Combinatorics, Probability and Computing (2020), **29**, pp. 113–127
doi:10.1017/S0963548319000129

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

On Komlós' tiling theorem in random graphs

Rajko Nenadov[1,*] and Nemanja Škorić[2]

[1]Department of Mathematics, Monash University, ETH Zurich, 8092 Zürich, Switzerland and [2]Institute of Theoretical Computer Science, ETH Zurich, 8092 Zürich, Switzerland, Email: nskoric@inf.ethz.ch
*Corresponding author. Email: rajkon@gmail.com

(Received 26 January 2017; revised 5 November 2018; first published online 25 July 2019)

Abstract

Given graphs G and H, a family of vertex-disjoint copies of H in G is called an *H-tiling*. Conlon, Gowers, Samotij and Schacht showed that for a given graph H and a constant $\gamma > 0$, there exists $C > 0$ such that if $p \geqslant Cn^{-1/m_2(H)}$, then asymptotically almost surely every spanning subgraph G of the random graph $\mathcal{G}(n, p)$ with minimum degree at least

$$\delta(G) \geqslant \left(1 - \frac{1}{\chi_{cr}(H)} + \gamma\right)np$$

contains an H-tiling that covers all but at most γn vertices. Here, $\chi_{cr}(H)$ denotes the *critical chromatic number*, a parameter introduced by Komlós, and $m_2(H)$ is the 2-*density* of H. We show that this theorem can be bootstrapped to obtain an H-tiling covering all but at most $\gamma(C/p)^{m_2(H)}$ vertices, which is strictly smaller when $p > Cn^{-1/m_2(H)}$. In the case where $H = K_3$, this answers the question of Balogh, Lee and Samotij. Furthermore, for an arbitrary graph H we give an upper bound on p for which some leftover is unavoidable and a bound on the size of a largest H-tiling for p below this value.

2010 MSC Codes: Primary 05C80; Secondary 05C35

1. Introduction

Given graphs G and H, a family of vertex-disjoint copies of H in G is called an *H-tiling*. This generalizes the notion of matchings from edges ($H = K_2$) to arbitrary graphs. The study of sufficient degree conditions of G which enforce the existence of a *perfect* H-tiling (an H-tiling that covers all vertices of G), usually referred to as an *H-factor*, dates back to the seminal work of Corrádi and Hajnal [8] and Hajnal and Szemerédi [11]. In particular, it was shown in [11] that every graph with $n = \ell k$ vertices and minimum degree at least $(\ell - 1)n/\ell$ contains an K_ℓ-factor. Such a bound on the minimum degree is easily seen to be best possible.

Progress towards generalizing this result to an arbitrary graph H was made in [3, 4, 18]. The approximate result was obtained by Komlós [17], who determined the best possible bound on the minimum degree which enforces an H-tiling covering all but at most $o(n)$ vertices. In particular, he showed that the main parameter which governs the existence of such a tiling is the so-called *critical chromatic number* $\chi_{cr}(H)$, defined as

$$\chi_{cr}(H) = \frac{(\chi(H) - 1)v(H)}{v(H) - \sigma(H)},$$

where $\sigma(H)$ denotes the smallest size of a colour class in a colouring of H with $\chi(H)$ colours.

Theorem 1.1. (tiling theorem [17]). *For every graph H and a constant $\gamma > 0$, there exists $n_0 \in \mathbb{N}$ such that if G is a graph with $n \geqslant n_0$ vertices and*

$$\delta(G) \geqslant \left(1 - \frac{1}{\chi_{cr}(H)}\right)n,$$

then G contains an H-tiling that covers all but at most γn vertices.

Theorem 1.1 was further strengthened by Shokoufandeh and Zhao [22] and the problem was fully solved only recently by Kühn and Osthus [20]. We refer the reader to [19, 20] for a detailed survey of the history of the problem and results not mentioned here.

1.1 Tiling in random graphs

In this paper we are interested in the degree to which the stated theorems hold in random graphs. In particular, we consider the binomial random graph model $\mathcal{G}(n,p)$. The obvious question is for which p does $\mathcal{G}(n,p)$ a.a.s.[1] contain an H-factor? The case where $H = K_2$, which corresponds to a perfect matching, has already been answered by Erdős and Rényi [9]. The best known bounds in the general case come from the work of Johansson, Kahn and Vu [13]. In particular, [13] resolves the case where H satisfies certain balancedness conditions (this includes, among others, the case where H is a complete graph) and in all other cases leaves a small gap between the obtained and the best possible value of of p. Some further progress was made by Gerke and McDowell [10]. Once this is (almost) settled, in the spirit of previously mentioned results it is natural to study whether subgraphs of random graphs with sufficiently large minimum degree contain a perfect (or almost-perfect) H-tiling.

It turns out that, once we have the right tools, Komlós's tiling theorem transfers to random graphs in a 'straightforward' way. The right tools are the sparse version of Szemerédi's regularity lemma observed by Kohayakawa [15] and Rödl (unpublished) together with the KŁR conjecture, first stated in [16] and proved much later by Balogh, Morris and Samotij [6] and, independently, Saxton and Thomason [21]. A somewhat different version was obtained by Conlon, Gowers, Samotij and Schacht [7], and in the same paper the authors gave the following theorem as an application. They only stated it for $H = K_\ell$ and remarked that the same proof works for any H.

Theorem 1.2. *For any constant $\gamma > 0$ and a graph H which contains a cycle, there exist constants $b, C > 0$ such that if $p \geqslant Cn^{-1/m_2(H)}$, where*

$$m_2(H) = \max\left\{\frac{e(H')-1}{v(H')-2} : H' \subseteq H \text{ and } v(H') \geqslant 3\right\},$$

then with probability at least $1 - e^{-bn^2p}$ the random graph $\Gamma \sim \mathcal{G}(n,p)$ has the property that every spanning subgraph $G \subseteq \Gamma$ with minimum degree

$$\delta(G) \geqslant \left(1 - \frac{1}{\chi_{cr}(H)} + \gamma\right)np$$

contains an H-tiling that covers all but at most γn vertices.

It is known that for $p \ll n^{-1/m_2(H)}$ a.a.s. there exists a spanning graph $G \subseteq \mathcal{G}(n,p)$ with $\delta(G) = (1 - o(1))np$ which does not contain a copy of H. Therefore, the bound on p in Theorem 1.2 is best possible even if we only want to cover a constant fraction of all the vertices. Moreover, the constructions which show the optimality of $\delta(G)$ in Theorem 1.1 also show that if

[1] Asymptotically almost surely, *i.e.* with probability going to 1 as $n \to \infty$.

https://doi.org/10.1017/S0963548319000129 Published online by Cambridge University Press

we weaken the minimum degree condition to $\delta(G) \geqslant (1 - 1/\chi_{cr}(H) - c)n$ in Theorem 1.2, for any constant $c > 0$, then one cannot hope to cover more than $(1 - c)n$ vertices.

Getting rid of the linear leftover in Theorem 1.2 seems to be a difficult task. Huang, Lee and Sudakov [12] showed that a.a.s. for constant p and minimum degree at least $(1 - 1/\chi(H) + \gamma)np$, there exists a perfect H-tiling if H contains a vertex which does not belong to K_3, and otherwise there exists an H-tiling covering all but at most $O(p^{-2})$ vertices. Moreover, they showed that the bound on the number of leftover vertices in the latter case is optimal up to the constant factor. Significantly improving the bound on p, Balogh, Lee and Samotij [5] showed that for $p \geqslant (C \log n/n)^{1/2}$ a.a.s. every spanning subgraph $G \subseteq \mathcal{G}(n, p)$ with minimum degree $\delta(G) \geqslant (2/3 + \gamma)np$ contains a K_3-tiling that covers all but at most $O(p^{-2})$ vertices. The authors further suggested that the $\sqrt{\log n}$ factor in the bound on p is not needed, which we confirm in Theorem 1.3. Recently, Allen, Böttcher, Ehrenmüller and Taraz [1], relying on a sparse version of the *blow-up lemma* [2], announced that the result of Huang *et al.* [12] holds for $p \geqslant (C \log n/n)^{1/\Delta}$ in a slightly weaker form (in particular, the number of leftover vertices is of order $O(\max\{p^{-2}, p^{-1} \log n\})$ if all vertices of H belong to K_3), where Δ is the maximum degree of a given graph H.

