Distinct degrees and homogeneous sets

Long, Eoin; Ploscaru, Laurentiu

DOI:
10.1016/j.jctb.2022.11.004

Citation for published version (Harvard):
Long, E & Ploscaru, L 2023, 'Distinct degrees and homogeneous sets', *Journal of Combinatorial Theory. Series B*, vol. 159, pp. 61-100. https://doi.org/10.1016/j.jctb.2022.11.004

Journal of Combinatorial Theory, Series B 159 (2023) 61–100

Contents lists available at ScienceDirect

Journal of Combinatorial Theory, Series B

journal homepage: www.elsevier.com/locate/jctb

Distinct degrees and homogeneous sets

Eoin Long, Laurenţiu Ploscaru [1]

School of Mathematics, University of Birmingham, UK

ARTICLE INFO

Article history:
Received 28 April 2022
Available online 13 December 2022

Keywords:
Distinct degrees
Homogeneous number
Ramsey
Probabilistic method

ABSTRACT

In this paper we investigate the extremal relationship between two well-studied graph parameters: the order of the largest homogeneous set in a graph G and the maximal number of distinct degrees appearing in an induced subgraph of G, denoted respectively by $\hom(G)$ and $f(G)$.

Our main theorem improves estimates due to several earlier researchers and shows that if G is an n-vertex graph with $\hom(G) \geq n^{1/2}$ then $f(G) \geq \big(n/\hom(G)\big)^{1-o(1)}$. The bound here is sharp up to the $o(1)$-term, and asymptotically solves a conjecture of Narayanan and Tomon. In particular, this implies that $\max\{\hom(G), f(G)\} \geq n^{1/2-o(1)}$ for any n-vertex graph G, which is also sharp.

The above relationship between $\hom(G)$ and $f(G)$ breaks down in the regime where $\hom(G) < n^{1/2}$. Our second result provides a sharp bound for distinct degrees in biased random graphs, i.e. on $f\big(G(n,p)\big)$. We believe that the behaviour here determines the extremal relationship between $\hom(G)$ and $f(G)$ in this second regime.

Our approach to lower bounding $f(G)$ proceeds via a translation into an (almost) equivalent probabilistic problem, and it can be shown to be effective for arbitrary graphs. It may be of independent interest.

E-mail addresses: e.long@bham.ac.uk (E. Long), ixp090@student.bham.ac.uk (L. Ploscaru).

[1] The second author is grateful for support through an EPSRC DTP studentship (EP/V520275/1).

https://doi.org/10.1016/j.jctb.2022.11.004
0095-8956/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The focus of this paper is on the extremal relationship between the order of the largest homogeneous set in a graph G and the maximal number of distinct degrees which appear in some induced subgraph of G. More precisely, let $\hom(G)$ denote the *homogeneous number* of a graph G, given by:

$$\hom(G) := \max\left\{|U| : U \subset V(G) \text{ with } G[U] \text{ a complete or empty graph}\right\}.$$

We also let $f(G)$ denote the *distinct degree number of* G, given by:

$$f(G) := \max\left\{k \in \mathbb{N} : G[S] \text{ has } k \text{ distinct degrees for some } S \subset V(G)\right\}.$$

These quantities have been well-studied in the literature. Indeed, $\hom(G)$ arises as a key parameter in a variety of settings, including extremal graph theory, graph Ramsey theory and perfect graph theory (see for example [7], [18], [11], [29]). On the other hand, a wide range of results aims to study the possible degree distributions of (induced) subgraphs of a graph, for example [24], [1], [28], [31], [17], [20], and $f(G)$ arises very naturally in this context.

Erdős, Faudree and Sós were the first to investigate the relationship between $\hom(G)$ and $f(G)$, focusing in particular on the Ramsey setting, where $\hom(G)$ is (essentially) minimal. Recall that Ramsey's theorem [30], [15] guarantees that every n-vertex graph G satisfies the relation $\hom(G) = \Omega(\log n)$. Erdős [14] showed, in what is one of the earliest instances of the probabilistic method [4], that there are n-vertex graphs G with $\hom(G) = \Theta(\log n)$ and so the logarithmic order is sharp here. However, the existence of all such graphs G has only been demonstrated indirectly via some random process and it is a major open problem to give explicit examples of such graphs (see [5], [23]). Motivated by this, a large body of research has developed concerning the structure of Ramsey graphs [2], [16], [27], [32], [22], [21], [25], aiming to show that they must behave similarly to appropriate random graphs.

In this context, Erdős, Faudree and Sós [13] noticed that the random graph $G(n, 1/2)$ has $f(G(n, 1/2)) = \Omega(n^{1/2})$ with high probability. They conjectured this property must be shared by Ramsey graphs: if G is an n-vertex graph with $\hom(G) = O(\log n)$ then $f(G) = \Omega(n^{1/2})$. Bukh and Sudakov confirmed this conjecture in [9] with an elegant and influential argument. Furthermore, they noted that there still appeared to be some flexibility here:

(a) Although $f(G(n, 1/2)) = \Omega(n^{1/2})$ forms a natural lower bound, they observed that it did not have a matching upper bound, as they proved that $f(G(n, 1/2)) = O(n^{2/3})$ whp.

(b) They conjectured that $\hom(G) = n^{o(1)}$ already implies that $f(G) \geq n^{1/2-o(1)}$.

It was later shown by Conlon, Morris, Samotij and Saxton [10], thus matching the upper bound given in (a), that in fact $f(G(n,1/2)) = \Omega(n^{2/3})$ whp. Recently, Jenssen, Keevash, Long and Yepremyan [19] proved that the same lower bound applies in the Ramsey context, giving a tight bound for the original Ramsey question of Erdős, Faudree and Sós.

In [26], Narayanan and Tomon solved the conjecture from (b) above, proving that actually $f(G) = \Omega\big((n/\hom(G))^{1/2}\big)$ for all n-vertex graphs G. They also provided an interesting construction, which suggested a tight bound between the following parameters: if $k \leq n^{1/2}$ then the n-vertex k-partite Turán graph T (see e.g. [7]) satisfies both $\hom(T) = n/k$ and $f(T) = k$. Narayanan and Tomon conjectured that a similar dependence must hold in general: if G is an n-vertex graph satisfying $\hom(G) \geq n^{1/2}$ then $f(G) = \Omega(n/\hom(G))$. Supporting their conjecture, the authors proved that indeed $f(G) = \Omega(n/\hom(G))$ when $\hom(G) = \Omega(n/\log n)$. Jenssen et al. [19] improved this bound to $\hom(G) \geq n^{9/10}$, noting that there were significant obstacles to obtaining $\hom(G) \geq n^{1/2}$.

Our main result here confirms the Narayanan–Tomon conjecture up to a logarithmic loss.

Theorem 1.1. *Let $m \geq n^{1/2}$. Then every n-vertex graph G with* $\hom(G) \leq m$ *satisfies:*

$$f(G) = \Omega\left(\frac{n/m}{\log^{7/2}(n/m)}\right).$$

As an immediate corollary of Theorem 1.1 we obtain the following result, which strengthens the bounds of Bukh and Sudakov [9] and of Narayanan and Tomon [26].

Corollary 1.2. *Every n-vertex graph G satisfies* $\max\big\{\hom(G), f(G)\big\} \geq n^{1/2-o(1)}$.

Again, note that the $n^{1/2}$-partite Turán graph on n vertices shows that this bound is essentially sharp. However, as discussed below, there is a large and varied collection of graphs which are close to extremal value here.

Our second result focuses on the regime where $\hom(G) < n^{1/2}$. The Turán construction given above begins to break down here, and in fact the above relationship between the parameters no longer holds; e.g. by our discussion above $\hom(G(n,1/2)) \cdot f(G(n,1/2)) = \Theta(n^{2/3}\log n) \ll n$ whp. Motivated by this, we prove sharp bounds on $f(G(n,p))$ for general values of p, extending the results of Bukh and Sudakov [9] and of Conlon, Morris, Samotij and Saxton [10].

Theorem 1.3. *Let $n \in \mathbb{N}$ and let $p := p(n) \in [0,1/2]$. Then whp the random graph $G(n,p)$ satisfies the following:*

(i) $f\big(G(n,p)\big) = \Theta\left(\sqrt[3]{pn^2}\right)$ *for $p \in [n^{-1/2}, 1/2]$;*
(ii) $f\big(G(n,p)\big) = \Theta\big(\Delta(G(n,p))\big)$ *for $p \leq n^{-1/2}$.*

Remark. As $f(G) = f(\overline{G})$ for any graph G, we see that $f(G(n,p))$ and $f(G(n,1-p))$ follow identical distributions, so Theorem 1.3 determines the behaviour of $f(G(n,p))$ for all $p \in [0,1]$.

Together with the known estimates on the homogeneous number of sparse random graphs, Theorem 1.3 suggests a natural extremal relationship between $\hom(G)$ and $f(G)$ when the hypothesis of Theorem 1.1 fails, i.e. when $\hom(G) < n^{1/2}$. We further examine this relationship in the concluding remarks in Section 7.

