Distinct degrees and homogeneous sets II

Eoin Long[*]   Laurențiu Ploscaru[†]

September 24, 2024

Abstract

Given an n-vertex graph G, let $\mathrm{hom}(G)$ denote the size of a largest homogeneous set in G and let $f(G)$ denote the maximal number of distinct degrees appearing in an induced subgraph of G. The relationship between these parameters has been well studied by several researchers over the last 40 years, beginning with Erdős, Faudree and Sós in the Ramsey regime when $\mathrm{hom}(G) = O(\log n)$.

Our main result here proves that any n-vertex graph G with $\mathrm{hom}(G) \leq n^{1/2}$ satisfies

$$f(G) \geq \sqrt[3]{\frac{n^2}{\mathrm{hom}(G)}} \cdot n^{-o(1)}.$$

This confirms a conjecture of the authors from a previous work, in which we addressed the $\mathrm{hom}(G) \geq n^{1/2}$ regime. Together, these provide the complete extremal relationship between these parameters (asymptotically), showing that any n-vertex graph G satisfies

$$\max\left(f(G) \cdot \mathrm{hom}(G), \sqrt{f(G)^3 \cdot \mathrm{hom}(G)} \right) \geq n^{1-o(1)}.$$

This relationship is tight (up to the $n^{-o(1)}$ term) for all possible values of $\mathrm{hom}(G)$, from $\Omega(\log n)$ to n, as demonstrated by appropriately generated Erdős–Renyi random graphs.

1   Introduction

The homogeneous number of an n-vertex graph G is defined as

$$\mathrm{hom}(G) := \max_{U \subset V(G)} \big\{ |U| : G[U] \text{ is a complete or empty graph} \big\}.$$

_____

[*]School of Mathematics, University of Birmingham, UK. Email: `e.long@bham.ac.uk`.
  The research of the first author is supported by an EPSRC New Investigator Award [EP/Y006399/1].
[†]School of Mathematics, University of Birmingham, UK. Email: `ixp090@student.bham.ac.uk`.
  The second author is grateful for support through an EPSRC DTP studentship.

This parameter is of central interest in Graph Theory and, in particular, in Ramsey Theory. One version of Ramsey's theorem [35], which is a cornerstone result in Combinatorics, states that $\hom(G) \to \infty$ as $n \to \infty$. A more quantitative bound, due to Erdős and Szekeres [17], gives that $\hom(G) \geq \log_2 n/2$. Random graphs show that this logarithmic growth is best possible, as proven by Erdős [16] in one of the earliest applications of the probabilistic method [4]. However, no explicit constructions of n-vertex graphs with $\hom(G) = O(\log n)$, also called *Ramsey graphs*, are known, despite carrying a \$100 Erdős prize. There are several compelling results proving that graphs produced via certain natural constructive procedures cannot attain this level of behaviour, along with outstanding conjectures asserting even broader restrictions.

A guiding belief in Ramsey theory is that large graphs with small homogeneous number must resemble the behaviour of random graphs and, as such, on might view the homogeneous number of a graph as a measure of quasirandomness. However, while several other similar measures are known to be equivalent (see for example [39], [10], [38]), this measurement is generally far weaker. Over time, Erdős and his coauthors made several influential conjectures [11], asserting that n-vertex Ramsey graphs resemble the typical behaviour of the random graph $G(n, 1/2)$. The vast majority of these questions have been affirmatively answered: Erdős and Szemerédi [18] proved that Ramsey graphs have density bounded away from 0 and 1, Shelah [37] showed they must have exponentially many distinct induced subgraphs, Rödl and Prömel [32] proved them to be $\Omega(\log n)$-universal, Kwan and Sudakov [25], [26] confirmed they share similar subgraph statistics with $G(n, 1/2)$, while recently Kwan, Sah, Sauermann and Sawhney [24] have shown that they have an induced subgraph of each size in the interval $[0, \Omega(n^2)]$.

The focus of this paper is on the relationship between the homogeneous number of a graph G and its distinct degree number. Given a graph G we let

$$f(G) := \max\big\{k \in \mathbb{N} : G[U] \text{ has } k \text{ vertices with distinct degrees, for some } U \subset V(G)\big\}.$$

Understanding the relationship between $\hom(G)$ and $f(G)$ was first raised in the Ramsey regime by Erdős, Faudree and Sós [15]. They conjectured that every n-vertex graph G with $\hom(G) = O(\log n)$ satisfies $f(G) = \Omega(n^{1/2})$, a lower bound which was easily observed to be true for $G(n, 1/2)$ with high probability (whp). This conjecture was proved by Bukh and Sudakov in an influential paper [9]. They also highlighted that that while $\Omega(n^{1/2})$ is a natural lower bound for $f(G(n, 1/2))$, there was no matching upper bound, and they proved instead that $f(G(n, 1/2)) = O(n^{2/3})$ whp. Conlon, Morris, Samotij and Saxton [12] later provided a matching lower bound for the random graph, showing that $f(G(n, 1/2)) = \Theta(n^{2/3})$ whp. Jenssen, Keevash, Long and Yepremyan later resolved the Ramsey setting in [22], proving that if G is an n-vertex Ramsey graph then $f(G) = \Omega(n^{2/3})$.

In their paper, Bukh and Sudakov raised the broader question of understanding $f(G)$ based on $\hom(G)$ when $\hom(G) \gg \log n$. In particular, they asked whether $\hom(G) = o(n)$ already guarantees that $f(G) \geq n^{1/2-o(1)}$. This was proven by Narayanan and Tomon [31], who showed that in fact $f(G) \geq \Omega\big((n/\hom(G))^{1/2}\big)$ for all graphs G. In the same paper, they also conjec-

tured that $f(G) \geq \Omega(n/\hom(G))$ provided $\hom(G) \geq n^{1/2}$, which is sharp, as demonstrated by the $\hom(G)$-partite Turán graph. Later, Jenssen, Keevash, Long and Yepremyan [22] proved that the Turán graph is indeed extremal for $f(G)$ when $\hom(G) \geq n^{9/10}$, improving an earlier bound of $\hom(G) \geq \Omega(n/\log n)$ given by Narayanan and Tomon.

In [28] the authors recently resolved the Narayanan-Tomon conjecture, up to a logarithmic loss, proving an essentially sharp result for $\hom(G) \geq n^{1/2}$.

Theorem 1.1. *There is an absolute constant $C > 0$ such that the following holds true. Every n-vertex graph G with $\hom(G) \geq n^{1/2}$ satisfies*

$$f(G) \geq \frac{n}{\hom(G)} \cdot \log^{-C} n.$$

Theorem 1.1 thus gives a sharp dependency between these parameters when $\hom(G) \geq n^{1/2}$. However, as $\hom(G)$ descends beyond the above $n^{1/2}$ threshold, the lower bound here begins to fail − this is illustrated by the random graph $G := G(n, 1/2)$, as $f(G) = O(n^{2/3})$ and $\hom(G) = O(\log n)$ whp. Bearing this example in mind, it is natural to investigate the behaviour of general random graphs $G(n,p)$ for $p \in [0,1]$. The authors gave a complete analysis in Section 6 of [28], which we summarise below. Noting that both $\hom(G)$ and $f(G)$ are preserved under complement, it suffices to restrict to $p \leq 1/2$. Then observe:

(i) $\hom(G(n,p)) = O(p^{-1} \log n) = p^{-1} n^{o(1)}$ whp, by a first moment argument;

(ii) $f(G(n,p)) \leq \widetilde{O}(np)$ whp, as $f(G) \leq \Delta(G) + 1$ for any graph G. For $p \leq n^{-1/2}$ it is quite easy to observe that this bound is sharp, up to a $\log^C n$ loss[1];

(iii) A more subtle upper bound on $f(G(n,p))$ is obtained as follows. Let H be an induced subgraph of $G(n,p)$ which has d vertices of distinct degrees in H; then there is a set S made of $d/4$ of these vertices and another vertex set W disjoint to S such that the event $\mathcal{E}_{S,W}$ that $\left|e_{G(n,p)}(S,W) - p|S||W|\right| \geq d^2/16$ holds. If $d \leq pn/4$, by Chernoff's inequality we obtain:

$$\mathbb{P}(\mathcal{E}_{S,W}) \leq \exp\left(-(d/4)^4 \cdot 4/d \cdot (4np)^{-1}\right) = \exp\left(-\Omega(d^3/pn)\right).$$

When $d = \Omega\left(\sqrt[3]{n^2 p}\right)$ above, we deduce that

$$\mathbb{P}\left(f(G(n,p)) \geq d\right) \leq \mathbb{P}(\text{some } \mathcal{E}_{S,W} \text{ holds}) \leq (2^n)^2 \cdot \exp\left(-\Omega(d^3/np)\right) \ll 1.$$

This gives that $f\left(G(n,p)\right) \leq \sqrt[3]{n^2 p} \cdot n^{o(1)}$ whp for $p \in [n^{-1/2}, 1/2]$.

In [28] the authors proved that the final inequality in (iii) is tight for $p \in [n^{-1/2}, 1/2]$ which, combined with (i), shows that $G := G(n,p)$ satisfies

$$f(G) = \left(n^2 p\right)^{1/3} \cdot n^{o(1)} = \sqrt[3]{\frac{n^2}{\hom(G)}} \cdot n^{o(1)} \qquad \text{whp.} \tag{1}$$

---

[1]This also follows from Theorem 1.1.

3

We also conjectured the right-hand expression in (1) provides a lower bound on $f(G)$ for all n-vertex graphs G with $\hom(G) \leq n^{1/2}$, the complementary regime to Theorem 1.1.

Our main theorem here confirms this conjecture.

Theorem 1.2. *There is an absolute constant $C > 0$ such that the following holds true. Every n-vertex graph G with $\hom(G) \leq n^{1/2}$ satisfies*

$$f(G) \geq \sqrt[3]{\frac{n^2}{\hom(G)}} \cdot e^{-C(\log n)^{2/3}}.$$

We note that by combining Theorems 1.1 and 1.2 this gives the following immediate corollary, which provides the complete extremal relationship between $\hom(G)$ and $f(G)$.

Corollary 1.3. *Given an n-vertex graph G, we have*

$$\max\left(f(G) \cdot \hom(G), \sqrt{f(G)^3 \cdot \hom(G)}\right) \geq n^{1-o(1)}.$$

We highlight that Corollary 1.3 is sharp in both regimes for appropriate random graphs. Thus the result could be interpreted as saying that random graphs determine the (asymptotic) extremal relation between $\hom(G)$ and $f(G)$. In particular, as mentioned above, the guiding belief that random graphs indicate the behaviour of graphs with small homogeneous numbers applies here throughout the entire regime, and appears to be one of the first known results of this character.

To illustrate another interesting instance of Theorem 1.2, we return to Narayanan and Tomon's result [31] that any n-vertex graph with $\hom(G) = n^{o(1)}$ satisfies $f(G) \geq n^{1/2-o(1)}$, which solved a question of Bukh and Sudakov [9]. Theorem 1.2 shows that this condition in fact produces essentially the same behaviour as was proven in [22] for Ramsey graphs.

Corollary 1.4. *Every n-vertex graph G with $\hom(G) = n^{o(1)}$ satisfies $f(G) \geq n^{2/3-o(1)}$.*

Lastly, we highlight a result of a more general form, which may be of independent interest, but which will also be helpful when discussing our proof approach. Two vertices u, v of a graph G are said to have *diversity* D to a set $S \subset V(G)$ if $|N^S(u) \triangle N^S(v)| \geq D$. A subset $U \subset V(G)$ is said to be *D-diverse* to S if $|N^S(u_1) \triangle N^S(u_2)| \geq D$ for all distinct $u_1 \neq u_2 \in U$.

In [9], diversity was identified as a key parameter in analysing $f(G)$, and this connection has had a large impact on the study of several properties of Ramsey graphs (e.g. degree and size distributions of induced subgraphs [26],[31],[30],[22]). In particular, in [9] Bukh and Sudakov proved that if $|U| = k$ and $D \geq k^2$ above, then $f(G) = \Omega(k)$. Using results from [28], one can show that $D \geq k^{3/2}$ suffices to give this conclusion, which turns out to be essentially sharp.

4

It is still natural however to suspect this result could be further improved if there was less *pressure* on the vertices of S here. For instance, at an extreme, if each vertex in S only appears in one $N^S(u)$ with $u \in U$ then it is easy to see that $D \geq k$ already implies $f(G) = \Omega(k)$. Our next theorem provides a general bound, depending on the pressure vertices from U to S.

Theorem 1.5. *Let $D \geq 1$ and $\gamma \in (0,1]$. Suppose G is a graph with $U, S \subset V(G)$, such that:*

(i) $|N^S(u_1) \triangle N^S(u_2)| \geq D$ for all distinct $u_1, u_2 \in U$;

(ii) $|N^U(s)| \leq \gamma|U|$ for all $s \in S$.

Then the following relation holds:

$$f(G) = \Omega\left(\frac{|U|}{\left(1 + \sqrt{\gamma|U|^3/D^2}\right) \cdot \log^2 |U|}\right).$$

Note in particular that Theorem 1.5 gives $f(G) = |U|^{1-o(1)}$ if either:

- $D = \Omega\left(|U|^{3/2}\right)$ and $\gamma = 1$ (trivially satisfied), or
- $D = \Omega(|U|)$ and $\gamma = |U|^{-1}$ – see the example discussed above.

It is also easy to verify that whp the hypothesis of Theorem 1.5 holds for the random graph $G(n,p)$ when $p \in [n^{-1/2}, 1/2]$, with $|U| = c(n^2p)^{1/3}$, $D = np/4$, $S = V(G)$ and $\gamma = 2p$. Therefore, we recover the essentially sharp bound $f(G(n,p)) \geq |U|^{1-o(1)} = (n^2p)^{1/3-o(1)}$ whp.