1.2 Our contribution

We give a short proof of the theorem which replaces γn in Theorem 1.2 with $\gamma(C/p)^{m_2(H)}$, which is clearly smaller for all $p > Cn^{-1/m_2(H)}$. The proof is based on simple bootstrapping of Theorem 1.2, which might be of independent interest.

Theorem 1.3. *For any constant $\gamma > 0$ and graph H which contains a cycle, there exists $C > 0$ such that if $Cn^{-1/m_2(H)} \leqslant p \leqslant (\log n)^{-1/(m_2(H)-1)}$, then $\Gamma \sim \mathcal{G}(n, p)$ a.a.s. has the property that every spanning subgraph $G \subseteq \Gamma$ with*

$$\delta(G) \geqslant \left(1 - \frac{1}{\chi_{cr}(H)} + \gamma\right)np$$

contains an H-tiling that covers all but at most $\gamma(C/p)^{m_2(H)}$ vertices.

Let us briefly compare our result to that of Allen *et al.* [1]. On the one hand, for large values of p our theorem gives a weaker bound on the size of a largest H-tiling whenever $m_2(H) > 2$. This difference is the most drastic if H contains a vertex that does not belong to a triangle, in which case the result from [1] gives a perfect H-tiling. On the other hand, our theorem is stronger in the sense that it applies for the whole range of p for which the problem is sensible. The result from [1] requires $p \gg (\log n/n)^{1/\Delta}$, and it is easy to check that for all connected graphs H which contain a cycle, other than $H = K_3$, we have $m_2(H) < \Delta$. In particular, this leaves a gap in the range of p covered by the result from [1] for all such graphs.

As a corollary we answer the question of Balogh, Lee and Samotij [5] about the case where $H = K_3$. As already mentioned, the following result is optimal with respect to all parameters, except for the technical upper bound on p.

Corollary 1.4. *Given a constant $\gamma > 0$, there exists $C > 0$ such that if $Cn^{-1/2} \leqslant p \leqslant (\log n)^{-1}$, then $\Gamma \sim \mathcal{G}(n, p)$ a.a.s. has the property that every spanning subgraph $G \subseteq \Gamma$ with $\delta(G) \geqslant (2/3 + \gamma)np$ contains a K_3-tiling that covers all but at most $\gamma(C/p)^2$ vertices.*

As mentioned earlier, for every graph H that contains a vertex which does not belong to a triangle, a result of Allen *et al.* [1] shows that if $p \gg (\log n/n)^{1/\Delta(H)}$, then a.a.s. every spanning subgraph $G \subseteq \mathcal{G}(n, p)$ with large minimum degree contains a perfect H-tiling. In Section 3 (Theorem 3.2)

https://doi.org/10.1017/S0963548319000129 Published online by Cambridge University Press

we derive an upper bound on p for which some leftover is unavoidable and, for p below this value, an upper bound on the size of a largest H-tiling one can guarantee. The obtained bounds suggest that both Theorem 1.3 (in terms of the size of a largest guaranteed H-tiling) and the result from [1] (in terms of a lower bound on p) are in general not optimal.

Notation

Given a graph $G=(V,E)$, we let $v(G)$ and $e(G)$ denote the size of its vertex and edge set, respectively. For a subset $S \subseteq V$ we use the standard notation $G[S]$ to denote the subgraph of G induced by S, that is, the graph with the vertex set S consisting of the edges of G with both endpoints in S. Given graphs G and H, we say that a function $f\colon V(H) \to V(G)$ is an *embedding* of H into G ($f\colon H \hookrightarrow G$ for short) if it is injective and for every $\{v,w\} \in E(H)$ we have $\{f(v), f(w)\} \in E(G)$.

A partition of a set is a family of pairwise *disjoint* subsets which cover the whole set. Whenever the use of floors and ceilings is not crucial it will be omitted. We use standard asymptotic notation $O, \Omega, \Theta, o, \omega$, and furthermore add $\sim$ on top of each to suppress logarithmic factors (*i.e.* $\tilde{O}$, $\tilde{\Omega}$, *etc.*).

2. Proof of Theorem 1.3

The proof of Theorem 1.3 is based on iterated application of the following corollary of Theorem 1.2.

Lemma 2.1. *For any constant $\gamma > 0$ and a graph H which contains a cycle, there exists $C > 0$ such that if $Cn^{-1/m_2(H)} \leqslant p \leqslant (\log n)^{-1/(m_2(H)-1)}$, then $\Gamma \sim \mathcal{G}(n,p)$ a.a.s. has the property that every subgraph $G \subseteq \Gamma$ with $v(G) \geqslant (C/p)^{m_2(H)}$ and minimum degree*

$$\delta(G) \geqslant \left(1 - \frac{1}{\chi_{\mathrm{cr}}(H)} + \gamma\right) v(G)p$$

contains an H-tiling that covers all but at most $\gamma v(G)$ vertices.

Proof. Let C and b be constants given by Theorem 1.2 applied with H and γ. We may assume that $C > 2/b$. We show that for every subset $S \subseteq V(\Gamma)$ of size $|S| = s \geqslant (C/p)^{m_2(H)}$, the induced subgraph $\Gamma[S]$ has the property that every spanning subgraph $G \subseteq \Gamma[S]$ with minimum degree $\delta(G) \geqslant (1 - 1/\chi_{\mathrm{cr}}(H) + \gamma)sp$ contains an H-tiling that covers all but at most γs vertices.

From $s \geqslant (C/p)^{m_2(H)}$ we have $p \geqslant Cs^{-1/m_2(H)}$, thus by Theorem 1.2 the induced subgraph $\Gamma[S] \sim \mathcal{G}(s,p)$ has the desired property with probability at least $1 - e^{-bs^2p}$. From the upper bound on p we further get

$$s \geqslant \left(\frac{C}{p}\right)^{m_2(H)} \geqslant Cp^{-1}p^{-m_2(H)+1} \geqslant Cp^{-1}\log n.$$

Therefore, $\Gamma[S]$ has the described property with probability at least $1 - n^{-2s}$, which is good enough to handle a union bound over all possible sets S. □

Having Lemma 2.1 at hand we describe our proof strategy. First, we partition the vertex set of G into subsets $V_1 \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} V_q$ of gradually decreasing size, with $V_q = \Theta(p^{-m_2(H)})$ being large enough to satisfy the requirement of Lemma 2.1. By doing this at random we make sure that every vertex has 'good' degree into every such subset. Now apply Lemma 2.1 on the largest subset V_1 to cover all but at most $\gamma|V_1|$ vertices, denoted by U_1. Even though the subgraph $G[U_1]$ might be empty, we know that every vertex in $U_1 \cup V_2$ has good degree into V_2. Crucially, if U_1 is much smaller than V_2, the second largest subset, then the number of neighbours of each $v \in U_1 \cup V_2$ relative to

https://doi.org/10.1017/S0963548319000129 Published online by Cambridge University Press

the size of $U_1 \cup V_2$ is negligibly smaller than relative to the size of V_2. Since the latter is sufficiently large, by carefully choosing the constants we obtain the required minimum degree of $G[U_1 \cup V_2]$ in order to apply Lemma 2.1. In this way we obtain an H-tiling of $G[U \cup V_2]$ that covers all but at most $\gamma(|U_1| + |V_2|) \leqslant 2\gamma|V_2|$ vertices, denoted by U_2, and recall that all the vertices in $V_1 \setminus U$ are already covered. Now we repeat the same on the subgraph $G[U_2 \cup V_3]$ to obtain an H-tiling of $G[V_1 \cup V_2 \cup V_3]$ that covers all but at most $2\gamma|V_3|$ vertices, and so on until we cover all but at most $2\gamma|V_q|$ vertices in G. We now make this precise.

