Dense induced bipartite subgraphs in triangle-free graphs

Matthew Kwan [*] Shoham Letzter[†] Benny Sudakov[‡] Tuan Tran[§]

Abstract

The problem of finding dense induced bipartite subgraphs in H-free graphs has a long history, and was posed 30 years ago by Erdős, Faudree, Pach and Spencer. In this paper, we obtain several results in this direction. First we prove that any H-free graph with minimum degree at least d contains an induced bipartite subgraph of minimum degree at least $c_H \log d/\log\log d$, thus nearly confirming one and proving another conjecture of Esperet, Kang and Thomassé. Complementing this result, we further obtain optimal bounds for this problem in the case of dense triangle-free graphs, and we also answer a question of Erdős, Janson, Łuczak and Spencer.

1 Introduction

The Max Cut problem seeks to find the largest bipartite subgraph of a graph G. This problem has been studied extensively both from the algorithmic perspective in computer science and optimisation, and from the extremal perspective in combinatorics. Let $b(G)$ denote the size of the largest bipartite subgraph of a graph G with n vertices and e edges. The extremal aspect of the Max Cut problem asks for lower bounds on $b(G)$ in terms of n and e. A classical result of Erdős from 1965 (see [9]) states that $b(G) \geq e/2$, and the complete graph on n vertices shows that the constant $1/2$ in this bound is asymptotically tight. As such, much of the focus has been on determining the error term $b(G)-e/2$. For instance, Erdwards [11] proved $b(G)-e/2 \geq (\sqrt{8e+1}-1)/8$, and noted that equality holds for complete graphs. We refer the reader to [2, 6, 8] and the references therein for further results in this direction. Following these works, there has been a lot of research concerning maximum bipartite subgraphs of various restricted classes of graphs. One class which has drawn most of the attention is H-free graphs, with work dating back to Erdős and Lovász (see [10]). Despite great effort, there are only a few graphs H for which we know (asymptotically) optimal lower bounds on $b(G) - e/2$, as G ranges over all H-free graph on e edges (see [2, 3, 5]).

In the dense case (i.e. $e = \Omega(n^2)$), there is a longstanding conjecture of Erdős [10] which posits that given any n-vertex triangle-free graph G, one can delete at most $n^2/25$ edges to make it bipartite (see [12, 13] for partial results and [28] for the analogous question in the H-free setting). In order to make progress on Erdős's problem, Erdős, Faudree, Pach and Spencer [12] initiated the study of the size of the largest *induced* bipartite subgraphs in triangle-free graphs, and proved some results in

[*]Department of Mathematics, Stanford University, CA 94305. Email: `mattkwan@stanford.edu`. Research supported in part by SNSF project 178493.

[†]ETH Institute for Theoretical Studies, ETH Zurich, Switzerland. Email: `shoham.letzter@math.ethz.ch`. Research supported by Dr. Max Rössler, the Walter Haefner Foundation and the ETH Zurich Foundation.

[‡]Department of Mathematics, ETH Zurich, Switzerland. Email: `benjamin.sudakov@math.ethz.ch`. Research supported in part by SNSF grant 200021-175573.

[§]Department of Mathematics, ETH Zurich, Switzerland. Email: `manh.tran@math.ethz.ch`. Research supported by the Humboldt Research Foundation.

this direction. Similar results were also the main subject of a later paper by Erdős, Janson, Łuczak and Spencer [14].

In addition to maximising the number of edges, it is also very natural to study induced bipartite subgraphs of large minimum degree. This more local approach was taken recently by Esperet, Kang and Thomassé in [16], where they made a number of intriguing conjectures.

Conjecture 1.1.

(i) *Any graph with minimum degree at least d contains either a complete subgraph on $\omega_{d\to\infty}(1)$ vertices or an induced bipartite subgraph with minimum degree $\omega_{d\to\infty}(1)$.*

(ii) *There is a constant $c > 0$ such that any triangle-free graph with minimum degree at least d contains an induced bipartite subgraph of minimum degree at least $c \log d$.*

We remark that Esperet, Kang and Thomassé were motivated by the concept of *separation choosability*; see [16] for more details.

Part (ii) is a quantitative version of a special case of Part (i). Esperet, Kang and Thomassé [16, Theorem 3.8] showed that the second part, if true, is sharp up to a constant factor, by considering an appropriate binomial random graph. On the other hand, they mentioned that proving the existence of an induced bipartite subgraph with minimum degree at least 3 seems difficult even in the case of graphs with large minimum degree and large girth (see [16, Conjecture 1.5]). As our main result, we prove a slight weakening of Part (ii) of Conjecture 1.1 and settle Part (i) in a strong form.

Theorem 1.2. *For every graph H, there is a constant $c_H > 0$ such that every H-free graph with minimum degree at least d contains an induced bipartite subgraph of minimum degree at least $c_H \log d/ \log\log d$.*

We remark that this result is essentially tight: if H is a graph that is not a forest (note that the statement is vacuous when H is a forest, because the minimum degree condition forces the appearance of any forest on d vertices), one can show using random graphs that there is an H-free graph with minimum degree d which does not contain an induced bipartite subgraph with minimum degree at least $c_H \log d$ for some constant c_H and for every large enough d. We show how to obtain such a construction for triangle-free graphs (see Section 5.1), but it is not hard to obtain a construction in general. We remark also that for triangle-free graphs which are reasonably dense (i.e. $d = n^{\Omega(1)}$), one can get rid of the $\log\log d$ factor in Theorem 1.2 (see Section 3.3 for details). We first prove Theorem 1.2 for triangle-free graphs in Section 3.1. We then prove a reduction to the triangle-free case in Section 4. Although we make no effort to optimise the constant c_H, we remark that our proof gives a bound of the form $c_H \geq \varepsilon^{|H|}$ for some constant $\varepsilon > 0$.

Next, we turn our attention to the more refined problem of finding induced bipartite subgraphs with large minimum degree in terms of both the number of vertices and the minimum degree of a host triangle-free graph. Let $g(n, d)$ be the maximum g such that every triangle-free graph with n vertices and minimum degree at least d contains an induced bipartite subgraph with minimum degree at least g. We observe an interesting 'phase transition' behaviour of this function.

Theorem 1.3. *With $g(n, d)$ as defined above,*

(i) *for $d \leq \sqrt{n}$ we have $\Omega(\log d/ \log\log d) \leq g(n, d) \leq O(\log d)$;*

(ii) *for $\sqrt{n} < d \leq n/2$ we have $d^2/(2n) \leq g(n, d) \leq O((d^2 \log d)/n)$.*

Furthermore,

(iii) *for* $n^{\Omega(1)} \le d \le \sqrt{n}$ *we have* $g(n,d) = \Theta(\log d)$*;*

(iv) *for* $n^{2/3} < d \le n/2$ *we have* $g(n,d) = \Theta(d^2/n)$*.*

Part (ii) of Theorem 1.3 was also proved independently by Cames van Batenburg, de Joannis de Verclos, Kang and Pirot [7, Theorem 1.2], though our proofs of both the lower and upper bounds are different to theirs. We prove all the lower bounds for Theorem 1.3 in Section 3, and we prove the upper bounds in Section 5.

Finally, let us return to the work of Erdős, Faudree, Pach and Spencer [12], and make some remarks. These authors defined $f(n,m)$ to be the maximum f such that every triangle-free graph with n vertices and at least m edges contains an induced bipartite subgraph with at least f edges, and they studied the approximate behaviour of $f(n,m)$ as n and m vary. They showed $\Omega(m^{1/3}) \le f(n,m) \le O(m^{1/3}\log^2 m)$ for $m \le n^{3/2}$, and $\Omega(m^3 n^{-4}) \le f(n,m) \le O(m^3 n^{-4}\log^2(n^2/m))$ if $m \ge n^{3/2}$. By leveraging some results concerning discrepancy and chromatic number of triangle-free graphs, we can find the correct order of magnitude of $f(n,m)$ for all n, m.

Proposition 1.4. *With* $f(n,m)$ *as defined above,*

(i) *for* $m \le n^{3/2}\sqrt{\log n}$ *we have* $f(n,m) = \Theta\Big(m^{1/3}\log^{4/3} m\Big)$*;*

(ii) *for* $m \ge n^{3/2}\sqrt{\log n}$ *we have* $f(n,m) = \Theta(m^3/n^4)$*.*

Proposition 1.4 resolves a problem raised in a paper of Erdős, Janson, Łuczak and Spencer [14]. We present its short proof in Section 2.