Proof overview. Before closing the introduction, we briefly discuss the proof of Theorem 1.2. In [28] we introduced a general approach to establishing lower bounds on $f(G)$ for a graph G, which we used to prove Theorem 1.1. Roughly speaking, instead of finding sets $U \subset W \subset V(G)$ such that U has many distinct degrees in $G[W]$, it suffices instead to find such a set U and a probability distribution $\mathcal{D}$ on $[0.1, 0.9]^{V(G)}$ such that certain small-ball like quantities associated with U and $\mathcal{D}$ are well-controlled (see Section 3 for more details). This resulted in a more technical task, but with the benefit that it allowed us to blend behaviour from different neighbourhoods of the graph together to find distinct degrees, and that it was more robust to an inductive approach, which seems forced upon us in this type of problem.

Our proof of Theorem 1.2 follows the same approach, although several substantial new ideas are required in the current regime. To begin, observe that if we could find sets U and S with $|U|$ large, which have diverse neighbourhoods and such that the pressure from U to S is small, then one could use Theorem 1.5 to prove the theorem, as illustrated in the remark for $G(n,p)$ above. We thus restrict our attention to graphs G in which one of the conditions (i) or (ii) from Theorem 1.5 fails for such sets.

Should (i) fail then there are vertices in G whose neighbourhoods cluster or significantly correlate. We investigate this correlation using the notion of 'cluster neighbourhood', which

allows us to separate vertices which are strongly correlated from others which have a smaller correlation. Under natural assumptions, we obtain a partial decomposition of a significant proportion of the graph using cluster neighbourhoods (see Lemma 4.7). We might now hope to find distinct degrees within each of these clusters by induction (as in [28]), and then combine them together to obtain U using the weak correlation between the clusters in place of (i) in Theorem 1.5.

Although this sounds plausible, it results in a new difficulty. Indeed, observe that this partition into clusters naturally complicates achieving (ii) from Theorem 1.5. Indeed, the 'pressure parameter' γ is necessarily forced to be much larger if U has many vertices of a similar structure. To bypass this new restriction, we instead build our probability distributions $\mathcal{D}$ based on the coarse structure of the cluster neighbourhoods, rather through vertex neighbourhoods which were used previously. This enables us to work with a measure of pressure based on the number of *clusters*, rather than the sought number distinct degrees (as in Theorem 1.5). We make this connection by exploiting the metric structure of the neighbourhoods (Lemma 3.5), which will in turn be controlled through our definition of clusters.

Lastly, with these tools in hand, we prove Theorem 1.2 via a delicate inductive argument. In particular, we were only able to guarantee a low enough level of pressure for the cluster partition in certain circumstances (see Case 3 of the proof), but luckily a trivial level of pressure ($\gamma = 1$) turned out to be sufficient outside of this (see Case 2 of the proof).

Structure of the paper. In the next section we collect some classic tools from Graph Theory and Probabilistic Combinatorics, which will be useful in proving our theorem. In the beginning of Section 3 we recap the key elements of our approach from [28] to lower bounding $f(G)$. We furthermore provide a crucial lemma that allows us to find distributions which separate large parts of the graph with relatively low overall pressure. Next, in Section 4 we introduce cluster neighbourhoods and study their properties. In particular, we prove a central result which helps us find a partial decomposition of a graph into highly correlated pieces. Finally, in Section 5 we combine these components together to prove Theorem 1.2.

Notation. Given a graph G and $u, v \in V(G)$, we write $u \sim v$ if u and v are adjacent vertices in G and $u \nsim v$ if they are not. The neighbourhood of u is given by $N_G(u) = \{v \in V(G) : u \sim v\}$ and given $S \subset V(G)$ we let $N_G^S(u) := N_G(u) \cap S$; we will omit the subscript G when the graph is clear from the context. We write $d_G^S(u) = |N_G^S(u)|$. The maximum and the minimum degree of G, denoted by $\Delta(G)$ and $\delta(G)$ respectively, are the largest and the smallest possible degree of a vertex of G. The average degree of G, denoted usually by $\bar{d}(G)$, is simply the average of all vertex degrees of G, i.e. $\bar{d}(G) := |V(G)|^{-1} \cdot \sum_{v \in V(G)} d_G(v)$.

Given a vertex u, we will also represent the neighbourhood of u by a vector $\mathbf{u} \in \{0,1\}^{V(G)}$ defined such that $\mathbf{u}_v = 1$ if and only if $u \sim v$. Given a set $U \subset V$ and a vector $\mathbf{u} \in \mathbb{R}^V$, we will denote the projection of $\mathbf{u}$ onto the coordinate set S by $\text{proj}_S(\mathbf{u})$, i.e. for any $v \in S$

we have $\text{proj}_S(\mathbf{u})_v = \mathbf{u}_v$. Given $u, v \in V(G)$ we write $\text{div}_G(u, v)$ for the symmetric difference $N(u) \triangle N(v)$. Thus $|\text{div}_G(u, v)|$ is simply the Hamming distance between $\mathbf{u}$ and $\mathbf{v}$.

We will write $\overline{G}$ for the complement of the graph G. It is easy to note that for any graph G we have $\text{hom}(G) = \text{hom}(\overline{G})$ and $f(G) = f(\overline{G})$ since $\text{div}_G(u, v) = \text{div}_{\overline{G}}(u, v)$ for any $u, v \in V(G)$.

Given $n \in \mathbb{N}$ and $p \in (0, 1)$, the Erdős−Rényi random graph $G(n, p)$ is the n-vertex graph in which each edge is included in the graph with probability p independently of every other edge. We say that an event that depends on n occurs *with high probability* (whp) if its probability tends to 1 as $n \to \infty$.

Throughout this paper we will omit floor and ceiling signs when they are not crucial, for the sake of clarity of presentation.

2 Tools

In this short section we introduce some tools required for the rest of the paper.

2.1 Graph theory tools

We will need the following version of Turán's theorem (see for example Chapter 6 in [6]).

Theorem 2.1. *Let G be a n-vertex graph with average degree d. Then G has an independent set of size at least $n/(d+1)$.*

Besides this, the following result (Lemma 5.7 from [28]) will be quite advantageous, as it will enable us to move to a large induced subgraph that is reasonably regular. Comparable results were proved by Alon, Krivelevich and Sudakov in [3].

Lemma 2.2. *Given a graph G on n vertices there is $A \subset V(G)$ with $|A| \geq n/30 \log n$ such that the induced subgraph $H = G[A]$ satisfies $\Delta(H) \leq 5 \log n \cdot \delta(H)$.*

2.2 Probabilistic Tools

We will require some classic concentration inequalities. See e.g. appendix A in [4].

Theorem 2.3 (Chernoff Inequality). *Let X be a random variable with binomial distribution and let $\mu = \mathbb{E}[X]$. Then, for $0 \leq \delta \leq 1$, the following inequalities hold:*

$$\mathbb{P}\big(X \leq (1 - \delta)\mu\big) \leq \exp\left(-\frac{\delta^2 \mu}{2}\right).$$

$$\mathbb{P}\big(X \geq (1+\delta)\mu\big) \leq \exp\left(-\frac{\delta^2\mu}{4}\right).$$

Theorem 2.4 (**Hoeffding's Inequality**)**.** *Let* $X_1, X_2, \ldots, X_n$ *be independent random variables such that* $a_i \leq X_i \leq b_i$ *for each* $i \in [n]$*, where* $a_i, b_i \in \mathbb{R}$*. Then given* $t > 0$*, the random variable* $S_n = X_1 + \cdots + X_n$ *satisfies:*

$$\mathbb{P}\left(|S_n - \mathbb{E}[S_n]| \geq t\right) \leq 2 \cdot \exp\left(\frac{-2t^2}{\sum_{i \in [n]}(b_i - a_i)^2}\right).$$

The next result will be very useful when dealing with large deviations.

Theorem 2.5. *Let* $n \in \mathbb{N}$*,* $p \in [0,1]$*,* $L > 0$ *an let* $X \sim Bin(n,p)$ *be a random variable. Then:*

$$\mathbb{P}(X \geq L) \leq \binom{n}{L}p^L \leq \left(\frac{enp}{L}\right)^L.$$

We end this section with a very useful lower bound − due to Greenberg and Mohri [21] − for the probability that a binomial random variable exceeds is mean.

Theorem 2.6. *Let* $n \in \mathbb{N}$*,* $p \in (n^{-1}, 1)$ *and let* $X \sim Bin(n,p)$ *be a random variable. Then:*

$$\mathbb{P}(X \geq np) > \frac{1}{4}.$$

3 Distinct degrees through probability distributions

As mentioned in the Introduction, in [28] we introduced a general framework which provided lower bounds on $f(G)$ for a graph G. Roughly speaking, in using this approach, one aims to find a set $U \subset V(G)$ together with a probability distribution $\mathcal{D}$ on $[0,1]^{V(G)}$ with the property that certain small-ball type quantities associated to U are well controlled. We will summarise what we require here, but a more extensive presentation can be found in [28], where the authors introduced this set up. In particular, see the conclusion of the introduction of [28] for some of the motivation for this approach along with Section 3 for further details and proofs.

Given a graph G and a probability vector $\underline{\mathbf{p}} = (p_v)_{v \in V(G)} \in [0.1, 0.9]^{V(G)}$ we will write $G(\underline{\mathbf{p}})$ to denote the probability space on the set of induced subgraphs of G, determined by including each vertex $v \in V(G)$ independently with probability p_v. Equivalently, given $S \subset V(G)$, the induced subgraph $G[S]$ is selected with probability $\prod_{v \in S} p_v \prod_{v \in V(G) \setminus S}(1 - p_v)$. Abusing notation slightly[2], we will usually write $G(\underline{\mathbf{p}})$ to denote a random graph $G[S] \sim G(\underline{\mathbf{p}})$.

---

[2]As with the Erdős–Renyi random graph $G(n,p)$.

Throughout the paper, given a vertex $u \in V(G)$, we will we write $\mathbf{u} \in \{0,1\}^{V(G)}$ to denote the *neighbourhood vector* of u, which is given by

$$(\mathbf{u})_v = \begin{cases} 1 & \text{if } uv \in E(G); \\ 0 & \text{otherwise.} \end{cases}$$

Note that, considering the standard inner product on $\mathbb{R}^{V(G)}$, given by $\mathbf{x} \cdot \mathbf{y} = \sum_{v \in V(G)} x_v y_v$, this notation leads us to the following useful representation:

$$\mathbb{E}\big[d_{G(\mathbf{p})}(u)\big] = \mathbf{u} \cdot \underline{\mathbf{p}}. \tag{2}$$

Expected degrees in suitable $G(\underline{\mathbf{p}})$'s prove to be useful in finding bounds for $f(G)$, however this notion is still too rigid for our purposes. The parameter we are about to define turns out to be robust enough.

Let G be a graph and let $\mathcal{D}$ be a probability distribution on $[0.1, 0.9]^{V(G)}$. Given distinct vertices $u, v \in V(G)$ and a set $S \subset V(G)$, we define

$$\mathrm{bad}_{\mathcal{D}}^S(u,v) := \max_{c \in \mathbb{R}} \ \mathbb{P}_{\underline{\mathbf{p}} \sim \mathcal{D}}\big(|\mathbb{E}[d_{G(\underline{\mathbf{p}})}^S(u)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}^S(v)] - c| \leq 1\big). \tag{3}$$

This quantity can be viewed as a small ball probability $-$ a measurement for two vertices $u, v \in V(G)$ of how likely the expected degrees to S in $G(\underline{\mathbf{p}})$ are to differ by an (almost) fixed amount. Given sets $U, S \subset V(G)$, we also set

$$\mathrm{bad}_{\mathcal{D}}^S(U) := \sum_{\{u,v\} \subset U} \mathrm{bad}_{\mathcal{D}}^S(u,v).$$

Given another set $V \subset V(G)$ we can also write

$$\mathrm{bad}_{\mathcal{D}}^S(U,V) := \sum_{(u,v) \in U \times V} \mathrm{bad}_{\mathcal{D}}^S(u,v).$$

We will sometimes suppress the superscript when $S = V(G)$, e.g. $\mathrm{bad}_{\mathcal{D}}(U) = \mathrm{bad}_{\mathcal{D}}^{V(G)}(U)$. Lastly, let us remark that in (3) we do not need $\mathcal{D}$ to be defined on all vertex coordinates of the set $[0.1, 0.9]^{V(G)}$; any vertex set T with $S \subseteq T \subseteq V(G)$ is enough so that we can define $\mathcal{D}$ on $[0.1, 0.9]^T$, as we can see by looking at the RHS of (3).

As highlighted in [28], in order to find many distinct degrees in a graph G it suffices to find a large set $U \subset V(G)$ and a probability distribution $\mathcal{D}$ such that $\mathrm{bad}_{\mathcal{D}}(U)$ is small.

Theorem 3.1. *Let G be a graph, let $\mathcal{D}$ be a probability distribution on $[0.1, 0.9]^{V(G)}$ and let $U \subset V(G)$ with $\mathrm{bad}_{\mathcal{D}}(U) = \alpha \cdot |U|$. Then*

$$f(G) = \Omega\left(\frac{|U|}{(1+\alpha) \cdot \log^{3/2} |U|}\right).$$

9

Thus producing a set U and a distribution $\mathcal{D}$ with strong 'bad' control results in a lower bound for $f(G)$. We also proved in [28] that one can reverse this process, to obtain a distribution with strong bad control from a lower bound on $f(G)$ giving that the two approaches are comparable, although we will not need this reverse direction here.

For the remainder of this section we will collect a number of results together, which will be used in combination to exhibit distributions $\mathcal{D}$ with bad control. As indicated earlier, the key benefit of working in this generalised setting is the flexibility of combining distributions on different sets while maintaining bad control. This is illustrated by the following result.

