Induced subdivisions in $K_{s,s}$-free graphs of large average degree

Daniela Kühn Deryk Osthus

Abstract

We prove that for every graph H and for every s there exists $d = d(H, s)$ such that every graph of average degree at least d contains either a $K_{s,s}$ as a subgraph or an induced subdivision of H.

1 Introduction

A classical theorem of Mader states that for every graph H there exists $d = d(H)$ such that every graph G of average degree at least d contains a subdivision of H. Obviously, the result becomes false if we ask for an *induced* subdivision of H. Here we prove that this stronger assertion holds if G is 'locally sparse' in the sense that it fails to contain some complete bipartite graph $K_{s,s}$:

Theorem 1. *For every graph H and every $s \in \mathbb{N}$ there exists $d = d(H, s)$ such that every graph G of average degree at least d contains either a $K_{s,s}$ as a subgraph or an induced subdivision of H.*

Of course, one cannot replace 'subdivision' by 'subgraph', as for example there exist graphs which have both arbitrarily large average degree and arbitrarily large girth. On the other hand, Kierstead and Penrice [10] proved that if H is a tree then one can indeed find it as an induced subgraph in any $K_{s,s}$-free graph of sufficiently large average degree. They used this result to prove a special case of the conjecture of Gyárfás [7] and Sumner [17] that given a tree T and $s \in \mathbb{N}$, every K_s-free graph of sufficiently large chromatic number contains an induced copy of T. Scott [16] proved that this conjecture becomes true if we only require an induced subdivision of T. In [16] he also proposed a conjecture which is analogous to Theorem 1—replacing 'average degree' by 'chromatic number' and $K_{s,s}$ by K_s. In fact, Theorem 1 was motivated by this conjecture. A simple proof of Theorem 1 for the case when H is a cycle can be found in [12]. Elsewhere we also obtained related results about $K_{s,s}$-free graphs of large average degree: we proved that such graphs contain rather large cliques as (not necessarily induced) subdivisions [13] and minors [14].

We now briefly outline the organization of this paper and the strategy of our proof of Theorem 1. Consider a $K_{s,s}$-free graph G of large average degree. In Section 2 we prepare the ground for the proof by collecting some tools which we will need later on. In particular, it turns out that in order to find an induced subdivision of H in G, it suffices to prove the following

Theorem 2. *For all $k, s \in \mathbb{N}$ there exists $d = d(s,k)$ such that every $K_{s,s}$-free graph G of average degree at least d contains an induced subdivision of some graph H^* where the average degree of H^* is at least k and every edge of H^* is subdivided exactly once.*

We will call such a subdivision an induced *1-subdivision* of H^*. Note that both the set B of branch vertices and the set S of subdividing vertices have to be independent in G. The first step towards finding such a 1-subdivision of H^* is to find a large independent set I in G (Section 3). Ideally, we would like to find another independent set B^* such that the bipartite subgraph between I and B^* has large average degree. In this case, one can find B in the smaller of B^* and I and S in the larger of the two. Unfortunately, we cannot guarantee that such a set B^* always exists. However, in Section 4 we will show that one can come fairly close: we will find sets $I^* \subseteq I$ and B^* such that the bipartite subgraph between I^* and B^* has large average degree and $G[B^*]$ has small chromatic number. In Section 5, which constitutes the core of our proof, we then show how to find our induced 1-subdivision of H^* within $G[I^* \cup B^*]$. In Section 6 we then put everything together to complete the proof of Theorem 2 (and thus of Theorem 1). In the final section we mention some open problems.

Theorem 2 also implies induced analogues of a result of Thomassen on subdivisions and of a result of Häggkvist and Scott on cycles in graphs: Thomassen [18] proved that for every $k, \ell \in \mathbb{N}$ there exists $f = f(k,\ell)$ such that every graph of minimum degree at least f contains a subdivision of some graph H with minimum degree at least k in which every edge is subdivided exactly ℓ times. Combined with Theorem 2 this gives the following analogue for odd integers ℓ:

Corollary 3. *For all $k, s \in \mathbb{N}$ and every odd integer ℓ there exists $g = g(k, \ell, s)$ such that every $K_{s,s}$-free graph of minimum degree at least g contains an induced subdivision of some graph H with minimum degree at least k in which every edge is subdivided exactly ℓ times.*

Häggkvist and Scott [8] proved that every graph of minimum degree at least $300k^2$ contains k cycles of consecutive even lengths. (Verstraëte [19] improved the bound on the minimum degree to a linear one.) Applying this result to the graph H^* provided by Theorem 2 we obtain k induced cycles in G which are twice as long. In particular, we have

Corollary 4. *For all $k, s \in \mathbb{N}$ there exists $g = g(k,s)$ such that every $K_{s,s}$-free graph of minimum degree at least g contains k induced cycles whose lengths form an arithmetic progression.*

2 Notation and tools

In this paper, all logarithms are base two. We write $e(G)$ for the number of edges of a graph G, $d(G) := 2e(G)/|V(G)|$ for its *average degree*, $\delta(G)$ for its minimum degree and $\chi(G)$ for its chromatic number. Given a vertex x of G, we denote by $d(x)$ or $d_G(x)$ the degree of x and by $N(x)$ or $N_G(x)$ the set of neighbours of x. Given graphs G and H we say that G is *H-free* if G does not contain H as a subgraph. A *subdivision* of a graph H is a graph G obtained from H by replacing the edges of H with internally disjoint paths between their endvertices. We view $V(H)$ as a subset of $V(G)$ and call these vertices the *branch vertices* of G. A *1-subdivision* of a graph H is the graph obtained from H by replacing the edges of H with internally disjoint paths of length two.

For disjoint sets $A, B \subseteq V(G)$ we write $e(A, B)$ for the number of A–B edges in G and $(A, B)_G$ for the bipartite subgraph of G whose vertex classes are A, B and whose edges are the A–B edges in G. If we say that a bipartite graph (A', B') is a subgraph of (A, B) then we tacitly assume that $A' \subseteq A$ and $B' \subseteq B$. We shall frequently consider the following class of graphs.