Our proofs to both Theorem 1.1 and Theorem 1.3 build upon earlier approaches from [9] and [19], but there are many extra challenges in this regime, which require several key new ingredients and ideas. For instance, although Turán graphs represent an example of n-vertex graphs G with $\hom(G) = n^{1/2}$ and $f(G) = \Theta(n^{1/2})$, there are several very different looking graphs which exhibit (essentially) the same behaviour, including the random graph $G(n, n^{-1/2})$.

One interesting class of examples was given by Narayanan and Tomon, which we call 'iterated Turán graphs': take $b < n^{1/2}$ vertex disjoint sets $V_1, \ldots, V_b$ of size n/b, and on each set V_i put a copy of the complement of the $n^{1/2}$-partite n/b-vertex Turán graph, and join all pairs lying in distinct V_i and V_j by an edge. It can be checked that such a graph G has n-vertices, that $\hom(G) = n^{1/2}$ and that $f(G) = n^{1/2}$ (any set V_i contains at most $n^{1/2}/b$ vertices with distinct degrees). Noting that the degree in each such graph is $n - n/b + n^{1/2}/b$, we see these graphs are non-isomorphic for different values of b, and so there are many distinct extremal situations.

Standing back from this, consider (1) starting with many vertex disjoint copies of the same graph, (2) complementing the edges of each, and (3) joining all vertices between different classes by an edge. One can observe that, starting with the graph on a single point and running these steps we can obtain a Turán graph (applying the process once), and the iterated Turán graph (applying it twice). One could furthermore iterate more times, and this leads to graphs with very limited neighbourhood diversity (see the definition before Lemma 4.3 below), which was a key parameter in many earlier approaches. One of our results below (see Theorem 3.2) allows us to prove lower bounds on $f(G)$ by instead lower bounding auxiliary parameters (see Theorem 3.2) and this connection crucially works without diversity assumptions, unlike in earlier approaches.

The above process also highlights a more significant challenge, which arises naturally for this problem. To find a large set U of vertices with distinct degrees in general, these iterated graphs show that sometimes we *must* first find sets U_i of distinct degrees locally in smaller graphs and then combine the results into a larger set $U = \cup U_i$. Combining such sets together can work very well for iterated graphs, but even small changes to the structure here can break the condition – at an extreme, it could be that the sets U_i have distinct degrees in $G[V_i]$ for $i = 1, 2$ with V_1 and V_2 disjoint, but that all vertices of $U_1 \cup U_2$ have the same degree when combined in $G[V_1 \cup V_2]$. We avoid this kind of difficulty by moving to a more general probabilistic setting, where we instead find probability distributions with certain well-controlled small ball probabilities.

Lastly, our approach in Sections 3 and 4 is quite applicable to the general problem of lower bounding $f(G)$ in an arbitrary graph – see Theorem 3.2 and Lemma 3.4 below.

The paper is organised as follows. In the next section we present a number of tools which will be required in our proof. In Section 3 we present a probabilistic analogue of the problem of finding many distinct degrees in a graph. In Section 4 we extend this approach to a more robust variant and develop a variety of tools and estimates for studying the distinct degree problem. In Section 5 we prove Theorem 1.1 as follows: we first deal with a slightly weaker version in Section 5.1, which applies when $\hom(G) \geq n^{3/5+o(1)}$, and then build upon this in subsection 5.2 to prove Theorem 1.1. In Section 6 we present the proof of Theorem 1.3. Finally, in Section 7 we conclude with a discussion of the case when $f(G) < n^{1/2}$.

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 the set $N_G(u) = \{v \in V(G) : u \sim v\}$ 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 $d_G^S(u) = |N_G^S(u)|$.

Given a vertex $u \in V(G)$, we will also represent the neighbourhood of u by a vector $\mathbf{u} \in \{0,1\}^{V(G)}$ defined such that $u_v = 1$ 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 $\mathrm{proj}_S(\mathbf{u})$, i.e. for any $v \in S$ we have $\mathrm{proj}_S(\mathbf{u})_v = \mathbf{u}_v$. Given $u, v \in V(G)$ we write $\mathrm{div}_G(u,v)$ for the symmetric difference $N(u) \triangle N(v)$. Thus $|\mathrm{div}_G(u,v)|$ 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 $\hom(G) = \hom(\overline{G})$ and $f(G) = f(\overline{G})$ since $\mathrm{div}_G(u,v) = \mathrm{div}_{\overline{G}}(u,v)$ 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. We will use the following version of Turán's theorem (see for example Chapter 6 in [7]).

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

Secondly, we require the following 'anticoncentration' theorem for the well-known Littlewood–Offord problem, which is due to Erdős [12]:

Theorem 2.2 *(Erdős–Littlewood–Offord). Let S be a set of n real numbers of absolute value at least 1. Then, for each $\alpha \in \mathbb{R}$, there are at most $\binom{n}{\lfloor n/2 \rfloor} = \Theta(2^n n^{-1/2})$ subsets of S whose sum of elements lie in the interval $[\alpha, \alpha+1)$.*

We now give a probabilistic interpretation of the previous theorem, which can also be found in [19]. For the sake of completeness, we include a proof of this result.

Theorem 2.3. *Fix non-zero parameters $a_1, a_2, \ldots, a_n \in \mathbb{R}$ and $p_1, p_2, \ldots, p_n \in [0.1, 0.9]$. Suppose $X_1, X_2, \ldots, X_n$ are independent Bernoulli random variables with $X_i \sim Be(p_i)$. Then:*

$$\max_{x \in \mathbb{R}} \ \mathbb{P}\left(\sum_{i=1}^{n} a_i X_i = x \right) = O(n^{-1/2}).$$

Proof. For each $i \in [n]$ choose $w_i, z_i \in [0,1]$ such that $p_i = w_i/2 + (1-w_i)z_i$. Then write X_i as $X_i = W_i Y_i + (1-W_i)Z_i$, where $W_i \sim Be(w_i),\ Z_i \sim Be(z_i)$ and $Y_i \sim Be(0.5)$ are independent random variables. We want to make this choice so that each $w_i \geq 0.2$ and we can do this by letting $z_i = 0, w_i = 2p_i$ if $p_i \leq 1/2$ and by letting $z_i = 1, w_i = 2(1-p_i)$ if $p_i > 1/2$.

We now condition on any choice $\mathcal{C}$ of the W_i's and Z_i's. Let $I = \{i \in [n] : W_i = 1\}$ and suppose that we have $Z_i = z_i$ after the conditioning. On one hand, if $|I| \geq n/10$ then $\mathbb{P}\big(\sum_{i=1}^n a_i X_i = x \mid \mathcal{C}\big) = \mathbb{P}\big(\sum_{i \in I} a_i Y_i + \sum_{i \notin I} a_i z_i = x \mid \mathcal{C}\big)$ becomes $\mathbb{P}\big(\sum_{i \in I} a_i Y_i = x_{\mathcal{C}}\big)$ where $x_{\mathcal{C}} = x - \sum_{i \notin I} z_i$ is a constant, which by Theorem 2.2 (eventually with a scaling argument) is at most $O(n^{-1/2})$. On the other hand, let $\overline{W} = (W_1 + W_2 + \cdots + W_n)/n$ and observe that $|I| = n\overline{W}$. Moreover, we have $\mathbb{E}[\overline{W}] \geq 0.2$ since $w_i \geq 0.2$ for each i. Therefore, we can deduce that $\mathbb{P}(|I| \leq n/10) = \mathbb{P}(\overline{W} \leq 0.1) \leq \mathbb{P}(\overline{W} - \mathbb{E}[\overline{W}] \leq -0.1) \leq \mathbb{P}(|\overline{W} - \mathbb{E}[\overline{W}]| \leq -0.1)$. So by Chebyshev's Inequality we get $\mathbb{P}(|I| \leq n/10) = O(n^{-1})$.

The conclusion follows by combining these results in the Law of Total Probability. □

The following optimization results will be very useful along the way.

Lemma 2.4. *Let $b \geq a > 0$ and let $0 < \alpha < 1$. Then the function $f : [0,a) \to \mathbb{R}$ given by $f(x) := (b+x)^\alpha + (a-x)^\alpha$ is strictly decreasing. In particular, for all $t \in (0,a)$ we have:*

$$b^\alpha + a^\alpha > (b+a-t)^\alpha + t^\alpha.$$

Proof. Note that $f'(x) = \alpha(b+x)^{\alpha-1} - \alpha(a-x)^{\alpha-1} < 0$ on the interval $[0,a)$ since $\alpha - 1 < 0$ and $b + x \geq a - x > 0$. This gives us the first part, whereas the second one is just $f(0) > f(a-t)$. □

Lemma 2.5. *Let $a, b > 0$. Then $a \log_2 a + b \log_2 b + 2\min\{a,b\} \leq (a+b)\log_2(a+b)$.*

Proof. We may assume that $a \leq b$. Let us now define $x := a+b$ and $a := tx$ with $0 < t \leq 1/2$. Upon dividing by x, the inequality we need to prove can be rewritten as

$2t + t\log_2(tx) + (1-t)\log_2((1-t)x) \leq \log_2 x$, where $0 < t \leq 1/2$, which is equivalent to $2t + t\log_2 t + (1-t)\log_2(1-t) \leq 0$.