Lemma 3.2. *Let G be a graph with vertex partition $V(G) = \bigsqcup_{i=1}^{L} V_i$ and for each $i \in [L]$ let $\mathcal{D}_i$ be a probability distribution on $[0,1]^{V_i}$. Then taking $\mathcal{D}$ to denote the product distribution $\Pi_{i \in [L]} \mathcal{D}_i$ on $[0,1]^{V(G)}$, for any distinct vertices $u, v \in V(G)$ and any set $S \subset V(G)$, one has:*

$$\text{bad}_{\mathcal{D}}^{S}(u,v) \leq \min_{i \in [L]} \text{bad}_{\mathcal{D}_i}^{S \cap V_i}(u,v).$$

Proof. See Lemma 4.1 in [28]. $\qquad \square$

Our second lemma gives a simple situation in which we can obtain 'bad' control. Let G be a graph and let $S \subset V(G)$. Let $\mathcal{U}_S$ denote the **uniformly constant distribution** on $[0.1, 0.9]^S$, given by selecting $\alpha \in [0.1, 0.9]$ uniformly at random and setting $\underline{\mathbf{p}} = \alpha \mathbf{1}_S \in [0.1, 0.9]^S$.

It is also convenient to write $\mathcal{T}_S$ to denote the **trivial S-induced probability distribution**, the distribution on $[0.1, 0.9]^S$ which simply selects the vector $\underline{\mathbf{p}}_0 = \frac{1}{2} \cdot \mathbf{1}_S$ with probability 1.

Lemma 3.3. *Let G be a graph, $S \subset V(G)$ and $u, v \in V(G)$ such that $d^S(u) \geq d^S(v) + D$ for some $D > 0$. Suppose that $\mathcal{U}_S$ denotes the uniform constant distribution on $[0.1, 0.9]^S$, that $\mathcal{D}'$ denotes a distribution on $[0.1, 0.9]^{V(G) \setminus S}$ and that $\mathcal{D}$ denotes the product distribution $\mathcal{U}_S \times \mathcal{D}'$ on $[0.1, 0.9]^{V(G)}$. Then $\text{bad}_{\mathcal{D}}(u,v) \leq 3D^{-1}$.*

Proof. See Lemma 4.2 from [28]. $\qquad \square$

We next seek to provide 'bad' control for a set by blending neighbourhood structures together − the idea here has some similarities to that of [22]. Let G be a graph, let $U, S \subset V(G)$, where $U := \{u_1, \ldots, u_k\}$, and let $\beta \in [0, 0.4]$. We now let $\mathcal{B}_\beta(U, S)$ denote the **blended probability distribution** on $[0.1, 0.9]^S$, which is defined as follows. First independently select $\alpha_i \in [-\beta, \beta]$ uniformly at random for each $i \in [k]$ and set:

$$\underline{\mathbf{p}}' := \frac{1}{2} \cdot \mathbf{1} + \sum_{i \in [k]} \alpha_i \cdot \text{proj}_S(\mathbf{u}_i) \in \mathbb{R}^S. \tag{4}$$

Having made these choices, the distribution then returns $\underline{\mathbf{p}}$, a truncated version of $\underline{\mathbf{p}}'$, where

$$\underline{\mathbf{p}}_v = \begin{cases} \underline{\mathbf{p}}'_v & \text{if } \underline{\mathbf{p}}'_v \in [0.1, 0.9]; \\ 0.9 & \text{if } \underline{\mathbf{p}}'_v > 0.9; \\ 0.1 & \text{if } \underline{\mathbf{p}}'_v < 0.1. \end{cases}$$

Our next lemma gives a useful extension of Lemma 4.3 from [28] which will allow us to obtain strong control on $\mathrm{bad}_{\mathcal{D}}(v_i, v_j)$. In particular, it is designed to be used in conjunction with the partial decomposition from cluster neighbourhoods given in Section 4 (see Lemma 4.7).

Definition 3.4. *Given a graph G, vertex sets $U, S \subset V(G)$ and a parameter $\gamma \in [0, 1]$, we say U is γ-**balanced** to S if for all $v \in S$ we have $d_G^U(v) \leq \gamma |U|$.*

Observe that U is always 1-*balanced to S.*

Lemma 3.5. *Let G be a graph, $\beta \in (0, 0.1)$, $\gamma \in (0, 1]$, $D \geq 1$ and let $U := \{u_1, u_2, \dots, u_t\}$ and S be subsets of $V(G)$ such that U is γ-balanced to S. Suppose there are $d_1, d_2, \dots, d_t > 0$ and pairwise disjoint sets $V_1, V_2, \dots, V_t \subset V(G)$ with $u_i \in V_i$ for all $i \in [t]$ such that:*

- *$|\mathrm{div}_G^S(u_i, v_i)| \leq d_i$ for each $v_i \in V_i$ and $i \in [t]$;*

- *$|\mathrm{div}_G^S(u_i, u_j)| \geq D + d_i + d_j$ for all distinct $i, j \in [t]$.*

Suppose further that $\mathcal{D}$ denote the product distribution $\mathcal{D} = \mathcal{B}_\beta(U, S) \times \mathcal{D}'$ on $[0.1, 0.9]^{V(G)}$ where $\mathcal{B}_\beta(U, S)$ denotes the blended probability distribution on $[0.1, 0.9]^S$ and $\mathcal{D}'$ is any distribution on $[0.1, 0.9]^{V(G) \setminus S}$. Then for all $v_i \in V_i$ and $v_j \in V_j$ with $i \neq j$ in $[t]$ one has

$$\mathrm{bad}_{\mathcal{D}}(v_i, v_j) \leq \frac{2}{\beta D} + 2 \cdot \max\left\{ d_G^S(u_i), d_G^S(u_j) \right\} \cdot \exp\left(\frac{-0.045}{\gamma \beta^2 |U|} \right). \tag{5}$$

Proof. For each $i \in [t]$, given the vector $\underline{\mathbf{p}}'$ on $\mathbb{R}^S$ from (4), define the random vector $\underline{\mathbf{q}}^i$ on $\mathbb{R}^S$ by $\underline{\mathbf{q}}^i := \underline{\mathbf{p}}' - \alpha_i \cdot \mathrm{proj}_S(\mathbf{u}_i)$. A key observation is that $\underline{\mathbf{q}}^i$ is independent of α_i. We will slightly abuse notation by writing $\underline{\mathbf{p}}$ for both a vector in $[0.1, 0.9]^{V(G)}$ and its projection $\mathrm{proj}_S(\underline{\mathbf{p}})$ onto the coordinate set S. Since $\mathcal{D}$ is the product distribution $\mathcal{B}_\beta(R, S) \times \mathcal{D}'$, we can do this without much concern due to Lemma 3.2.

Fix $i, j \in [t]$ with $i \neq j$ and $v_i \in V_i$, $v_j \in V_j$. Given $c \in \mathbb{R}$, let $E_{v_i, v_j}^{i,j}(c)$ denote the event that $\left| \mathbb{E}[d_{G(\underline{\mathbf{p}})}(v_i)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(v_j)] - c \right| \leq 1$. According to (3), it will suffice to show that:

$$\mathbb{P}\big(E_{v_i, v_j}^{i,j}(c)\big) \leq \frac{2}{\beta D} + 2 \max\{ d_G^S(u_i), d_G^S(u_j) \} \cdot \exp\left(\frac{-0.045}{\gamma \beta^2 |U|} \right).$$

Let us assume that $d_G^S(v_i) \geq d_G^S(v_j)$. Call a vertex $v \in N_G^S(u_i)$ *naughty* if $\underline{\mathbf{q}}_v^i \notin [0.2, 0.8]$. We call the vertex u_i *problematic* if $N_G^S(u_i)$ contains a naughty vertex and we denote this event by F_i. By the law of total probability we get that

$$\mathbb{P}\big(E_{v_i, v_j}^{i,j}(c)\big) = \mathbb{P}\big(E_{v_i, v_j}^{i,j}(c) | F_i\big) \cdot \mathbb{P}(F_i) + \mathbb{P}\big(E_{v_i, v_j}^{i,j}(c) | \overline{F_i}\big) \cdot \mathbb{P}(\overline{F_i}) \leq \mathbb{P}(F_i) + \mathbb{P}\big(E_{v_i, v_j}^{i,j}(c) | \overline{F_i}\big). \tag{6}$$

To upper bound $\mathbb{P}(F_i)$, let $v \in S$ and note that $\underline{\mathbf{q}}_v^i$ is a sum of $d_G^{U \setminus \{u_i\}}(v)$ uniform independent random variables, as the v-coordinate of $\mathbf{u}_i$ is non-zero when $v \sim u_i$. Therefore, we have

$$\mathbb{P}\big(\underline{\mathbf{q}}_v^i \notin [0.2, 0.8]\big) = \mathbb{P}\big(|\underline{\mathbf{q}}_v^i - 1/2| > 0.3\big) \leq 2 \exp\left(\frac{-2 \cdot 0.09}{4\beta^2 d_G^{U \setminus \{u_i\}}(v)}\right) \leq 2 \exp\left(\frac{-2 \cdot 0.09}{4\beta^2 \gamma |U|}\right),$$

where the first inequality follows from Hoeffding's inequality, whereas the second one uses that $d_G^{U \setminus \{u_i\}}(v) \leq d_G^U(v) \leq \gamma |U|$ as U is γ-balanced to S. By the union bound we then get

$$\mathbb{P}(F_i) \leq d_G^S(u_i) \cdot \mathbb{P}\big(\underline{\mathbf{q}}_v^i \notin [0.2, 0.8]\big) \leq 2 d_G^S(u_i) \cdot \exp\big(-0.045(\gamma \beta^2 |U|)^{-1}\big). \tag{7}$$

To compute $\mathbb{P}\big(E_{v_i, v_j}^{i,j}(c) | \overline{F_i}\big)$ we condition on any choice of $\boldsymbol{\alpha} := (\alpha_l)_{l \neq i}$ such that F_i does not hold. Given such a choice, let us first note that $\underline{\mathbf{p}}_v' = \underline{\mathbf{q}}_v^i + \alpha_i \mathbf{u}_{i_v} \in [0.1, 0.9]$ for all $v \in N_G^S(u_i)$ since $|\alpha_i| < 0.1$. So none of the $N_G^S(u_i)$-coordinates of $\underline{\mathbf{p}}'$ will get truncated and recall that α_i is independent of F_i. Given a choice of $\boldsymbol{\alpha}$, consider now the following expression as a map of α_i:

$$f_c(\alpha_i) := \mathbb{E}[d_{G(\underline{\mathbf{p}})}^S(v_i)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}^S(v_j)] - c = (\mathbf{v}_i - \mathbf{v}_j) \cdot \underline{\mathbf{p}} - c. \tag{8}$$

Having conditioned on $\boldsymbol{\alpha}$ above, note that the event $E_{v_i, v_j}^{i,j}(c)$ holds only if $f_c(\alpha_i)$ lies in an interval of length 2. To bound the probability of this happening we have to understand how f_c changes as α_i increases. Observe the contribution from each coordinate v of $\underline{\mathbf{p}}$ to the inner product on the right hand side of (8) is 0 if $v \notin N_G^S(u_i)$ and exactly α_i otherwise, since none of these coordinates were truncated from $\underline{\mathbf{p}}'$ when conditioning on $\overline{F_i}$. It follows that for $\varepsilon > 0$:

$$f_c(\alpha_i + \varepsilon) - f_c(\alpha_i) = \varepsilon \sum_{v \in S} \big((\mathbf{v}_i)_v - (\mathbf{v}_j)_v\big) \cdot \mathbb{1}_{v \sim u_i} = \varepsilon\big(|N_G^S(u_i) \cap N_G^S(v_i)| - |N_G^S(u_i) \cap N_G^S(v_j)|\big).$$

However, it is not hard to see that

$$2|N_G^S(u_i) \cap N_G^S(v_i)| = d_G^S(u_i) + d_G^S(v_i) - \text{div}_G^S(u_i, v_i),$$
$$2|N_G^S(u_i) \cap N_G^S(v_j)| = d_G^S(u_i) + d_G^S(v_j) - \text{div}_G^S(u_i, v_j).$$

It follows that

$$2|N_G^S(u_i) \cap N_G^S(v_i)| - 2|N_G^S(u_i) \cap N_G^S(v_j)| = d_G^S(v_i) - d_G^S(v_j) - \text{div}_G^S(u_i, v_i) + \text{div}_G^S(u_i, v_j)$$
$$\geq \text{div}_G^S(u_i, v_j) - \text{div}_G^S(u_i, v_i)$$
$$\geq \text{div}_G^S(u_i, u_j) - \text{div}_G^S(u_j, v_j) - \text{div}_G^S(u_i, v_i) \geq D,$$

by using the hypothesis, that $d_G^S(v_i) \geq d_G^S(v_j)$ and triangle's inequality. From here we can then deduce that $f(\alpha_i + \varepsilon) - f(\alpha_i) \geq \varepsilon D/2$.

Therefore, conditioned on $\boldsymbol{\alpha}$ as above, if $E_{v_i, v_j}^{i,j}(c)$ occurs then α_i lies in an interval of length at most $4/D$. This implies that $\mathbb{P}(E_{v_i, v_j}^{i,j}(c) | \overline{F_i}) \leq 2\beta^{-1} D^{-1}$. Combined with (6) and (7), this proves the lemma. $\qquad \square$

Let us note that by taking $E_i = 0$ and $W_i = \{u_i\}$ for each $i \in [|U|]$ in the previous lemma we recover the following result, which is similar to Lemma 4.3 from [28].

Corollary 3.6. *Let G be a graph, $D \geq 1$, $\beta \in (0, 0.1)$, $\gamma \in (0, 1]$ and $U, S \subset V(G)$ such that U is γ-balanced to S and $|div_G^S(u, v)| \geq D$ for all $u \neq v$ in U. Suppose that $\mathcal{D}$ is the product distribution $\mathcal{B}_\beta(U, S) \times \mathcal{D}'$ on $[0.1, 0.9]^{V(G)}$ where $\mathcal{B}_\beta(U, S)$ is the blended probability distribution on $[0.1, 0.9]^S$ and $\mathcal{D}'$ is any distribution on $[0.1, 0.9]^{V(G) \setminus S}$. Then for all $u, v \in U$ one has:*

$$\mathrm{bad}_\mathcal{D}(u, v) \leq \frac{2}{\beta D} + \max\{d_G^S(u), d_G^S(v)\} \cdot 2 \exp\left(\frac{-0.045}{\gamma \beta^2 |U|}\right). \tag{9}$$

As a quick application of the previous corollary, we obtain a proof of Theorem 1.5.