Definition. Given non-negative numbers d, i and $k \leq d/4$, we say that a bipartite graph (A, B) is a *(d, i, k)-graph* if $|A| \geq d^{12i}|B|$ and $d/4 - k \leq d(a) \leq 4d$ for all vertices $a \in A$. (Note that the order of A and B matters here.)

We now list some results which we need later on in the proof of Theorem 1. The following theorem of Mader (for a proof see e.g. [6, Thm. 3.6.1]) implies that Theorem 1 is a consequence of Theorem 2. Indeed, from Theorem 5 it follows that the graph H^* provided by Theorem 2 contains a subdivision of H; and it is easily checked that the corresponding subdivision of H in G is induced.

Theorem 5. *For every $r \in \mathbb{N}$ there exists $d = d(r)$ such that every graph of average degree at least d contains a subdivision of K_r.*

Bollobás and Thomason [5] as well as Komlós and Szemerédi [11] independently showed that the order of magnitude of $d(r)$ is r^2.

We shall frequently use the following simple observations. Proofs are for example included in [6, Prop. 1.2.2 resp. Cor. 5.2.3].

Proposition 6. *Every graph G contains an induced subgraph of average degree at least $d(G)$ and minimum degree at least $d(G)/2$.*

Proposition 7. *Every graph G contains an induced subgraph of minimum degree at least $\chi(G) - 1$.*

Clearly, it suffices to prove Theorem 2 for graphs G which do not have subgraphs of average degree $> d(G)$. So the propositions enable us to assume that $\delta(G) \geq d(G)/2$ and $\chi(G) \leq d(G) + 1$.

We shall also use the following well known upper bound for the average degree of $K_{s,s}$-free graphs (see e.g. [4, p. 74]).

Theorem 8. *The average degree of every $K_{s,s}$-free graph G is at most $c_s|G|^{1-1/s}$ where c_s is some constant depending on s.*

The next lemma is a special case of Chernoff's inequality (see e.g. [3, Thm. A.1.12 and A.1.13]).

Lemma 9. *Let $X_1, \dots, X_n$ be independent 0-1 random variables with $\mathbb{P}(X_i = 1) = p$ for all $i \leq n$, and let $X := \sum_{i=1}^n X_i$. Then $\mathbb{P}(X \geq 2\mathbb{E}X) \leq (4/\mathrm{e})^{-\mathbb{E}X}$ and $\mathbb{P}(X \leq \mathbb{E}X/2) \leq \mathrm{e}^{-\mathbb{E}X/8}$.*

One case which arises in our proof of Theorem 2 is that we first find an induced bipartite subgraph (A, B) of large average degree in G and then find an induced subdivision of H in (A, B). To carry out this second step, it will turn out to be useful if the vertices in A have almost the same degree and $|B|$ is much smaller than $|A|$. The following lemma shows that by replacing (A, B) with an induced subgraph we can always satisfy these two additional conditions. The lemma is a slight extension of [15, Lemma 2.4]. Although the proof is almost the same, we include it here for completeness.

Lemma 10. *Let $r \geq 2^6$, $s \geq 1$ and $d \geq 8r^{12s+1}$. Then every bipartite graph of average degree d contains an induced copy of an $(r, s, 0)$-graph.*

Proof. Clearly, we may assume that our given bipartite graph has no subgraph of average degree $> d$. So by Proposition 6 this graph contains an induced subgraph $G = (A, B)$ such that $\delta(G) \geq d/2$, $d(G) = d$ and $|A| \geq |B|$. Thus at least half of the vertices of A have degree at most $2d$ in G. So, writing A' for the set of all vertices in A of degree at most $2d$, we have $|A'| \geq |A|/2 \geq |B|/2$.

Let us now consider a random subset B_p of B which is obtained by including each vertex of B independently with probability $p := r/d$. For every $a \in A'$ let $X_a := |N_G(a) \cap B_p|$. Then $r/2 \leq \mathbb{E}X_a \leq 2r$. Given B_p, let us call $a \in A'$ *useful* if $r/4 \leq X_a \leq 4r$. Lemma 9 implies that

$$\mathbb{P}(a \text{ is not useful}) \leq \mathbb{P}(X_a \geq 2\mathbb{E}X_a) + \mathbb{P}(X_a \leq \mathbb{E}X_a/2) \leq (4/\mathrm{e})^{-r/2} + \mathrm{e}^{-r/16} \leq \frac{1}{4}.$$

Hence the expected number of vertices in A' which are not useful is at most $|A'|/4$. So Markov's inequality (which states that $\mathbb{P}(X \geq c\,\mathbb{E}X) \leq 1/c$ for every $c \geq 1$) implies that

$$\mathbb{P}(\text{at least half of the vertices in } A' \text{ are not useful}) \leq \frac{1}{2}.$$

Moreover, using Lemma 9 again,

$$\mathbb{P}(|B_p| \geq 2p|B|) = \mathbb{P}(|B_p| \geq 2\mathbb{E}|B_p|) \leq (4/\mathrm{e})^{-p|B|} \leq \frac{1}{4}.$$

So the probability that both $|B_p| \leq 2p|B|$ and that at least half of the vertices in A' are useful is at least $1/2 - 1/4 > 0$. Hence there exists a choice B^* for B_p which has these two properties. Let A^* be the set of useful vertices in A'. Then $r/4 \leq d_{(A^*, B^*)_G}(a) \leq 4r$ for every vertex $a \in A^*$ and

$$|A^*| \geq \frac{|A'|}{2} \geq \frac{|B|}{4} \geq \frac{|B^*|}{8p} = \frac{d|B^*|}{8r} \geq r^{12s}|B^*|.$$

Thus $(A^*, B^*)_G$ is an induced $(r, s, 0)$-subgraph of G. $\square$

3 Independent sets

Clearly, every graph G of maximum degree Δ has an independent set of size at least $|G|/\chi(G) \geq |G|/(\Delta+1)$. Lemma 11 shows that we obtain a small but significant improvement if G is $K_{s,s}$-free. The proof is based on Alon's elegant proof of the result that any triangle-free graph H of maximum degree Δ contains an independent set of size $c|H|\log\Delta/\Delta$ (see e.g. [3], the result itself is due to Ajtai, Komlós and Szemerédi [1]).