The map $f : (0,\infty) \to \mathbb{R}$ given by $f(y) = y\log_2 y$ is convex and can be continuously extended to $f(0) = 0$. Therefore the LHS in our last inequality above is convex, so we only need to check that the inequality holds for $t = 0$ and $t = 1/2$, which can be easily seen. □

Finally, we require some classic concentration inequalities. See e.g. appendix A in [4].

Theorem 2.6 *(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:*

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

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

The following bound will be useful for larger deviations.

Theorem 2.7. *Let $n \in \mathbb{N},\ p \in [0,1],\ L > 0$ an let $X \sim Bin(n,p)$ be a random variable. Then:*

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

Lastly, we will also require Hoeffding's inequality.

Theorem 2.8 *(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:*

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

3. Degrees and distributions on the continuous cube

3.1. Recasting the problem

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 $G[S] \sim G(\underline{\mathbf{p}})$.

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 the space $\mathbb{R}^{V(G)}$, which is given by $\mathbf{x} \cdot \mathbf{y} = \sum_{v \in V(G)} x_v y_v$, this notation leads us to the useful representation:

$$\mathbb{E}\left[d_{G(\underline{\mathbf{p}})}(u)\right] = \mathbf{u} \cdot \underline{\mathbf{p}}. \tag{1}$$

Our first lemma comes to show that two vertices whose expected degrees (under the distribution $G(\underline{\mathbf{p}})$) are separated are unlikely to have the same degree in an induced subgraph selected according to $G(\underline{\mathbf{p}})$.

Lemma 3.1. *Let G be a graph and let u, v be distinct vertices in G. Suppose that there is a probability vector $\underline{\mathbf{p}} \in [0.1, 0.9]^V$ such that $\left|\mathbb{E}[d_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(v)]\right| \geq D \geq 2$. Then:*

$$\mathbb{P}\left(d_{G(\underline{\mathbf{p}})}(u) = d_{G(\underline{\mathbf{p}})}(v)\right) = O\left(\frac{\sqrt{\log D}}{D}\right).$$

Proof. Set $W := \operatorname{div}(u, v)$ and $T = |W|$; by hypothesis we deduce that $T \geq 2$. Letting $X := d_{G(\underline{\mathbf{p}})}(u) - d_{G(\underline{\mathbf{p}})}(v)$, this random variable can be written as $X = \sum_{w \in W} \pm X_w$ where $X_w \sim \mathrm{Be}(p_w)$ are independent Bernoulli random variables. We seek to upper bound $\mathbb{P}\left(d_{G(\mathbf{p})}(u) = d_{G(\mathbf{p})}(v)\right) = \mathbb{P}(X = 0)$.

As $\left|\mathbb{E}[X]\right| \geq D$ by our hypothesis, one gets by Hoeffding's Inequality that:

$$\mathbb{P}(X = 0) \leq \mathbb{P}\left(|X - \mathbb{E}[X]| \geq D\right) \leq 2\exp(-D^2/4T),$$

since X is a sum of T independent random variables taking values in the interval $[-1, 1]$. On the other hand, by Theorem 2.3 we get $P(X = 0) = O\left(T^{-1/2}\right)$. Thus:

$$\mathbb{P}\left(X = 0\right) \leq \min\left\{O(T^{-1/2}), 2\exp(-D^2/4T)\right\}. \tag{2}$$

The map $x \mapsto 1/\sqrt{x}$ is decreasing on $(0, \infty)$, whereas $x \mapsto \exp(-D^2/4x)$ is increasing, and their intersection point satisfies the equation $\sqrt{x} = \exp(D^2/4x)$, i.e. $D^2 = 2x \log x$. This gives $x = \Theta(D^2/\log(D))$ and we get the conclusion by substituting this into (2). □

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

Given a graph G, a probability vector $\underline{\mathbf{p}} \in [0,1]^{V(G)}$ and $D > 0$, a set $U \subset V(G)$ is said to be *D-separated in $G(\underline{\mathbf{p}})$* if $|\mathbb{E}[d_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(v)]| \geq D$ for all distinct $u, v \in U$.

In analogy with $f(G)$, define:

$$f_{\underline{\mathbf{p}}}(G) := \max \big\{ |U| : U \subset V(G) \text{ such that } U \text{ is 1-separated in } G(\underline{\mathbf{p}}) \big\}.$$

The next result shows a lower bound for $f(G)$ follows from a lower bound for $f_{\underline{\mathbf{p}}}(G)$.

Theorem 3.2. *Given a graph G and a probability vector $\underline{\mathbf{p}} \in [0.1, 0.9]^{V(G)}$ with $f_{\underline{\mathbf{p}}}(G) \geq 2$, the following relation holds:*

$$f(G) = \Omega\left(\frac{f_{\underline{\mathbf{p}}}(G)}{\log^{3/2}\big(f_{\underline{\mathbf{p}}}(G)\big)} \right).$$

Proof. First note that $f(G) \geq 1$ for every non-empty graph G, therefore we may assume that $L := f_{\underline{\mathbf{p}}}(G) \geq C$ for some absolute constant C. As above, we will write $G[S]$ to denote a random induced subgraph $G[S] \sim G(\underline{\mathbf{p}})$. Let $U \subset V(G)$ be a 1-separated set in $G(\underline{\mathbf{p}})$ with $U = \{u_1, u_2 \dots, u_L\}$, so that the vertices are ordered with increasing expected degree in $G(\underline{\mathbf{p}})$. It follows that if $j - i \geq 2$ then $D_{i,j} := \mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_j)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_i)] \geq j - i \geq 2$ and so we can apply Lemma 3.1 to obtain that:

$$\mathbb{P}\big(d_{G(\underline{\mathbf{p}})}(u_j) = d_{G(\underline{\mathbf{p}})}(u_i)\big) \leq \frac{c\sqrt{\log(D_{i,j})}}{D_{i,j}} \leq \frac{c\sqrt{\log(j-i)}}{j-i},$$

where here $c > 0$ is an absolute constant. Here we used that $\sqrt{\log x}/x$ is decreasing for $x \geq 2$.

Now let us consider a random graph H on $U_1 = \{u_3, u_6, \dots, u_{3\lfloor L/3 \rfloor}\}$, where we build an edge between two vertices if they have the same degree in $G[S] \sim G(\underline{\mathbf{p}})$. The expected number of edges in H is given by:

$$\begin{aligned}
\mathbb{E}[e(H)] = \sum_{\{u_{3i}, u_{3j}\} \subset U_1} \mathbb{P}\big(d_{G(\underline{\mathbf{p}})}(u_{3j}) = d_{G(\underline{\mathbf{p}})}(u_{3i})\big) &\leq \sum_{\{u_{3i}, u_{3j}\} \subset U_1} \frac{c\sqrt{\log(3j-3i)}}{3(j-i)} \\
&\leq \frac{cL}{9}\sqrt{\log(L)} \cdot \left(\sum_{d=1}^{L/3} \frac{1}{d} \right) \\
&\leq \frac{cL}{9} \log^{3/2}(L).
\end{aligned}$$

It follows by Markov that $\mathbb{P}\big(e(H) \leq cL\log^{3/2}(L)/3\big) \geq 2/3$.

On the other hand, we have $\mathbb{E}[|S \cap U_1|] \geq |U_1|/10 \geq L/32$ and so by Chernoff's inequality, using that $L \geq C$, we have $\mathbb{P}(|S \cap U_1| \geq L/64) \geq 2/3$.

Combining these two bounds guarantees that there exists an induced subgraph $G[S]$ with the property that the set $U_2 := S \cap U_1$ satisfies the relations $|U_2| \geq L/64$ and

$e(H[U_2]) \leq e(H) \leq cL\log^{3/2}(L)/3$. By Turán's theorem the subgraph $H[U_2]$ contains an independent set of order $\Omega(3|U_2|^2/cL\log^{3/2}(L)) = \Omega\big(L/\log^{3/2}(L)\big)$. By definition of H such a set necessarily has distinct degrees in $G[S]$, thus completing the proof. □

3.2. Moving to distributions

The message we get from Theorem 3.2 is that a lower bound on $f(G)$ for *any* graph G (up to logarithmic factors) follows from a lower bound on:

$$\widetilde{f}(G) := \max_{\underline{\mathbf{p}}\in[0.1,0.9]^{V(G)}} f_{\underline{\mathbf{p}}}(G).$$

This second quantity can be perceived as a continuous relaxation of $f(G)$ – which trivially corresponds to maximizing over $\{0,1\}^{V(G)}$. However, from our point of view the second solution space is considerably richer, and in particular will allow different behaviours to be blended in a way that is not possible with vectors from the discrete cube; for example, one can take convex combinations of vectors in $[0,1]^{V(G)}$.

Although we would like to lower bound $\widetilde{f}(G)$, this quantity turns out to be just too rigid for certain inductive steps which we want to carry out later.[3] Instead, we introduce a generalised parameter, defined in terms of probability distributions on $[0.1, 0.9]^{V(G)}$, which turns out to be more robust in this respect.

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:

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

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:

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

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

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

We will sometimes suppress the superscript, e.g. write $\mathrm{bad}_{\mathcal{D}}(U) = \mathrm{bad}^{V(G)}_{\mathcal{D}}(U)$ when $S = V(G)$. Lastly, let us remark that in (3) we do not need $\mathcal{D}$ to be defined on all vertex

[3] See comment before Lemma 4.1 below.

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).