Notation. We use standard asymptotic notation throughout the paper. For functions $f = f(n)$ and $g = g(n)$ we write $f = O(g)$ to mean there is a constant c such that $|f| \le c|g|$, we write $f = \Omega(g)$ to mean there is a constant $c > 0$ such that $f \ge c|g|$ for sufficiently large n, we write $f = \Theta(g)$ to mean that $f = O(g)$ and $f = \Omega(g)$, and we write $f = o(g)$ or $g = \omega(f)$ to mean that $f/g \to 0$ as $n \to \infty$. All asymptotics are as $n \to \infty$ unless specified otherwise (specifically, notation of the form $o_k(1)$ indicates that asymptotics are as $k \to \infty$). Also, we write 'w.h.p.' (standing for 'with high probability') to indicate that an event holds with probability $1 - o(1)$.

2 Induced bipartite subgraphs with many edges

In this section we will prove Proposition 1.4. The lower bound in (ii) was already proved in [12, 14]. For the lower bound in (i), we recall the following easy observation, which appears in the proof of [12, Theorem 4].

Lemma 2.1. *Any graph* G *has an induced bipartite subgraph with at least* $e(G)/\binom{\chi(G)}{2}$ *edges.*

The desired lower bound in (i) is then an immediate consequence of the following bound on the chromatic number of triangle-free graphs, which seems to have appeared independently in [18, 26, 27].

Theorem 2.2. *Let* G *be a triangle-free graph with* m *edges. Then* $\chi(G) = O\Big(\frac{m^{1/3}}{\log^{2/3} m}\Big)$.

It remains to prove the upper bounds in (i) and (ii). For both of these we will take advantage of the following theorem proved by Guo and Warnke [19, Theorem 4] resulting from analysis of a certain random triangle-free process.

Theorem 2.3. *There exist $n_0, c, \beta > 0$ such that for all $n \geq n_0$, there is an n-vertex triangle-free graph G_n with the following property. For any pair of disjoint vertex-sets $A, B \subseteq V(G_n)$ with $|A| = |B| = \lceil c\sqrt{n \log n} \rceil$ we have*

$$\beta\sqrt{\frac{\log n}{n}}\,|A||B| \leq e(G_n[A,B]) \leq 2\beta\sqrt{\frac{\log n}{n}}\,|A||B|. \tag{1}$$

The reason Theorem 2.3 is useful for us is as follows.

Lemma 2.4. *Consider a graph G_n as in Theorem 2.3. Then*

(i) $e(G_n) = \Theta\big(n^{3/2}\sqrt{\log n}\big)$*, and;*

(ii) *every induced bipartite subgraph in G_n has $O(\sqrt{n \log n})$ vertices and $O\Big(\sqrt{n}\log^{3/2} n\Big)$ edges.*

Proof. For (i), note that in fact we have

$$\beta\sqrt{(\log n)/n}\binom{n}{2} \leq e(G_n) \leq 2\beta\sqrt{(\log n)/n}\binom{n}{2}.$$

Indeed, if $e(G_n)$ fell outside this range, then the average value of $e(G_n[A,B])$, over all disjoint pairs (A,B) of vertex subsets of size $\lceil c\sqrt{n\log n}\rceil$, would fall outside the range in (1).

For (ii), first note that every set of $2\lceil c\sqrt{n\log n}\rceil$ vertices contains at least one edge, because it can be partitioned into two sets A, B, which by (1) must have an edge between them. That is to say, $\alpha(G_n) = O\big(\sqrt{n\log n}\big)$. But for any induced bipartite subgraph, the two parts of that subgraph are each independent sets, and therefore such a subgraph consists of at most $2\alpha(G_n) = O\big(\sqrt{n\log n}\big)$ vertices. Then, the number of edges between the two parts is

$$O\Big(\sqrt{\log n/n}\big(\sqrt{n\log n}\big)^2\Big) = O\Big(\sqrt{n}\log^{3/2} n\Big).$$

(We can arbitrarily enlarge A, B if necessary, in order to use (1)). ☐

We deduce the following corollary by taking blowups of the graphs in Lemma 2.4.

Corollary 2.5. *For any $k \in \mathbb{N}$, there is an n-vertex triangle-free graph $G_n^{(k)}$ with the following properties.*

(i) $e\Big(G_n^{(k)}\Big) = \Theta\Big(n^{3/2}\sqrt{k\log(n/k)}\Big)$*, and;*

(ii) *every induced bipartite subgraph in $G_n^{(k)}$ has $O\Big(k^{3/2}\sqrt{n}\log^{3/2}(n/k)\Big)$ edges.*

Proof. We note that the corollary holds trivially when $k = \Omega(n)$ (take, say, $G_n^{(k)}$ to be the complete bipartite graph $K_{n/2,n/2}$). Thus, we may assume that $n/k \geq n_0$, where n_0 is as in Theorem 2.3.

Let $G_{\lceil n/k\rceil}$ be as in Lemma 2.4, and define the n-vertex graph $G_n^{(k)}$ as follows. The vertex set of $G_n^{(k)}$ consists of the disjoint union of $\lceil n/k\rceil$ sets S_v each indexed by a vertex v of $G_{\lceil n/k\rceil}$ and having sizes as equal as possible (so each S_v has size $\Theta(k)$). For every edge $uv \in E\big(G_{\lceil n/k\rceil}\big)$, we put a complete bipartite graph between each vertex of S_u and each vertex of S_v. This graph $G_n^{(k)}$ is called a *blowup* of $G_{\lceil n/k\rceil}$. Note that $G_n^{(k)}$ is triangle-free, and that $e\Big(G_n^{(k)}\Big)$ exceeds $e\big(G_{\lceil n/k\rceil}\big) = \Theta\Big((n/k)^{3/2}\sqrt{\log(n/k)}\Big)$ by a factor of $\Theta(k^2)$, implying (i).

For (ii), note that an induced bipartite subgraph of $G_n^{(k)}$ can only be obtained by 'blowing up' an induced bipartite subgraph of $G_{\lceil n/k\rceil}$. To be more precise, for any induced bipartite subgraph H in $G_n^{(k)}$, consider the set W of all $v \in V\big(G_{\lceil n/k\rceil}\big)$ such that a vertex in S_v appears in H. Then W induces a bipartite subgraph of $G_{\lceil n/k\rceil}$. It follows that every induced bipartite subgraph in $G_n^{(k)}$ has at most $O\Big(k\sqrt{(n/k)\log(n/k)}\Big) = O\Big(\sqrt{kn\log(n/k)}\Big)$ vertices, and at most $O\Big(k^2\sqrt{n/k}\log^{3/2}(n/k)\Big) = O\Big(k^{3/2}\sqrt{n}\log^{3/2}(n/k)\Big)$ edges. □

Now we are ready to prove the upper bounds in Proposition 1.4. For every n, let G_n be as in Lemma 2.4, so $e(G_n) = \Theta\big(n^{3/2}\sqrt{\log n}\big)$. We will show that if $m \le e(G_n)$ then $f(n,m) = \Theta\Big(m^{1/3}\log^{4/3} m\Big)$, and if $m > e(G_n)$ then $f(n,m) = \Theta\big(m^3/n^4\big)$.

First, suppose that $m \le e(G_n)$. Let n' be the least integer such that $m \le e(G_{n'})$. It follows that $n' \le n$ and $m = \Theta(e(G_{n'})) = \Theta\Big((n')^{3/2}\sqrt{\log n'}\Big)$, so $n' = \Theta\Big(m^{2/3}/\log^{1/3} m\Big)$. If necessary, add some isolated vertices to $G_{n'}$ to obtain a graph G with exactly n vertices and at least m edges. Now, consider any induced bipartite subgraph in G, and discard its isolated vertices. By Lemma 2.4, the number of edges in this bipartite subgraph is at most

$$O\Big(\sqrt{n'}\log^{3/2} n'\Big) = O\Big(m^{1/3}\log^{4/3} m\Big),$$

as desired.