Alternatively, we could have applied another result from [1]: for all ε there exists a constant c_0 so that every graph with maximum degree at most Δ which contains at most $|G|\Delta^{2-\varepsilon}$ triangles has an independent set of size at least $c_0|G|\log\Delta/\Delta$. But Theorem 8 implies that in a $K_{s,s}$-free graph G the neighbourhood of any vertex x can span at most $c_s d(x)^{2-1/s} \leq c_s\Delta^{2-1/s}$ edges and thus G contains at most $c_s|G|\Delta^{2-1/s}$ triangles. Although the proof of Lemma 11 given below yields a weaker bound, it is simpler and has the advantage of being self-contained.

Lemma 11. *For every $s \in \mathbb{N}$ there exists $c' = c'(s)$ such that for each $\Delta \geq 9$ every $K_{s,s}$-free graph G of maximum degree at most Δ has an independent set of size at least*

$$f := c'|G|\frac{(\log\Delta)^{1/s}}{\Delta\log\log\Delta}.$$

Proof. Let $n := |G|$. Let I be an independent set chosen uniformly at random from all independent sets of G. For every vertex $x \in G$ define

$$Z_x := \begin{cases} \Delta & \text{if } x \in I; \\ |N(x)\cap I| & \text{otherwise.} \end{cases}$$

Then

$$\sum_{x\in G} Z_x = \sum_{x\in I} Z_x + \sum_{x\notin I} Z_x \leq \Delta|I| + e(I, V(G)\setminus I) \leq 2\Delta|I|.$$

So it suffices to show that $\mathbb{E}(\sum_{x\in G} Z_x) \geq 2\Delta f$. Given any vertex $x \in G$, let $I_x := I \setminus (N(x)\cup\{x\})$. Rather than directly showing that $\mathbb{E}(\sum_{x\in G} Z_x)$ is large, we will show that $\mathbb{E}(Z_x|I_x)$ is large for every vertex x and every I_x.

Let N_x be the set of all neighbours of x which are not adjacent to a vertex in I_x. We will now show that if N_x is large then the average size of an independent subset of N_x is large as well. So suppose first that $|N_x| \geq 2$. Since $G[N_x]$ is $K_{s,s}$-free, it follows from Theorem 8 that every subgraph H of $G[N_x]$ has average degree at most $c_s|H|^{1-1/s} \leq c_s|N_x|^{1-1/s}$. Thus by Proposition 7 we have that $\chi(G[N_x]) \leq c_s|N_x|^{1-1/s} + 1 \leq 2c_s|N_x|^{1-1/s}$. So $G[N_x]$ has an independent set of size at least $|N_x|^{1/s}/(2c_s) =: \alpha$. Hence $G[N_x]$ contains at least $2^\alpha/2$ independent sets of size at least $\alpha/2$. Put $\beta := \alpha/(4\log|N_x|)$. Then the number of independent subsets of N_x of size at most β is at most

$$\binom{|N_x|}{0} + \cdots + \binom{|N_x|}{\lfloor\beta\rfloor} \leq |N_x|^{2\beta} = 2^{2\beta\log|N_x|} = 2^{\alpha/2}.$$

If $|N_x| \geq (4c_s)^s$ then $2^\alpha/2 \geq 2^{\alpha/2}$ and $\alpha/2 \geq 2\beta$; and so in this case the average size ℓ_x of an independent subset of N_x is at least β.

Now note that, writing k_x for the number of independent sets in N_x, for every $|N_x| \geq 0$ we have

$$\mathbb{E}(Z_x|I_x) \geq \frac{\Delta + k_x \ell_x}{1+k_x} \geq \frac{\Delta}{2k_x} + \frac{\ell_x}{2}.$$

Thus, if $|N_x| \geq (\log \Delta)/2$ and if c' is sufficiently small compared with s, then

$$\mathbb{E}(Z_x|I_x) \geq \frac{\ell_x}{2} \geq \frac{\beta}{2} \geq \frac{|N_x|^{1/s}}{16c_s \log |N_x|} \geq \frac{2c'(\log \Delta)^{1/s}}{\log\log \Delta},$$

while if $0 \leq |N_x| \leq (\log \Delta)/2$ then

$$\mathbb{E}(Z_x|I_x) \geq \frac{\Delta}{2 \cdot 2^{|N_x|}} \geq \frac{\Delta}{2 \cdot 2^{(\log \Delta)/2}} = \frac{\sqrt{\Delta}}{2} \geq \frac{2c'(\log \Delta)^{1/s}}{\log\log \Delta}.$$

Hence we have $\mathbb{E}(Z_x) \geq 2\Delta f/n$ and so $\mathbb{E}(\sum_{x \in G} Z_x) = \sum_{x \in G} \mathbb{E}(Z_x) \geq 2\Delta f$, which completes the proof. □

Corollary 12. *For every $s \in \mathbb{N}$ there exists $d_0 = d_0(s)$ such that every $K_{s,s}$-free graph G of average degree $d \geq d_0$ contains an independent set of size at least $|G|(\log d)^{1/(s+1)}/d$.*

Proof. Let G' be the subgraph of G induced by the vertices of degree at most $2d$. Clearly, $|G'| \geq |G|/2$. If d is sufficiently large, then by Lemma 11, G' (and thus G) has an independent set of size at least $|G|(\log d)^{1/(s+1)}/d$. □

4 Finding a 'nearly' induced bipartite subgraph of large average degree

As remarked in the introduction, we would like to find an induced bipartite subgraph of large average degree in our original graph G. The aim of this section is to prove that if G does not contain such a subgraph, we can still come close to it: by Corollary 12 we may assume that G contains a large independent set I. We will use this to find a subgraph (A, B) of large average degree so that $A \subseteq I$ (so A is independent) and B has small chromatic number and is much smaller than A. The following lemma shows how to construct one colour class of B.

Lemma 13. *Let I be an independent set in a graph G such that $d(x) \geq d/2$ for every $x \in I$ and $|I| = 2c|G|/d$ for some $c \geq 2$. Suppose that $\chi(G) \leq 3d$. Then G has one of the following properties.*

(i) G contains an induced bipartite subgraph of average degree at least $(\log c)/24$.