The following lemma shows that a lower bound on $f_{\underline{\mathbf{p}}}(G)$ (and consequently on $f(G)$ by Theorem 3.2) follows by finding a large subset $U \subset V(G)$ such that $\mathrm{bad}_{\mathcal{D}}(U)$ is bounded in terms of $|U|$.

Lemma 3.3. *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 there is a vector $\underline{\mathbf{p}} \in [0.1, 0.9]^{V(G)}$ which satisfies $f_{\underline{\mathbf{p}}}(G) \geq |U|/(1+\alpha)$.*

Proof. To see this, select $\underline{\mathbf{p}} \sim \mathcal{D}$ and let Y denote the random set:

$$Y(\underline{\mathbf{p}}) := \big\{\{u, v\} \subset U : \big|\mathbb{E}[d_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(v)]\big| \leq 1\big\}.$$

Note that:

$$\begin{aligned}\mathop{\mathbb{E}}_{\underline{\mathbf{p}}\sim\mathcal{D}}\big[|Y(\underline{\mathbf{p}})|\big] &= \sum_{\{u,v\}\subset U} \mathbb{P}\big(\big|\mathbb{E}[d_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(v)]\big| \leq 1\big) \\ &\leq \sum_{\{u,v\}\subset U} \mathrm{bad}_{\mathcal{D}}(u, v) = \mathrm{bad}_{\mathcal{D}}(U) = \alpha|U|.\end{aligned}$$

It follows that there is a choice of $\underline{\mathbf{p}} \in [0.1, 0.9]^{V(G)}$ such that $|Y(\underline{\mathbf{p}})| \leq \alpha|U|$. Viewing the pairs in $Y(\underline{\mathbf{p}})$ as the edges of a graph J on the vertex set U, again by Turán's theorem we can find an independent set in this graph which has order $|U|/(1+\alpha)$. By definition of J, this gives a lower bound on $f_{\underline{\mathbf{p}}}(G)$, as required. □

From Theorem 3.2, the quantity $f(G)$ is (essentially) lower bounded by $f_{\underline{\mathbf{p}}}(G)$. To close this subsection, and complete the circle, we show that this also holds in the reverse direction. In particular, up to logarithms the quantities $f(G)$ and $\widetilde{f}(G)$ are of the same order of magnitude.

Lemma 3.4. *Let G be a graph and let $U \subset S \subset V(G)$ be vertex subsets such that all vertices of U have distinct degrees in $G[S]$. Then there is a distribution $\mathcal{D}$ on $[0.1, 0.9]^{V(G)}$ such that $\mathrm{bad}_{\mathcal{D}}(U) = O\big(|U| \log |U|\big)$. In particular, there is $\underline{\mathbf{p}} \in [0.1, 0.9]^{V(G)}$ such that:*

$$f_{\underline{\mathbf{p}}}(G) = \Omega\bigg(\frac{f(G)}{\log f(G)}\bigg).$$

Proof. To see this, let $\mathbf{s} \in \{0, 1\}^{V(G)}$ denote the indicator vector of the set S and let $\mathbf{1}$ denote the constant 1 vector. Let $U := \{u_1, u_2, \ldots, u_{|U|}\}$ and assume that $d_{G[S]}(u_i)$ is increasing with i, which by (1) gives $(\mathbf{u}_j - \mathbf{u}_i) \cdot \mathbf{s} \geq j - i$ for all $1 \leq i < j \leq |U|$.

Select α uniformly at random in $[-0.4, 0.4]$ and consider the random vector:

$$\underline{\mathbf{p}} := \frac{1}{2} \cdot \mathbf{1} + \alpha \cdot \mathbf{s} \in [0.1, 0.9]^{V(G)}.$$

Write $\mathcal{D}$ for the resulting probability distribution on $[0.1, 0.9]^{V(G)}$. Given $\underline{\mathbf{p}}$ chosen from $\mathcal{D}$, by (1) we get:

$$\mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_j)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_i)] = (\mathbf{u}_j - \mathbf{u}_i) \cdot \underline{\mathbf{p}} = \alpha \cdot \big(\mathbf{u}_j - \mathbf{u}_i\big) \cdot \mathbf{s} + c',$$

for some fixed constant c'. As $\big(\mathbf{u}_j - \mathbf{u}_i\big) \cdot \mathbf{s} \geq j - i$ and α is uniformly chosen from $[-0.4, 0.4]$, it follows that $\mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_j)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_i)]$ is uniformly distributed over an interval of length at least $0.8(j-i)$. By definition (3), this then gives:

$$\text{bad}_{\mathcal{D}}(u_i, u_j) \leq \frac{2}{0.8(j-i)} \leq \frac{3}{j-i}.$$

It follows that $\text{bad}_{\mathcal{D}}(U) = \sum_{1 \leq i < j \leq |U|} \text{bad}_{\mathcal{D}}(u_i, u_j) \leq \sum_{d=1}^{|U|} \frac{3|U|}{d} \leq 6|U| \log |U|$, giving us the first bound. The second then follows immediately from Lemma 3.3. □

4. Building distributions for distinct degrees

From the previous section, via Theorem 3.2 and Lemma 3.3, we know that 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 $\text{bad}_{\mathcal{D}}(U)$ is small. In this section we will collect a number of results together, which will be used in combination to exhibit such distributions $\mathcal{D}$.

From the 'iterated' graph examples discussed in Section 1 we saw that occasionally we must first find distinct degree sets U_i in graphs $G[S_i]$ where $\{S_i\}_i$ are disjoint, and then combine these sets together so that $\bigcup_i U_i$ will have distinct degrees in $G[\bigcup_i S_i]$. Unfortunately, it is also not hard to see that vertices within U_i can easily agree in degree in the resulting union graph, even if we move from sets U_i to vectors $\underline{\mathbf{p}}_i$ as in Section 3.

While working with fixed sets or vectors can cause difficulties, our first lemma shows that the setting of distributions allows more flexibility here: we can combine distributions while maintaining 'bad' control. This flexibility was the key motivation for working in this more generalised setting (indicated in subsection 3.2).

Lemma 4.1. *Let G be a graph with a 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. To see this, take $c \in \mathbb{R}$ and define X to be the random variable:

$$X(\underline{\mathbf{p}}) := \mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(v)] - c.$$

It suffices to prove that $\underset{\mathbf{p}\sim\mathcal{D}}{\mathbb{P}}\big(|X| \leq 1\big) \leq \text{bad}^{S\cap V_i}_{\mathcal{D}_i}(u,v)$ for all $i \in [L]$ as the result will follow from our definition of $\text{bad}_{\mathcal{D}}(u,v)$. Let $W_i := V(G) \setminus V_i$ for each $i \in [L]$. Given $\underline{\mathbf{p}} \in [0,1]^{V(G)}$, we denote by $\underline{\mathbf{p}}_i$ and $\underline{\mathbf{q}}_i$ its projections on V_i and W_i, respectively, which are mutually independent.

It is easy to see that:

$$X(\underline{\mathbf{p}}) = \mathbb{E}[d^{S\cap V_i}_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d^{S\cap V_i}_{G(\underline{\mathbf{p}})}(v)] + \mathbb{E}[d^{S\cap W_i}_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d^{S\cap W_i}_{G(\underline{\mathbf{p}})}(v)] - c.$$

Conditioned on any choice for $\underline{\mathbf{q}}_i$, we see that $\mathbb{E}[d^{S\cap W_i}_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d^{S\cap W_i}_{G(\underline{\mathbf{p}})}(v)]$ becomes a constant, therefore we obtain that:

$$\underset{\mathbf{p}\sim\mathcal{D}}{\mathbb{P}}\big(|X| \leq 1|\ \underline{\mathbf{q}}_i\big) = \underset{\mathbf{p}_i\sim\mathcal{D}_i}{\mathbb{P}}\big(|\mathbb{E}[d^{S\cap V_i}_{G(\underline{\mathbf{p}}_i)}(u)] - \mathbb{E}[d^{S\cap V_i}_{G(\underline{\mathbf{p}}_i)}(v)] - c'| \leq 1\big) \leq \text{bad}^{S\cap V_i}_{\mathcal{D}_i}(u,v),$$

as $\underline{\mathbf{p}}_i$ and $\underline{\mathbf{q}}_i$ are independent. It follows that $\underset{\mathbf{p}\sim\mathcal{D}}{\mathbb{P}}\big(|X| \leq 1\big) \leq \text{bad}^S_{\mathcal{D}_i}(u,v)$, as desired. □

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 then simply setting $\underline{\mathbf{p}} = \alpha \mathbf{1}_S \in [0.1, 0.9]^S$.

Lemma 4.2. *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. First note that by Lemma 4.1 we have $\text{bad}_{\mathcal{D}}(u,v) \leq \text{bad}^S_{\mathcal{U}_S}(u,v)$ and so it suffices to upper bound this second quantity.