Next, suppose that $m > e(G_n)$. Let k be the maximum integer such that $m \ge e\Big(G_n^{(k)}\Big)$. Then $m = \Theta\Big(e\Big(G_n^{(k)}\Big)\Big) = \Theta\Big(n^{3/2}\sqrt{k\log(n/k)}\Big)$, so $k\log(n/k) = \Theta\big(m^2/n^3\big)\big)$. By Corollary 2.5, the number of edges in every induced bipartite subgraph in $G_n^{(k)}$ is at most

$$O\Big(k^{3/2}\sqrt{n}\log^{3/2}(n/k)\Big) = O\big(m^3/n^4\big),$$

as desired. This completes the proof of Proposition 1.4.

3 Lower bounds on minimum degree

In this section we prove the lower bounds in Theorem 1.3. The lower bound for Theorem 1.3 (i) (which also proves Theorem 1.2 for triangle-free graphs) appears in Section 3.1, the lower bound for (ii) and (iv) appears in Section 3.2, and the lower bound for (iii) appears in Section 3.3.

3.1 The sparse regime

Here we prove Theorem 1.2 for triangle-free graphs, which also provides the lower bound in Theorem 1.3 (i).

Let G be a triangle-free graph with minimum degree at least d. We may assume that G is minimal with respect to the minimum degree assumption, namely that every proper induced subgraph of G has minimum degree less than d. Under this assumption, note that G is d-degenerate. Indeed, by assumption every proper induced subgraph of G has a vertex of degree at most d. We can therefore order the vertices of G from left to right in such a way that every vertex has at most d neighbours to its right. Let $N^+(v)$ be the set of neighbours of v to its right, and let $\ell = \lfloor \log d/\log\log d\rfloor$. Throughout this section we assume that d is sufficiently large. The heart of the proof of Theorem 1.2 for triangle-free graphs is the following claim.

Claim 3.1. *There exist two non-empty disjoint sets $X, Y \subseteq V(G)$ with the following properties.*

(i) *every vertex in Y has at least ℓ neighbours in X;*

(ii) *there are at most $(\ell/7)|Y|$ edges between X and Y that are incident to a vertex x in X with $|N^+(x) \cap X| \geq 4$;*

(iii) $|X| < 3|Y|$*;*

(iv) $e(Y) < 3|Y|$.

Before proving Claim 3.1, we will show how it implies Theorem 1.2 for triangle-free graphs.

Proof of Theorem 1.2 for triangle-free graphs. Let X and Y be the sets given by Claim 3.1. Let X_0 be the set of vertices $x \in X$ for which $|N^+(x) \cap X| \geq 4$. The graph $G[X \setminus X_0]$ is 3-degenerate, hence it is 4-colourable. Let $\{X_1, \ldots, X_4\}$ be a partition of $X \setminus X_0$ into four independent sets. Let Y_0 be the set of vertices in Y that send at least $3\ell/7$ edges into X_0. The number of edges between X_0 and Y_0 is at least $(3\ell/7)|Y_0|$, and at most $(\ell/7)|Y|$ (by Property (ii) of the claim). So $|Y_0| \leq (1/3)|Y|$. Note that the number of edges in $Y \setminus Y_0$ is at most $e(Y) \leq 3|Y| \leq (9/2)|Y \setminus Y_0|$, due to Property (iv). In other words, the average degree of $G[Y \setminus Y_0]$ is at most 9. Thus by Turán's theorem [29], $Y \setminus Y_0$ contains an independent set Y' of size at least $(1/10)|Y \setminus Y_0| \geq (1/15)|Y|$. To prove Theorem 1.2, it suffices to show that $G[X_i, Y']$ has average degree at least $\Omega(\ell)$ for some $i \in [4]$.

From Property (i) and the definition of Y_0, we see that every vertex in $Y \setminus Y_0$ has at least $4\ell/7$ neighbours in $X \setminus X_0$. Moreover, as

$$\left(\tfrac{1}{4}|X \setminus X_0| + |Y'|\right) \cdot \sum_{1 \leq i \leq 4} e(X_i, Y') = \tfrac{1}{4} \sum_{1 \leq i \leq 4} (|X_i| + |Y'|) \cdot e(X \setminus X_0, Y'),$$

there exists $i \in [4]$ such that $\left(\frac{1}{4}|X \setminus X_0| + |Y'|\right) \cdot e(X_i, Y') \geq \frac{1}{4}(|X_i| + |Y'|) \cdot e(X \setminus X_0, Y')$. Therefore, the average degree of $G[X_i, Y']$ is

$$\frac{2e(X_i, Y')}{|X_i| + |Y'|} \geq \frac{e(X \setminus X_0, Y')}{(1/2)|X \setminus X_0| + 2|Y'|} \geq \frac{(4\ell/7)|Y'|}{(1/2)|X \setminus X_0| + 2|Y'|} \geq (8/343)\ell.$$

Here the last inequality holds since $|Y'| \geq (1/15)|Y|$ and $\frac{(4\ell/7)|Y'|}{(1/2)|X \setminus X_0| + 2|Y'|}$ is an increasing function in $|Y'|$, and since $|X \setminus X_0| \leq 3|Y|$ (by Property (iii)). Thus G contains an induced bipartite graph with minimum degree at least $(4/735)\ell = \Omega(\log d / \log\log d)$. $\square$

The rest of this section is devoted to the proof of Claim 3.1. Let $p = 1/d$. For every vertex $u \in V(G)$, define p_u to satisfy $(1-p) \cdot \mathbb{P}[\mathrm{Bin}(d(u), p) \geq \ell] \cdot p_u = p$. We define two random sets X and Y as follows. First, each vertex $x \in V(G)$ is included in X with probability p, independently. Second, for every vertex $y \in V(G)$, we put it in Y if it is not in X, it has at least ℓ neighbours in X, and a biased coin flip, whose probability of heads is p_y, turns out to be heads. (These coin flips are independent for each $y \in V(G)$.) Let Z be the set of edges xy where $x \in X$, $y \in Y$ and $|N^+(x) \cap X| \geq 4$. In order to prove Claim 3.1, we prove the following claim.

Claim 3.2. $\mathbb{E}[|Y| - (1/3)|X| - (1/3)e(Y) - (7/\ell)|Z|] > 0$.

Before proving Claim 3.2, we show that it implies Claim 3.1.

Proof of Claim 3.1 given Claim 3.2. By Claim 3.2, there is a choice of sets X and Y as above, with

$$|Y| - (1/3)|X| - (1/3)e(Y) - (7/\ell)|Z| > 0.$$

In particular, X and Y are disjoint; every vertex in Y sends at least ℓ edges into X; $|Z| < (\ell/7)|Y|$, i.e. there are at most $(\ell/7)|Y|$ edges that are incident with a vertex x in X with $|N^+(x) \cap X| \geq 4$; $|X| < 3|Y|$; and $e(Y) < 3|Y|$. Claim 3.1 follows. $\square$

To prove Claim 3.2 we shall need the following routine estimates concerning the binomial distribution. The reader may wish to skip their verification and go straight to the proof of Claim 3.2, given after Lemma 3.3. We remark that the need to satisfy (i) is the bottleneck in the choice of the value ℓ.

Lemma 3.3. *We have*

(i) $(1-p) \cdot \mathbb{P}[\mathrm{Bin}(d,p) \geq \ell] \geq p$,

(ii) $\mathbb{P}[\mathrm{Bin}(m-1,p) \geq \ell - 1] \leq (\ell+1) \cdot \mathbb{P}[\mathrm{Bin}(m,p) \geq \ell]$ *for every* $m \geq d$.