Proof of Theorem 1.5. To begin, we note some convenient assumptions between the parameters. Observe first that we may assume that $D \leq |U|^{3/2}$. Indeed the conclusion for larger values just asserts that $f(G) \geq \widetilde{\Omega}(|U|)$, which follows from the case $D = |U|^{3/2}$. We can also assume that $|S| \leq D|U|^2$ since for larger values of $|S|$ it is easily seen that if properties (i) and (ii) from the Theorem hold, then they also hold for a subset $S' \subset S$ with this bound. Thus, working with S' in place of S, we can assume this upper bound. Lastly, note that we can assume that $\gamma \geq 1/|U|$, as otherwise conditions (i) and (ii) cannot be satisfied.

Set $\beta^{-1} := 10\sqrt{\gamma |U| \log |U|}$ and apply Corollary 3.6. This gives

$$\mathrm{bad}_\mathcal{D}(u, v) \leq \frac{20\sqrt{\gamma |U| \log |U|}}{D} + |S| \cdot 2 \exp(-4.5 \log |U|) \leq \frac{40\sqrt{\gamma |U| \log |U|}}{D},$$

where the final inequality uses that $\gamma |U| \geq 1$ and that $D|S| \leq |U|^{3.5}$. Summing over all pairs of vertices in U, this gives $\mathrm{bad}_\mathcal{D}(U) \leq \alpha |U|$ where $\alpha := 20D^{-1}\sqrt{\gamma |U| \log |U|}$. An application of Theorem 3.1 then completes the proof. □

The last lemma of this section provides a hypothesis to maintain 'bad' control when combining multiple disjoint sets with 'bad' control under different well-behaved distributions.

Lemma 3.7. *Let $1 \leq m_0 \leq M \leq 2M \leq M_0$ and let $f : [m_0, M_0] \to [0, \infty)$ be a differentiable map such that f is increasing, while f' is decreasing on $[m_0, M_0]$. Suppose G is a graph which contains disjoint vertex sets $V_1, V_2, \ldots, V_t, S$ satisfying the following properties:*

(i) *There are sets $U_i \subset V_i$ such that $\sum_{i \in [t]} |U_i| \geq M$ with $\max_{i \in [t]} |U_i| \geq m_0$;*

(ii) *For each $i \in [t]$ there is a distribution $\mathcal{D}_i$ on $[0.1, 0.9]^{V_i}$ such that $\mathrm{bad}_{\mathcal{D}_i}^{V_i}(U_i) \leq |U_i| f(|U_i|)$;*

(iii) *There is a distribution $\mathcal{E}$ on $[0.1, 0.9]^S$ such that $\mathrm{bad}_\mathcal{E}^S(u_i, u_j) \leq f'(M_0)$ for all $u_i \in U_i$ and $u_j \in U_j$ with $i \neq j$.*

Then there is a set $U \subset V(G)$ with $|U| \geq M$ and a distribution $\mathcal{D}$ on $[0.1, 0.9]^{V(G)}$ such that

$$\mathrm{bad}_{\mathcal{D}}(U) \leq |U| f(|U|).$$

Proof. We first note that if $a \in [m_0, M]$ and $0 \leq b \leq a$ then $m_0 \leq a + b \leq 2M \leq M_0$ and

$$a \cdot f(a) + b \cdot f(b) + ab \cdot f'(M_0) \leq (a + b) \cdot f(a + b). \tag{10}$$

Indeed, by the mean value theorem there is $\xi \in [a, a + b] \subset [m_0, M_0]$ with the property that $f(a + b) = f(a) + bf'(\xi)$. As f' is decreasing on $[m_0, M_0]$, we have $f'(M_0) \leq f'(\xi)$ and so

$$\begin{aligned}
af(a) + bf(b) + ab \cdot f'(M_0) &= a\big(f(a) + b \cdot f'(M_0)\big) + bf(b) \\
&\leq a\big(f(a) + b \cdot f'(\xi)\big) + bf(b) \\
&\leq af(a + b) + bf(b) \leq (a + b) \cdot f(a + b),
\end{aligned}$$

where the final inequality uses that f is increasing and $b \leq a + b$.

We are now able to prove the lemma. Order the sets so that $|U_1| \geq |U_2| \geq \ldots \geq |U_t|$. Note that if $|U_1| > M$ then the conclusion is trivial taking $U = U_1$ and $\mathcal{D} = \mathcal{D}_1 \times \mathcal{T}$, where $\mathcal{T}$ is the trivial distribution on $[0.1, 0.9]^{V(G) \backslash U_1}$. Thus we may assume that $|U_1| \leq M$. Setting $U_{<j} := \bigcup_{i<j} U_i$ for all $j \in [2, t+1]$, by discarding sets $\{U_i\}_{i>t'}$ for some $t' \leq t$ we can additionally assume that $|U_{<j}| < M$ for all $j \leq t$ and that $U := U_{<t+1}$ satisfies $M \leq |U| \leq 2M < M_0$. Observing that $m_0 \leq |U_1| \leq |U_{<j}| \leq M$ for $j \in [2, t]$, we can apply (10) to obtain

$$|U_{<j}| \cdot f(|U_{<j}|) + |U_j| \cdot f(|U_j|) + |U_{<j}||U_j| \cdot f'(M_0) \leq |U_{<j+1}| \cdot f(|U_{<j+1}|). \tag{11}$$

Let $X := V(G) \backslash \big(S \cup \bigcup_{i=1}^{t} V_i\big)$ and take $\mathcal{D}$ to be the product distribution $\mathcal{D} := \prod_{i=1}^{t} \mathcal{D}_i \times \mathcal{E} \times \mathcal{T}$ on $\prod_{i=1}^{t} [0.1, 0.9]^{V_i} \times [0.1, 0.9]^{S} \times [0.1, 0.9]^{X} = [0.1, 0.9]^{V(G)}$, where $\mathcal{T}$ is the trivial X-induced distribution. By Lemma 3.2 and (i)-(iii) from the hypothesis we obtain

$$\begin{aligned}
\mathrm{bad}_{\mathcal{D}}(U) &= \sum_{i \in [t]} \mathrm{bad}_{\mathcal{D}}(U_i) + \sum_{\{i,j\} \subset [t]} \mathrm{bad}_{\mathcal{D}}(U_i, U_j) \\
&\leq \sum_{i \in [t]} \mathrm{bad}_{\mathcal{D}_i}^{V_i}(U_i) + \sum_{\{i,j\} \subset [t]} \mathrm{bad}_{\mathcal{E}}^{S}(U_i, U_j) \\
&\leq \sum_{i \in [t]} \mathrm{bad}_{\mathcal{D}_i}^{V_i}(U_i) + \sum_{\{i,j\} \subset [t]} |U_i| \cdot |U_j| \cdot \max_{(v_i, v_j) \in U_i \times U_j} \big\{ \mathrm{bad}_{\mathcal{B}}^{S}(v_i, v_j) \big\} \\
&\leq \sum_{i \in [t]} |U_i| \cdot f(|U_i|) + \sum_{\{i,j\} \subset [t]} |U_i| \cdot |U_j| \cdot f'(M_0) \\
&= \sum_{i \in [t]} |U_i| \cdot f(|U_i|) + f'(M_0) \cdot \sum_{j \in [t]} |U_{<j}| \cdot |U_j| \leq |U| \cdot f(|U|),
\end{aligned}$$

where the final inequality follows by repeatedly applying (11). This completes the proof. $\square$

14

4 Cluster neighbourhoods

In order to prove Theorem 1.2 it will be crucial to understand how vertex neighbourhoods correlate, or cluster, in the graph. The following definition gives a useful measurement of this neighborhood clustering.

Definition 4.1. *Let G be a graph, let v be a vertex of G and let $S \subset V(G)$. Given $M > 1$ and $t \in \mathbb{N} \cup \{0\}$, the (M, t)-**cluster neighbourhood** of v to S is defined as*

$$W_t^S(v; M) := \left\{ u \in V(G) : \left|N_G^S(u) \Delta N_G^S(v)\right| \leq \left(\frac{4^t}{M}\right) \cdot \left|N_G^S(v)\right| \right\}.$$

When S and M are clear from the context we simply write $W_t(v)$ or $W_t^S(v)$.

*Given $\lambda > 1$, setting $\Theta := (M, \lambda)$, the Θ-**moment of** v **to** S is given by*

$$T_\Theta^S(v) := \min \left\{ t \in \mathbb{N} \cup \{0\} : |W_t^S(v; M)| \leq \lambda \cdot |W_t^S(v; M)| \right\}.$$

*The Θ-**cluster** of v to S is then given by $W_{T_\Theta(v)}^S(v; M)$. Once again, when S, M are clear from the context we write $W_*(v) := W_{T_\Theta(v)}^S(v; M)$ and $W_+(v) := W_{T_\Theta(v)+1}^S(v; M)$.*

In the next subsection we present some useful properties of cluster neighbourhoods. The second subsection then proves a key lemma, which allows a partial decomposition of a graph using its cluster neighbourhoods. This result will be central in our proof of Theorem 1.2.

4.1 Simple properties of cluster neighbourhoods

The next lemma captures several useful facts about cluster neighbourhoods and Θ-clusters.

Lemma 4.2. *Given G, S and $\Theta = (M, \lambda)$ as above, the following hold:*

(i) $W_t^S(v) \subset W_{t+1}^S(v)$ *for all $t \geq 0$.*

(ii) $|W_*(v)| = |W_{T_\Theta(v)}^S(v)| \geq \lambda^{T_\Theta^S(v)}.$

(iii) $T_\Theta^S(v) \leq \log_\lambda |S| \leq \log_\lambda |G|.$

(iv) $|W_+(v)| \leq \lambda |W_*(v)|.$

Proof. Part (i) is clear from the definition of $W_t^S(v)$. Parts (ii) and (iv) hold by the definition of $T_\Theta^S(v)$, noting that $v \in W_0^S(v)$. Lastly, part (iii) follows from (ii) since $|W_*(v)| \leq |S| \leq |G|$. □

We next present the following simple, yet useful lemmas.

15

Lemma 4.3. *Let G, S and $\Theta = (M, \lambda)$ be as above. Then for any vertex v and non-negative integer $t < \log_4 M - 1$ one has $|W_t^S(v)| \leq 2\Delta(G)$.*

Proof. Fix $v \in V(G)$ and double count the edges (say there are e_v of them) with one endpoint in $W_t^S(v)$ and the other in $N^S(v)$. On one hand, each vertex in $N^S(v)$ contributes at most $\Delta(G)$ edges, hence $e_v \leq \Delta(G) \cdot d^S(v)$. On the other hand, each vertex in $W_t^S(v)$ must be adjacent to at least $d^S(v)(1 - 4^t M^{-1})$ vertices in $N^S(v)$, thus $e_v \geq |W_t^S(v)| \cdot d^S(v)(1 - 4^t M^{-1})$. The result follows by combining these two bounds and using that $4^t < M/2$. $\square$

Lemma 4.4. *Let G, S and $\Theta = (M, \lambda)$ be as above. Let $t_1, t_2 \in \mathbb{N}$ and suppose $v_1, v_2 \in V(G)$ such that $v_2 \notin W_{t_1+1}^S(v_1)$ and $3 \cdot 4^{t_1} d^S(v_1) \geq 4^{t_2} d^S(v_2)$. Then $W_{t_1}^S(v_1) \cap W_{t_2}^S(v_2) = \emptyset$.*

Proof. Assume by contradiction there is $u \in W_{t_2}^S(v_2) \cap W_{t_1}^S(v_1)$. Then, by triangle inequality

$$|\mathrm{div}_G^S(v_1, v_2)| \leq |\mathrm{div}_G^S(v_1, u)| + |\mathrm{div}_G^S(v_2, u)| \leq M^{-1}\left(4^{t_1} d^S(v_1) + 4^{t_2} d^S(v_2)\right) \leq 4M^{-1} 4^{t_1} d^S(v_1),$$

which means that $v_2 \in W_{t_1+1}^S(v_1)$, contradicting our hypothesis. Therefore our assumption is false and so $W_{t_1}^S(v_1) \cap W_{t_2}^S(v_2) = \emptyset$. $\square$

Corollary 4.5. *Let G, S and $\Theta = (M, \lambda)$ be as above. Let $t_1, t_2 \in \mathbb{N}$ and suppose $v_1, v_2 \in V(G)$ such that $v_{3-i} \notin W_{t_i+1}^S(v_i)$ for both $i \in \{1, 2\}$. Then $W_{t_2}^S(v_2) \cap W_{t_1}^S(v_1) = \emptyset$.*

Proof. Relabelling if necessary, we can assume that $4^{t_1} d^S(v_1) \geq 4^{t_2} d^S(v_2)$. The degree condition in Lemma 4.4 then holds, implying the result. $\square$

We note the following short lemma shows that having many clusters of small size and high enough degree in a graph guarantees a certain level of diversity. It also serves as a preamble to the next result, which states that otherwise vertices cluster around some centers of mass.