(ii) There are a set $I' \subseteq I$ and an independent set J in $G - I$ such that in G every vertex of I' has exactly one neighbour in J, $|J| \leq |I| \log c/c$ and $|I'| \geq |I|/4(\log c)^2$.

Proof. Put $n := |G|$, $\overline{I} := V(G) \setminus I$ and let Y be the set of all vertices in $\overline{I}$ which have at least $c/2$ neighbours in I. Then $e(I, \overline{I} \setminus Y) \leq c|\overline{I} \setminus Y|/2 \leq cn/2$. On the other hand the degree of every vertex in I is at least $d/2$, and so we have that $e(I, \overline{I}) \geq cn$. Thus $e(I, Y) \geq cn/2$. As $\chi(G) \leq 3d$, there exists an independent set $A \subseteq Y$ such that

$$e(I, A) \geq \frac{e(I, Y)}{3d} \geq \frac{cn}{6d} = \frac{|I|}{12}. \tag{1}$$

Note also that

$$\frac{c}{2} \cdot |A| \leq e(I, A). \tag{2}$$

We may assume that the average degree of $(I, A)_G$ is at most $(\log c)/2$ (otherwise $(I, A)_G$ would be as desired in (i)). Since every vertex in A has at least $c/2 \geq (\log c)/2$ neighbours in I, this implies that $|I| \geq |A|$. Therefore

$$\frac{c}{2} \cdot |A| \leq e(I, A) = \frac{1}{2} \cdot d((I, A)_G)(|I| + |A|) \leq \frac{\log c}{2} \cdot |I|,$$

and hence

$$|A| \leq \frac{|I| \log c}{c}. \tag{3}$$

Using a probabilistic argument, we will show that there exist sets $J \subseteq A$ and $I' \subseteq I$ as desired in (ii). To make this work, we first need to replace I with the set $I_1 \subseteq I$ of all vertices which have at least one and at most $\log c$ neighbours in A. So let us first estimate the size of I_1. Denote by I_2 the set of all vertices in I which have no neighbours in A and put $I_3 := I \setminus (I_1 \cup I_2)$. We will show that we may assume that both $e(I_1, A) \geq e(I, A)/2$ and $|I_1| \geq |I|/\log c$. Suppose to the contrary that $e(I_1, A) \leq e(I, A)/2$. Then $e(I_3, A) \geq e(I, A)/2$ and so (2) implies that $e(I_3, A) \geq c|A|/4$. Thus on average, a vertex in A has at least $c/4$ neighbours in I_3. As every vertex in I_3 has at least $\log c$ neighbours in A, it follows that $(I_3, A)_G$ is as desired in (i). Hence we may assume that $e(I_1, A) \geq e(I, A)/2$. Next suppose that $|I_1| \leq |I|/\log c$. Then

$$e(I_1, A) \geq e(I, A)/2 \overset{(1)}{\geq} |I|/24 \geq |I_1|(\log c)/24$$

and

$$e(I_1, A) \geq e(I, A)/2 \overset{(2)}{\geq} c|A|/4.$$

Thus $(I_1, A)_G$ is as desired in (i). Therefore we may also assume that $|I_1| \geq |I|/\log c$.

Let us now consider a random subset A_p of A which is obtained by including each $a \in A$ independently with probability $p := 1/(2 \log c)$. Call a vertex $x \in I_1$ *useful* if it has exactly one neighbour in A_p. Using the definition of I_1 it follows that for every $x \in I_1$

$$\begin{aligned}\mathbb{P}(x \text{ is useful}) &= |N(x) \cap A| \cdot p \cdot (1-p)^{|N(x) \cap A| - 1} \geq 1 \cdot p \cdot (1-p)^{\lfloor \log c \rfloor} \\ &\geq p(1 - p\lfloor \log c \rfloor) \geq p/2.\end{aligned}$$

(The second inequality can be easily proved by induction.) Hence the expected number of useful vertices in I_1 is at least $p|I_1|/2$. So there exists a choice J for A_p such that at least $p|I_1|/2$ vertices in I_1 are useful. Let I' be the set of these useful vertices. Then

$$|I'| \geq \frac{p|I_1|}{2} = \frac{|I_1|}{4\log c} \geq \frac{|I|}{4(\log c)^2}$$

and

$$|J| \leq |A| \overset{(3)}{\leq} \frac{|I|\log c}{c}.$$

So I' and J are as desired in (ii). □

By repeated applications of Lemma 13 we obtain the following result.

Lemma 14. *Let $c \geq 2^{512}$, $d > 2c$ and let G be a graph of minimum degree at least $d/2$. Suppose that $\chi(G) \leq d+1$ and that G has an independent set I of size $2c|G|/d$. Put $r := \lfloor \log\log c \rfloor$. Then G has one of the following properties.*

(i) G contains an induced bipartite subgraph of average degree at least $(\log c)/48$.

(ii) There are a set $I^* \subseteq I$ and disjoint independent subsets $J_1, \dots, J_r$ of $G - I^*$ such that every vertex of I^* has exactly one neighbour in each J_k, $|I^*| \geq |I|/4^r(\log c)^{2r}$ and $|J_k| \leq 4|I|\log c/c$ for every $k \leq r$.