Proof. For Part (i) we bound

$$\begin{aligned}(1-p) \cdot \mathbb{P}[\mathrm{Bin}(d,p) \geq \ell] &\geq \binom{d}{\ell} p^\ell (1-p)^{d-\ell+1} \geq \left(\frac{pd}{\ell}\right)^\ell e^{-2p(d-\ell+1)} \\ &\geq \ell^{-\ell} e^{-2} = e^{-\ell \log \ell - 2} \\ &\geq e^{-\frac{\log d}{\log\log d}(\log\log d - \log\log\log d) - 2} \\ &\geq e^{-\log d} = p.\end{aligned}$$

Here we used the inequalities $\binom{k}{t} \geq (k/t)^t$, which holds for $1 \leq t \leq k$, and $1 - x \geq e^{-2x}$, which holds for $0 < x < 1/2$. For Part (ii) we estimate

$$\begin{aligned}\frac{\mathbb{P}[\mathrm{Bin}(m-1,p) \geq \ell-1]}{\mathbb{P}[\mathrm{Bin}(m,p) \geq \ell]} &= \frac{\mathbb{P}[\mathrm{Bin}(m-1,p) = \ell-1]}{\mathbb{P}[\mathrm{Bin}(m,p) \geq \ell]} + \frac{\mathbb{P}[\mathrm{Bin}(m-1,p) \geq \ell]}{\mathbb{P}[\mathrm{Bin}(m,p) \geq \ell]} \\ &\leq \frac{\mathbb{P}[\mathrm{Bin}(m-1,p) = \ell-1]}{\mathbb{P}[\mathrm{Bin}(m,p) = \ell]} + 1 \\ &= \frac{\binom{m-1}{\ell-1} p^{\ell-1}(1-p)^{m-\ell}}{\binom{m}{\ell} p^\ell (1-p)^{m-\ell}} + 1 \\ &= \frac{\ell}{mp} + 1 \leq \ell + 1.\end{aligned}$$

$\square$

Proof of Claim 3.2. We shall estimate the expectations of the four quantities $|X|, |Y|, e(Y), |Z|$ separately, and show that they satisfy the desired inequality. Note that, by Lemma 3.3 (i) and the choice of p_u, we have $0 < p_u \leq 1$. In addition, the probability that a vertex y is in Y is $(1-p) \cdot \mathbb{P}[\mathrm{Bin}(d(y),p) \geq \ell] \cdot p_y = p$.

Since each vertex is in X with probability p, and similarly each vertex is in Y with probability p, we have

$$\mathbb{E}[|X|] = \mathbb{E}[|Y|] = np = n/d. \tag{2}$$

Let uv be an edge. Then the probability of the event $u, v \in Y$ is

$$\begin{aligned}
&p_u \cdot p_v \cdot \mathbb{P}\big[|N(u) \cap X|, |N(v) \cap X| \geq \ell \text{ and } u, v \notin X\big] \\
&= p_u \cdot p_v \cdot (1-p)^2 \cdot \mathbb{P}\big[|(N(u) \setminus \{v\}) \cap X|, |(N(v) \setminus \{u\}) \cap X| \geq \ell\big] \\
&= p_u \cdot p_v \cdot (1-p)^2 \cdot \mathbb{P}\big[|(N(u) \setminus \{v\}) \cap X| \geq \ell\big] \cdot \mathbb{P}\big[|(N(v) \setminus \{u\}) \cap X| \geq \ell\big] \\
&\leq p_u \cdot p_v \cdot (1-p)^2 \cdot \mathbb{P}[\mathrm{Bin}(d(u), p) \geq \ell] \cdot \mathbb{P}[\mathrm{Bin}(d(v), p) \geq \ell] = p^2.
\end{aligned}$$

In the second equality we used the fact that the sets $N(u) \setminus \{v\}$ and $N(v) \setminus \{u\}$ are disjoint (as G is triangle-free). This implies that the events $|(N(u) \setminus \{v\}) \cap X| \geq \ell$ and $|(N(v) \setminus \{u\}) \cap X| \geq \ell$ are independent. As G is d-degenerate, it follows that

$$\mathbb{E}[e(Y)] \leq e(G)p^2 \leq ndp^2 = n/d. \tag{3}$$

Given an edge xy, the probability that $xy \in Z$, with $x \in X, y \in Y$, is

$$\begin{aligned}
&p_y \cdot \mathbb{P}\big[x \in X,\ |N^+(x) \cap X| \geq 4,\ y \notin X,\ |N(y) \cap X| \geq \ell\big] \\
&= p_y \cdot p \cdot \mathbb{P}\big[|(N^+(x) \setminus \{y\}) \cap X| \geq 4\big] \cdot (1-p) \cdot \mathbb{P}\big[|(N(y) \setminus \{x\}) \cap X| \geq \ell - 1\big] \\
&\leq p_y \cdot p \cdot \binom{d}{4} p^4 \cdot (1-p) \cdot \mathbb{P}\big[\mathrm{Bin}(d(y)-1, p) \geq \ell - 1\big] \\
&\leq p_y \cdot p \cdot \frac{(pd)^4}{4!} \cdot (1-p) \cdot \mathbb{P}[\mathrm{Bin}(d(y), p) \geq \ell] \cdot (\ell + 1) \\
&= p^2 \cdot \frac{1}{24} \cdot (\ell + 1) \ \leq\ p^2 \ell / 23.
\end{aligned}$$

In the first equality we again used the assumption that G is triangle-free to deduce independence of events that depend on $N(x) \setminus \{y\}$ and $N(y) \setminus \{x\}$. In the second inequality we used the inequality $\binom{k}{t} \leq k^t/t!$ and Lemma 3.3 (ii) and in the last equality we used the definition of p_y. We deduce that

$$\mathbb{E}[|Z|] \leq e(G) \cdot (p^2\ell/23) \leq nd \cdot (p^2\ell/23) = n\ell/(23d). \tag{4}$$

Finally, it follows from (2), (3) and (4) that

$$\mathbb{E}[|Y| - (1/3)|X| - (1/3)e(Y) - (7/\ell)|Z|] > n/d - n/(3d) - n/(3d) - n/(3d) = 0,$$

finishing the proof of Claim 3.2. □

3.2 The dense regime

In this subsection we will provide two different proofs of the fact that $g(n, d) = \Omega(d^2/n)$. We note that a different proof of the bound $g(n, d) \geq d^2/2n$ (which is the bound obtained in the first proof) is given in [7]. This will establish the lower bound in parts (ii) and (iv) of Theorem 1.3. Let G be a triangle-free graph on n vertices with minimum degree at least d.

First proof. For each vertex $v \in V(G)$, we fix a subset $A_v \subseteq N(v)$ of size d. Let $X = \{x_1, x_2\}$ be a uniformly random pair of vertices of $V(G)$, and define a vertex set Y as follows: $v \in V(G)$ is included in Y if and only if $A_v \cap X \neq \emptyset$. Since G is triangle-free, $G[Y]$ is a bipartite graph with bipartition $\{N(x_1) \cap Y, (N(x_2) \setminus N(x_1)) \cap Y\}$. Thus, in order to prove the theorem, it suffices to show that $G[Y]$ has average degree at least d^2/n (and therefore has an induced subgraph of minimum degree at least $d^2/(2n)$) for some choice of X.

For each vertex $v \in V(G)$, we have $\mathbb{P}(v \in Y) = 1 - \binom{n-d}{2}/\binom{n}{2} < \frac{2d}{n}$. Given an edge $uv \in E(G)$, we have $N(u) \cap N(v) = \emptyset$ as G is triangle-free, and so $\mathbb{P}(u, v \in Y) = \frac{|A_u||A_v|}{\binom{n}{2}} = \frac{d^2}{\binom{n}{2}} > \frac{2d^2}{n^2}$. Therefore, using the fact that $e(G) \geq nd/2$, we obtain

$$\mathbb{E}\left[e(Y) - \frac{d^2}{2n}|Y|\right] > e(G) \cdot \frac{2d^2}{n^2} - \frac{d^2}{2n} \cdot n \cdot \frac{2d}{n} \geq 0.$$