Lemma 4.6. *Let G, S and $\Theta = (M, \lambda)$ be as above. Let $n := |V(G)|$ and suppose there is a set $A \subset V(G)$ with $|A| = m \geq \delta n \geq 1$ such that $|W_t^S(v)| \leq n/m$ and $d_G^S(v) \geq d$ for all $v \in A$. Then there is $U \subset A$ with $|U| \geq \delta m/2$ such that $|\mathrm{div}^S(u, v)| \geq 4^t d/M$ for all $u \neq v$ in U.*

Proof. Let us build a new graph H on the vertex set A by drawing an edge between two vertices u and v if $|\mathrm{div}^S(u, v)| < 4^t d/M$. Now, given $v \in A$, let us observe that $N_H(v) \subset W_t^S(v)$, therefore the average degree of J is at most n/m. By Turán's theorem we find an independent set U in H with $|U| \geq |A|(n/m + 1)^{-1} \geq m|A|/(n + m) \geq \delta m/2$, which is the desired set. $\square$

4.2 Partial decomposition through cluster neighbourhoods

This subsection is dedicated to a central result in our proof of Theorem 1.2. Although the statement is technical, roughly speaking, it gives conditions under which we are able to use cluster neighbourhoods in a graph G to obtain disjoint sets $V_1, \ldots, V_t, S$, such that:

- $\bigcup_{i \in [t]} V_i$ covers a substantial portion of $V(G)$,
- the neighbourhoods of vertices within each V_i to S are similar,
- the neighbourhoods of vertices within distinct V_i to S differ substantially,
- the overall contribution of degrees from V_i to S is balanced (in a certain technical sense).

We proceed to the statement.

Lemma 4.7. *Let G be an n-vertex graph with $n \geq 2^{32}$. Suppose that $k \in \mathbb{N}$, $\lambda, M \geq 2$, $\alpha > 0$ such that $\log_2^2 n \leq k \leq n \log_2^{-3} n$ and $0 \leq \alpha \leq (\lambda \log^5 n)^{-1}$, and that $\Theta = (M, \lambda)$. Suppose that*

$$A \subset \left\{ v \in V(G) : M \log^2 n \leq d_G(v) \leq n/2 \text{ and } \left| W_{T_\Theta(v)}^{V(G)}(v; M) \right| \leq \alpha n \right\} \subset V(G)$$

satisfies $|A| \geq n/8$. Then there is a set $U = \{u_1, \ldots, u_t\} \subset A$, a collection of pairwise-disjoint sets $V_1, V_2, \ldots, V_t, S \subset V(G)$ and a set $\{d_i\}_{i \in [t]} \subset [0, n]$ satisfying the following properties:

(i) $u_i \in V_i$ and $|V_i| \leq \alpha n$ for all $i \in [t]$;

(ii) $t \leq k$. Furthermore, if $t < k$ then $\sum_{i \in [t]} |V_i| \geq \frac{n}{1000 \lambda \log^2 n}$;

(iii) $\mathrm{div}_G^S(u_i, v_i) \leq d_i$ for all $v_i \in V_i$ and $i \in [t]$;

(iv) $\mathrm{div}_G^S(u_i, u_j) \geq (5M)^{-1} \cdot \max \left(d_G(u_i), d_G(u_j) \right) + d_i + d_j$ for all $i \neq j$ in $[t]$;

(v) The set U is γ-balanced to S, where $\gamma := \log^5 n \cdot \max \left(n^{-1} \cdot \Delta(G), \ t^{-1} \right)$.

Proof. To begin with, write $W_*(v) = W_{T_\Theta(v)}^{V(G)}(v; M)$ and $W_+(v) = W_{T_\Theta(v)+1}^{V(H)}(v; M)$ for all $v \in A$. By the pigeonhole principle there are $L \in [n]$, $T \in [\log_\lambda(n)] \subset [\log_2 n]$ and a set $B \subset A$ with $|B| \geq |A|/\log^2 n$ such that $|W_*(v)| \in [L, 2L]$ and $T_\Theta(v) = T$ for all $v \in B$. Also $L \leq \alpha n$ by the definition of A, whereas Lemma 4.2 (iv) gives $|W_+(v)| \leq \lambda |W_*(v)| \leq 2\lambda L$.

Next, we take a random partition $V(G) = R \cup S$ by adding each $v \in V(G)$ independently to R with probability $3/4$ and to S with probability $1/4$. Set $p := \min \left(32k|B|^{-1}, (4\lambda L)^{-1} \right) \leq 0.2$ and select a random subset U' of R by including each vertex independently with probability $p' := 4p/3 < 0.3$. Stepping back, it is easy to see that elements of U' are selected from $V(G)$ independently with probability p and, moreover, $U' \cap S = \emptyset$.

We now consider the following set

$$U := \left\{ u \in B \cap U' : W_+(u) \cap U' = \{u\} \text{ and } |W_*(u) \cap R| \geq |W_*(u)|/2 \right\},$$

and the following three related events:

- $\mathcal{A}_1$ is the event that $|U| \in \left[2^{-5} p|B|, 2p|B| \right]$;

- $\mathcal{A}_2$ is the event that $|N_H(v) \cap U| \leq \log^2 n \cdot \max \left(1, p \cdot \Delta(G) \right)$ for all $v \in V(G)$;

17

- $\mathcal{A}_3$ is the event that $\operatorname{div}_G^S(u,v) = \left(\frac{1}{4} \pm 0.05\right)\operatorname{div}_G(u,v)$ for all vertices $u,v \in V(G)$ with $\operatorname{div}_G(u,v) \geq 2^{15}\log n$.

We will now show that these events can occur simultaneously.

Claim: $\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3) > 0$.

To see this, let us first look at $\mathbb{P}(\mathcal{A}_1)$ and note that $|U'| \sim \operatorname{Bin}(|B|,p)$, thus:

$$\mathbb{E}\big[|U'|\big] = p|B| \geq \min\big(32k, |B|\cdot(4\lambda L)^{-1}\big) \geq \min\big(32k, n(32\lambda L\log^2 n)^{-1}\big) \geq \log^2 n,$$

where we have used that $k \geq \log^2 n$ and that $L \leq \alpha n \leq n(\lambda\log^5 n)^{-1}$. Since $U \subset U'$, by applying Theorem 2.3 (Chernoff's inequality) we get:

$$\mathbb{P}(|U| > 2|B|p) \leq \mathbb{P}(|U'| > 2|B|p) \leq \exp(-|B|p/4) \leq \exp(-\log^2 n/4) < n^{-3} < p/64. \quad (12)$$

Next, for each $v \in B$ let $\mathcal{E}_v$ denote the event that $\big|R \cap (W_*(v)\setminus\{v\})\big| \geq \big|W_*(v)\setminus\{v\}\big|/2$. As $\big|R \cap (W_*(v)\setminus\{v\})\big| \sim \operatorname{Bin}\big(|W_*(v)|-1, 3/4\big)$ we have $\mathbb{P}(\mathcal{E}_v) \geq 1/4$ by Theorem 2.6. This gives

$$\mathbb{P}\big(v \in U\big) \geq \mathbb{P}\big(v \in U'\big)\cdot\mathbb{P}\big(U' \cap (W_+(v)\setminus\{v\}) = \emptyset \mid \mathcal{E}_v\big)\cdot\mathbb{P}(\mathcal{E}_v)$$

$$\geq p\cdot(1-p)^{|W_*(v)|}\cdot\frac{1}{4} \geq \frac{p(1-p)^{2\lambda L}}{4} \geq \frac{pe^{-4\lambda Lp}}{4} \geq \frac{pe^{-1}}{4} \geq \frac{p}{16},$$

since $1-x \geq e^{-2x}$ for $x \in [0,1/2]$ and $p\lambda L \leq 1/4$. Therefore $\mathbb{E}\big[|U|\big] \geq p|B|/16$. As $|U| \leq |B|$, by Markov we get $\mathbb{P}\big(|U| \geq p|B|/32\big) \geq p/32$. Combined with (12), this gives

$$\mathbb{P}(\mathcal{A}_1) = \mathbb{P}\Big(|U| \in \big[2^{-5}p|B|, 2p|B|\big]\Big) \geq p/64.$$

To estimate $\mathbb{P}(\mathcal{A}_2)$, given $v \in V(G)$ we note that $|N_G(v) \cap U'| \sim \operatorname{Bin}(|N_G(v)|,p)$ and that $\mathbb{E}\big[|N_G(v) \cap U'|\big] = p\cdot|N_G(v)| \leq p\cdot\Delta(G)$. By Theorem 2.5 we obtain that

$$\mathbb{P}\big(|N_G(v) \cap U'| \geq K\big) \leq (e\Delta(G)p/K)^K \leq (e/\log^2 n)^{\log^2 n} \leq n^{-4},$$

where $K := \log^2 n \cdot \max\big(1, p\cdot\Delta(G)\big)$. In particular, as $|N_G(v) \cap U| \leq |N_G(v) \cap U'|$, we get

$$\mathbb{P}(\mathcal{A}_2) \geq \mathbb{P}\big(|N_G(v) \cap U'| \leq K \text{ for all } v \in V(G)\big) \geq 1 - n\cdot n^{-4} \geq 1 - p/256.$$

Lastly, to lower bound $\mathbb{P}(\mathcal{A}_3)$, note that given $u,v \in V(G)$ with $\operatorname{div}_G(u,v) \geq 2^{15}\log n$, we have $\operatorname{div}_G^S(u,v) \sim \operatorname{Bin}\big(\operatorname{div}_G(u,v), 1/4\big)$ and $\mathbb{E}\big[\operatorname{div}_G^S(u,v)\big] = \operatorname{div}_G(u,v)/4 \geq 2^{13}\log n$. Thus

$$\mathbb{P}\big(|\operatorname{div}_G^S(u,v) - \operatorname{div}_G(u,v)/4| \geq 0.05\cdot\operatorname{div}_G(u,v)\big) \leq \exp\big(-0.05^2(2^{13}\log n)/4\big) \leq n^{-5},$$

by using Chernoff's inequality. We then deduce that $\mathbb{P}(\mathcal{A}_3) \geq 1 - n^2(n^{-5}) \geq 1 - p/256$.

Combining our estimates, we get $\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3) \geq \mathbb{P}(\mathcal{A}_1) - \mathbb{P}\left(\overline{\mathcal{A}_2}\right) - \mathbb{P}\left(\overline{\mathcal{A}_3}\right) > p/128 > 0$, as claimed. We thus fix a choice of R, S, U and U' such that $\mathcal{A}_1 \cap \mathcal{A}_3 \cap \mathcal{A}_3$ holds.

We are now in a position to select $U, V_1, \ldots, V_t$ and S as in the statement of the Lemma. We have already chosen the set S. As $\mathcal{A}_1$ holds, we may also assume, by discarding some elements, that $U = \{u_1, \ldots, u_t\}$ where $t = \lceil 2^{-5} p |B| \rceil$. For all $i \in [t]$ we let $V_i := W_*(u_i) \cap R$. Note that, by definition of U we have $u_i \notin W_+(u_j)$ for all $i \neq j$ in $[t]$, which by Corollary 4.5 implies that $V_i \cap V_j \subset W_*(u_i) \cap W_*(u_j) = \emptyset$. Therefore the sets $V_1, \ldots, V_t$ are pairwise disjoint. As $V_i \subset R$ for all $i \in [t]$, we also have $V_i \cap S = \emptyset$.

We can finally prove that (i)-(v) hold for these sets. As $u_i \in A$ we have $|V_i| \leq |W_*(u_i)| \leq \alpha n$, giving (i). Recalling the definition of p we see that $t = \lceil p|B|/32 \rceil = \min\left(k, \lceil |B|(32\lambda L)^{-1} \rceil\right) \leq k$. Furthermore, if $t < k$ then $t \geq |B| \cdot (32\lambda L)^{-1}$. Besides, $|V_i| = |W_*(u_i) \cap R| \geq |W_*(u_i)|/2 \geq L/2$ since $u_i \in U$. We therefore deduce the following inequality, which shows that (ii) holds:

$$\sum_{i \in [t]} |V_i| \geq \frac{tL}{2} \geq \frac{|B|}{32\lambda L} \cdot \frac{L}{2} \geq \frac{|A|}{64\lambda \log^2 n} \geq \frac{n}{1000\lambda \log^2 n}.$$

To confirm (iii) and (iv), fix $i \neq j$ with $d_G(u_i) \geq d_G(u_j)$. As $u_i, u_j \in U$, from the definition of U we deduce that $u_j \notin W_+(u_i)$. Since $u_i \in A$, it follows that

$$\mathrm{div}_G(u_i, u_j) \geq \frac{4^{T+1}}{M} \cdot d_G(u_i) \geq \frac{4}{M} \cdot M \log^2 n \geq 2^{15} \log n.$$

Moreover, given $v_i \in V_i$ and $v_j \in V_j$, we have $v_i \in W_*(u_i)$ and $v_j \in W_*(u_j)$. Thus

$$\mathrm{div}_G(u_i, v_i) \leq \frac{4^T}{M} \cdot d_G(u_i) \quad \textbf{and} \quad \mathrm{div}_G(u_j, v_j) \leq \frac{4^T}{M} \cdot d_G(u_j) \leq \frac{4^T}{M} \cdot d_G(u_i).$$

Moving to S, since $d_G(u_i) \geq M \log^2 n$ and $\mathcal{A}_2$ holds, given $i \in [t]$ and $v_i \in V_i$ we have

$$\mathrm{div}_G^S(u_i, v_i) \leq \max\left((0.25 + 0.05) \cdot \mathrm{div}_G(u_i, v_i),\ 2^{15} \log n\right) \leq 0.3 \cdot \frac{4^T}{M} \cdot d_G(u_i),$$

Continuing along these lines, as $\mathrm{div}_G(u_i, u_j) \geq 2^{15} \log n$ for $i \neq j$, we obtain

$$\mathrm{div}_G^S(u_i, u_j) \geq (0.25 - 0.05) \cdot \mathrm{div}_G(u_i, u_j) \geq 0.8 \cdot \frac{4^T}{M} \cdot d_G(u_i)$$

$$\geq \frac{0.2}{M} \cdot d_G(u_i) + \mathrm{div}_G^S(u_i, v_i) + \mathrm{div}_G^S(u_j, v_j).$$

Thus (iii) and (iv) hold by taking $d_i = 0.3 \cdot 4^T \cdot d_G(u_i)/M$ and $d_j = 0.3 \cdot 4^T \cdot d_G(u_j)/M$.