Proof. The proof follows from r applications of Lemma 13. Indeed, let $I_0 := I$ and suppose inductively that for some $0 \leq \ell < r$ we already have obtained a set $I_\ell \subseteq I$ and disjoint independent sets $J_1, \dots, J_\ell$ in $G - I_\ell$ such that every vertex of I_ℓ has exactly one neighbour in each J_k, $|I_\ell| = \lceil |I|/4^\ell(\log c)^{2\ell} \rceil$ and $|J_k| \leq 4|I|\log c/c$ for every $1 \leq k \leq \ell$. Put $n := |G|$, $G' := G - (J_1 \cup \dots \cup J_\ell)$, $n' := |G'|$ and $d' := d/2$. Thus $d_{G'}(x) \geq d/2 - \ell \geq d/4 = d'/2$ for every $x \in I_\ell$. Moreover, since $|J_k| \leq 4n\log c/c$, we have that $n' \geq n/2$. Let c' be defined by $|I_\ell| = 2c'n'/d'$. Using $|I_\ell| \leq |I|$ it follows that $c' \leq c$. On the other hand

$$\frac{|I|}{4^\ell(\log c)^{2\ell}} \leq |I_\ell| = \frac{2c'n'}{d'} \leq \frac{4c'n}{d},$$

and so

$$c' \geq \frac{c}{2 \cdot 4^\ell(\log c)^{2\ell}} = \frac{c}{2(2\log c)^{2\ell}}.$$

In particular, $c' \geq 2$. Since also $\chi(G_\ell) \leq d+1 \leq 3d'$, we may apply Lemma 13 to the graph G' and the independent set I_ℓ. As

$$\frac{\log c'}{24} \geq \frac{\log c - 1 - \log((2\log c)^{2\ell})}{24} \geq \frac{\log c - 1 - 2r\log(2\log c)}{24} \geq \frac{\log c}{48}$$

we may assume that we have $I_{\ell+1} \subseteq I_\ell$ and $J_{\ell+1}$ satisfying condition (ii) of Lemma 13. Hence

$$|I_{\ell+1}| \geq \frac{|I_\ell|}{4(\log c')^2} \geq \frac{|I_\ell|}{4(\log c)^2} \geq \frac{|I|}{4^{\ell+1}(\log c)^{2(\ell+1)}},$$

and

$$|J_{\ell+1}| \le \frac{|I_\ell|\log c'}{c'} \le \frac{2\cdot 4^\ell\cdot|I_\ell|(\log c)^{2\ell+1}}{c} \le \frac{4|I|\log c}{c}.$$

Making $I_{\ell+1}$ smaller if necessary, we may assume that $|I_{\ell+1}| = \lceil |I|/4^{\ell+1}(\log c)^{2(\ell+1)}\rceil$, which completes the induction step. □

Corollary 15. *For every $s \in \mathbb{N}$ there exists $c(s)$ such that the following holds. Let $c \ge c(s)$, $d > 2c$ and let G be a graph of minimum degree at least $d/2$. Suppose that G has an independent set I of size $2c|G|/d$ and that $\chi(G) \le d+1$. Put $r := \lfloor \log\log c\rfloor$. Then G has one of the following properties.*

(i) G contains an induced bipartite subgraph of average degree at least $(\log c)/48$.

(ii) There are disjoint vertex sets $A, B \subseteq V(G)$ such that A is independent, $\chi(G[B]) \le r$ and $(A,B)_G$ is an $(r,s,0)$-graph.

Proof. Applying Lemma 14 we may assume that G contains independent sets I^* and $J_1,\dots J_r$ satisfying condition (ii) of Lemma 14. Let $A := I^*$ and $B := J_1 \cup \dots \cup J_r$. Clearly, every vertex of A has degree r in the bipartite graph $(A,B)_G$ and $\chi(G[B]) \le r$. Thus it remains to show that $|A| \ge r^{12s}|B|$. But

$$\frac{|A|}{|B|} \ge \frac{c}{4^{r+1}r(\log c)^{2r+1}} \ge r^{12s},$$

if c is sufficiently large. □

5 Finding an induced 1-subdivision of a graph of large average degree

In the previous section we showed that we may assume that our original graph G contains a bipartite subgraph (A,B) of large average degree such that A is independent in G and $G[B]$ has small chromatic number (or is possibly independent as well). In this section we will show that this (A,B) contains a 1-subdivision of some graph H^* where H^* has large average degree and this 1-subdivision is induced in G.

To accomplish this, we first find a 1-subdivision of some graph H' of large average degree in (A,B) (Corollary 17). The branch vertices of this 1-subdivision are vertices in B, its subdivided edges are paths of length two in (A,B) and so the midpoints of the subdivided edges are vertices in A. In Lemma 18 we then show how to find a subgraph H'' of H' for which every midpoint of a subdivided edge is joined in G only to the two endpoints of this edge and to no other branch vertex. As A is independent, it follows that every edge of G which prevents the 1-subdivision of H'' from being induced must join two branch vertices, i.e. two vertices in B. So if B is also independent then this 1-subdivision is induced in G, as desired. The case when B is not independent is more difficult and dealt with in Lemma 20.

Let us now introduce some notation. A path P of length two in a bipartite graph (A, B) is called a *hat* of G if it begins and ends in B. A set $\mathcal{H}$ of hats of (A, B) is *uncrowded* if any two hats in $\mathcal{H}$ join distinct pairs of vertices and have distinct midpoints. (So the sets of subdivided edges of the 1-subdivisions of the graphs H' and H'' described above are both uncrowded; and conversely, an uncrowded set of hats can serve as the set of subdivided edges of a 1-subdivision whose set of branch vertices is B.)

Lemma 16. *Let $r, i \geq 1$ and $0 \leq k \leq r/8$. Let $G = (A, B)$ be an (r, i, k)-graph. Then either G has an uncrowded set of at least $r^{11}|B|/2^8$ hats or there are a vertex $b' \in B$ and an induced copy (A', B') of an $(r, i-1, k+1)$-graph in $G - b'$ such that $\emptyset \neq A' \subseteq N_G(b')$.*

Proof. Let us first suppose that every vertex $b \in B$ satisfies

$$|N^2(b)| \geq d(b)/r^{12(i-1)},$$

where $N^2(b)$ is the set of all vertices with distance two from b. In other words, for each $b \in B$ there is a set $\mathcal{H}_b$ of at least $d(b)/r^{12(i-1)}$ hats in G which have b as one endvertex, but whose other endvertices are distinct. Note that every pair of vertices in B belongs to at most two hats in $\bigcup_{b \in B} \mathcal{H}_b$. Hence there are at least $e(G)/2r^{12(i-1)}$ hats with distinct pairs of endpoints. Since the degree of every vertex $a \in A$ is at most $4r$, at most $(4r)^2$ of these hats have a as their midpoint. Thus G has a uncrowded set of at least

$$\frac{e(G)}{2 \cdot 16 r^{12(i-1)+2}} \geq \frac{(r/4-k)|A|}{2^5 r^{12(i-1)+2}} \geq \frac{(r/4-k)r^{12i}|B|}{2^5 r^{12(i-1)+2}} \geq \frac{r^{11}|B|}{2^8}$$

hats, as required.