Proof of Theorem 1.3. Let C be a constant given by Lemma 2.1 applied with H and $\gamma/20$ (as γ). We show that if $\Gamma \sim \mathcal{G}(n, p)$ satisfies the property of Lemma 2.1 with these parameters, then every spanning subgraph $G \subseteq \Gamma$ with the required minimum degree contains an H-tiling that covers all but at most $\gamma(C/p)^{m_2(H)}$ vertices. Since the above happens a.a.s. for p as stated, this proves the theorem. For the rest of the proof we let $G \subseteq \Gamma$ be an arbitrary spanning subgraph with $\delta(G) \geqslant (1 - 1/\chi_{\mathrm{cr}}(H) + \gamma)np$.

Let $q \in \mathbb{N}$ be the largest integer such that $n/2^{q-1} > \lceil (C/p)^{m_2(H)} \rceil$ and consider a random partition of $V(G)$ into subsets $V_1, \ldots, V_q$ with $|V_i| = \lfloor n/2^i \rfloor$ for $i \in \{1, \ldots, q-1\}$. Observe that

$$|V_q| = n - \sum_{i<q} \left\lfloor \frac{n}{2^i} \right\rfloor \geqslant n - n \sum_{i<q} 2^{-i} = \frac{n}{2^{q-1}},$$

and similarly $|V_q| \leqslant n/2^{q-1} + q$. Therefore, from $p \leqslant (\log n)^{-1/(m_2(H)-1)}$ we obtain

$$|V_i| \geqslant \frac{n}{2^{q-1}} - 1 \geqslant \left(\frac{C}{p}\right)^{m_2(H)} \geqslant Cp^{-1} \log n \tag{2.1}$$

for every $i \in [q]$. The expected number of neighbours of each vertex $v \in V(G)$ in V_i is at least $(1 - 1/\chi_{\mathrm{cr}}(H) + \gamma)|V_i|p$, thus it follows from the Chernoff's inequality for hypergeometric distributions that

$$\mathbb{P}\left[\deg_G(v, V_i) \leqslant \left(1 - \frac{1}{\chi_{\mathrm{cr}}(H)} + \frac{\gamma}{2}\right)|V_i|p\right] = e^{-\Omega(|V_i|p)} \overset{(2.1)}{<} \frac{1}{n^2}.$$

(In the last inequality we assumed C is sufficiently large.) A simple application of a union bound shows that there exists a partition $V(G) = V_1 \cup \cdots \cup V_q$ with sizes as stated above such that for each $v \in V(G)$ and each V_i we have

$$\deg_G(v, V_i) \geqslant \left(1 - \frac{1}{\chi_{\mathrm{cr}}(H)} + \frac{\gamma}{2}\right)|V_i|p. \tag{2.2}$$

Our plan is to inductively find an H-tiling of $G[V_1 \cup \cdots \cup V_i]$ for $1 \leqslant i \leqslant q$ that covers all but at most $\gamma|V_i|/10$ vertices. A calculation similar to the one in (2.1) shows that

$$|V_q| \leqslant \frac{n}{2^{q-1}} + q \leqslant 2\left(\frac{C}{p}\right)^{m_2(H)} + \log n \leqslant 3\left(\frac{C}{p}\right)^{m_2(H)},$$

where in the second inequality we used the maximality of q and an implicit assumption that n is sufficiently large. Therefore, such an H-tiling for $i = q$ covers all but at most $\gamma|V_q|/10 \leqslant \gamma(C/p)^{m_2(H)}$ vertices of G, which proves the theorem.

For $i = 1$ we get the desired tiling by simply applying Lemma 2.1 on $G[V_1]$. This is indeed possible since $|V_1| \geqslant (C/p)^{m_2(H)}$ (see (2.1)) and the minimum degree condition holds by (2.2). Note that we obtain a slightly larger tiling than needed (*i.e.* we cover all but at most $\gamma|V_1|/20$ vertices).

Next, let us suppose that there exists such an H-tiling for some $i < q$ and let $U \subseteq V_1 \cup \cdots \cup V_i$ denote the subset of vertices which are not covered. Then $|U| \leqslant \gamma |V_i|/10 \leqslant \gamma |V_{i+1}|/4$, and for every vertex $v \in V(G)$ we have

$$\begin{aligned} \deg_G(v, U \cup V_{i+1}) &\geqslant \deg_G(v, V_{i+1}) \\ &\overset{(2.2)}{\geqslant} \left(1 - \frac{1}{\chi_{\mathrm{cr}}(H)} + \frac{\gamma}{2}\right)|V_{i+1}|p \\ &\geqslant \left(1 - \frac{1}{\chi_{\mathrm{cr}}(H)} + \frac{\gamma}{2}\right)\frac{|U| + |V_{i+1}|}{1 + \frac{\gamma}{4}}p \\ &\geqslant \left(1 - \frac{1}{\chi_{\mathrm{cr}}(H)} + \frac{\gamma}{4}\right)|U \cup V_{i+1}|p. \end{aligned}$$

In the last inequality we implicitly assumed that γ is sufficiently small such that $\gamma/4 \leqslant 1/\chi_{\mathrm{cr}}(H)$. In particular, this implies

$$\delta(G[U \cup V_{i+1}]) \geqslant \left(1 - \frac{1}{\chi_{\mathrm{cr}}(H)} + \frac{\gamma}{4}\right)|U \cup V_{i+1}|p,$$

and, as $|V_{i+1}| \geqslant (C/p)^{m_2(H)}$ (see (2.1)), we can apply Lemma 2.1 to obtain an H-tiling of $G[U \cup V_{i+1}]$ that covers all but at most $\gamma|U \cup V_{i+1}|/20 \leqslant \gamma|V_{i+1}|/10$ vertices. Since all vertices in $\bigcup_{j \leqslant i} V_j \setminus U$ are already covered by the tiling obtained for i, this gives the desired H-tiling of $G[V_1 \cup \cdots \cup V_{i+1}]$. □

3. A lower bound on the number of leftover vertices

The upper bound on the number of leftover vertices in Theorem 1.3 asymptotically matches the lower bound obtained by Huang *et al.* [12] in the case of triangles. As we will see shortly, the situation is quite different for arbitrary graphs. The main result of this section is a general bound on p for which some leftover is unavoidable and a lower bound on the size of a leftover in any H-tiling below this value of p, for an arbitrary graph H that contains a cycle (Theorem 3.2). In Section 3.1 we explicitly calculate these bounds in the special cases where H is a cycle and a complete graph, and compare them with Theorem 1.3 and the result from [1]. Theorem 3.2 is then proved in Section 3.2.