Hence $e(Y) > \frac{d^2}{2n}|Y|$ for some choice of X, as desired. □

Second proof, with a slightly worse bound. Since the statement holds trivially for $d \leq 2\sqrt{n}$, we may assume that $d > 2\sqrt{n}$. Given $uv \in E(G)$, let $c(u, v)$ be the number of 4-cycles passing through $\{u, v\}$. Since G is triangle-free, we see that $N(u) \setminus \{v\}$ and $N(v) \setminus \{u\}$ are two disjoint independent sets, and that there are exactly $c(u, v)$ edges between $N(u) \setminus \{v\}$ and $N(v) \setminus \{u\}$. Hence the subgraph of G induced by $(N(u) \setminus \{v\}) \cup (N(v) \setminus \{u\})$ is a bipartite graph with average degree at least $2c(u, v)/(d(u) + d(v) - 2)$, and thus has an induced bipartite subgraph with minimum degree at least $c(u, v)/(d(u) + d(v) - 2)$. In order to prove that $g(n, d) \geq d^2/(4n)$, it thus suffices to show that $q \geq d^2/(4n)$, where

$$q := \max_{uv \in E(G)} \frac{c(u, v)}{d(u) + d(v) - 2}.$$

From the definition of q, we see that $q \cdot \sum_{uv \in E(G)}(d(u) + d(v) - 2) \geq \sum_{uv \in E(G)} c(u, v)$. Letting c_4 denote the number of 4-cycles in G, we then obtain

$$q \geq \frac{\sum_{uv \in E(G)} c(u, v)}{\sum_{uv \in E(G)}(d(u) + d(v) - 2)} = \frac{2c_4}{\sum_{u \in V(G)} \binom{d(u)}{2}}.$$

By Jensen's inequality,

$$\sum_{x,y \in V(G)} \binom{|N(x) \cap N(y)|}{2} \geq \binom{n}{2} \binom{\sum_{x,y \in V(G)} \binom{|N(x) \cap N(y)|}{2} / \binom{n}{2}}{2}.$$

Noting that $\sum_{x,y \in V(G)} |N(x) \cap N(y)| = \sum_{u \in V(G)} \binom{d(u)}{2}$, we get

$$\begin{aligned} 2c_4 &= \sum_{x,y \in V(G)} \binom{|N(x) \cap N(y)|}{2} \\ &\geq \binom{n}{2} \binom{\sum_{u \in V(G)} \binom{d(u)}{2} / \binom{n}{2}}{2} \\ &\geq \frac{1}{3\binom{n}{2}} \left(\sum_{u \in V(G)} \binom{d(u)}{2} \right)^2, \end{aligned}$$

where the last inequality holds since $\sum_{u \in V(G)} \binom{d(u)}{2} / \binom{n}{2} \geq n\binom{d}{2}/\binom{n}{2} > 3$ (using our assumption that $d > 2\sqrt{n}$). Therefore,

$$q \geq \frac{\sum_{u \in V(G)} \binom{d(u)}{2}}{3\binom{n}{2}} \geq \frac{n\binom{d}{2}}{3\binom{n}{2}} \geq \frac{d^2}{4n},$$

where the second inequality follows from the assumption that $d(u) \geq d$ for every $u \in V(G)$, and the last inequality holds since $d > 2\sqrt{n}$. This finishes the proof. $\square$

3.3 Graphs with polynomially-large minimum degree

In this subsection we prove that Conjecture 1.1 holds for graphs with minimum degree at least $n^{\Omega(1)}$, which also proves the lower bound in Theorem 1.3 (iii). The following lemma appeared as [16, Theorem 3.4], and is a corollary of Johansson's theorem [22] for colouring triangle-free graphs and a connection between the fractional chromatic number and dense bipartite induced subgraphs (see [16, Theorem 3.1] and a discussion in Section 6).

Lemma 3.4. *There is a constant $c > 0$ such that any triangle-free graph with minimum degree d and maximum degree Δ contains a bipartite induced subgraph of average degree at least $\frac{cd}{\Delta} \log \Delta$.*

Every graph with average degree d has an induced subgraph with minimum degree at least $d/2$, so one can deduce a version of Lemma 3.4 which guarantees a bipartite induced subgraph with *minimum* degree at least $\frac{cd}{2\Delta} \log \Delta$.

The other ingredient we will need is a theorem of Erdős and Simonovits [15] (see also [21, Lemma 2.6]).

Lemma 3.5. *Consider $\varepsilon \in (0, 1)$, and let n be sufficiently large relative to ε. Let G be an n-vertex graph with $e(G) \geq n^{1+\varepsilon}$. Then G contains an induced subgraph H on $m \geq n^{\varepsilon(1-\varepsilon)(1+\varepsilon)}$ vertices such that $e(H) \geq \frac{2}{5} m^{1+\varepsilon}$ and $\Delta(H)/\delta(H) \leq 10 \cdot 2^{1/\varepsilon^2+1}$.*

Now, consider $d = n^{\Omega(1)}$ and let G be an n-vertex graph with minimum degree at least d. Then G has $n^{1+\Omega(1)}$ edges, so we may apply Lemma 3.5 to deduce that it has an induced subgraph, all of whose degrees are in the range between d' and Kd', for some $d' = n^{\Omega(1)}$ and some $K = O(1)$. By Lemma 3.4, inside this induced subgraph there is a bipartite induced subgraph with minimum degree at least $\frac{c}{2K} \log d' = \Omega(\log n)$, as desired.

4 Reducing from H-free graphs to triangle-free graphs

In this section we show that in order to prove Theorem 1.2 it suffices to consider the case of triangle-free graphs.

Proof of Theorem 1.2 under the assumption that it holds for $H = K_3$. We note that it suffices to prove the theorem for K_t-free graphs, as every H-free graph is $K_{|H|}$-free. We proceed by induction. Fix $t > 3$ and assume that for every $3 \leq s < t$, every K_s-free graph with minimum degree q has an induced bipartite subgraph with minimum degree $\Omega(\log q / \log \log q)$.

Let G be a K_t-free graph on n vertices with minimum degree at least d. As in Section 3.1, we may assume that G is d-degenerate, and fix an ordering of the vertices from left to right such that every vertex has at most d neighbours to its right. Let $N^+(x)$ be the set of neighbours of a vertex x to its right.

Note that each $G[N^+(v)]$ is K_{t-1}-free (otherwise, appending v to a copy of K_{t-1} would give a copy of K_t in G). We may assume that each $G[N^+(v)]$ has at most $d^{7/6}$ edges. Indeed, otherwise $G[N^+(v)]$ would have average degree at least $2d^{1/6}$, and therefore have an induced subgraph with minimum degree at least $d^{1/6}$. By the induction hypothesis it would therefore have an induced bipartite subgraph with minimum degree $\Omega\big(\log d^{1/6} / \log \log d^{1/6}\big) = \Omega(\log d / \log \log d)$.

Now, let U be a random subset of the vertices of G, containing each vertex with probability $p := d^{-2/3}$, independently. Then, let $W \subseteq U$ be obtained by removing the leftmost vertex of every triangle in $G[U]$. From the definition of W, we know that $G[W]$ is triangle-free. We wish to show that with positive probability $G[W]$ has at least $|W|d^{1/6}$ edges, which will imply that it has a subgraph with minimum degree at least $d^{1/6}$ and therefore has an induced bipartite subgraph with minimum degree $\Omega(\log d/\log\log d)$, by the induction hypothesis.

Let X_1 be the number of edges in $G[U]$, let X_2 be the number of triangles in $G[U]$ and let X_3 be the number of edges e in $G[U]$ which are incident to the leftmost vertex of a triangle in $G[U]$ (such that e is not actually in the triangle). Then $G[W]$ has at least $X_1 - 2X_2 - X_3$ edges.

The number of edges in G is at least $nd/2$ by the minimum degree assumption, and the number of triangles in G is $\sum_{v\in V(G)} e(G[N^+(v)]) \le nd^{7/6}$ (counting triangles by their leftmost vertex). The number of edge-triangle pairs (e,T) such that e is not in T and e is incident to the leftmost vertex of T is at most

$$\sum_{v\in V(G)} d(v)e(G[N^+(v)]) \le \sum_{v\in V(G)} d(v)d^{7/6} = 2e(G)d^{7/6} \le 2nd^{13/6}.$$

Therefore

$$\mathbb{E}|U| = np, \quad \mathbb{E}X_1 \ge \tfrac{1}{2}ndp^2, \quad \mathbb{E}X_2 \le nd^{7/6}p^3, \quad \mathbb{E}X_3 \le 2nd^{13/6}p^4,$$

and

$$\begin{aligned}\mathbb{E}\Big[(X_1 - 2X_2 - X_3) - d^{1/6}|U|\Big] &\ge \tfrac{1}{2}ndp^2 - 2nd^{7/6}p^3 - 2nd^{13/6}p^4 - d^{1/6}np \\ &= n\Big(\tfrac{1}{2}d^{-1/3} - 2d^{-5/6} - 2d^{-1/2} - d^{-1/2}\Big) > 0\end{aligned}$$

for sufficiently large d. It follows that there is an outcome of U with $(X_1 - 2X_2 - X_3) - d^{1/6}|U| > 0$. Then, $G[W]$ is an induced subgraph with at least $X_1 - 2X_2 - X_3 > d^{1/6}|U| \ge d^{1/6}|W|$ edges, as desired. □

5 Upper bounds on minimum degree

In this section we describe both probabilistic and explicit constructions of triangle-free graphs which have no high-degree bipartite induced subgraphs. These will be used for the proofs of the upper bounds in Theorem 1.3. To be specific, the upper bounds for (i)-(iii) are proved in Section 5.1, and the upper bound for (iv) is proved in Section 5.2.