So we may assume that there is a vertex $b' \in B$ with

$$|N^2(b')| < d(b')/r^{12(i-1)}.$$

Let $A' := N(b')$ and $B' := N^2(b')$. Then $(A', B')_G$ has the required properties. $\square$

The proof of the preceding lemma shows that in the case where we failed to find a large set of uncrowded hats (i.e. a 1-subdivision of some graph of large average degree), there must be a vertex b' so that the set of vertices with distance two from b' is much smaller that the neighbourhood of b'. However, if this happens we can reapply the lemma to the bipartite graph induced by these sets. In case of renewed failure, we can iterate the process—but if we encounter i successive failures, then this means that G contains contains a $K_{i,i}$:

Corollary 17. *Let $s \in \mathbb{N}$ and let $r \geq 8s$. Let $G = (A, B)$ be a $K_{s,s}$-free $(r, s, 0)$-graph. Then there exists $0 \leq i \leq s$ such that G contains an induced copy (A', B') of an $(r, s-i, i)$ graph which has an uncrowded set of at least $r^{11}|B'|/2^8$ hats.*

Proof. Applying Lemma 16 repeatedly, assume that there are sequences $(A, B) = (A_0, B_0) \supseteq (A_1, B_1) \supseteq \cdots \supseteq (A_s, B_s)$ of induced subgraphs of G and $b_1, b_2, \ldots, b_s$ of distinct vertices in B such that, for each $0 < i \leq s$, (A_i, B_i) is an $(r, s-i, i)$-graph and $\emptyset \neq A_i \subseteq N_G(b_i)$. Note that every vertex in A_s has degree at least $r/4 - s \geq r/8$ and so

$$s \leq \frac{r}{8} \leq |B_s| = r^{12(s-s)}|B_s| \leq |A_s|.$$

Thus together with any s vertices from A_s the vertices $b_1, \ldots, b_s$ induce a $K_{s,s}$ in G, a contradiction. $\square$

We say that an uncrowded set $\mathcal{H}$ of hats of a bipartite graph (A, B) is *induced* if $\bigcup \mathcal{H}$ is an induced subgraph of (A, B), i.e. if every midpoint of a hat in $\mathcal{H}$ has degree two in (A, B).

Lemma 18. *Let $r \geq 1$ and let $G = (A, B)$ be a bipartite graph with $d(a) \leq 4r$ for every vertex $a \in A$. Suppose that G has an uncrowded set $\mathcal{H}$ of at least $r^{11}|B|/2^8$ hats. Then there is an induced subgraph $G' = (A', B')$ of G which has an induced uncrowded set $\mathcal{H}'$ of at least $r^9|B'|/2^{15}$ hats.*

Proof. We may assume that A consists only of midpoints of hats in $\mathcal{H}$. Since $\mathcal{H}$ is uncrowded, every vertex $a \in A$ is the midpoint of exactly one hat in $\mathcal{H}$, and we say that a *owns* the endvertices of these hat. So every vertex in A owns exactly two vertices in B and

$$|A| = |\mathcal{H}| \geq \frac{r^{11}|B|}{2^8}.$$

Let us consider a random subset B_p of B which is obtained by including each vertex of B independently with probability $p := 1/(8r)$. Given B_p, let us call a vertex $a \in A$ *useful* if $N(a) \cap B_p$ consists precisely of the two vertices owned by a. Thus

$$\mathbb{P}(a \text{ is useful}) = p^2(1-p)^{d(a)-2} \geq p^2(1-p)^{\lfloor 4r \rfloor} \geq p^2(1 - \lfloor 4r \rfloor p) \geq p^2/2,$$

and so the expected number of useful vertices is at least $p^2|A|/2$. Hence there exists a choice B' for B_p such that at least $p^2|A|/2$ vertices in A are useful. Let A' denote the set of these vertices, and let $\mathcal{H}'$ be the set consisting of all hats in $\mathcal{H}$ whose midpoints lie in A'. Then

$$|\mathcal{H}| = |A'| \geq \frac{|A|}{2^7 r^2} \geq \frac{r^9|B|}{2^{15}} \geq \frac{r^9|B'|}{2^{15}},$$

and so $(A', B')_G$ and $\mathcal{H}'$ have the required properties. $\square$

Corollary 19. *Let $s \in \mathbb{N}$ and $r \geq 8s$. Let $G = (A, B)$ be an $(r, s, 0)$ graph. Then either G contains a $K_{s,s}$ or an induced 1-subdivision of some graph H with $d(H) \geq r^9/2^{14}$.*

Proof. We may apply Corollary 17 and Lemma 18 to obtain an induced bipartite graph $G' = (A', B') \subseteq G$ and a set $\mathcal{H}'$ of hats as in Lemma 18. Let H be the graph whose vertex set is B' and in which $b, b' \in B'$ are joined by an edge if there is a hat in $\mathcal{H}'$ whose endvertices are b and b'. So every edge of H corresponds to a hat in $\mathcal{H}'$. As $\mathcal{H}'$ is induced, the 1-subdivision of H is induced in G' (and thus in G). Moreover $e(H) = |\mathcal{H}'| \geq r^9|B'|/2^{15}$, as desired. $\square$

Lemma 20. *Let $r \geq 2^{25}$. Let A, B be a vertex partition of a graph G such that A is independent, $\chi(G[B]) \leq r$ and $d(G') \leq r^3$ for every $G' \subseteq G[B]$. Suppose that $(A, B)_G$ has an induced uncrowded set $\mathcal{H}$ of at least $r^9|B|/2^{15}$ hats. Then G contains an induced 1-subdivision of some graph H with $d(H) \geq r$.*

Proof. Let H_0 be the graph whose vertex set is B and in which $b, b' \in B$ are joined by an edge if they are the endpoints of a hat in $\mathcal{H}$. Hence G contains a 1-subdivision of H_0. Note that $e(H_0) = |\mathcal{H}|$ and so $d(H_0) \geq r^9/2^{14}$. Let H_1 be a subgraph of H_0 with

$$\delta(H_1) \geq \frac{r^9}{2^{15}}, \tag{4}$$

and put $B_1 := V(H_1)$ (where B_1 is thought of as a subset of B). Let G^* be the 1-subdivision of H_1 contained in G. Note that every edge which prevents G^* from being induced must join two branch vertices of G^*, i.e. vertices in B_1. Using a probabilistic argument, we will show that H_1 contains a subgraph H_2 of average degree at least r whose 1-subdivision in G is induced. In other words, we are given two graphs H_1 and $F := G[B_1]$ on the same vertex set such that H_1 has large average degree while every subgraph of F has small average degree. The desired subgraph H_2 of H_1 must avoid all edges of F.