Taking $c \in \mathbb{R}$, we seek to upper bound, when $\underline{\mathbf{p}} \sim \mathcal{U}_S$, the probability of the event that $|\mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(v)] - c| \leq 1$. To analyse this, note that:

$$\mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(v)] = (\text{proj}_S(\mathbf{u}) - \text{proj}_S(\mathbf{v})) \cdot \underline{\mathbf{p}}.$$

Since $\underline{\mathbf{p}} = \alpha \mathbf{1}_S$ where α is selected uniformly at random from $[0.1, 0.9]$, this gives:

$$\mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(v)] = (\text{proj}_S(\mathbf{u}) - \text{proj}_S(\mathbf{v})) \cdot \alpha \mathbf{1}_S = \alpha\big(d^S(u) - d^S(v)\big).$$

As α varies uniformly over $[0.1, 0.9]$ and $d^S(u) - d^S(v) \geq D$ by hypothesis, the quantity $\mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(v)]$ varies uniformly over an interval of length at least $0.8D$, hence the probability that $|\mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(v)] - c| \leq 1$ is at most $2/(0.8D) \leq 3D^{-1}$. □

We next seek to provide 'bad' control for a set by blending neighbourhood structures together – the idea here has some similarities to that of [19]. Let G be a graph, let $U, S \subset V(G)$, where $U := \{u_1, \dots, u_k\}$, 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:

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

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 final lemma in this section provides 'bad' control for blended distributions under certain well-behaved situations. Given $D > 0$, $\gamma \in [0, 1]$ and sets U and S as above we say that:

- U is *D-diverse to S* if for all distinct $u, v \in U$ we have $|N_G^S(u) \triangle N_G^S(v)| \geq D$.
- U is *γ-balanced to S* if for all $v \in S$ we have $d_G^U(v) \leq \gamma |U|$.

Let us quickly remark that U is always 1-*balanced to S*.

Lemma 4.3. *Let G be a graph, $D > 0$, $\beta \in (0, 0.1)$, $\gamma \in (0, 1]$ and $U, S \subset V(G)$ such that U is both D-diverse and γ-balanced to S. Suppose that $\mathcal{D}'$ denotes a distribution on $[0.1, 0.9]^{V(G) \setminus S}$, that $\mathcal{B}_\beta(U, S)$ is the blended probability distribution on $[0.1, 0.9]^S$ and that $\mathcal{D}$ is the product distribution $\mathcal{B}_\beta(U, S) \times \mathcal{D}'$ on $[0.1, 0.9]^{V(G)}$. Then for all $u, v \in U$ one has:*

$$\mathrm{bad}_{\mathcal{D}}(u, v) \leq \frac{2}{\beta D} + D \exp\left(\frac{-0.045}{\gamma \beta^2 |U|} \right). \tag{5}$$

Proof. Suppose $U = \{u_1, u_2, \dots, u_{k+1}\}$ and that for each $i \in [k+1]$, given the vector $\underline{\mathbf{p}}'$ on $\mathbb{R}^S$ from (4), we define the random vector $\underline{\mathbf{q}}^i$ on $\mathbb{R}^S$ by $\underline{\mathbf{q}}^i := \underline{\mathbf{p}}' - \alpha_i \cdot \mathrm{proj}_S(\mathbf{u}_i)$. The 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. We can do this without much of a worry since $\mathcal{D}$ is the product distribution $\mathcal{B}_\beta(U, S) \times \mathcal{D}'$.

From now on, fix a constant $c \in \mathbb{R}$ and $i, j \in [k+1]$, then let $E_{i,j}(c)$ denote the event that $\left| \mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_i)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_j)] - c \right| \leq 1$. According to (3), to prove the lemma it will suffice to show that:

$$\mathbb{P}(E_{i,j}(c)) \le \frac{2}{\beta D} + D\exp\left(\frac{-0.045}{\gamma\beta^2|U|}\right).$$

To upper bound $\mathbb{P}(E_{i,j}(c))$, we assume that $|N^S(u_i) \setminus N^S(u_j)| \ge |N^S(u_j) \setminus N^S(u_i)|$, so that $|N^S(u_i) \setminus N^S(u_j)| \ge D/2$. Pick a subset $Y_{i,j} \subset N^S(u_i) \setminus N^S(u_j)$ of size $D/2$. We call a vertex $v \in Y_{i,j}$ *naughty* if $\underline{\mathbf{q}}_v^i \notin [0.2, 0.8]$. We say the set $Y_{i,j}$ is *naughty* if it contains a naughty vertex and we let $F_{i,j}$ denote this event. By the law of total probability we get that:

$$\mathbb{P}\big(E_{i,j}(c)\big) = \mathbb{P}\big(E_{i,j}(c)|F_{i,j}\big)\cdot\mathbb{P}(F_{i,j}) + \mathbb{P}\big(E_{i,j}(c)|\overline{F_{i,j}}\big)\cdot\mathbb{P}(\overline{F_{i,j}}) \le \mathbb{P}(F_{i,j}) + \mathbb{P}\big(E_{i,j}(c)|\overline{F_{i,j}}\big).$$

Let $v \in S$. Note that $\underline{\mathbf{q}}_v^i$ is a sum of $d_G^{U\setminus\{u_i\}}(v)$ uniform independent random variables, as the coordinates $\mathbf{u}_{i_v}$ are non-zero when $v \sim u_i$. Thus by Hoeffding Inequality we get:

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

where we have used that $d_G^{U\setminus\{u_i\}}(v) \le d_G^U(v) \le \gamma|U|$ as U is γ-balanced to S. By the union bound we get that $\mathbb{P}(F_{i,j}) \le |Y_{i,j}|\mathbb{P}\big(\underline{\mathbf{q}}_v^i \notin [0.2, 0.8]\big) \le D\exp\big(-0.045(\gamma\beta^2|U|)^{-1}\big)$.

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

$$f_c(\alpha_i) := \mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_i)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_j)] - c = (\mathbf{u}_i - \mathbf{u}_j)\cdot\underline{\mathbf{p}} - c. \tag{6}$$

Having conditioned on $\boldsymbol{\alpha}$ above, note that the event $E_{i,j}(c)$ holds only if $f(\alpha_i)$ lies in an interval of length 2. However, as α_i increases, the contribution from each coordinate of $\underline{\mathbf{p}}$ to the inner product on the right hand side of (6) is non-decreasing. Furthermore, the contribution of all of the $Y_{i,j}$-coordinates is exactly α_i, since none of these coordinates were truncated from $\underline{\mathbf{p}}'$ as we have conditioned on $\overline{F_{i,j}}$. It follows that for $\varepsilon > 0$:

$$f(\alpha_i + \varepsilon) - f(\alpha_i) = \sum_{v\in V(G)} \big((u_i)_v - (u_j)_v\big)(u_i)_v \cdot g_{\varepsilon,v} \ge \varepsilon|Y_{i,j}| = \varepsilon D/2,$$

where $g_{\varepsilon,v} \ge 0$ for all $v \in V(G)$ and $g_{\varepsilon,v} = \varepsilon$ for $v \in Y_{i,j}$. Therefore, conditioned on $\boldsymbol{\alpha}$ as above, if $E_{i,j}(c)$ occurs then α_i lies in an interval of length $4/D$. This happens with probability at most $2\beta^{-1}D^{-1}$ and the result in (5) quickly follows from the law of total probability. □

Before we end this section, we define a simple but convenient distribution. Given a graph G and a set $S \subset V(G)$, let $\mathcal{T}_S$ denote the ***trivial S-induced probability distribution***, which is simply the distribution on $[0.1, 0.9]^S$ which selects the vector $\underline{\mathbf{p}}_0 = \frac{1}{2} \cdot \mathbf{1}_S$ with probability 1.

5. The Narayanan–Tomon conjecture

In this section we will prove Theorem 1.1, our approximate version of the Narayanan–Tomon conjecture. From Theorem 3.2 and Lemma 3.3 it will suffice to prove the following theorem.

Theorem 5.1. *Let $n \in \mathbb{N}$ and $k \geq 1$ with $n \geq 20000k^2$ and suppose that G be an n-vertex graph with $\hom(G) \leq n/25k$. Then there is a set $U \subset V(G)$ and a probability distribution $\mathcal{D}$ on $[0.1, 0.9]^{V(G)}$ such that:*

$$|U| = \Omega\left(\frac{k}{\log_2^2(k+1)}\right) \qquad \textit{and} \qquad \mathrm{bad}_{\mathcal{D}}(U) = O\big(|U| \log |U|\big).$$

The proof will split into two regimes. The first deals with the case where $n = \Omega(k^{5/2})$ and the more difficult second case focuses on the regime $k^2 \leq n = O(k^{5/2})$.

To begin, we first present a quick application of Lemma 4.3 that guarantees ‘bad’ control for a set which is $\Omega(k^{3/2})$-diverse.