Let us first give a heuristic argument for a value of p below which some leftover is unavoidable. For this it will be convenient to assume that H is a labelled graph and we always consider labelled copies of H in $\Gamma \sim \mathcal{G}(n, p)$. Our goal is to remove some edges from Γ, not too many touching each vertex, such that one particular vertex $v \in V(\Gamma)$ does not belong to a copy of H in the resulting graph G. We do this using the following strategy. First, for each vertex $x \in V(H)$ choose an edge $e_x = \{y_x, z_x\} \in E(H)$ such that $x \notin e_x$. Note that such an edge must exist as H contains a cycle. Set $G = \Gamma$ and, sequentially, for each embedding $\phi\colon H \hookrightarrow G$ which uses v, let $x = \phi^{-1}(v)$ and remove the edge $\{\phi(y_x), \phi(z_x)\}$ from G. In other words, we always 'destroy' a copy of H which maps x onto v by removing the edge corresponding to $\{y_x, z_x\}$. The following justifies the decision not to remove e_x which is incident to x: for $p \gg n^{-1/m_2(H)}$ we expect every edge to belong to a copy of H, thus this process would remove almost all edges incident to v.

We now give an estimate of the (expected) largest number of removed edges κ_w incident to some vertex $w \in V(\Gamma) \setminus \{v\}$. Let $\kappa_w(x)$ denote the number of edges incident to w that are removed because of copies of H which mapped x onto v and y_x onto w. The expectation of a random variable $\kappa_w(x)$ is easily seen to be of order at most $n^{v(H)-2}p^{e(H)}$, which is simply the expected number of such copies of H. However, this potentially over-counts the number of deleted edges significantly: if $H' \subseteq H$ is a subgraph which contains x, y_x and z_x and, in expectation, every copy of H' in $\mathcal{G}(n, p)$ extends to at least, say, $K > 1$ copies of H, then we have counted each removed

https://doi.org/10.1017/S0963548319000129 Published online by Cambridge University Press

edge at least K times. In other words, it is not the expected number of copies of H which sit on v and w that we are interested in, but rather the expected number of copies of the 'densest' subgraph H' which contains x, y_x and z_x. This gives the following corrected estimate:

$$\mathbb{E}[\kappa_w(x)] = \Theta\big(\min\big\{n^{v(H')-2}p^{e(H')} \colon H' \subseteq H \text{ such that } x, y_x, z_x \in V(H')\big\}\big). \tag{3.1}$$

However, the problem now is that $\kappa_w(x)$ is not concentrated, that is, there could be a vertex w for which $\kappa_w(x)$ is much larger than its expectation. For example, if there is an edge between x and y_x, then by conditioning on w being a neighbour of v, we get

$$\begin{aligned}&\mathbb{E}[\kappa_w(x) \mid w \text{ and } v \text{ are neighbours}]\\ &\quad = \Theta\big(\min\big\{n^{v(H')-2}p^{e(H')-1} \colon H' \subseteq H \text{ such that } x, y_x, z_x \in V(H')\big\}\big).\end{aligned} \tag{3.2}$$

Note that in both (3.1) and (3.2), any subgraph H' which attains the minimum is necessarily induced (assuming $p \ll 1$, which will indeed be the case).

Before we continue, let us first compare (3.1) and (3.2) on a concrete example. If H is a triangle then, for each $x \in H$, the edge $\{y_x, z_x\}$ is the unique edge that is not adjacent to x. The quantity in (3.1) amounts to np^3, while the quantity in (3.2) is np^2. On the other hand, if a vertex w is not adjacent to v, then $\kappa_w(x) = 0$. As there are $\Theta(np)$ vertices adjacent to v, we see that the average value of $\kappa_w(x)$ is np^3, as given by (3.1), while the maximum is significantly larger.

Of course, there is nothing special about the assumption that x and y_x are adjacent in H, which after all might not always be the case, and we could instead just condition on w and v being in a correctly labelled copy of H. Indeed, if w and v do not lie in a copy of H then $\kappa_w(x) = 0$, thus such a conditioning is in a way necessary, and gives

$$\begin{aligned}&\mathbb{E}[\kappa_w(x) \mid w \text{ and } v \text{ belong to a copy of } H]\\ &\quad = \Theta\left(\min_{H'} \max_{R} \left\{ n^{v(H')-v(R)}p^{e(H')-e(R)} \colon \begin{array}{l} R \subseteq H' \subseteq H \text{ such that} \\ x, y_x, z_x \in V(H') \text{ and } x, y_x \in V(R) \end{array} \right\}\right).\end{aligned} \tag{3.3}$$

Note that the right-hand side of (3.3) is determined by induced subgraphs H' and R, and moreover the same calculations hold if we replace y_x with z_x.

Finally, different choices for an edge e_x might give a different maximum number of removed edges. As we require this to be at most a tiny fraction of np, we need that for every $x \in H$ there exists an edge $e_x = \{y_x, z_x\}$ such that

$$\min_{S} \max_{R} \left\{ n^{|S|-|R|}p^{e(H[S])-e(H[R])} \colon \begin{array}{l} R \subseteq S \subseteq V(H) \text{ such that } x, y_x, z_x \in S, \\ x \in R \text{ and } |R \cap e_x| = 1 \end{array} \right\} \leqslant \varepsilon np, \tag{3.4}$$

for a sufficiently small constant $\varepsilon > 0$ (recall that the value in (3.3) is attained for some induced subgraphs H' and R of H). This motivates the following definition.

Definition 3.1. Let $H = (V, E)$ be a graph. We define the *H-removal-density* $r(H)$ as follows:

$$r(H) = \max_{x \in V} \min_{\substack{e_x \in E \\ x \notin e_x}} \min_{S} \max_{R} \left\{ \frac{|S| - |R| - 1}{e(H[S]) - e(H[R]) - 1} \colon \begin{array}{l} R \subseteq S \subseteq V(H) \text{ such that } x, y_x, z_x \in S, \\ x \in R \text{ and } |R \cap e_x| = 1 \end{array} \right\},$$

where we define $a/0 = \infty$ for every $a \geqslant 0$.

It is easy to see that if H contains a cycle then $r(H) < \infty$ (for example, for each x choose e_x to be an edge on a cycle not incident to x and $S = V(H)$), thus the convention $a/0 = \infty$ is used just to make the function $r(H)$ well-defined. Moreover, taking e_x and S which minimize the part of $r(H)$ associated with $x \in V(H)$, we see that the inequality (3.4) holds provided $p \leqslant \beta n^{-r(H)}$ for some small constant $\beta = \beta(\varepsilon)$ (we shall explain this in more detail after stating Theorem 3.2).

https://doi.org/10.1017/S0963548319000129 Published online by Cambridge University Press

The following theorem shows that the intuition behind the described removal strategy and its analysis is indeed correct, at least when we replace β with a function which slowly goes to 0. Additionally, it gives a lower bound on the number of vertices one can *isolate* from copies of H without harming the minimum degree significantly.

Theorem 3.2. *Given $\varepsilon > 0$ and a graph $H = (V, E)$ which contains a cycle, there exists $c, K > 0$ such that if $n^{-1/m_2(H)} < p < n^{-r(H)}(K(\log n)^{2e(H)})^{-1}$, then $\Gamma \sim \mathcal{G}(n, p)$ a.a.s. contains a spanning subgraph $G \subseteq \Gamma$ with $\delta(G) \geqslant (1-\varepsilon)np$ such that at least $\lfloor q_p(H) \rfloor$ vertices do not belong to a copy of H in G,*

$$q_p(H) = \min_{x \in V} \max \left\{ \frac{c}{n^{|S|-3} p^{e(H[S])-1}} : S \subseteq V, x \in S \text{ and } S \text{ is } (x)\text{-admissible} \right\},$$

where a subset S is (x)-admissible *if there exists an edge $e \in H[S]$ such that $x \notin e$ and*

$$\max\left\{ n^{|S|-|R|} p^{e(H[S])-e(H[R])} : R \subseteq S, x \in R \text{ and } |R \cap e_x| = 1 \right\} < \frac{np}{K(\log n)^{2e(H)}}.$$

To see that $q_p(H)$ is well-defined it is enough to show that for each x there exists an (x)-admissible subset S. This can be seen by taking an edge e_x and a set S which minimize the part of $r(H)$ associated with x:

$$\max_{\substack{x \in R \subset S \\ |R \cap e| = 1}} n^{|S|-|R|} p^{e(H[S])-e(H[R])} = \max_{\substack{x \in R \subset S \\ |R \cap e| = 1}} n^{|S|-|R|-1} p^{e(H[S])-e(H[R])-1} \cdot np \leqslant \frac{np}{K(\log n)^{2e(H)}}, \tag{3.5}$$

where in the last inequality we used an upper bound on p, the fact that

$$\frac{|S| - |R| - 1}{e(H[S]) - e(H[R]) - 1} \leqslant r(H)$$

and that for such e_x and S and any $R \subseteq S$ satisfying required conditions we have $e(H[S]) - e(H[R]) - 1 \geqslant 1$. Indeed, if the last observation fails then $r(H) = \infty$, which we have already ruled out.