It only remains to check that (v) is satisfied when $\gamma = \log^5 n \cdot \max(\Delta/n, 1/t)$. The statement follows since $\mathcal{A}_1$ and $\mathcal{A}_3$ hold, because for all $v \in S \subset V(G)$ we have

$$|N_H(v) \cap U| \leq \log^2 n \cdot \max\left(1, p \cdot \Delta(G)\right) = \log^2 n \cdot \max\left(|U|^{-1}, p \cdot \Delta(G) \cdot |U|^{-1}\right) \cdot |U|$$

$$\leq \log^5 n \cdot \max\left(t^{-1}, n^{-1} \cdot \Delta(G)\right) \cdot |U| = \gamma |U|,$$

where the final inequality uses $|U| \geq p|B|/32 \geq 2^{-5}|A|p \log^{-2} n \geq 2^{-8} pn \log^{-2} n \geq pn \log^{-3} n$. With this, we have completed the proof of the lemma. $\qquad\square$

5 Main Theorem

The focus of this final section will be on the proof of Theorem 5.1 below. Our main result, Theorem 1.2, follows immediately from combining this result with Theorem 3.1.

Theorem 5.1. *There is $C > 1$ such that the following holds with $g_1(x) := \exp(C \cdot (\log_2 x)^{2/3})$. Suppose that G is an n-vertex graph and let $k \in [1, \infty)$ such that $n \leq k^2$ and $\mathrm{hom}(G) \leq n^2/k^3$. Then there is a set $U \subset V(G)$ with $|U| \geq k/g_1(k)$ and a probability distribution $\mathcal{D}$ on $[0.1, 0.9]^{V(G)}$ with $\mathrm{bad}_{\mathcal{D}}(U) \leq |U| \cdot g_1(|U|)$.*

As mentioned in the Introduction, our proof will be inductive. Here we will also sometimes require the following alternative result, which appeared[3] as Theorem 5.1 in [28].

Theorem 5.2. *There is $C > 1$ such that the following holds with $g_2(x) := C(\log_2 x)^2$. Suppose that G is an n-vertex graph and let $k \in [1, \infty)$ such that $n \geq k^2$ and $\mathrm{hom}(G) \leq n/k$. Then there is a set $U \subset V(G)$ with $|U| \geq k/g_2(k)$ and a probability distribution $\mathcal{D}$ on $[0.1, 0.9]^{V(G)}$ with $\mathrm{bad}_{\mathcal{D}}(U) \leq |U| \cdot g_2(|U|)$.*

5.1 Proof of Theorem 5.1

We will prove the theorem by induction on n. Note that we can assume $n \geq k^{3/2}$, as otherwise the hypothesis gives $\mathrm{hom}(G) \leq n^2/k^3 < 1$ which is never satisfied if $n \geq 1$.

For the induction it is convenient to prove the theorem taking $g_1(x) := C_1 \exp(C_2(\log x)^{2/3})$ for constants $C_1, C_2 > 1$ instead; it is easily seen the statement given in Theorem 5.1 follows quickly from this by letting $C := C_2 \log C_1$. Observe that g_1 is increasing on $[1, \infty)$ and that its derivative

$$g_1'(x) = \frac{2C_1 C_2}{3} \cdot \frac{\exp\left(C_2(\log x)^{2/3}\right)}{x(\log x)^{1/3}}$$

is decreasing on $[x_0, \infty)$, where $x_0 := e^{C_2^3}$. By taking C_1 sufficiently large we can assume n (and hence k, as $k^2 \geq n$) is sufficiently large for our estimates below, and that $\log k \geq \frac{1}{2} \log n > x_0$, which will be helpful in applying Lemma 3.7 below.

As a hypothesis for our induction, we will assume the result holds for all $G[V]$ with $V \subsetneq V(G)$. To take advantage of this, we introduce the function $k : \mathcal{P}(V(G)) \to [0, \infty)$ given by

$$k(V) := \begin{cases} k \cdot \left(\frac{|V|}{n}\right)^{2/3} & \text{if } |V|^{1/2} \geq n^2/k^3, \\ |V| \cdot \left(\frac{k^3}{n^2}\right) & \text{if } |V|^{1/2} \leq n^2/k^3. \end{cases}$$

As any non-trivial subset $V \subsetneq V(G)$ satisfies $\mathrm{hom}(G[V]) \leq \mathrm{hom}(G) \leq n^2/k^3$, it follows that:

---

[3]The presentation here differs very slightly from Theorem 5.1 in [28] in that the hypothesis there was that $k \geq 1$, $n \geq 20000k^2$ and $\mathrm{hom}(G) \leq n/25k$. However, the statement in Theorem 5.2 follows from this by applying this result with $k' := k/200$ in place of k (taking C large enough to cover the small regime when $k' < 1$).

(O1) If $|V| \in [n^4/k^6, n)$ then $|V|^{1/2} \geq n^2/k^3$, hence $k(V) = k(|V|/n)^{2/3}$. Thus $|V| \leq \left(k(V)\right)^2$ and $\hom(G[V]) \leq n^2/k^3 = |V|^2/k(V)^3$, so by induction there is $U \subset V$ and a distribution $\mathcal{E}$ on $[0.1, 0.9]^V$ such that $|U| \geq k(V)/g_1(k(V))$ and $\mathrm{bad}_{\mathcal{E}}(U) \leq |U| \cdot g_1(|U|)$.

(O2) If $|V| \leq n^4/k^6$ then $|V|^{1/2} \leq n^2/k^3$, hence $k(V) = |V|(k^3/n^2)$. Thus $|V| \geq \left(k(V)\right)^2$ and $\hom(G[V]) \leq n^2/k^3 = |V|/k(V)$, so by Theorem 5.2 there is $U \subset V$ and a distribution $\mathcal{E}$ on $[0.1, 0.9]^V$ such that $|U| \geq k(V)/g_2(k(V))$ and $\mathrm{bad}_{\mathcal{E}}(U) \leq |U| \cdot g_2(|U|)$.

We need to show that there is a set $U \subset V(G)$ with $|U| \geq k/g_1(k)$ and a distribution $\mathcal{D}$ on $[0.1, 0.9]^{V(G)}$ with $\mathrm{bad}_{\mathcal{D}}(U) \leq |U| \cdot g_1(|U|)$. The following strategy will be key in the proof.

Strategy: First find disjoint sets $V_1, V_2, \ldots, V_t, S \subset V(G)$ and a distribution $\mathcal{E}$ on $[0.1, 0.9]^S$. Given $i \in [t]$, we apply either **(O1)** or **(O2)** to V_i (depending on $|V_i|$) to obtain sets $U_i \subset V_i$ and distributions $\mathcal{D}_i$ on $[0.1, 0.9]^{V_i}$ such that:

- $|U_i| \geq k(V_i)/g_1(k(V_i))$ and $\mathrm{bad}_{\mathcal{D}_i}^{V_i}(U_i) \leq |U_i| \cdot g_1(|U_i|)$ if **(O1)** applies;
- $|U_i| \geq k(V_i)/g_2(k(V_i))$ and $\mathrm{bad}_{\mathcal{D}_i}^{V_i}(U_i) \leq |U_i| \cdot g_2(|U_i|)$ if **(O2)** applies.

This naturally results in a partition $[t] = I_1 \sqcup I_2$ according to which **(Oa)** with $a \in \{1, 2\}$ applies. We can then complete the proof with Lemma 3.7 (with parameters in that lemma as $m_0 = x_0$, $M = k/g_1(k)$ and $M_0 = 2k$) provided we can ensure that:

(A) There is a distribution $\mathcal{E}$ on $[0.1, 0.9]^S$ such that $\mathrm{bad}_{\mathcal{E}}^S(u_i, u_j) \leq g_1'(2k)$ for all $u_i \in U_i$ and $u_j \in U_j$ with $i \neq j$;

(B) $\sum_{i \in I_j} |U_i| \geq k/g_1(k)$ for some $j \in \{1, 2\}$ with $\max_{i \in I_j} |U_i| \geq x_0$.

Before proceeding, we fix the following parameters, as well as $\Theta = (M, \lambda)$:

$$\log_2 \lambda = (\log_2 k)^{\frac{4}{9}}, \quad \log_2 T = (\log_2 k)^{\frac{5}{9}}, \quad \log_2 M = (\log_2 k)^{\frac{2}{3}}. \tag{13}$$

To begin with, let $A = \{v \in V(G) : d_G(v) < n/2\}$. Note that we may assume that $|A| \geq n/2$; otherwise we simply work with $\overline{G}$ instead, as the hypothesis and the conclusion of Theorem 5.1 do not change under taking complement. Given Θ as above, we will write $W_*(v)$ to denote the Θ-cluster of v in G. We consider the partition $A = A_1 \sqcup A_2 \sqcup A_3$ given by:

$$A_1 := \left\{v \in V(G) : d_G(v) \in (k^{3/2}/T, n/2) \text{ with } |W_*(v)| \geq 4k\right\},$$
$$A_2 := \left\{v \in V(G) : d_G(v) \in (k^{3/2}/T, n/2) \text{ with } |W_*(v)| < 4k\right\},$$
$$A_3 := \left\{v \in V(G) : d_G(v) < k^{3/2}/T\right\}.$$

Our proof proceeds according to the size of these sets.

Case 1: There is $v_0 \in A_1$.

Let $t = T_\Theta(v_0)$. We start by taking a $4k$-element subset $S_0 \subset W_*(v_0) = W_t(v_0)$, which is possible as $v_0 \in A_1$, and define:

$$Y := \{v \in N(v_0) : d_G^{S_0}(v) \leq 3k\} \quad \text{and} \quad \overline{Y} := \{v \notin N(v_0) : d_G^{S_0}(v) \geq k\}.$$

We will show that neither of these sets is too large and then apply induction to both sets $V_1 := N(v_0) \setminus (Y \cup S_0)$ and $V_2 := V(H) \setminus (N(v_0) \cup \overline{Y} \cup S_0)$, combining them together and using concavity to show that (despite the small loss of vertices) we get a set of the desired size.

To see this, we double count the non-edges between $N(v_0)$ and S_0. Recall that $S_0 \subset W_t(v_0)$, so there are at most $|S_0|(4^t d_0/M) = (4k)(4^t d_0/M)$ such non-edges. However each $v \in Y$ sends at least k non-edges to $N(v_0)$, so this is at least $|Y|k$, giving $|Y| \leq 4^{t+1}d_0/M$. By similarly double counting edges between $V(H) \setminus N(v_0)$ and S we get $|\overline{Y}| \leq 4^{t+1}d_0/M$. As a consequence,

$$|V_1| \geq d_0 - 4^{t+1}d_0/M - 4k \geq 15d_0/16 \quad \text{and} \quad |V_2| \geq n - d_0 - 4^{t+1}d_0/M - 4k \geq n - 17d_0/16,$$

using $4^{t+1}/M \leq 2^{-5}$, which is true since $t = T_\Theta(v) \leq \log_\lambda n \leq 2\log_\lambda k \leq \log_4 M - 6$ by Lemma 4.2 (iii) and also $d_0/2^5 \geq k^{3/2}/2^5 T \geq 4k$.

First observe that $d_G^S(u_1) \geq d_G^S(u_2) + 2k$ for all $u_1 \in U_1$ and $u_2 \in U_2$ by construction, thus, by Lemma 3.3 taking $\mathcal{E} = \mathcal{U}_S$ we obtain $\mathrm{bad}_\mathcal{E}^S(u_1, u_2) \leq 3/2k \leq \exp(\frac{C_2}{2}(\log k)^{2/3})/k \leq g_1'(2k)$ for all $u_1 \in U_1$ and $u_2 \in U_2$. This provides **(A)** from the **Strategy**.

It remains to show that **(B)** also holds. To do this, we will split the argument into two subcases according to the value of $d_G(v_0)$.

Case 1(a): $d_G(v_0) = d_0 \geq 2n^4/k^6$.

Since $d_0 \leq n/2$ we have $n - d_0 \geq d_0$ and so $|V_i| \geq d_0 - (4^{t+1}/M)d_0 - 4k \geq d_0/2 \geq n^4/k^6$ for $i \in \{1,2\}$. Setting $k_i := k(V_i)$ for $i = 1, 2$ we find via the induction in **(O1)** a subset $U_i \subset V_i$ and a distribution $\mathcal{D}_i$ on $[0.1, 0.9]^{V_i}$ such that $|U_i| \geq k_i/g_1(k_i)$ and $\mathrm{bad}_{\mathcal{D}_i}(U_i) \leq |U_i| \cdot g_1(|U_i|)$.

Next, note that $|V_2|/n \geq (n - 17d_0/16)/n > 0.46$. As the map $x \to x^{2/3} - (1 + 3x)/4$ is positive for $x \in (0.46, 1)$, we deduce that

$$\left(\frac{|V_2|}{n}\right)^{\frac{2}{3}} \geq \frac{1}{4}\left(1 + \frac{|V_2|}{n}\right) = 1 - \frac{3}{4}\left(1 - \frac{|V_2|}{n}\right) \geq 1 - \frac{3}{4}\left(\frac{|V_1|}{n} + \frac{4^{t+2}d_0}{Mn}\right) \geq 1 - \frac{|V_1|}{n},$$

where the final inequality holds since $\dfrac{4^{t+2}d_0}{Mn} < \dfrac{|V_1|}{3n}$. Therefore, we get, as desired, that

$$|U_1| + |U_2| \geq \frac{k_1}{g_1(k_1)} + \frac{k_2}{g_1(k_2)} = \frac{k}{g_1(k_1)} \cdot \left(\frac{|V_1|}{n}\right)^{\frac{2}{3}} + \frac{k}{g_1(k_2)} \cdot \left(\frac{|V_2|}{n}\right)^{\frac{2}{3}}$$

22

$$\geq \frac{k}{g_1(k)} \cdot \frac{|V_1|}{n} + \frac{k}{g_1(k)} \cdot \left(1 - \frac{|V_1|}{n}\right) \geq \frac{k}{g_1(k)}.$$

Noting also that $\max_{i \in [2]} |U_i| \geq k/2g_1(k) \geq k^{1/2} \geq x_0$ this gives **(B)**.