Lemma 5.2. *Let G be a n-vertex graph and suppose that $U = \{v_1, v_2, \ldots, v_{k+1}\}$ is a set of vertices of G such that $|N(v_i) \triangle N(v_j)| \geq k^{3/2} + k$ for all $i \neq j$ in $[k+1]$. Then there is a probability distribution $\mathcal{D}$ on $[0.1, 0.9]^{V(G)}$ such that $\mathrm{bad}_{\mathcal{D}}(U) \leq 8|U| \log_2 |U|$.*

Proof. Let $S := V(G)$ so that $|N^S(v_i) \triangle N^S(v_j)| = |\mathrm{div}(v_i, v_j)| \geq k^{3/2} + k$ for all $i \neq j$ in $[k+1]$. Therefore U is both $(k^{3/2}+k)$-diverse and 1-balanced to S. We let $\mathcal{D} := \mathcal{B}_\beta(U, S)$, where $\beta^{-1} := \sqrt{56(k+1)\log(k+1)}$, and apply Lemma 4.3 to obtain for all $i \neq j$ that:

$$\begin{aligned} \mathrm{bad}_{\mathcal{D}}(v_i, v_j) &\leq 4\sqrt{14} \cdot \frac{\log^{1/2}(k+1)}{k} + (k^{3/2} + k) \cdot \exp\big(-2.52 \log(k+1)\big) \\ &\leq k^{-1}\big(4\sqrt{14} \log^{1/2}(k+1) + 1\big). \end{aligned}$$

As the map $f : [1, \infty) \to \mathbb{R}$ given by $f(x) := 16 \log_2(x+1) - 4\sqrt{14} \log^{1/2}(x+1) - 1$ is increasing and positive at 1, we can now easily deduce that for all $i \neq j$ one has:

$$\mathrm{bad}_{\mathcal{D}}(v_i, v_j) \leq k^{-1}\big(4\sqrt{14} \log^{1/2}(k+1) + 1\big) \leq 16k^{-1} \log_2(k+1).$$

By summing over all $i \neq j$ in $[k+1]$ we finally deduce that:

$$\mathrm{bad}_{\mathcal{D}}(U) \leq \frac{16 \log_2(k+1)}{k} \cdot \binom{k+1}{2} = 8|U| \log_2 |U|,$$

as required. □

5.1. The case when $n = \Omega(k^{5/2})$

The next result controls 'bad' under the assumption that G has bounded maximum degree.

Lemma 5.3. *Let $n \in \mathbb{N}, x \in [1, \infty)$ and suppose G is an n-vertex graph with $n \geq 25x \cdot \Delta(G)$ and $\hom(G) \leq n/5x$. Then there is a probability distribution $\mathcal{D}$ on $[0.1, 0.9]^{V(G)}$ and a vertex set $U \subset V(G)$ with $|U| = \lceil x \rceil + 1$ such that $\mathrm{bad}_{\mathcal{D}}(U) \leq |U|$.*

Proof. Set $k = \lceil x \rceil$, noting that $x \leq k < 2x$. We will first select $U = \{u_1, \ldots, u_{k+1}\}$ step by step over a series of rounds. To do so, we are going to select a 'control' set Y_i for each $u_i \in U$, so that u_i is strongly joined to Y_i, but any $u_j \neq u_i$ in U with $j > i$ is quite weakly joined to Y_i. This property will allow us to separate the expected degrees of vertices in U and build the distribution $\mathcal{D}$.

We inductively build vertex sets $U_i = \{u_1, u_2, \ldots, u_i\}$, V_i and Y_i for $i \in [k]$ so that:

(i) the sets U_i, $\{Y_j\}_{j \leq i}$ and V_i are all pairwise disjoint;
(ii) $u_i \in V_{i-1}$ for all $i \in [2, k]$;
(iii) $d^{Y_i}(u_i) = |Y_i| = 2k$;
(iv) $d^{Y_i}(v) \leq k/2$ for all vertices $v \in V_i$;
(v) $|V_i| \geq n - 5i\Delta(G)$.

To begin, we set $U_0 = Y_0 = \emptyset$ and $V_0 := V(G)$. Suppose now $i \in [k]$ and that we have found U_{i-1}, V_{i-1} and $\{Y_j\}_{j<i}$ as above and wish to find these sets for i. We look at $G_i := G[V_{i-1}]$ and see that it must have a vertex u_i with $d_{G_i}(u_i) \geq 2k$; in particular $\Delta(G) = \Delta(G_1) \geq 2k$. If not, then $\Delta(G_i) \leq 2k - 1$ and so by Turán's Theorem we obtain an independent set in G_i which has size at least $|V_{i-1}|/2k \geq (n - 5(i-1)\Delta(G))/(2k) > (n - 5x\Delta(G))/(4x) \geq n/5x$, contradicting the $\hom(G)$ condition from our hypothesis. We now let $U_i := U_{i-1} \cup \{u_i\}$ and we pick a subset $Y_i \subset N_{G_i}(u_i)$ of size $2k$. We then define the set $Z_i := \{v \in V_{i-1} : d^{Y_i}_{G_i}(v) \geq k/2\}$ and note that $u_i \in Z_i$. We now let $V_i := V_{i-1} \setminus (Y_i \cup Z_i)$. Observe that by construction (i)-(iv) hold above, and it just remains to show (v).

As $|V_i| = |V_{i-1}| - |Y_i \cup Z_i|$, by induction it is enough to show that $|Y_j \cup Z_j| \leq 5\Delta(G)$. Clearly $|Y_i| \leq 2k$. We bound $|Z_i|$ by double counting the number of edges between Z_i and Y_i. From each $z \in Z_i$ there are at least $k/2$ edges going to Y_i, hence $e(Y_i, Z_i) \geq k|Z_i|/2$. However, $d^{Z_i}_{G_i}(y) \leq \Delta(G)$ for each $y \in Y_i$, thus $e(Y_i, Z_i) \leq 2k\Delta(G)$. It follows that $|Z_i| \leq 4\Delta(G)$ and so $|Z_j \cup Y_j| \leq |Z_j| + 2k \leq 4\Delta(G) + 2k \leq 5\Delta(G)$, as required.

To complete the proof, we set $i := k$ and take $u_{k+1} \in V_k \neq \emptyset$. By using (i)-(iv) above we get disjoint sets $U = U_k \cup \{u_{k+1}\} = \{u_1, \ldots, u_{k+1}\}$ and $\{Y_j\}_{j \in [k]}$ such that:

$$d_{Y_i}(u_i) \geq d_{Y_i}(u_j) + 3k/2 \quad \text{for all } i < j. \tag{7}$$

For each $i \in [k]$ we let $\mathcal{D}_i$ denote the uniformly constant distribution on the set Y_i, i.e. $\mathcal{D}_i := \mathcal{U}_{Y_i}$. Taking $Y_0 := V(G) \setminus (\cup_i Y_i)$, we also let $\mathcal{D}_0 := \mathcal{T}_{Y_0}$ denote the trivial distribution induced by the set $Y_0 := V(G) \setminus (\cup_i Y_i)$ (as defined at the end of Section 4). Lastly, we take $\mathcal{D}$ to be the product distribution $\mathcal{D} := \prod_{i \in [0,k]} \mathcal{D}_i$ on $[0.1, 0.9]^{V(G)}$. Note that from Lemma 4.1, equation (7) and Lemma 4.2, for all $i < j$ we obtain that:

$$\mathrm{bad}_{\mathcal{D}}(u_i, u_j) \le \mathrm{bad}_{\mathcal{D}_i}^{Y_i}(u_i, u_j) \le \frac{3}{(3k/2)} = \frac{2}{k}.$$

It follows that $\mathrm{bad}_{\mathcal{D}}(U) \le \binom{k+1}{2}\left(\frac{2}{k}\right) = |U|$, as desired. □

We are now in a position to prove Theorem 5.1 for $n = \Omega(k^{5/2})$.

Theorem 5.4. *Let $n \in \mathbb{N}$ and $x \ge 1$ with $n \ge 1000x^{5/2}$. Suppose that G is an n-vertex graph with $\mathrm{hom}(G) \le n/20x$. Then there is a probability distribution $\mathcal{D}$ on $[0.1, 0.9]^{V(G)}$ and a vertex set $U \subset V(G)$ with $|U| \ge x + 1$ such that $\mathrm{bad}_{\mathcal{D}}(U) \le 8|U| \log_2 |U|$.*

Proof. Let $k := \lceil x \rceil$. We will prove the theorem by induction on $|V(G)|$. To start with, observe that there is nothing to prove when $k \le 4$ as we can set $\mathcal{D}$ to be any distribution on $[0.1, 0.9]^{V(G)}$ and the requirements are trivially satisfied by any $(k+1)$-vertex set U, since $\mathrm{bad}_{\mathcal{D}}(u, v) \le 1$ for any pair u, v of vertices; such a set U exists as $k+1 \le 1000x^{3/2}$. In particular, this proves that the theorem holds for the smallest possible case, when $n = 1000$ (where x must equal 1). We will proceed with the induction step and assume that $k > 4$.

Let V_0 be a largest vertex set of G such that $|\mathrm{div}(u, v)| \ge 2k^{3/2}$ for all $u, v \in V_0$. If $|V_0| \ge k+1$ then we are done by Lemma 5.2, otherwise assume that $V_0 = \{v_1, v_2, \dots, v_L\}$ for some $L \le k$ and for each $i \in [L]$ define the set $V_i := \{v \in V(G) : |\mathrm{div}(v, v_i)| < 2k^{3/2}\}$. Due to the maximality of S_0 we get $V(G) = \bigcup_{i=1}^{L} V_i$. The proof splits into two cases:

Case I: Every $j \in [L]$ with $d_G(v_j) \in [10k^{3/2}, n - 1 - 10k^{3/2}]$ satisfies $|V_j| \le 3k$.