5.1 Probabilistic construction

In this subsection we prove the upper bounds in Theorem 1.3 (i)-(iii). This essentially boils down to the following fact.

Lemma 5.1. *For any n, there exists a triangle-free graph G_n on between $n/2$ and n vertices such that every vertex of G_n has degree $\Theta(\sqrt{n})$, and every induced bipartite subgraph of G has minimum degree at most $O(\log n)$.*

Proof of the upper bounds in Theorem 1.3 (i)-(iii) *given Lemma 5.1.* Let G_n be a graph as in the statement of Lemma 5.1. By blowing up each of the vertices into an independent set of size 1 or 2 (as in the proof of Proposition 1.4), in such a way that the total number of verties in the blown-up graph is n, we obtain a graph on n vertices, with minimum degree $\Theta(\sqrt{n})$, whose every induced bipartite subgraph has minimum degree $O(\log n)$. By replacing G_n with the blown-up graph, we may thus assume that G_n itself has exactly n vertices.

Let $\delta(G_n) = \Theta(\sqrt{n})$ be the minimum degree of the graph G_n. We divide our analysis into two cases: $\delta(G_n) \le d \le n/2$, and $d \le \delta(G_n)$.

Suppose that $\delta(G_n) \le d \le n/2$. We will show that $g(n,d) = O((d^2 \log d)/n)$ by taking blowups as in the proof of Proposition 1.4. Let n' be maximal such that $n' \le n$ and $\lfloor n/n' \rfloor \delta(G_{n'}) \ge d$. (There exists such an n' because $n' = 2$ satisfies the required inequalities.) We claim that $n' = \Theta((n/d)^2)$. Indeed, recall that $\delta(G_{n'}) = \Theta(\sqrt{n'})$. We thus have $d = O(n/\sqrt{n'})$, and so $n' = O((n/d)^2)$. For the lower bound, note that if $n'' = \lfloor c(n/d)^2 \rfloor$, for a sufficiently small constant $c > 0$, then $\lfloor n/n'' \rfloor \delta(G_{n''}) \ge d$. It follows that $n' \ge n'' = \Omega((n/d)^2)$. Blow up the vertices of $G_{n'}$ into independent sets with almost-equal sizes (about n/n') to obtain an n-vertex graph with minimum degree at least d, in which every induced bipartite subgraph has minimum degree at most $O((n/n') \log(n')) = O((d^2 \log d)/n)$. This shows that $g(n,d) = O((d^2 \log d)/n)$ in the regime where $\delta(G_n) \le d \le n/2$.

Finally, we consider the case $d \le \delta(G_n)$. Let n' be minimal such that $\delta(G_{n'}) \ge d$, so $n' = \Theta(d^2)$. Consider the graph which is a disjoint union of $\lfloor n/n' \rfloor$ copies of $G_{n'}$. This graph has between $n/2$ and n vertices, it has minimum degree at least d, and all of its induced bipartite subgraphs have minimum degree $O(\log n') = O(\log d)$ (because every induced bipartite subgraph is a disjoint union of induced bipartite subgraphs of the copies of $G_{n'}$). As above, by blowing up each vertex into an independent set of size 1 or 2, we can obtain a graph on exactly n vertices, with minimum degree at least d, whose every induced bipartite subgraph has minimum degree $O(\log d)$. It follows that $g(n,d) = O(\log d)$ when $d \le \delta(G_n)$. $\square$

To prove Lemma 5.1 we will follow the approach of Krivelevich [23]. Our main tools are the following large deviation inequalities.

Lemma 5.2. *The following hold.*

(a) *Let $n \in \mathbb{N}$ and $p, \varepsilon \in [0,1]$. Then $\mathbb{P}(|\operatorname{Bin}(n,p) - np| \ge \varepsilon np) \le 2 \cdot e^{-\varepsilon^2 np/3}$.*

(b) *Consider a finite set Γ. Let $\{X_i : i \in \Gamma\}$ be a set of independent random variables each supported on $\{0,1\}$, and let $\mathcal{F} \subset 2^\Gamma$ be a collection of subsets of Γ. Given $F \in \mathcal{F}$, write X_F for the random variable $\prod_{i \in F} X_i$. Now, define $X = \sum_{F \in \mathcal{F}} X_F$, and*

$$X_0 = \max\{m : \exists \text{ pairwise disjoint sets } F_1, \dots, F_m \in \mathcal{F} \text{ such that } X_{F_1} = \dots = X_{F_m} = 1\}.$$

Then we have $\mathbb{P}(X_0 \ge 5\mathbb{E}X) \le e^{-\mathbb{E}X}$.

Part (a) of Lemma 5.2 is the well-known Chernoff bound (see, e.g., [17, Corollary 21.7]), while Part (b) follows from [23, Claim 1]. We will use Lemma 5.2 to prove the following statement about triangle-free subgraphs of a random graph.

Lemma 5.3. *Consider the binomial random graph $G \sim G(n,p)$, with $p = cn^{-1/2}$ for some absolute constant $c \in (0, \frac{1}{20})$. Set $a = \max\{10^6 c^{-1}, 12c^{-3}, 28\}$. Then w.h.p. G contains a triangle-free subgraph H satisfying the following properties.*

$$\Delta(H) \le 1.01np, \tag{5}$$

$$e_H(A,B) \le a|A| \log n \text{ for every disjoint sets } A, B \subset V(G) \text{ with } |A| = |B| \le a\sqrt{n} \log n, \tag{6}$$

$$e(H) \ge 0.9p\binom{n}{2}, \tag{7}$$

$$\alpha(H) \le a\sqrt{n} \log n. \tag{8}$$

Before proving Lemma 5.3, we show how to deduce Lemma 5.1 from it.

Proof of Lemma 5.1. Let H be a triangle-free graph which satisfies (5)–(8). Write H' for the graph obtained from H by iteratively removing vertices of degree at most $pn/30$. Obviously H' is triangle-free. We have $e(H') \geq e(H) - pn^2/30 \geq 0.83p\binom{n}{2}$ by (7). Combining this with (5) we deduce that $|H'| \in [0.8n, n]$, and each vertex of H' has degree $\Theta(pn) = \Theta(\sqrt{n})$. Moreover, from (6) and (8) we learn that every induced bipartite subgraph of H' has minimum degree at most $O(\log n)$. $\square$

Proof of Lemma 5.3. We follow the method of Krivelevich in [23]. Let $\mathcal{T}$ be any maximal family of edge-disjoint triangles in G. Let H be the triangle-free graph obtained from G by removing all edges in $\mathcal{T}$. We will use Lemma 5.2 to show that w.h.p. H satisfies (5)–(8). For ease of notation, set $k = \lfloor a\sqrt{n}\log n \rfloor$.

As $d_G(v) \sim \mathrm{Bin}(n-1, p)$ for every $v \in V(G)$, using Lemma 5.2 (a) and the union bound we get

$$\mathbb{P}(\exists v \in V(G) \text{ with } d_G(v) \geq 1.01np) \leq n \cdot 2e^{-\Omega(np)} = n \cdot e^{-\Omega(\sqrt{n})} = o(1),$$

Thus w.h.p. $d_G(v) \leq 1.01np$ for every $v \in V(G)$. This implies that (5) holds w.h.p.