The following corollary of Theorem 3.2 verifies our heuristic argument that below $n^{-r(H)}$ it is not possible to guarantee an H-factor. Note that the only thing we need to show is that for such p we have $q_p(H) \geqslant 1$.

Corollary 3.3. *Let H be a graph which contains a cycle. Then there exists $K > 0$ such that if $p < n^{-r(H)}(K(\log n)^{2e(H)})^{-1}$, then $\Gamma \sim \mathcal{G}(n, p)$ a.a.s. contains a spanning subgraph $G \subseteq \Gamma$ with $\delta(G) \geqslant (1-\varepsilon)np$ which does not contain an H-factor.*

Proof. Consider an arbitrary $x \in V$. Let S be an (x)-admissible set and $e \in H[S]$ be a witness for that. Then by the definition we have

$$n^{|S|-3} p^{e(H[S])-1} \leqslant \max_{\substack{x \in R \subset S \\ |R \cap e| = 1}} n^{|S|-|R|} p^{e(H[S])-e(H[R])} \leqslant \frac{1}{K(\log n)^{2e(H)}},$$

where the left-hand side corresponds to $R = \{x, y_x\}$. Thus $q_p(H) \geqslant 1$ for sufficiently large n. □

Unfortunately, the values of both $r(H)$ and $q_p(H)$ are difficult to estimate. In the next section we do this in the case of cycles and complete graphs. We finish this part of the section with a couple of results which, even though not always optimal, give a quick estimate of $r(H)$ and $q_p(H)$.

https://doi.org/10.1017/S0963548319000129 Published online by Cambridge University Press

Lemma 3.4. *Let H be a graph which contains a cycle. Then*

$$r(H) \leqslant \max_{x \in V(H)} \min\left\{ \frac{|S|-3}{e(H[S])-2} : S \subseteq V(H) \text{ such that } H[S] \text{ contains a cycle} \right\} \leqslant \frac{v(H)-3}{e(H)-2}.$$

Proof. Consider some $x \in V$ and let $H_x \subseteq H$ be a subgraph which minimizes

$$\frac{v(H_x)-3}{e(H_x)-2} \tag{3.6}$$

subject to $x \in V(H_x)$ and H_x containing a cycle. Note that then there exists an edge $e_x \in H[S]$ such that $x \notin e_x$ and H_x is necessarily a subgraph induced by some subset S. Consider a subset $R \subset S$ such that $x \in R$ and $|R \cap e_x| = 1$, and let H' be the graph obtained by adding the edge e_x to the induced graph $H[R]$. As R contains only one endpoint of e_x, we have $|R| = v(H') - 1$ and $e(H[R]) = e(H') - 1$. Note that H' satisfies conditions under which we minimized (3.6), thus

$$\frac{|S|-(|R|+1)}{e(H[S])-(e(H[R])+1)} = \frac{v(H_x)-v(H')}{e(H_x)-e(H')} = \frac{(v(H_x)-3)-(v(H')-3)}{(e(H_x)-2)-(e(H')-2)} \leqslant \frac{v(H_x)-3}{e(H_x)-2}.$$

This implies the desired bound on $r(H)$ by its definition. □

The following corollary of Theorem 3.2 gives a simpler estimate of the lower bound on the number of vertices which can be isolated from copies of H.

Corollary 3.5. *Given $\varepsilon > 0$ and a graph H which contains a cycle, there exists $c, K > 0$ such that if $n^{-1/m_2(H)} \leqslant p \leqslant n^{-r(H)}(K(\log n)^{2e(H)})^{-1}$, then $\Gamma \sim \mathcal{G}(n,p)$ a.a.s. contains a spanning subgraph G with minimum degree $\delta(G) \geqslant (1-\varepsilon)np$ such that at least*

$$\left\lfloor \frac{c}{n^{v(H)-3}p^{e(H)-1}} \right\rfloor$$

vertices do not belong to a copy of H.

Proof. Consider a vertex $x \in V(H)$. By Theorem 3.2, it suffices to show that there exist $S \subseteq V(H)$ and $e_x \in E(H[S])$ with $x \notin e_x$, such that

$$\max_{\substack{x \in R \subset S \\ |R \cap e_x| = 1}} n^{|S|-|R|} p^{e(H[S])-e(H[R])} < \frac{np}{K(\log n)^{2e(H)}} \tag{3.7}$$

and

$$n^{|S|} p^{e(H[S])} \leqslant n^{v(H)} p^{e(H)}. \tag{3.8}$$

By the discussion following the statement of Theorem 3.2, there exists a pair (S, e_x) which satisfies (3.7). If it also satisfies (3.8) then we are done. Otherwise, let $S \subset S' \subseteq V(H)$ be a subset which minimizes $n^{|S'|}p^{e(H[S'])}$, so

$$n^{|S|} p^{e(H[S])} > n^{v(H)} p^{e(H)} \geqslant n^{|S'|} p^{e(H[S'])}. \tag{3.9}$$

It remains to show that (S', e_x) also satisfies (3.7).

To prove (S', e_x) satisfies (3.7), it suffices to show that for each $R' \subseteq S'$ such that $x \in R' \subset S'$ and $|R' \cap e_x| = 1$, there exists $R \subseteq S$ with $x \in R \subset S$ and $|R \cap e_x| = 1$ such that

$$n^{|S'|-|R'|} p^{e(H[S'])-e(H[R'])} \leqslant n^{|S|-|R|} p^{e(H[S])-e(H[R])}. \tag{3.10}$$

Consider a subset $R' \subseteq S'$. If $R' \subseteq S$ then it follows from (3.9) that $R = R'$ satisfies this property. Otherwise, let $S'' = S \cup R' \subseteq S'$ and $R = R' \cap S$. Then

$$n^{|S''|-|R'|} p^{e(H[S''])-e(H[R'])} = n^{|S|-|R|} p^{e(H[S])-e(H[R])},$$

so (3.10) again follows from the choice of S'. □

https://doi.org/10.1017/S0963548319000129 Published online by Cambridge University Press

3.1 Cycles and complete graphs

In the following we use C_t to denote a cycle on t vertices.

Corollary 3.6. (cycles). *Given $t \in \{2k, 2k+1\}$ for some integer $k \geqslant 2$ and a constant $\varepsilon > 0$, there exist $c, K > 0$ such that if*

$$n^{-1/m_2(C_t)} \leqslant p \leqslant n^{-(k-1)/k}(K(\log n)^{2t})^{-1},$$

then $\Gamma \sim \mathcal{G}(n,p)$ a.a.s. contains a spanning subgraph G of minimum degree at least $(1-\varepsilon)np$ such that at least

$$\left\lfloor \frac{c}{n^{t-3}p^{t-1}} \right\rfloor \geqslant 1$$

vertices do not belong to a copy of C_t in G.

