$ with the same number of edges as I . Hence, we assume that $E(K'') \cap E(I) \neq \emptyset$, which implies V contains at least two of the vertices of K'' . Observe that $K^{(1)} \neq K'$. Then, replace i_1 by i' . In this way, since i_1 is the first vertex from $K^{(1)}$, we remove $r - 1$ edges, but at the same time we add at least r edges. But this contradicts our choice of V .

Assume now that i' is light and let K be the unique clique K_r from $K_{r,2,n}$ with $i' \in V(K)$. Since i' is light, by its definition we must have $|V(K) \setminus V| \leq r - 2$. We replace the first $t := |V(K) \setminus V| \leq r - 2$ vertices from $K^{(1)}$ by adding $V(K) \setminus V$ to V . In this way, we remove $\sum_{i=1}^t (r - i)$ edges, but at the same time we add at least $\sum_{i=1}^t (r - i)$ edges. Since $i_2 > i'$, this means the new set lies in a shorter segment but induces at least as many edges.

In all described situations we managed to make the segment containing V shorter while not decreasing the number of edges. Hence, we eventually find a segment V' inducing at least as many edges as V . \square

The following now follows directly by combining Fact B.1, Proposition B.2 and Lemmas B.3 and B.4 (and noting that the bound holds trivially if $|V| \leq r$).

Corollary B.5. *Let $r \geq 4$ and $n \in \mathbb{N}$ with $(r - 2) \mid n$. Let $V \subseteq V(K_{r,2,n})$ with $|V| \leq n/(2r) + 1$. Then,*

$$e(K_{r,2,n}[V]) \leq \frac{r + 1}{2}|V| - \frac{r + 2}{2}.$$

Combining the different results above, we can obtain the proofs of Lemmas 3.7 and 3.8.

Proof of Lemma 3.7. Let I_1, \dots, I_c be the components of I with at least one edge, and let v_1, \dots, v_c be the number of vertices spanned by I_1, \dots, I_c , respectively. For each $j \in [c]$, since $|I_j| \leq |I| \leq n/(2r)$, by Corollary B.5 we have the following easy bound:

$$|I_j| \leq \frac{r+1}{2} v_j - \frac{r+2}{2}.$$

Summing over all $j \in [c]$, we obtain that

$$\ell = |I| \leq \frac{r+1}{2} |V(I)| - \frac{r+2}{2} c = \frac{r+1}{2} (|V(I)| - c) - \frac{c}{2},$$

and the claim follows by reordering. □

Proof of Lemma 3.8. The vertex set of $K_{r,2,n}$ consists of $t := n/(r-2)$ segments of heavy vertices and t segments of light vertices, which alternate as we traverse the vertex set. The segments of heavy vertices have length 2, and the segments of light vertices have length $r-4$ (the case $r=4$ is special: here the segments of light vertices are empty). For each $i \in [t]$, let h_i and ℓ_i denote the number of heavy vertices and light vertices of I in the i -th segment of heavy or light vertices, respectively. For notational purposes, let $h_{t+1} := h_1$. Then, we can bound the number of edges of I as follows:

$$\begin{aligned} |I| &\leq \sum_{i=1}^t \left(\binom{h_i}{2} + \binom{\ell_i}{2} + h_i \ell_i + h_{i+1} \ell_i + h_{i+1} h_i \right) \\ &\leq \sum_{i=1}^t \left(\binom{h_i}{2} + \binom{\ell_i}{2} + h_i \ell_i + 2\ell_i + 2h_i \right) = \sum_{i=1}^t \left(\binom{h_i + \ell_i}{2} + 2(\ell_i + h_i) \right). \end{aligned}$$

Next, observe that, for each $i \in [t]$,

$$\binom{h_i + \ell_i}{2} + 2(\ell_i + h_i) = (h_i + \ell_i) \left(\frac{h_i + \ell_i - 1}{2} + 2 \right) \leq \frac{r+1}{2} (h_i + \ell_i),$$

where the inequality holds since $h_i + \ell_i \leq r-2$. The conclusion follows by adding over all $i \in [t]$. □