Since $e_G(A,B) \sim \mathrm{Bin}(|A||B|, p)$ for every pair of disjoint sets $A, B \subseteq V(G)$, applying Lemma 5.2 (a) and the union bound we find that

$$\begin{aligned}
&\mathbb{P}(\exists \text{ disjoint vertex sets } A, B \text{ with } |A| = |B| \leq k \text{ and } e_G(A,B) \geq a|A|\log n) \\
&\qquad \leq \sum_{\ell=1}^{k} n^{2\ell} \cdot \mathbb{P}\big(\mathrm{Bin}(\ell^2, p) \geq a\ell\log n\big) \\
&\qquad \leq \sum_{\ell=1}^{k} n^{2\ell} \cdot \mathbb{P}\big(\mathrm{Bin}\big(\lfloor \tfrac{1}{2}ap^{-1}\ell\log n \rfloor, p\big) \geq a\ell\log n\big) \\
&\qquad \leq \sum_{\ell=1}^{k} n^{2\ell} \cdot 2e^{-(a\ell\log n)/7} \leq 2\sum_{\ell=1}^{k} n^{-2\ell} = o(1).
\end{aligned}$$

The second inequality holds since $\ell^2 \leq \frac{1}{2}ap^{-1}\ell\log n$ for $\ell \leq a\sqrt{n}\log n$ and $p \leq \frac{1}{20}n^{-1/2}$, while the last inequality follows from the assumption that $a \geq 28$. We have proved that (6) holds w.h.p. It remains to consider (7) and (8).

For every subset $S \subset V(G)$ of size k, denote by Y_S the number of triangles that have at least two vertices in S, and by Z_S the maximum number of pairwise edge-disjoint triangles with at least two vertices in S. Clearly $Z_S \leq Y_S$. We claim that w.h.p.

$$e_G(S) \geq 34\, Z_S \text{ for every subset } S \subset V(G) \text{ of size } k. \tag{9}$$

Before proving (9), we will show how it implies (7) and (8). Since $e_G(S) \sim \mathrm{Bin}(\binom{k}{2}, p)$ for every subset $S \subset V(G)$ of size k, using Lemma 5.2 (a) and the union bound we get that

$$\mathbb{P}\left(\exists \text{ a vertex set } S \text{ of size } k \text{ with } e_G(S) \leq 0.99p\binom{k}{2}\right) \leq n^k \cdot 2e^{-10^{-5}p\binom{k}{2}} \leq 2n^{-k} = o(1).$$

In the second inequality we used the facts that $p = cn^{-1/2}$, $k = \lfloor a\sqrt{n}\log n \rfloor$ and $a \geq 10^6c^{-1}$. Hence w.h.p. every size-k vertex set S satisfies

$$e_G(S) \geq 0.99p\binom{k}{2}. \tag{10}$$

Now, from the definition of H and Z_S we see that w.h.p. every size-k subset $S \subset V(G)$ spans at least

$$e_G(S) - 3Z_S \overset{(9)}{>} 0.91e_G(S) \overset{(10)}{>} 0.9p\binom{k}{2}$$

edges in H. It follows that w.h.p. $\alpha(H) \leq k$ and $e(H) \geq 0.9p\binom{n}{2}$, as required.

We now return to the proof of (9). Given a size-k vertex set S, we will bound $\mathbb{P}(e_G(S) < 34Z_S)$ using Lemma 5.2 (b). Note that $\frac{\mathbb{E}[e_G(S)]}{\mathbb{E}Y_S} \geq \frac{p\binom{k}{2}}{n\binom{k}{2}p^3} \geq 400$, assuming $p \leq \frac{1}{20}n^{-1/2}$. Hence

$$\begin{aligned}\mathbb{P}(e_G(S) < 34Z_S) &\leq \mathbb{P}(e_G(S) \leq \mathbb{E}[e_G(S)]/2) + \mathbb{P}(34Z_S \geq \mathbb{E}[e_G(S)]/2)\\ &\leq \mathbb{P}(e_G(S) \leq \mathbb{E}[e_G(S)]/2) + \mathbb{P}(Z_S \geq 5\mathbb{E}Y_S).\end{aligned}$$

Since $e_G(S) \sim \mathrm{Bin}(\binom{k}{2}, p)$, applying Lemma 5.2 (a) we get that $\mathbb{P}(e_G(S) \leq \mathbb{E}[e_G(S)]/2) \leq 2e^{-p\binom{k}{2}/12}$. Moreover, Lemma 5.2 (b) implies that $\mathbb{P}(Z_S \geq 5\mathbb{E}Y_S) \leq e^{-\mathbb{E}Y_S} = e^{-(1+o(1))n\binom{k}{2}p^3}$. Therefore, using the facts $p = cn^{-1/2} \leq \frac{1}{20}n^{-1/2}$ and $k = \omega(1)$ we obtain

$$\mathbb{P}(e_G(S) < 34Z_S) \leq 2e^{-p\binom{k}{2}/12} + e^{-(1+o(1))n\binom{k}{2}p^3} \leq 3e^{-\frac{c^2}{3}pk^2}.$$

With the union bound, we deduce that

$$\mathbb{P}((9)\text{ does not hold}) \leq n^k \cdot 3e^{-\frac{c^2}{3}pk^2} \leq 3e^{k\log n - \frac{c^2}{3}pk^2} \leq 3e^{-k\log n} = o(1),$$

where the last inequality holds since $p = cn^{-1/2}$, $k = \lfloor a\sqrt{n}\log n\rfloor$ and $a \geq 12c^{-3}$. This finishes our proof of Lemma 5.3. □

5.2 Explicit construction

In this section we prove the upper bound in Theorem 1.3 (iv), by blowing up a sequence of triangle-free quasirandom graphs constructed by Alon [1]. A graph G is called an (n, d, λ)-graph if it is d-regular, has n vertices, and all eigenvalues of its adjacency matrix, save the largest, are smaller in absolute value than λ. Alon's construction implies the following lemma.

Lemma 5.4. *For every integer n of the form $n = 2^{3k}$ for some $k \in \mathbb{N}$, one can explicitly construct a triangle-free (n, d, λ)-graph G_n with $d = 2^{2k-2} - 2^{k-1} = \Theta(n^{2/3})$ and $\lambda = O(n^{1/3})$.*

The expander mixing lemma (see, for example, [24, Theorem 2.11]) asserts that (n, d, λ)-graphs with small λ have low discrepancy, as follows.

Lemma 5.5 (Expander mixing lemma)**.** *Suppose G is an (n, d, λ)-graph. Then for any disjoint vertex sets $A, B \subset V(G)$, we have*

$$\left|e(A, B) - \frac{d}{n}|A||B|\right| \leq \lambda\sqrt{|A||B|}.$$

In much the same way as we used Theorem 2.3 in the proof of Proposition 1.4, we can use the expander mixing lemma to prove that the graphs in Lemma 5.4 do not have bipartite induced subgraphs with high minimum degree.

Corollary 5.6. *Let $n = 2^{3k}$, and let G_n be a graph coming from Lemma 5.4. Then every induced bipartite subgraph of G_n has minimum degree at most $O(n^{1/3})$.*

Proof. From the expander mixing lemma, we learn that $\alpha(G_n) \le 2(n\lambda/d+1) = O(n^{2/3})$. Moreover, for any two sets $A, B \subset V(G_n)$ of size $O(n^{2/3})$, the expander mixing lemma gives

$$e(A,B) \le n^{-1/3}|A||B| + n^{1/3}\sqrt{|A||B|} = O(n^{1/3}\max\{|A|,|B|\}).$$

Thus any induced bipartite subgraph of G_n has minimum degree at most $O(n^{1/3}) = O(d^2/n)$. □

Now, we can prove the upper bound in Theorem 1.3 (iv) by taking blowups as in previous proofs.