Let B_1' denote the set of all vertices $b \in B_1$ with $d_F(b) \leq 2r^3$. Then

$$2r^3|B_1 \setminus B_1'| \leq 2e(F) = d(F)|F| \leq r^3|B_1|$$

and thus

$$|B_1'| \geq \frac{|B_1|}{2}. \tag{5}$$

Consider a random subset B_p of B_1 which is obtained by including each vertex of B_1 independently with probability $p = 1/(4r^3)$. Given B_p, call a vertex $b \in B_1'$ *useful* if

(a) $b \in B_p$,

(b) $N_F(b) \cap B_p = \emptyset$,

(c) $|(N_{H_1}(b) \setminus N_F(b)) \cap B_p| \geq pr^9/2^{17}$.

Thus every useful vertex is isolated in $G[B_p]$ and in the graph H_1 it has many neighbours which are contained in B_p. The aim now is to show that with non-zero probability the set I_0 of useful vertices is large. As the chromatic number of $G[B_p]$ is small compared to $|N_{H_1}(b) \cap B_p|$ for any useful vertex b, there will be an independent set in $B_p \setminus I_0$ which together with I_0 induces a subgraph H_2

of H_1 with large average degree. Observe that the 1-subdivision of H_2 in G will be induced.

To prove that with non-zero probability B_1' contains many useful vertices, first note that for every $b \in B_1'$ the random variable $X := |(N_{H_1}(b) \setminus N_F(b)) \cap B_p|$ is binomially distributed with

$$\mathbb{E}X = p|N_{H_1}(b) \setminus N_F(b)| \geq p|\delta(H_1) - d_F(b)| \overset{(4)}{\geq} \frac{pr^9}{2^{16}} \geq 8.$$

So Lemma 9 implies that

$$\mathbb{P}(X \leq \frac{pr^9}{2^{17}}) \leq \mathbb{P}(X \leq \frac{\mathbb{E}X}{2}) \leq \mathrm{e}^{-\mathbb{E}X/8} \leq \frac{1}{2}.$$

Moreover, note that the events (a), (b) and (c) are mutually independent. Thus for every vertex $b \in B_1'$ we have that

$$\mathbb{P}(b \text{ is useful}) \geq p \cdot (1-p)^{d_F(b)} \cdot \frac{1}{2} \geq p \cdot (1-p)^{\lfloor 2r^3 \rfloor} \cdot \frac{1}{2} \geq \frac{p(1 - \lfloor 2r^3 \rfloor p)}{2} \geq \frac{p}{4}.$$

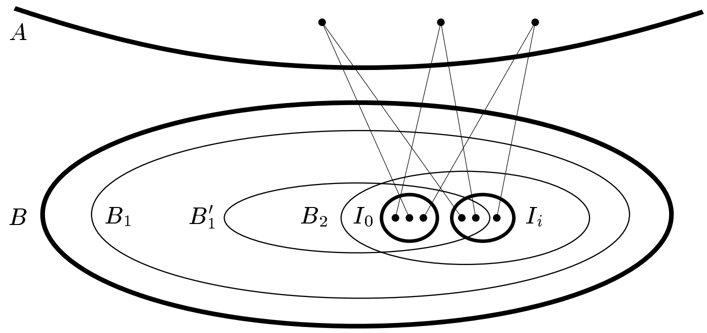

Figure 1: Finding an independent set of vertices in F which induces many hats

Hence by (5) the expected number of useful vertices is at least $p|B_1'|/4 \geq p|B_1|/8$. So there exists a choice B_2 for B_p such that at least $p|B_1|/8$ vertices in B_1' are useful. Let I_0 denote the set of these vertices. Every useful vertex is contained in B_2 and has at least $pr^9/2^{17}$ neighbours in H_1 which are contained in B_2. Thus there are at least

$$\frac{1}{2} \cdot \frac{pr^9}{2^{17}} \cdot \frac{p|B_1|}{8} = \frac{r^3|B_1|}{2^{25}}$$

edges of H_1 which emanate from vertices contained in I_0. Since $\chi(G[B]) \leq r$, we may partition $G[B_2 \setminus I_0]$ into r independent sets, $I_1, \dots, I_r$ say. Then there exists $0 \leq i \leq r$ such that at least a $1/(r+1)$th of the edges of H_1 emanating from I_0 ends in I_i (see Fig. 1). But then the subgraph H_2 of H_1 induced by $I_0 \cup I_i$ has at least

$$\frac{1}{r+1} \cdot \frac{r^3|B_1|}{2^{25}} \geq \frac{r|B_1|}{2}$$

edges and so it has average degree at least r. Moreover, since in F both I_0 and I_i are independent and no vertex in $B_2 \supseteq I_i$ is joined to a vertex in I_0, it follows that $I_0 \cup I_i$ is independent in G. As mentioned above, this implies that the 1-subdivision of H_2 is induced in G. $\square$

By successively applying Corollary 17 and Lemmas 18 and 20 we obtain the following result.

Corollary 21. *Let $s \in \mathbb{N}$ and $r \geq \max\{8s, 2^{25}\}$. Let G be a $K_{s,s}$-free graph and let $A, B \subseteq V(G)$ be disjoint sets of vertices such that A is independent, $\chi(G[B]) \leq r$, $d(G') \leq r^3$ for every $G' \subseteq G[B]$ and so that $(A, B)_G$ is an $(r, s, 0)$ graph. Then G contains an induced 1-subdivision of some graph H with $d(H) \geq r$.* $\square$

6 Proof of Theorem 2

We can now put everything together.