It is easy to see that at most $3k^2$ vertices of G do not lie in a set V_j of size at least $3k$. Moreover, $d_G(v_i) - 2k^{3/2} < d_G(v) < d_G(v_i) + 2k^{3/2}$ for all $i \in [L]$ and $v \in V_i$. Thus, at least $n - 3k^2$ vertices $v \in V(G)$ have their degree satisfy $d_G(v) \notin [12k^{3/2}, n - 1 - 12k^{3/2}]$. Therefore, for all such vertices we have $d_G(v) \le 12k^{3/2}$ or $d_G(v) \ge n - 1 - 12k^{3/2}$. We will assume that at least half of these vertices fulfil the first condition, as otherwise we can follow an identical argument by working with the complement $\overline{G}$ instead. Consequently, we find a set $V \subset V(G)$ of size $|V| \ge (n - 3k^2)/2 \ge 450x^{5/2}$ with $\Delta(G[V]) \le 12k^{3/2}$. Thus $|V| \ge 25x\Delta(G[V])$ and $\mathrm{hom}(G[V]) \le \mathrm{hom}(G) \le (n - 6k^2)/10x \le |V|/5x$, hence we can apply Lemma 5.3 to $G[V]$ to obtain a distribution $\mathcal{D}_1$ on $[0.1, 0.9]^V$ and a vertex set $U \subset V$ of size $\lceil x \rceil + 1 = k + 1$ with $\mathrm{bad}_{\mathcal{D}_1}(U) \le |U|$. We also take $\mathcal{D}_0 := \mathcal{T}_{V(G) \setminus V}$ to be the trivial distribution induced by $V(G) \setminus V$, and let $\mathcal{D} := \mathcal{D}_0 \times \mathcal{D}_1$ denote the product

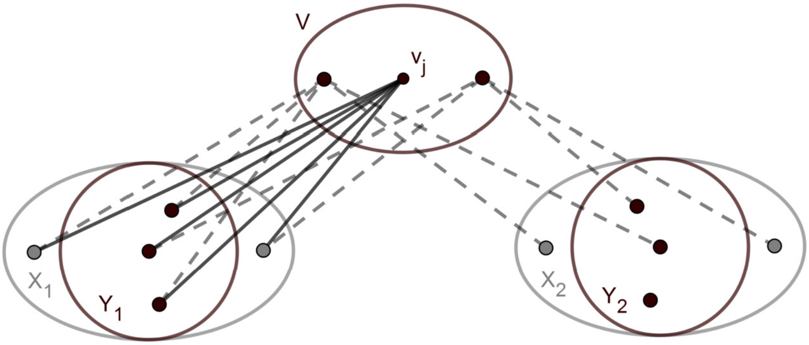

Fig. 1. The clustering behaviour of vertices according to their degree in V.

distribution on $[0.1, 0.9]^{V(G)}$. By Lemma 4.1 we obtain $\mathrm{bad}_{\mathcal{D}}(U) \leq \mathrm{bad}^{V}_{\mathcal{D}_1}(U) \leq |U|$, as required.

Case II: There is $j \in [L]$ such that $d_G(v_j) \in [10k^{3/2}, n-1-10k^{3/2}]$ and $|V_j| \geq 3k$.

We pick a subset V of V_j of size $3k$ such that $v_j \in V$. Next, we set $X_1 := N(v_j)\backslash V$ and $X_2 =: V(G) \setminus (V \cup N(v_j))$. By the choice of v_j note that both $|X_1|, |X_2| \geq 10k^{3/2} - 3k$. Our aim is to show that most vertices in X_1 have big degree in V, whereas most vertices in X_2 have small degree in V. This will allow us to separate the distinct degrees we get in $G[X_1]$ from those we get in $G[X_2]$. This clustering behaviour is illustrated in Fig. 1.

To show that this split occurs, we double count the edges in $\overline{G}$ between X_1 and V. Recall that $X_1 \subset N(v_j)$ and for each $v \in V$ we have $|\operatorname{div}(v, v_j)| \leq 2k^{3/2}$, so each $v \in V$ gives at most $2k^{3/2}$ edges from itself to X_1. Hence $e_{\overline{G}}(X_1, V) \leq (3k)(2k^{3/2}) = 6k^{5/2}$. It follows that there are at most $6k^{3/2}$ vertices of X_1 that are connected to less than $2k$ vertices in V. Thus, if we let $Y_1 := \{u \in X_1 : d^V_G(u) \geq 2k\}$ we can easily observe that $t_1 := |Y_1| \geq |X_1| - 6k^{3/2} \geq 4k^{3/2} - 3k > 10$.

Similarly, we double count the edges in G between X_2 and V to see that there are at most $6k^{5/2}$ of them. It follows that at most $6k^{3/2}$ vertices of X_2 that are connected to more than k vertices in V. Therefore, if we let $Y_2 := \{u \in X_2 : d^V_G(u) \leq k\}$, then we can also see that $t_2 := |Y_2| \geq |X_2| - 6k^{3/2} > 4k^{3/2} - 3k > 10$. Recalling that $V(G) = V \cup X_1 \cup X_2$ is a partition, this shows that $Z := V(G) \setminus (Y_1 \cup Y_2)$ satisfies $|Z| \leq 2 \cdot 6k^{3/2} + 3k \leq 15k^{3/2}$.

To complete the proof, we apply the induction hypothesis to both Y_1 and Y_2. For $i \in \{1, 2\}$ let $x_i := x(t_i/n) \leq x$. This gives $\hom(G[Y_i]) \leq \hom(G) \leq n/20x = t_i/20x_i$. Furthermore:

$$\frac{t_i}{x_i^{5/2}} = \frac{1}{x_i^{3/2}} \cdot \frac{t_i}{x_i} = \frac{1}{x_i^{3/2}} \cdot \frac{n}{x} \geq \frac{n}{x^{5/2}} \geq 1000.$$

Thus, for $i \in \{1, 2\}$, provided $x_i \geq 1$ holds, we can apply the induction hypothesis to $G[Y_i]$ to find a probability distribution $\mathcal{D}_i$ on $[0.1, 0.9]^{Y_i}$ and a set $U_i \subset Y_i$ satisfying $|U_i| \geq x_i + 1$ and:

$$\mathrm{bad}_{\mathcal{D}_i}(U_i) \le |U_i| \cdot f(|U_i|), \quad \text{where } f(y) := 8\log_2 y. \tag{8}$$

Note that if instead $x_i < 1$ above, then as $|Y_i| = t_i \ge 10$, we can take any set $U_i \subset Y_i$ of order $\lceil x_i \rceil + 1 = 2$ and any distribution $\mathcal{D}_i$ on $[0.1, 0.9]^{Y_i}$, so (8) holds in all cases.

We can also assume that $\max\{|U_1|, |U_2|\} < k + 1$, as otherwise taking U to simply be one of these sets proves the theorem.

Next, let $\mathcal{D}_0 := \mathcal{U}_V$ denote the uniformly constant distribution on $[0.1, 0.9]^V$ and pick $\mathcal{D}_3 := \mathcal{T}_Z$, the trivial Z-induced distribution. Setting $U := U_1 \cup U_2$, we let $\mathcal{D}$ denote the product distribution $\prod_{i=0}^{3} \mathcal{D}_i$ on $[0.1, 0.9]^V \times \prod_{i\in[2]}[0.1, 0.9]^{Y_i} \times [0.1, 0.9]^Z = [0.1, 0.9]^{V(G)}$.

Note that $d_G^V(u) \ge 2k \ge d_G^V(v) + k$ for all $u \in Y_1$ and $v \in Y_2$, by the definition of Y_1 and Y_2. It then follows from Lemma 4.2 that for all such vertices we have:

$$\mathrm{bad}^V_{\mathcal{D}_0}(u, v) \le \frac{3}{k}. \tag{9}$$

As $n \ge 1000x^{5/2}$ and $|Z| \le 15k^{3/2}$, we can now lower bound the size of U:

$$\begin{aligned}|U| = |U_1| + |U_2| &\ge (x_1 + 1) + (x_2 + 1) \ge x\left(\frac{t_1}{n}\right) + x\left(\frac{t_2}{n}\right) + 2\\ &\ge \frac{(n - |Z|)x}{n} + 2 \ge x - \frac{15x \cdot k^{3/2}}{n} + 2 \ge x + 1,\end{aligned}$$

which gives $|U| \ge x + 1$. Finally, we are able to estimate $\mathrm{bad}_{\mathcal{D}}(U)$ as follows:

$$\begin{aligned}\mathrm{bad}_{\mathcal{D}}(U) &= \sum_{\{u,v\}\subset U_1} \mathrm{bad}_{\mathcal{D}}(u, v) + \sum_{\{u,v\}\subset U_2} \mathrm{bad}_{\mathcal{D}}(u, v) + \sum_{(u,v)\in U_1\times U_2} \mathrm{bad}_{\mathcal{D}}(u, v)\\ &= \mathrm{bad}_{\mathcal{D}}(U_1) + \mathrm{bad}_{\mathcal{D}}(U_2) + |U_1||U_2| \cdot \max_{(u,v)\in U_1\times U_2} \left\{ \mathrm{bad}_{\mathcal{D}}(u, v)\right\}\\ &\le \mathrm{bad}_{\mathcal{D}_1}(U_1) + \mathrm{bad}_{\mathcal{D}_2}(U_2) + |U_1||U_2| \cdot \max_{(u,v)\in U_1\times U_2} \left\{ \mathrm{bad}^V_{\mathcal{D}_0}(u, v)\right\}\\ &\le |U_1| \cdot f(|U_1|) + |U_2| \cdot f(|U_2|) + \frac{3}{k} \cdot |U_1||U_2| \le |U| \cdot f(|U|).\end{aligned}$$

The final three inequalities here respectively follow from Lemma 4.1, then from (8) and (9), and lastly from $\max\{|U_1|, |U_2|\} < k + 1$ and Lemma 2.5. This completes the proof. □

5.2. The case when $n = O(k^{5/2})$

Before we move to the case when $n = O(k^{5/2})$, we present two results which will allow us to move to a large induced subgraph, which is reasonably regular. Comparable results, with a different range of parameters, were proved by Alon, Krivelevich and Sudakov in [3] (Section 2). The next lemmas follow their approach. We first introduce the following notion.