Proof of the upper bound in Theorem 1.3 (iv). Note that the upper bound holds trivially when $d \ge n/32$. Thus, we may assume that $n^{2/3} \le d \le n/32$. Let k be maximal such that $2^{3k} \le n$ and $\lfloor n/2^{3k} \rfloor d(G_{2^{3k}}) \ge d$, where $G_{2^{3k}}$ is the graph given by Lemma 5.4. We claim that such k exists. Indeed, let ℓ be an integer such that $n/(32d) \le 2^\ell \le n/(16d)$. Then $2^{3\ell} \le (n/(16d))^3 \le n/2$, and

$$\left\lfloor \frac{n}{2^{3\ell}} \right\rfloor d(G_{2^{3\ell}}) \ge \frac{n}{2^{3\ell+1}}\left(2^{2\ell-2} - 2^{\ell-1}\right) = \frac{n}{2^\ell}\left(\frac{1}{8} - \frac{1}{4\cdot 2^\ell}\right) \ge \frac{n}{16\cdot 2^\ell} \ge d.$$

This implies that there exists k that satisfies both inequalities, as claimed. As $d(G_{2^{3k}}) = 2^{2k-2} - 2^{k-1} = \Theta((2^{3k})^{2/3})$, this means that $n' := 2^{3k} = \Theta((n/d)^3)$. Blow up the vertices of $G_{n'}$ into independent sets with almost-equal sizes (about n/n') to obtain an n-vertex graph with minimum degree at least d, in which every induced bipartite subgraph has minimum degree at most $O((n/n')(n')^{1/3}) = O(d^2/n)$. □

6 Concluding remarks

In this paper we have proved that for every fixed graph H, any H-free graph with minimum degree d contains an induced bipartite subgraph of minimum degree at least $\Omega(\log d/\log\log d)$. It would be very interesting to improve this to $\Omega(\log d)$, and thus fully confirm Conjecture 1.1 of Esperet, Kang and Thomassé. We note that by the methods we employed in Section 4, an improvement to $\Omega(\log d)$ when H is a triangle, would imply the same for general H.

Given a fixed graph H, let $g_H(n,d)$ be the maximum g such that every H-free graph with n vertices and minimum degree at least d contains an induced bipartite subgraph with minimum degree at least g. When $H = K_3$, we gave quite accurate estimates on $g_H(n,d)$. It would be interesting to study this function further for other forbidden subgraphs H. Some of our methods can be generalised, but the overall picture is not clear, since the behaviour of $g_H(n,d)$ seems closely related to the (hard) Ramsey problem of bounding independence number of H-free graphs. See [7, Section 6] for more explicit problems and conjectures regarding the function $g_H(n,d)$.

There is an interesting connection between the existence of induced bipartite graphs with large minimum degree and the *fractional chromatic number*. Recall that a *fractional colouring* of a graph G is an assignment of non-negative weights to the independent sets of G, in such a way that for each vertex u, the sum of weights of independent sets that contain u is at least 1. The *fractional chromatic number* of G is the minimum sum of weights of independent sets, over all possible fractional colourings. Esperet, Kang and Thomassé [16, Theorem 3.1] proved that a graph with minimum degree d and fractional chromatic number at most k has a bipartite induced subgraph of average degree at least d/k. Cames van Batenburg, de Joannis de Verclos, Kang and Pirot [7] exploited this connection in order to prove Theorem 1.3 (ii) by proving an upper bound on the fractional chromatic number of triangle-free graphs on n vertices with minimum degree d, which is tight when d is sufficiently large with respect to n. It is plausible that a similar approach could be used to obtain an alternative proof of Theorem 1.3 (i), or even to improve our bound. This raises the following question: how large can the fractional chromatic number of a d-degenerate triangle-free graph be? (One can assume that the graph is d-degenerate, as in Section 3.) More precisely, is it true that the fractional chromatic number of such a graph is at most $O(d/\log d)$? Harris [20] conjectured that the answer to the latter question is 'yes'. An affirmative answer to this question would be tight and would prove Conjecture 1.1. We note that under the stronger assumption that the graph has maximum degree d it was proved by Johansson [22] (see also Molloy [25]) that already the chromatic number is bounded by $O(d/\log d)$. On the other hand, this is no longer the case for d-degenerate graphs, as shown in [4].

Acknowledgements

We would like to thank the anonymous referees for their helpful comments and suggestions.

References

[1] N. Alon, *Explicit Ramsey graphs and orthonormal labelings*, Elec. J. Combin. **1** (1994), R12.

[2] N. Alon, *Bipartite subgraphs*, Combinatorica **16** (1996), no. 3, 301–311.

[3] N. Alon, B. Bollobás, M. Krivelevich and B. Sudakov, *Maximum cuts and judicious partitions in graphs without short cycles*, J. Combin. Theory Ser. B **88** (2003), 329–346.

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

[5] N. Alon, M. Krivelevich and B. Sudakov, *MaxCut in H-free graphs*, Comb. Prob. Comput. **14** (2005), 629–647.

[6] B. Bollobás and A. D. Scott, *Better bounds for max cut*, in Contemporary combinatorics, 185–246, Bolyai Soc. Math. Stud., 10, János Bolyai Math. Soc., Budapest, 2002.

[7] W. Cames van Batenburg, R. de Joannis de Verclos, R. J. Kang and F. Pirot, *Bipartite induced density in triangle-free graphs*, arXiv preprint arXiv:1808.02512.

[8] D. Conlon, J. Fox, M. Kwan and B. Sudakov, *Hypergraph cuts above the average*, Israel J. Math., to appear.

[9] P. Erdős, *On some extremal problems in graph theory*, Israel J. Math. **3** (1965), 113–116.

[10] P. Erdős, *Problems and results in graph theory and combinatorial analysis*, (1976), 169–192. Congressus Numerantium, No. XV.

[11] C. S. Edwards, *Some extremal properties of bipartite subgraphs*, Canad. J. Math. **3** (1973), 475–485.

[12] P. Erdős, R. Faudree, J. Pach, and J. Spencer, *How to make a graph bipartite*, J. Combin. Theory Ser. B **45** (1988), 86–98.

[13] P. Erdős, and M. Simonovits, *How many edges should be deleted to make a triangle-free graph bipartite?*, in: Sets, graphs and numbers, Colloq. Math. Soc. János Bolyai **60**, North-Holland, Amsterdam, 1992, 239–263.

[14] P. Erdős, S. Janson, T. Łuczak, and J. Spencer, *A note on triangle-free graphs*, Random discrete structures (Minneapolis, MN, 1993), IMA Vol. Math. Appl., vol. 76, Springer, New York, 1996, pp. 117–119.

[15] P. Erdős and M. Simonovits, *Some extremal problems in graph theory*, Combinatorial theory and its applications, I (Proc. Colloq., Balatonfüred, 1969), pp. 377–390, North-Holland, Amsterdam, 1970.

[16] L. Esperet, R. J. Kang and S. Thomassé, *Separation choosability and dense bipartite induced subgraphs*, Comb. Prob. Comput., doi:10.1017/S0963548319000026, arXiv preprint arXiv:1802.03727.

[17] A. Frieze A and M. Karoński, *Introduction to random graphs*, Cambridge University Press, 2016.

[18] J. Gimbel and C. Thomassen, *Coloring triangle-free graphs with fixed size*, Discrete Math. **219** (2000), no. 1-3, 275–277.

[19] H. Guo and L. Warnke, *Packing nearly optimal Ramsey $R(3,t)$ graphs*, Combinatorica, accepted.

[20] D. G. Harris, *Some results on chromatic number as a function of triangle count*, SIAM J. Discrete Math. **33** (2019), 546–563.

[21] T. Jiang and R. Seiver, *Turán numbers of subdivided graphs*, SIAM J. Discrete Math. **26** (2012), no. 3, 1238–1255.

[22] A. Johansson, *Asymptotic choice number for triangle-free graphs*, Technical Report 91-5, DIMACS, 1996.

[23] M. Krivelevich, *Bounding Ramsey numbers through large deviation inequalities*, Random Struct. Algor. **7** (1995), 145-155.

[24] M. Krivelevich and B. Sudakov, *Pseudo-random graphs*, in More sets, graphs and numbers, 199–262, Springer, Berlin, Heidelberg, 2006.

[25] M. Molloy, *The list chromatic number of graphs with small clique number*, J. Combin. Theory Ser. B, **134** (2019), 264–184.

[26] A. Nilli, *Triangle-free graphs with large chromatic numbers*, Discrete Math. **211** (2000), no. 1-3, 261–262.

[27] S. Poljak and Z. Tuza, *Bipartite subgraphs of triangle-free graphs*, SIAM J. Discrete Math. **7** (1994), no. 2, 307–313.

[28] B. Sudakov, *Making a K_4-free graph bipartite*, Combinatorica **27** (2007), 509–518.

[29] P. Turán, *On an external problem in graph theory*, Mat. Fiz. Lapok **48** (1941), 436–452.