Case 1(b): $d_G(v_0) = d_0 \in (k^{3/2}/T, 2n^4/k^6)$.

As d_0 is smaller here than in Case 1(a), the bound on $|V_2|$ from Case 1(a) still applies, and in particular $|V_2| \geq n^4/k^6$. Thus V_2 falls into **(O1)**. Setting $k_2 := k(V_2)$, this gives a set $U_2 \subset V_2$ with $|U_2| \geq k/g_1(k_2)$ and a distribution $\mathcal{D}_2$ on $[0.1, 0.9]^{V_2}$ with $\mathrm{bad}_{\mathcal{D}_2}(U) \leq |U_2| \cdot g_1(|U_2|)$.

Next, we again have $d_0/2 \leq |V_1| \leq d_0 \leq 2n^4/k^6$. Take a set $V_1' \subset V_1$ of size $\min(|V_1|, n^4/k^6)$, so that $n^4/k^6 \geq |V_1'| \geq |V_1|/2$. Therefore by **(O2)**, taking $k_1 = k(V_1')$, there is a set $U_1 \subset V_1'$ with $|U_1| \geq k_1/g_2(k_1)$ and a distribution $\mathcal{D}_1$ on $[0.1, 0.9]^{V_1}$ with $\mathrm{bad}_{\mathcal{D}_1}(U_1) \leq |U_1| \cdot g_2(|U_1|)$.

As the bound $\left(|V_2|/n\right)^{2/3} \geq 1 - |V_1|/n$ still applies here, it follows that

$$|U_1| + |U_2| \geq \frac{k_1}{g_2(k_1)} + \frac{k_2}{g_1(k_2)} = \frac{k}{g_2(k_1)} \cdot \frac{|V_1'|}{n} \cdot \left(\frac{k^2}{n}\right) + \frac{k}{g_1(k_2)} \cdot \left(\frac{|V_2|}{n}\right)^{2/3}$$

$$\geq \frac{k}{2 \cdot g_2(k)} \cdot \frac{|V_1|}{n} + \frac{k}{g_1(k)} \cdot \left(1 - \frac{|V_1|}{n}\right) \geq \frac{k}{g_1(k)},$$

using that $g_1(k) \geq 2g_2(k)$. Noting also that $\max_{i \in [2]} |U_i| \geq k/2g_1(k) \geq k^{1/2} \geq x_0$ this gives **(B)** and completes the proof in Case 1.

Case 2: $|A_2| \geq n/4$.

In this case we apply Lemma 4.7, taking A_2 as A, $\alpha := 4k/n$ and $\Theta = (M, \lambda)$. To check that all the conditions apply, we note that $\log^2 n \leq n^{1/2} \leq k \leq n^{2/3} \leq n/\log^3 n$ and that $\alpha = (4k/n) \leq 4n^{-1/3} \leq (\lambda \log^5 n)^{-1}$. Moreover, the set A_2 satisfies

$$A_2 \subset \left\{v \in V(G) : M\log^2 n \leq T^{-1} \cdot k^{3/2} \leq d_G(v) \leq n/2 \text{ and } |W_*(v)| \leq \alpha n\right\} \subset V(G).$$

Thus the conditions of Lemma 4.7 hold. Let $U = \{u_1, \ldots, u_t\} \subset A_2$ and $V_1, \ldots, V_t, S \subset V(G)$ be sets and let $\{d_i\}_{i \in [t]}$ be the values given by the Lemma, recalling that (ii) gives us $t \leq k$.

We now show the **Strategy**, applied to some of the $\{V_i\}_{i \in [t]}$, holds. To see that **(A)** holds, we take $\mathcal{E} = \mathcal{B}_\beta(U, S)$ to be a blended distribution on $[0.1, 0.9]^S$. By Lemma 4.7 (iii) and (iv), the hypothesis of Lemma 3.5 is true for $D = (5M)^{-1} \cdot \min_{i \in [t]} d_G(u_i) \geq (5TM)^{-1} \cdot k^{3/2}$. Since U is trivially 1-balanced to S, we take $\gamma = 1$ and $\beta^{-1} = 10\sqrt{k \log k}$, so that for all $v_i \in V_i$ and $v_j \in V_j$ with $i \neq j$ we have the following 'bad' control:

$$\mathrm{bad}_{\mathcal{E}}^S(v_i, v_j) \leq \frac{2}{\beta D} + 2 \cdot \max\left\{d_G^S(u_i), d_G^S(u_j)\right\} \cdot \exp\left(\frac{-0.045}{\gamma \beta^2 |U|}\right)$$

23

$$\leq \frac{10TM \cdot 10\sqrt{k \log k}}{k^{3/2}} + 2 \cdot \frac{k^2}{2} \cdot \exp\left(\frac{-4.5 \cdot k \log k}{t}\right)$$

$$\leq \frac{100TM\sqrt{\log k}}{k} + \frac{1}{k^2} \leq \frac{200TM\sqrt{\log(k)}}{k}$$

$$\leq \frac{\exp\left(\frac{C_2}{2}(\log k)^{2/3}\right)}{k} \leq g_1'(2k). \tag{14}$$

This confirms **(A)** from the **Strategy**, provided that C_2 is large enough.

Case 2(a): $k \geq |U| = t \geq k/\lambda^3$.

Here we can proceed directly using (14). We have $U \subset V(G)$ with $|U| = k/\lambda^3 > k/g_1(k)$ and a distribution $\mathcal{D} = \mathcal{E} \times \mathcal{T}$ on $[0.1, 0.9]^{V(G)}$, where $\mathcal{T}$ is the trivial distribution on $[0, 1, 0.9]^{V(G)\setminus S}$, which by (14) satisfies $\mathrm{bad}_{\mathcal{D}}(U) \leq \binom{|U|}{2} \cdot 200 \cdot TM \cdot (\sqrt{\log(k)}/k) \leq |U| \cdot g_1(|U|)$, using $|U| \leq k$, as required.

Case 2(b): $|U| = t < k/\lambda^3$.

Here we continue with the **Strategy**, showing that **(B)** also holds. Consider the partition $[t] = I_1 \cup I_2$, where $I_1 = \{i \in [t] : |V_i| \geq n^4/k^6\}$ and $I_2 = \{i \in [t] : |V_i| < n^4/k^6\}$. As $t < k$, by the 'Furthermore' statement in Lemma 4.7 (ii) we have

$$\sum_{i \in I_1} |V_i| + \sum_{i \in I_2} |V_i| \geq \frac{n}{1000\lambda \log^2 n} \geq 2\left(\frac{n}{\lambda^2}\right).$$

It follows from above that for some $j \in \{1, 2\}$ we have $\sum_{i \in I_j} |V_i| \geq n/\lambda^2$. As $|I_j| \leq t \leq k/\lambda^3$, this gives $\max_{i \in I_j} |V_i| \geq \lambda(n/k)$. Let us also observe[4] that $|V_i| \leq |W_*(u_i)| \leq 4k \leq n/\lambda^6$.

Claim: Suppose we have sets $\{V_i\}_{i \in I_j}$ with $\sum_{i \in I_j} |V_i| \geq n/\lambda^2$ and $\lambda(n/k) \leq \max_{i \in I_j} |V_i| \leq n/\lambda^6$. Then the sets $\{U_i\}_{i \in I_j}$ given by the **Strategy** satisfy **(B)**, i.e.

$$\sum_{i \in I_j} |U_i| \geq k/g_1(k) \quad \text{and} \quad \max_{i \in I_j} |U_i| \geq x_0.$$

To see this, write $k_i := k(V_i)$ for all $i \in I_j$ and let $U_0 = \bigcup_{i \in I_j} U_i$. Then $|U_0| = \sum_{i \in I_j} |U_i|$. If $j = 1$ then the sets V_i fall into **(O1)** for $i \in J_1$. Then $k_i = k \cdot (|V_i|/n)^{2/3}$ and hence

$$|U_0| \geq \sum_{i \in I_1} \frac{k_i}{g_1(k_i)} = \sum_{i \in I_1} \frac{k}{g_1(k_i)} \cdot \left(\frac{|V_i|}{n}\right)^{2/3} = \frac{k}{g_1(k)} \cdot \sum_{i \in I_1} \frac{|V_i|}{n} \cdot \left(\frac{n}{|V_i|}\right)^{1/3}$$

---

[4]While $|V_i| \leq n/\lambda^6$ is a much weaker than our known bound of $|V_i| \leq 4k$, our current analysis will also apply Case 3 below, where only a weaker bound can be achieved.

$$\geq \frac{k}{g_1(k)} \cdot \left(\sum_{i \in I_1} \frac{|V_i|}{n} \right) \cdot \left(\frac{n}{n/\lambda^6} \right)^{1/3} \geq \frac{k}{g_1(k)} \cdot \frac{1}{\lambda^2} \cdot \lambda^2 = \frac{k}{g_1(k)}.$$

Besides, taking $i \in I_j$ such that $|V_i| \geq \lambda(n/k)$, we have $k_i \geq k(|V_i|/n)^{2/3} \geq k^{1/3}$. Therefore $|U_i| \geq k_i/g_1(k_i) \geq k^{1/6} \geq x_0$.

If $j = 2$ instead then the sets V_i fall into **(O2)** for $i \in J_2$. Then $k_i = |V_i|(k^3/n^2)$ and so

$$|U_0| \geq \sum_{i \in I_2} \frac{k_i}{g_2(k_i)} = \sum_{i \in I_2} \frac{k}{g_2(k_i)} \cdot \left(\frac{|V_i|}{n} \right) \cdot \left(\frac{k^2}{n} \right) \geq \frac{k}{g_2(k)} \cdot \left(\sum_{i \in I_2} \frac{|V_i|}{n} \right) \cdot 1 \geq \frac{k}{\lambda^2 \cdot g_2(k)}.$$

Also, taking $i \in I_j$ such that $|V_i| \geq \lambda(n/k)$, we have $k_i \geq |V_i|(k^3/n^2) \geq \lambda(k^2/n) \geq \lambda$. Therefore $|U_i| \geq k_i/g_1(k_i) \geq \lambda^{1/2} \geq \log k \geq x_0$.

By combining these two bounds we obtain, as required, that

$$\sum_{i \in I_j} |U_i| \geq \min \left(\frac{k}{g_1(k)}, \frac{k}{\lambda^2 \cdot g_2(k)} \right) = \frac{k}{g_1(k)}.$$

Therefore **(B)** holds for $\{U_i\}_{i \in I_j}$, completing the proof of the **Claim**. Since **(A)** also holds, this means that we can apply the **Strategy** for the set U_0, thus completing Case 2.

Case 3: $|A_3| \geq n/4$.

This case has some similarities with the approach from Case 2 above, though there are also several new elements and the analysis is more subtle at certain points.

To begin, note that $G[A_3]$ is an induced subgraph of G with $|G[A_3]| \geq n/4$ and, by definition of A_3, we have $\Delta(G[A_3]) \leq k^{3/2}/T$. We apply Theorem 2.2 to $G[A_3]$ to find an induced subgraph H of $G[A_3]$ such that $m := |H| \geq |A_3|/30 \log n \geq n/120 \log n$ with the property that there is D such that $D \leq d_G(v) \leq 5D \log n$ for all $v \in V(H)$. Since $\hom(H) \leq \hom(G) \leq n^2/k^3$, we note that we must have $D \geq |H| \cdot (n^2/k^3 + 1)^{-1} \geq k^3 \cdot (250n \log n)^{-1} \geq k^{1/2}$ by Turán's theorem. Clearly also $D \leq \Delta(H) \leq T^{-1} \cdot k^{3/2} < m/2$.

We now set $\Theta = (M, \lambda)$ and consider the Θ-clusters, *this time in H*. We first show that Lemma 4.7 applies for $A = V(H)$ and $\alpha := (Tm)^{-1} \cdot 2k^{3/2}$. We have $\log^2 m \leq k \leq m \log^{-3} m$ since $|A| = m \geq 2n/\log^2 n$, and $\alpha = 2k^{3/2}/(Tm) \leq k^{3/2} \log^2 n/(Tn) \leq \log^2 n/T \leq (\lambda \log^5 m)^{-1}$. Given $v \in V(H)$, using Lemma 4.2 (iii) we have $T_\Theta^{V(H)}(v) \leq \log_\lambda(n) < \log_4 M - 1$, and so by Lemma 4.3 we have $|W_*(v)| \leq 2 \cdot \Delta(H) \leq 2k^{3/2}/T < \alpha m$. To check the remaining condition, observe that $M \log^2 m \leq k^{1/2} \leq D \leq d_H(v) \leq \Delta(H) < m/2$, hence:

$$A \subset \left\{ v \in V(H) : M \log^2 m \leq d_H(v) \leq m/2 \text{ and } |W_*(v)| \leq \alpha m \right\} \subset V(H).$$

Thus Lemma 4.7 applies with these parameters to give sets $U, V_1, \ldots, V_t, S$ and values $\{d_i\}_{i \in [t]}$ which satisfy conditions (i)-(v) (with n replaced by m in these statements). As advised by (v), we also let $\gamma := \log^5 m \cdot \max(\Delta(G) \cdot m^{-1}, t^{-1}) \leq \log^8 n \cdot \max(D/n, 1/t)$.