Proof of Theorem 2. Suppose that G is a $K_{s,s}$-free graph with $d(G) = d \geq d_0$ where d_0 is sufficiently large compared to k and s. Put $n := |G|$. Clearly, we may assume that G has no subgraph of average degree $> d$. So Propositions 6 and 7 enable us to assume that $\delta(G) \geq d/2$ and $\chi(G) \leq d + 1$. Also Lemma 10 and Corollary 19 imply that Theorem 2 holds if G contains an induced bipartite subgraph of large average degree—we will make use of this fact twice in what follows.

Turning to the proof itself, we first apply Corollary 12 to G, which gives us an independent set I of size $2cn/d$ where $c \geq (\log d)^{1/(s+1)}/2$. We then apply Corollary 15 to obtain (without loss of generality) disjoint sets $A, B \subseteq V(G)$ as in condition (ii) of the corollary. In other words, A is independent, $\chi(G[B]) \leq r$ and $(A, B)_G$ is an $(r, s, 0)$-graph, where $r = \lfloor \log \log c \rfloor$. Now if $G[B]$ has an (induced) subgraph G' whose average degree is at least r^3 then, as $\chi(G') \leq r$, there must be two disjoint independent sets B_1, B_2 of G' such that

$$e((B_1, B_2)_{G'}) \geq \frac{e(G')}{\binom{r}{2}} \geq \frac{d(G')|G'|}{r^2} \geq r|G'| \geq r(|B_1| + |B_2|).$$

Hence $(B_1, B_2)_G$ is an induced bipartite subgraph of average degree at least $2r$. So we may assume that $d(G') \leq r^3$ for every $G' \subseteq G[B]$. But then Corollary 21 implies that G contains an induced 1-subdivision of some graph H^* which has average degree at least k, as desired. $\square$

7 Open problems

An obvious question is that of the growth of $d(s, k)$ in Theorem 2. The bounds which follow from our proof are quite large: k is about the 3-fold logarithm of d even for the case $s = 2$. Also, we are not aware of any nontrivial lower bound on d.

Our proof of Theorem 2 becomes easier if G contains an induced bipartite subgraph of large average degree. This raises the question whether there exists $d(s,k)$ such that every $K_{s,s}$-free graph of average degree at least $d(s,k)$ contains an induced bipartite subgraph with average degree at least k. The following result implies that much more is true for regular graphs: using a theorem of Johansson [9], Alon, Krivelevich and Sudakov [2, Corollary 2.4] proved that every $K_{s,s}$-free graph G with maximum degree Δ has chromatic number at most $c\Delta/\log\Delta$ for some constant c depending on s (and thus if G is regular, the largest colour class together with another one induce a bipartite graph of average degree at least $(\log\Delta)/c$). Of course the result of Alon, Krivelevich and Sudakov does not hold if we replace maximum degree by average degree: just consider a $K_{s,s}$-free graph G whose chromatic number is large and add sufficiently many isolated vertices to G.

References

[1] M. Ajtai, J. Komlós and E. Szemerédi, A dense infinite Sidon sequence, *Europ. J. Comb.* **2** (1981), 1–11.

[2] N. Alon, M. Krivelevich and B. Sudakov, Coloring graphs with sparse neighbourhoods, *J. Combin. Theory B* **77** (1999), 73–82.

[3] N. Alon and J. Spencer, *The Probabilistic Method* (2nd edition), Wiley-Interscience 2000.

[4] B. Bollobás, *Graph Theory*, Graduate Texts in Mathematics 63, Springer-Verlag 1979.

[5] B. Bollobás and A. Thomason, Proof of a conjecture of Mader, Erdős and Hajnal on topological complete subgraphs, *Europ. J. Comb.* **19** (1998), 883–887.

[6] R. Diestel, *Graph Theory*, Graduate Texts in Mathematics 173, Springer-Verlag 1997.

[7] A. Gyárfás, On Ramsey covering-numbers, *Coll. Math. Soc. János Bolyai*, in: *Infinite and Finite Sets*, North Holland/American Elsevier, New York 1975.

[8] R. Häggkvist and A. Scott, Arithmetic progressions of cycles, preprint.

[9] A.R. Johansson, Asymptotic choice number for triangle free graphs, DIMACS technical report 91-5.

[10] H.A. Kierstead and S.G. Penrice, Radius two trees specify χ-bounded classes, *J. Graph Theory* **18** (1994), 119–129.

[11] J. Komlós and E. Szemerédi, Topological cliques in graphs II, *Combin. Probab. Comput.* **5** (1996), 79–90.

[12] D. Kühn, *Cycles, Minors and Trees*, PhD thesis, Universität Hamburg 2001.

[13] D. Kühn and D. Osthus, Large topological cliques in graphs without a 4-cycle, *Combin. Probab. Comput.*, to appear.

[14] D. Kühn and D. Osthus, Complete minors in $K_{s,s}$-free graphs, submitted.

[15] L. Pyber, V. Rödl and E. Szemerédi, Dense graphs without 3-regular subgraphs, *J. Combin. Theory B* **63** (1995), 41–54.

[16] A. Scott, Induced trees in graphs of large chromatic number, *J. Graph Theory* **24** (1997), 297–311.

[17] D.P. Sumner, Subtrees of a graph and chromatic number, in: *The Theory and Applications of Graphs*, G. Chartrand (ed.), John Wiley, New York 1981.

[18] C. Thomassen, Subdivisions of graphs with large minimum degree, *J. Graph Theory* **8** (1984), 23–28.

[19] J. Verstraëte, On arithmetic progressions of cycle lengths in graphs, *Combin. Probab. Comput.* **9** (2000), 369–373.

Daniela Kühn
Mathematisches Seminar
Universität Hamburg
Bundesstraße 55
20146 Hamburg
Germany
E-mail address: kuehn@math.uni-hamburg.de

Deryk Osthus
Institut für Informatik
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin
Germany
E-mail address: osthus@informatik.hu-berlin.de