Proof. In order to apply Corollary 3.5 we need to show $r(C_t) \leqslant (k-1)/k$. Note that C_t is vertex-transitive, so we only need to consider arbitrary vertex $x \in V(C_t)$.

Let $S = V(C_t)$ and choose $e_x = \{y, z\}$ to be the edge such that x is at distance k from y, and at distance $t-k-1$ from z. We show that for every $R \subseteq S$ which contains x and exactly one vertex from e_x we have

$$\frac{t-|R|-1}{t-e(C_t[R])-1} \leqslant \frac{k-1}{k}.$$

If $C_t[R]$ is not connected, then $e(C_t[R]) \leqslant |R|-2$ and

$$\frac{t-|R|-1}{t-e(C_t[R])-1} \leqslant \frac{t-|R|-1}{t-|R|+1}.$$

If the right-hand side of the above inequality is larger than $(k-1)/k$, then $(t-|R|-1)k > (t-|R|+1)(k-1)$, which after rearranging gives $t-|R| > 2k-1$. This is a contradiction as $|R| \geqslant 2$ and $t \leqslant 2k+1$.

It remains to consider the case where $C_t[R]$ is connected. Note that then $C_t[R]$ is a path, thus $(t-|R|-1)/(t-e(C_t[R])-1) = (t-|R|-1)/(t-|R|)$. As R needs to contain either y or z, the path it induces has to be of size at least $\min\{k, t-k-1\}$. Then $|R| \geqslant \min\{k+1, t-k\}$, and plugging both values gives a desired bound. To conclude, $S = V(C_t)$ and the chosen edge e_x are a witness for $r(C_t) \leqslant (k-1)/k$. □

Tedious analysis shows that Corollary 3.6 is the best one can get out of Theorem 3.2. The case when $t \in \{4,5\}$ gives some evidence that at least the upper bound on p in Corollary 3.6 is best possible up to a logarithmic factor: Corollary 3.6 shows that some leftover is unavoidable when $p = \tilde{O}(n^{-1/2})$ and the result of Allen *et al.* [1] shows that if $p = \tilde{\Omega}(n^{1/2})$, then every subgraph of $\mathcal{G}(n,p)$ with minimum degree at least $(1/2+o(1))np$ and $(2/3+o(1))np$ contains a C_4- and C_5-factor, respectively (assuming the divisibility constraint). It remains a challenging open problem to determine if this is indeed the case for all other cycles, that is, to show that if $p = \tilde{\Omega}(n^{-r(C_t)})$, then a.a.s. every subgraph of $\mathcal{G}(n,p)$ with minimum degree $(1-1/\chi(C_t)+o(1))np$ contains a C_t-factor.

In the regime of p covered by Corollary 3.6, Theorem 1.3 gives a C_t-tiling that covers all but at most $O(p^{-m_2(C_t)})$ vertices. For $t=4$ this is of order $p^{-3/2}$ whereas Corollary 3.6 shows that one cannot guarantee a leftover smaller than $\Omega((np^3)^{-1})$. In particular, this leaves a gap in the whole range $n^{-2/3} \ll p \leqslant n^{-1/2}$. Similarly, for $t=5$ we have an upper bound on the leftover of order $p^{-4/3}$ whereas Corollary 3.6 only gives $\Omega(1/(n^2p^4))$. As we take larger cycles the discrepancy becomes bigger.

The next claim is a corollary of Theorem 3.2 applied to the case of complete graphs.

https://doi.org/10.1017/S0963548319000129 Published online by Cambridge University Press

Corollary 3.7. (complete graphs). *Given an integer $t \geqslant 3$ and a constant $\varepsilon > 0$, there exists $c > 0$ such that if*

$$n^{-1/m_2(K_t)} \leqslant p \ll 1,$$

then $\Gamma \sim \mathcal{G}(n,p)$ a.a.s. contains a spanning subgraph G of minimum degree at least $(1-\varepsilon)np$ such that at least

$$\left\lfloor \max_{\ell \in \{3,\dots,t\}} \left\{ \frac{c}{n^{\ell-3} p^{\binom{\ell}{2}-1}} \right\} \right\rfloor \tag{3.11}$$

vertices are not contained in a copy of K_t.

Proof. Note that for $p \geqslant n^{-1/(t-1)}$ we have $n^3p^3 \leqslant n^\ell p^{\binom{\ell}{2}}$ for every $\ell \in \{4,\dots,t\}$, thus the maximum in (3.11) is achieved for $\ell = 3$. As $t \geqslant 3$, we can apply a result of Huang *et al.* [12] to conclude that one can 'isolate' $\Omega(p^{-2})$ vertices from copies of K_3.

We treat the remaining regime of p, that is, $n^{-1/m_2(K_t)} \leqslant p < n^{-1/(t-1)}$, using Theorem 3.2. As K_t contains a triangle it is easy to see $r(K_t) = 0$, so we can indeed apply it. We now estimate $q_p(K_t)$.

Every complete graph is trivially vertex-transitive, so we can consider an arbitrary vertex $x \in V(K_t)$. Consider a subset $S \subseteq V(K_t)$ which maximizes (3.11) conditioned on $|S| \geqslant 3$. Then such an S also minimizes $n^{|S|}p^{\binom{|S|}{2}}$. We show that for an arbitrary edge $e \in K_t[S]$ which does not contain x, we have

$$\max_{\substack{x \in R \subset S \\ |R \cap e| = 1}} n^{|S|-|R|} p^{e(K_t[S]) - e(K_t[R])} \leqslant \frac{np}{K(\log n)^{2e(H)}}, \tag{3.12}$$

which shows that

$$q_p(K_t) \geqslant \frac{c}{n^{|S|-3} p^{\binom{|S|}{2}-1}},$$

as desired. Let $R \subseteq S$ be a subset of size $|R| \geqslant 3$. By the choice of S we have

$$n^{|R|} p^{\binom{|R|}{2}} \geqslant n^{|S|} p^{\binom{|S|}{2}},$$

so (3.12) clearly holds. Suppose now that $|R| = 2$. By the choice of S we have

$$n^{|S|} p^{e(K_t[S])} \leqslant n^3 p^3.$$

As $K_t[R]$ contains exactly one edge, this is easily seen to imply (3.12). □

Recall that the bound of Huang *et al.* [12] implies that for any $\varepsilon > 0$ and $t \geqslant 4$ there exists a constant C such that if $p \geqslant Cn^{-m_2(K_t)}$, then $\mathcal{G}(n,p)$ a.a.s. contains a spanning subgraph with minimum degree $(1-\varepsilon)np$ such that $\Omega(p^{-2})$ vertices do not belong to a copy of K_3 and therefore K_t. The corollary above guarantees a larger set of 'isolated' vertices in the certain range of p. For example, in the case of K_4 and p in the interval $n^{-1/m_2(K_4)} = n^{-2/5} \leqslant p \ll n^{-1/3}$, we obtain a spanning subgraph which contains $c/(np^5) \gg 1/p^2$ vertices that are not contained in a copy of K_4. On the other hand, Theorem 1.3 gives the existence of a K_4-tiling that covers all but at most $O(p^{-m_2(K_4)}) = O(p^{-5/2})$ vertices. In particular, this leaves the gap in such a range of p. In the case of larger complete graphs the situation becomes more complicated.

3.2 Proof of Theorem 3.2

We make use of a concentration inequality by Kim and Vu [14], slightly rephrased for our particular application.