25

We once again show the **Strategy** holds when applied to some of the $\{V_i\}_{i\in[t]}$. To see that **(A)** holds, recall from Lemma 4.7 (v) that U is γ-balanced to S. Set $\beta^{-1} = 10(\gamma|U|\log^4 k)^{1/2}$ and take $\mathcal{E} = \mathcal{B}_\beta(U,S)$ to be the blended distribution on $[0.1, 0.9]^S$. By recalling the definition of the Θ-cluster, Lemma 3.5 for D/M gives that for all $v_i \in V_i$ and $v_j \in V_j$ with $i \neq j$ we have

$$
\begin{aligned}
\mathrm{bad}_{\mathcal{E}}^S(v_i, v_j) &\leq \frac{2M}{\beta D} + 2\cdot(5D\log n)\cdot\exp\left(\frac{-0.045}{\gamma\beta^2|U|}\right)\\
&\leq \frac{20\cdot M\sqrt{\gamma|U|\log^4 k}}{D} + \frac{10\cdot k^{3/2}\log n}{T}\cdot\exp\left(-4.5\cdot\log^4 k\right)\\
&\leq \frac{20M\log^6 k}{D}\cdot\max\left(\sqrt{\frac{D|U|}{n}},1\right) + k^{-3} \leq \frac{42M\log^7 k}{k} \leq g_1'(2k).
\end{aligned}
\tag{15}
$$

The final inequality above holds since $D \geq k^3/(n\log^2 n) \geq k^3/(4n\log^2 k)$, which implies that $\sqrt{|U|/Dn} \leq \sqrt{4\log^2 k \cdot k^{-2}} = 2\log k/k$. Thus part **(A)** of the **Strategy** holds.

<u>**Case 3(a)**</u>: $k \geq |U| = t \geq k/\lambda^3$.

As before, in this case we proceed directly since $|U| = k/\lambda^3 \geq k/g_1(k)$ and there is already a distribution $\mathcal{D} = \mathcal{E} \times \mathcal{T}$ on $[0.1, 0.9]^{V(G)}$, where $\mathcal{T}$ is the trivial distribution on $[0,1,0.9]^{V(G)\setminus S}$. From (15) we deduce that $\mathrm{bad}_{\mathcal{D}}(U) \leq \binom{|U|}{2}\cdot 42Mk^{-1}\log^7 k \leq |U|\cdot g_1(|U|)$, using $|U| = t \leq k$, as required.

<u>**Case 3(b)**</u>: $|U| = t < k/\lambda^3$.

Here we continue with the **Strategy**, verifying that **(B)** holds. Let $[t] = I_1 \cup I_2$ denote again the partition with $I_1 = \{i \in [t] : |V_i| \geq n^4/k^6\}$ and $I_2 = \{i \in [t] : |V_i| < n^4/k^6\}$. As $t < k$, by Lemma 4.7 (ii) we obtain that

$$
\sum_{i\in I_1}|V_i| + \sum_{i\in I_2}|V_i| \geq \frac{m}{10^3\lambda\log^2 m} \geq \frac{n}{10^3\lambda\log^4 n} \geq 2\left(\frac{n}{\lambda^2}\right).
$$

By using the pigeonhole principle, there is $j \in \{1,2\}$ such that

$$
\sum_{i\in I_j}|V_i| \geq \frac{n}{\lambda^2}.
$$

Moreover, $\lambda(n/k) \leq \sum_{i\in I_j}|V_i|/t \leq \max_{i\in I_j}|V_i| \leq \alpha n \leq k^{3/2}\log^2 n/T \leq n/\lambda^6$. It follows that the **Claim** from Case 2 above can be applied to the current sets $\{V_i\}_{i\in I_j}$ as well, so **(B)** holds. Therefore the **Strategy** once again applies, completing the proof of Case 3.

We finish the proof by noting that our three cases exhaust all possibilities, since otherwise $|A| = |A_1| + |A_2| + |A_3| < 0 + (n/4) + (n/4) = n/2$, contrary to the assumption that $|A| \geq n/2$. Therefore one of these cases must apply and so the theorem is proved. $\qquad\square$

References

[1] N Alon, S Friedland, and G Kalai. "Regular subgraphs of almost regular graphs". In: *Journal of Combinatorial Theory, Series B* 37.1 (1984), pp. 79–91. ISSN: 0095-8956. DOI: https://doi.org/10.1016/0095-8956(84)90047-9. URL: https://www.sciencedirect.com/science/article/pii/0095895684900479.

[2] Noga Alon, Michael Krivelevich, and Benny Sudakov. "Induced subgraphs of prescribed size". In: *Journal of Graph Theory* 43.4 (2003), pp. 239–251. DOI: https://doi.org/10.1002/jgt.10117. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jgt.10117. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jgt.10117.

[3] Noga Alon, Michael Krivelevich, and Benny Sudakov. "Large nearly regular induced subgraphs". English (US). In: *SIAM Journal on Discrete Mathematics* 22.4 (2008), pp. 1325–1337. ISSN: 0895-4801. DOI: 10.1137/070704927.

[4] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Second. New York: Wiley, 2004. ISBN: 0471370460.

[5] Boaz Barak, Anup Rao, Ronen Shaltiel, and Avi Wigderson. "2-source dispersers for $n^{o(1)}$-entropy, and Ramsey graphs beating the Frankl-Wilson construction". In: *Annals of Mathematics* (2012), pp. 1483–1543.

[6] B. Bollobás. *Extremal Graph Theory*. USA: Dover Publications, 2004. ISBN: 0486435962.

[7] Béla Bollobás. *Random Graphs*. 2nd ed. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001. DOI: 10.1017/CBO9780511814068.

[8] Béla Bollobás and Paul Erdős. "Cliques in random graphs". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 80.3 (1976), pp. 419–427. DOI: 10.1017/S0305004100053056.

[9] Boris Bukh and Benny Sudakov. "Induced subgraphs of Ramsey graphs with many distinct degrees". In: *Journal of Combinatorial Theory, Series B* 97.4 (2007), pp. 612–619. ISSN: 0095-8956. DOI: https://doi.org/10.1016/j.jctb.2006.09.006. URL: https://www.sciencedirect.com/science/article/pii/S0095895606001080.

[10] F. Chung, R. Graham, and R. Wilson. "Quasi-random graphs". In: *Proceedings of the National Academy of Sciences* 85 (4 1988), pp. 969–970. DOI: 10.1073/pnas.85.4.969.

[11] Fan Chung and Ron Graham. *Erdős on Graphs: His Legacy of Unsolved Problems*. A K Peters, Ltd., July 2020. ISBN: 9781003075950. DOI: 10.1201/9781003075950.

[12] D. Conlon, R. Morris, W. Samotij, and D. Saxton. "The number of distinct degrees in an induced subgraph of a random graph". In: *unpublished* ().

[13] David Conlon, Jacob Fox, and Benny Sudakov. "Recent developments in graph Ramsey theory". In: *Surveys in Combinatorics 2015*. Ed. by Artur Czumaj, Agelos Georgakopoulos, Daniel Král, Vadim Lozin, and Oleg (Editors) Pikhurko. London Mathematical Society Lecture Note Series. Cambridge University Press, 2015, pp. 49–118. DOI: `10.1017/CBO9781316106853.003`.

[14] P. Erdős. "On a lemma of Littlewood and Offord". In: *Bulletin of the American Mathematical Society* 51.12 (1945), pp. 898–902. DOI: `bams/1183507531`. URL: `https://doi.org/`.

[15] P. Erdős. "Some of my favourite problems in various branches of Combinatorics". In: *Le mathematiche* 47.2 (1992), pp. 231–240. URL: `https://lematematiche.dmi.unict.it/index.php/lematematiche/article/view/587`.

[16] P. Erdős. "Some remarks on the theory of graphs". In: *Bulletin of the American Mathematical Society* 53.4 (1947), pp. 292–294. DOI: `bams/1183510596`.

[17] P. Erdős and G. Szekeres. "A Combinatorial Problem in Geometry". In: *Classic Papers in Combinatorics*. Ed. by Ira Gessel and Gian-Carlo Rota. Boston, MA: Birkhäuser Boston, 1987, pp. 49–56. ISBN: 978-0-8176-4842-8. DOI: `10.1007/978-0-8176-4842-8_3`. URL: `https://doi.org/10.1007/978-0-8176-4842-8_3`.

[18] P. Erdős and A. Szemerédi. "On a Ramsey-type theorem". In: *Period Math Hung* 2.1 (1972), pp. 295–299. DOI: `https://doi.org/10.1007/BF02018669`. URL: `https://link.springer.com/article/10.1007%5C%2FBF02018669#citeas`.

[19] Asaf Ferber and Michael Krivelevich. *Every graph contains a linearly sized induced subgraph with all degrees odd*. 2021. arXiv: `2009.05495 [math.CO]`.

[20] R.L. Graham, B.L. Rothschild, and J.H. Spencer. *Ramsey Theory*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 1991. ISBN: 9780471500469.

[21] Spencer Greenberg and Mehryar Mohri. "Tight lower bound on the probability of a binomial exceeding its expectation". In: *Statistics & Probability Letters* 86 (2014), pp. 91–98. ISSN: 0167-7152. DOI: `https://doi.org/10.1016/j.spl.2013.12.009`. URL: `https://www.sciencedirect.com/science/article/pii/S0167715213004082`.

[22] Matthew Jenssen, Eoin Long, Peter Keevash, and Liana Yepremyan. "Distinct degrees in induced subgraphs". In: *Proceedings of the American Mathematical Society* 148.9 (Sept. 2020), pp. 3835–3846. ISSN: 0002-9939. DOI: `10.1090/proc/15060`.

[23] Ross Kang, Eoin Long, Viresh Patel, and Guus Regts. "On a Ramsey-type problem of Erdős and Pach". English. In: *Bulletin of the London Mathematical Society* 49.6 (Dec. 2017), pp. 991–999. ISSN: 0024-6093. DOI: `10.1112/blms.12094`.

[24] Matthew Kwan, Ashwin Sah, Lisa Sauermann, and Mehtaab Sawhney. "Anticoncentration in Ramsey graphs and a proof of the Erdős–McKay conjecture". In: *Forum of Mathematics, Pi* 11 (2023), e21. DOI: `10.1017/fmp.2023.17`.

[25] Matthew Kwan and Benny Sudakov. "Proof of a conjecture on induced subgraphs of Ramsey graphs". In: *Transactions of the American Mathematical Society* 372 (Dec. 2017). DOI: 10.1090/tran/7729.

[26] Matthew Kwan and Benny Sudakov. "Ramsey Graphs Induce Subgraphs of Quadratically Many Sizes". In: *International Mathematics Research Notices* 2020 (Nov. 2017). DOI: 10.1093/imrn/rny064.

[27] Xin Li. "Improved Non-Malleable Extractors, Non-Malleable Codes and Independent Source Extractors". In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2017. Montreal, Canada: Association for Computing Machinery, 2017, pp. 1144–1156. ISBN: 9781450345286. DOI: 10.1145/3055399.3055486. URL: https://doi.org/10.1145/3055399.3055486.

[28] Eoin Long and Laurenţiu Ploscaru. "Distinct degrees and homogeneous sets". In: *Journal of Combinatorial Theory, Series B* 159 (2023), pp. 61–100. ISSN: 0095-8956. DOI: https://doi.org/10.1016/j.jctb.2022.11.004. URL: https://www.sciencedirect.com/science/article/pii/S0095895622001216.

[29] L. Lovász. *Combinatorial Problems and Exercises*. AMS/Chelsea publication. North-Holland Publishing Company, 1993. ISBN: 9780821869475. URL: https://books.google.co.uk/books?id=e99fXXYx9zcC.

[30] Bhargav Narayanan, Julian Sahasrabudhe, and István Tomon. "Ramsey Graphs Induce Subgraphs of Many Different Sizes". In: *Combinatorica* 39 (Sept. 2016). DOI: 10.1007/s00493-017-3755-0.

[31] Bhargav Narayanan and István Tomon. "Induced Subgraphs With Many Distinct Degrees". In: *Combinatorics, Probability and Computing* 27.1 (2018), pp. 110–123. DOI: 10.1017/S0963548317000256.

[32] H. J. Prömel and V. Rödl. "Non-Ramsey Graphs Are clogn-Universal". In: *Journal of Combinatorial Theory, Series A* 88.2 (1999), pp. 379–384. ISSN: 0097-3165. DOI: https://doi.org/10.1006/jcta.1999.2972. URL: https://www.sciencedirect.com/science/article/pii/S0097316599929722.

[33] L. Pyber, V. Rödl, and E. Szemerédi. "Dense Graphs without 3-Regular Subgraphs". In: *Journal of Combinatorial Theory, Series B* 63.1 (1995), pp. 41–54. ISSN: 0095-8956. DOI: https://doi.org/10.1006/jctb.1995.1004. URL: https://www.sciencedirect.com/science/article/pii/S0095895685710040.

[34] J.L. Ramírez-Alfonsín and B.A. Reed. *Perfect Graphs*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 2001. ISBN: 9780471489702.

[35] F. P. Ramsey. "On a Problem of Formal Logic". In: *Proceedings of the London Mathematical Society* s2-30.1 (1930), pp. 264–286. DOI: https://doi.org/10.1112/plms/s2-30.1.264. URL: https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/plms/s2-30.1.264.

[36] A. D. Scott. "Large Induced Subgraphs with All Degrees Odd". In: *Combinatorics, Probability and Computing* 1.4 (1992), pp. 335–349. DOI: 10.1017/S0963548300000389.

[37] S. Shelah. "Erdős and Rényi Conjecture". In: *J. Comb. Theory, Ser. A* 82 (1998), pp. 179–185.

[38] Miklós Simonovits and Vera T. Sós. "Szemerédi's partition and quasirandomness". In: *Random Structures & Algorithms* 2.1 (1991), pp. 1–10. DOI: https://doi.org/10.1002/rsa.3240020102. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.3240020102.

[39] Andrew Thomason. "Pseudo-Random Graphs". In: *Annals of Discrete Mathematics (33)*. Vol. 144. North-Holland Mathematics Studies. North-Holland, 1987, pp. 307–331. DOI: https://doi.org/10.1016/S0304-0208(08)73063-9. URL: https://www.sciencedirect.com/science/article/pii/S0304020808730639.