Theorem 3.8. (Kim–Vu polynomial concentration). *Let $\mathcal{H} = (V(\mathcal{H}), E(\mathcal{H}))$ be a k-uniform hypergraph on N vertices and let $\{t_v : v \in V(\mathcal{H})\}$ be a set of mutually independent Bernoulli random variables with $\mathbb{E}[t_v] = p$. For any subset of vertices $T \subseteq V(\mathcal{H})$, let us denote*

$$Y_T = \sum_{\substack{h \in E(\mathcal{H}) \\ T \subseteq h}} \prod_{v \in h \setminus T} t_v.$$

For $i \in \{0, 1, \ldots, k\}$ we set

$$E_i = \max_{T \subseteq V(\mathcal{H}), |T| = i} \mathbb{E}[Y_T],$$

and let $E' = \max_{i \geqslant 1} E_i$ and $E_{\mathcal{H}} = \max_{i \geqslant 0} E_i$. Then for any $\lambda > 1$ we have

$$\mathbb{P}\Big[|Y_\emptyset - E_0| > a_k \sqrt{E' E_{\mathcal{H}}} \lambda^k\Big] = O\big(\exp(-\lambda + (k-1)\log n)\big),$$

where $a_k = 8^k (k!)^{1/2}$.

Proof of Theorem 3.2. Let $x \in V(H)$ and consider a subset $S_x \subseteq V(H)$ and an edge $e_x \in E(H[S_x])$ with $x \notin e_x$ such that

$$\max_{\substack{x \in R \subset S_x \\ |R \cap e_x| = 1}} n^{|S_x| - |R|} p^{e(H[S_x]) - e(H[R])} \leqslant \frac{np}{K(\log n)^{2e(H)}}, \tag{3.13}$$

for a sufficiently large constant K. Such a pair (S_x, e_x) exists by the definition of $r(H)$ and an upper bound on p (see the discussion following Theorem 3.2). We show that for a given subset $Q \subseteq V(\Gamma)$ of size

$$|Q| \leqslant q_x = \frac{c}{n^{|S_x| - 3} p^{e(H[S_x]) - 1}},$$

$\Gamma \sim \mathcal{G}(n, p)$ a.a.s. has the property that for every $w \in V(\Gamma)$ there are at most

$$4 q_x n^{|S_x| - 2} p^{e(H[S_x])} < \varepsilon np / v(H)$$

copies of $H[S_x]$ that map x to Q and an endpoint of e_x to w. Deleting an edge corresponding to e_x from each copy of $H[S_x]$ that maps x to Q removes at most $\varepsilon np/v(H)$ edges touching each vertex and clearly destroys all the copies of $H[S_x]$, and therefore H, that map x to Q. Once we show this, the theorem follows by choosing the same Q of size $q_p(H)$ for every x (where S_x is chosen such that $q_x \geqslant q_p(H)$).

First, we may suppose that

$$n^{|S_x|} p^{e(H[S_x])} < n^{|S'|} p^{e(H[S'])} \tag{3.14}$$

for every $S' \subset S_x$ that contains x and both endpoints of e_x. If this is not the case then consider any such S' that violates (3.14). The pair (S', e_x) satisfies (3.13) since any $R \subseteq S'$ is also a subset of S_x, and furthermore allows for a larger set Q. Thus we may prove the claim for $S_x := S'$ instead.

Without loss of generality, we may assume $|Q| = q_x$ (ignoring the possible rounding error). For the rest of the proof consider some vertex $w \in V(\Gamma)$ and the copies of $H[S_x]$ which map x to Q and an endpoint of e_x to w. Let us restate what we aim to prove in the language of Theorem 3.8. Define $\mathcal{H}$ to be an $e(H_x)$-uniform hypergraph, where $H_x = H[S_x]$, such that the vertex set corresponds to edges of the complete graph on $V(\Gamma)$ and a set of edges $E \subseteq V(\mathcal{H})$ forms a hyperedge if and only if the edges in E induce a graph isomorphic to H_x such that x is mapped to some vertex in Q and an endpoint of e_x to w. For each vertex $v_e \in V(\mathcal{H})$ we have $\mathbb{E}[t_{v_e}] = p$ (as v_e corresponds to an edge of the underlying complete graph and $\Gamma \sim \mathcal{G}(n, p)$) and $Y = Y_\emptyset$ is exactly the number of all desired copies of H_x. The expected value of Y is of order

$$E_0 = \mathbb{E}[Y] = \Theta\big(q_x n^{|S_x| - 2} p^{e(H_x)}\big)$$

https://doi.org/10.1017/S0963548319000129 Published online by Cambridge University Press

and at most $E_0 \leqslant 2q_x n^{|S_x|-2} p^{e(H[S_x])}$: there are two choices for which endpoint of e_x we map onto w, q_x choices for x and for every other vertex at most n choices. By choosing c such that $E_0 \leqslant \varepsilon np/2v(H)$, it suffices to show $\mathbb{P}[Y_0 > 2|E_0|] = 1/n^2$ as it allows us to take a union bound over all vertices $w \in V(\Gamma)$. This will follow from Theorem 3.8 once we give good estimates of E'.

Note that for every $T \subseteq V(\mathcal{H})$ such that T is not a subset of any $h \in E(\mathcal{H})$, we have $Y_T = 0$. Therefore we can partition all the relevant sets $\{T \subseteq h\}_{h \in E(\mathcal{H})}$ into groups depending on the subgraph of H they induce. More precisely, for each $F \subseteq H_x$ let $\mathcal{T}_F$ denote the set of those $T \subseteq V(\mathcal{H})$ that induce a copy of F with the restriction that if $x \in F$ then it is mapped to Q, and if an endpoint of e_x belongs to G then it has to be mapped to w (if both endpoints are present either of them can be mapped). For each two $T, T' \in \mathcal{T}_F$ we have $\mathbb{E}[Y_T] = \mathbb{E}[Y_{t'}]$ which, by the definition, denotes the expected number of desired copies of H_x which extend some fixed copy of F. Let us denote this quantity by $Z_F = \mathbb{E}[Y_T]$ (for any $T \in \mathcal{T}_F$). We now have

$$E' = \max_{\substack{T \subseteq V(\mathcal{H}) \\ |T| \geqslant 1}} \mathbb{E}[Y_T] = \max_{F \subseteq H_x} Z_F.$$

In order to apply Theorem 3.8 to deduce $\mathbb{P}[Y_0 > 2|E_0|] = 1/n^2$, it suffices to show

$$\mathbb{E}[Y] \geqslant (\log n)^{e(H)} \sqrt{C\mathbb{E}[Y] Z_F} \tag{3.15}$$

for every $F \subseteq H_x$ and a sufficiently large constant C. Note that this immediately gives $E_{\mathcal{H}} = E_0$, so we can take $\lambda = C' \log n$ for some C' which grows with C, and Theorem 3.8 gives the desired probability. Equivalently, we show $\mathbb{E}[Y]/Z_F \geqslant C(\log n)^{2e(H)}$.

If $V(F) = S_x$, then from $F \subseteq H_x$ we have $Z_F \leqslant 1$ and we are done (it follows from the lower bound on p). Next, suppose that F is not an induced subgraph. Then the induced graph $F' = H_x[V(F)]$ contains more edges than F, so $Z_F < Z_{F'}$ and it suffices to show (3.15) for F'. To summarize, we only need to consider the case where F is a subgraph of H_x induced by a subset $R \subset S_x$.

If $x \notin R$ then $Z_F = O(q_x n^{|S_x|-|R|-1} p^{e(H[S_x])-e(H[R])})$ (if $R \cap e_x \neq \emptyset$ then we can improve this bound further, but this is not necessary), which gives

$$\mathbb{E}[Y]/Z_F = \Theta\big(n^{|R|-1} p^{e(H[R])}\big) = \Omega(np).$$

The last inequality follows from $p \geqslant n^{-1/m_2(H)}$. Otherwise, if $x \in R$ and $|R \cap e_x| = 1$ then (3.15) follows from (3.13) for some $K = K(C)$, and if $x \in R$ and $|R \cap e_x| = 2$ then it follows from (3.14). Finally, if $x \in R$ and $|R \cap e_x| = 0$ then $Z_F = \Theta(n^{|S|-|R|-1} p^{e(H[S])-e(H[R])})$, which is of the same order as $Z_{F'}$ where F' is obtained from F by adding either of the endpoints from e_x (and no edges incident to such a vertex). We have already covered this case, which concludes the proof. □

4. Concluding remarks

Using a simple bootstrapping approach, we showed that if $p \geqslant Cn^{-1/m_2(H)}$ then $\mathcal{G}(n,p)$ a.a.s. has the property that every spanning subgraph with the minimum degree at least $(1 - 1/\chi_{\mathrm{cr}}(H) + \gamma)np$ contains an H-tiling that covers all but at most $O(p^{-m_2(H)})$ vertices. As observed in [12] this is the best one can hope for in the case where $H = K_3$, since $\mathcal{G}(n,p)$ contains a spanning subgraph with minimum degree $(1 - o(1))np$ and a set of $\Omega(p^{-m_2(K_3)})$ vertices which do not belong to a copy of K_3.

In the case of larger cliques the discrepancy between our lower (Corollary 3.7) and upper bounds (Theorem 1.3) on the number of leftover vertices becomes more significant. It would be interesting to reduce this gap, but the fact that there will always have to be some leftover makes this question somewhat less appealing. We refer the reader to Section 3.1 for further discussion. We believe the following question is worth considering.

https://doi.org/10.1017/S0963548319000129 Published online by Cambridge University Press

Question 1. Let $t \geqslant 3$ be an integer and $\varepsilon > 0$, and suppose $p \gg n^{-1/m_2(K_t)}$ and $t \mid n$. Is it true that $\Gamma \sim \mathcal{G}(n,p)$ a.a.s. has the property that every spanning subgraph $G \subseteq \Gamma$ such that

- the minimum degree of G is at least $(1 - 1/t + o(1))np$, and
- every vertex is contained in $\varepsilon n^{t-1} p^{\binom{t}{2}}$ copies of K_t

contains a K_t-factor?

Let us now turn our attention to the question of when an H-factor is possible without such extra assumptions. Note that if $p > Cn^{-r(H)}$ (see Definition 3.1) for some large constant C, then, at least in expectation, the deletion process described in Section 3 removes almost all the edges incident to some vertex. However, this just shows that our removing strategy does not work well past $n^{-r(H)}$; for example, one can easily destroy *all* copies of H in $\mathcal{G}(n,p)$ by removing $(1+\varepsilon)np/(\chi(H)-1)$ edges incident to each vertex. We leave the question of whether one can destroy all copies of H containing some fixed v by removing at most $(1-\varepsilon)np/\chi(H)$ edges incident to each vertex when $p = Cn^{-1/r(H)}$ for future research. It is tempting to conjecture that this is not the case. Assuming that the main obstacle towards making an H-factor is if, trivially, there exists a vertex which does not belong to a copy of H, this prompts the following question.

Question 2. Let H be a graph which contains a cycle and let $\varepsilon > 0$. Is it true that if $p = \tilde{\Omega}(n^{-r(H)})$, then $\Gamma \sim \mathcal{G}(n,p)$ a.a.s. has the property that every spanning subgraph $G \subseteq \Gamma$ with minimum degree at least $(1 - 1/\chi(H) + \varepsilon)np$ contains an H-factor?

As pointed out in Section 3.1, a result of Allen *et al.* [1] answers this question in the affirmative for $H = C_4$ and $H = C_5$. If this is true, it would be interesting to further relax the minimum degree condition akin to the one from [20].

Acknowledgement

The authors would like to thank Angelika Steger and Miloš Trujić for pointing out that an upper bound on p in Corollary 3.6 stated in an earlier version of the paper can be improved.

References

[1] Allen, P., Böttcher, J., Ehrenmüller, J. and Taraz, A. (2015) Local resilience of spanning subgraphs in sparse random graphs. *Electron. Notes Discrete Math.* **49** 513–521.
[2] Allen, P., Böttcher, J., Hàn, H., Kohayakawa, Y. and Person, Y. (2016) Blow-up lemmas for sparse graphs. arXiv:1612.00622
[3] Alon, N. and Yuster, R. (1992) Almost H-factors in dense graphs. *Graphs Combin.* **8** 95–102.
[4] Alon, N. and Yuster, R. (1996) H-factors in dense graphs. *J. Combin. Theory Ser. B* **66** 269–282.
[5] Balogh, J., Lee, C. and Samotij, W. (2012) Corrádi and Hajnal's theorem for sparse random graphs. *Combin. Probab. Comput.* **21** 23–55.
[6] Balogh, J., Morris, R. and Samotij, W. (2015) Independent sets in hypergraphs. *J. Amer. Math. Soc.* **28** 669–709.
[7] Conlon, D., Gowers, W., Samotij, W. and Schacht, M. (2014) On the KŁR conjecture in random graphs. *Israel J. Math.* **203** 535–580.
[8] Corrádi, K. and Hajnal, A. (1963) On the maximal number of independent circuits in a graph. *Acta Math. Acad. Sci. Hungar.* **14** 423–439.
[9] Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** 17–61.
[10] Gerke, S. and McDowell, A. (2015) Nonvertex-balanced factors in random graphs. *J. Graph Theory* **78** 269–286.
[11] Hajnal, A. and Szemerédi, E. (1970) Proof of a conjecture of P. Erdős. In *Combinatorial Theory and its Applications, II (Proc. Colloq., Balatonfüred, 1969)*, North-Holland, pp. 601–623.
[12] Huang, H., Lee, C. and Sudakov, B. (2012) Bandwidth theorem for random graphs. *J. Combin. Theory Ser. B* **102** 14–37.
[13] Johansson, A., Kahn, J. and Vu, V. (2008) Factors in random graphs. *Random Struct. Alg.* **33** 1–28.

[14] Kim, J. H. and Vu, V. H. (2000) Concentration of multivariate polynomials and its applications. *Combinatorica* **20** 417–434.

[15] Kohayakawa, Y. (1997) Szemerédi's regularity lemma for sparse graphs. In *Foundations of Computational Mathematics*, Springer, pp. 216–230.

[16] Kohayakawa, Y., Łuczak, T. and Rödl, V. (1997) On K_4-free subgraphs of random graphs. *Combinatorica* **17** 173–213.

[17] Komlós, J. (2000) Tiling Turán theorems. *Combinatorica* **20** 203–218.

[18] Komlós, J., Sárközy, G. and Szemerédi, E. (2001) Proof of the Alon–Yuster conjecture. *Discrete Math.* **235** 255–269.

[19] Kühn, D. and Osthus, D. (2009) Embedding large subgraphs into dense graphs. In *Surveys in Combinatorics 2009*, Vol. 365 of London Mathematical Society Lecture Note Series, Cambridge University Press, pp. 137–167.

[20] Kühn, D. and Osthus, D. (2009) The minimum degree threshold for perfect graph packings. *Combinatorica* **29** 65–107.

[21] Saxton, D. and Thomason, A. (2015) Hypergraph containers. *Inventio. Math.* **201** 925–992.

[22] Shokoufandeh, A. and Zhao, Y. (2003) Proof of a tiling conjecture of Komlós. *Random Struct. Alg.* **23** 180–205.

Cite this article: Nenadov R and Škorić N (2020). On Komlós' tiling theorem in random graphs. *Combinatorics, Probability and Computing* **29**, 113–127. https://doi.org/10.1017/S0963548319000129