Definition 5.5. For every n-vertex graph G, its average degree, denoted by $\overline{d}(G)$, is given by the formula $\overline{d}(G) := n^{-1}\sum_{v\in V(G)} d_G(v)$.

Lemma 5.6. *Every n-vertex graph G contains an induced subgraph H of order at least $n/3$ such that $\Delta(H) \leq 2\log_2 n \cdot \overline{d}(H)$.*

Proof. We set $G_0 := G$ and for $i = 0$ to $i = \log_2 n$ we repeat the following algorithm: first set $n_i := |V(G_i)|$, $\Delta_i := \Delta(G_i)$ and $d_i := \overline{d}(G_i)$. Then, if $\Delta_i \leq 2d_i \log_2 n$ we simply stop the process. Otherwise we repeatedly delete from G_i all vertices of degree at least $d_i \log_2 n$ to create a new graph G_{i+1}. Let H be the graph we obtain after we complete the algorithm.

Observe that at the i^{th} iteration we delete at most $e(G_i)/(d_i \log_2 n) = n_i/(2\log_2 n)$ vertices, therefore $n_{i+1} \geq n_i(1 - (2\log_2 n)^{-1})$. It follows that the subgraph H has at least $n \cdot (1-(2\log_2 n)^{-1})^{\log_2 n}$ vertices. As $1-x \geq e^{-2x}$ for $0 < x \leq 1/2$, we deduce that $|V(H)| \geq n/e > n/3$.

If H was created because at some point $\Delta_i \leq 2d_i \log_2 n$ then we are done. Otherwise H was obtained after $\log_2 n$ iterations and at each step i we have $\Delta_{i+1} \leq d_i \log_2 n$ and $2d_i \log_2 n \leq \Delta_i$. Therefore we can see that $\Delta_{i+1} \leq \Delta_i/2$. It follows inductively that $\Delta(H) \leq \Delta(G) \cdot 2^{-\log_2 n} < n \cdot n^{-1} = 1$. We then get that $\Delta(H) = \overline{d}(H) = 0$, which also ends the solution. □

Lemma 5.7. *Every n-vertex graph G contains an induced subgraph H that is of order at least $n/30\log_2 n$ with $\Delta(H) \leq 5\log_2 n \cdot \delta(H)$.*

Proof. By the previous lemma we can find an induced subgraph G_0 of G of order $m \geq n/3$ such that $\Delta(G_0) \leq 2\log_2 n \cdot \overline{d}(G_0)$. We now perform the following algorithm: starting with $i = 0$, let $d_i := \overline{d}(G_i)$ and delete a vertex v of G_i if $5d_{G_i}(v) < 2d_i$. Let now G_{i+1} be the resulting graph and increment i. Note that at each step we remove from G_i at most $2d_i/5$ edges, which implies that $d_{i+1}|G_{i+1}| \geq d_i|G_i| - 4d_i/5 > d_i(|G_i| - 1)$, thus $(d_i)_{i\geq 0}$ is an increasing sequence. Therefore we stop before deleting all the vertices and we let H be the resulting graph.

We can now observe that $\Delta(H) \leq \Delta(G_0)$ and $\delta(H) \geq 2d_0/5$, which immediately implies that $\Delta(H) \leq \Delta(G_0) \leq 2d_0 \log_2 n \leq 5\log_2 n \cdot \delta(H)$. We finally have to lower bound the number t of vertices that are left in H. When we created H from G_0 we deleted less than $2(m-t)d_0/5$ edges, hence $2td_0 \log_2 n \geq t\Delta(H) \geq t\overline{d}(H) \geq md_0 - 4(m-t)d_0/5$. By rearranging the last inequality we obtain $t \geq m/(10\log_2 n) \geq n/(30\log_2 n)$ and so H is the required induced subgraph. □

We are interested in finding sets that have many diverse pairs of vertices as they will give us the freedom required to select vertices with distinct degrees. We thus make the following:

Definition 5.8. Given a graph G and $\varepsilon > 0$, its *diversity graph* $J_\varepsilon(G)$ is the graph on $V(G)$ with an edge between vertices u and v if $|N_G(u) \triangle N_G(v)| \leq \varepsilon \min\{|N_G(u)|, |N_G(v)|\}$.

The following theorem is the main component of our proof in this case. We note that our earlier results from Subsection 5.1 will be crucial here.

Theorem 5.9. *Let G be a n-vertex graph and let $k \in \mathbb{N}$ with $1000k^{5/2} \geq n \geq 8000k^2$, $hom(G) \leq n/12k$ and $\Delta(G) \leq 4nk^{-1/3}$. Then there exists a probability distribution $\mathcal{D}$ on $[0.1, 0.9]^{V(G)}$ and a vertex set $U \subset V(G)$ of order $|U| = \Omega\big(k \log_2^{-2}(k+1)\big)$ which satisfies $\mathrm{bad}_{\mathcal{D}}(U) \leq 8|U| \log_2 |U|$.*

Proof. We first note that if k is small then there is nothing to prove, so we can assume $k > 2^{40}$. Moreover, together with the hypothesis this gives:

$$20 + 4\log_2 k \leq 2\log_2 n \leq 20 + 5\log_2 k \leq (5.5)\log_2 k \leq k/100. \tag{10}$$

Next, by Lemma 5.7 we find an induced subgraph H of G of order $m \geq n \log_2^{-1} n/30$ with $\Delta(H) \leq 5\log_2 n \cdot \delta(H)$. From now on we will only work with this subgraph H. Notice that $\Delta(H) \geq k \log_2^{-1} n/10$, as otherwise by Turán's Theorem, combined with (10), we find an independent set in H (and so in G) of order at least $m/(\Delta(H) + 1) \geq n(3k + 30\log_2 n)^{-1} > n/4k$, contradicting the hypothesis.

Take J to denote the diversity graph $J := J_\varepsilon(H)$, where $\varepsilon = 1/48$. We then set:

$$S_1 := \left\{ v \in V(H) : d_J(v) \leq \frac{m}{600k} \right\} \quad \text{and} \quad S_2 := V(H) \setminus S_1.$$

Our proof will split according to the sizes of S_1 and S_2.

CASE 1: $|S_1| \geq m/2$.

We will show that in this scenario we can take the desired set $U \subset S_1$. We select a set $W \subset S_1$ by including every element of S_1 independently with probability $p := 8k/|S_1|$.

We now claim that each of the following events holds with probability at least 3/4:

(i) $|W| \geq 4k$;
(ii) $e(J[W]) \leq k$;
(iii) $d_H^W(v) \leq 2\log_2 n \cdot m_\Delta$ for all $v \in V(H)$, where $m_\Delta := \max\{1, 240 \cdot \Delta(H) \cdot k/n\}$.

To prove the claim for (i)−(iii) above, let us first denote by $\mathcal{A}_i, \mathcal{A}_{ii}$ and $\mathcal{A}_{iii}$ the events that $|W| \leq 4k$, $e(J[W]) \leq 2k$ and $d_H^W(v) \geq 2\log n \cdot m_\Delta$, respectively.

Starting with (i), recall $|W| \sim Bin(|S_1|, p)$ with $\mathbb{E}[|W|] = p|S_1| = 8k$, therefore by Chernoff's Inequality we get $\mathbb{P}(\mathcal{A}_i) = \mathbb{P}\big(|W| \leq 4k\big) \leq \exp(-k) < 1/4$, proving it for (i).

For (ii) note that $\mathbb{E}[e(J[U])] \leq p^2 e(J[S_1]) \leq p^2|S_1|(m/600k) \leq 64km/600|S_1| \leq k/4$. From Markov's inequality we get $\mathbb{P}(\mathcal{A}_{ii}) = \mathbb{P}(e_J[U] \geq k) \leq 1/4$, which gives us (ii).

Lastly, to prove it for (iii) take any $v \in V(H)$ and let $n_v := d^{S_1}(v)$. Then observe that $d_H^W(v) \sim Bin(n_v, p)$. Now Theorem 2.7 gives $\mathbb{P}\left(d_H^W(v) \geq 2m_v \log_2 n\right) \leq 2^{-2\log_2 n} = n^{-2}$, where $m_v := \max\{1, 240n_v k/n\}$. As $m_\Delta \geq m_v$ for all $v \in V(H)$, the union bound implies that $\mathbb{P}(\mathcal{A}_{iii}) \leq n^{-1} < 1/4$.

Combining the above bounds gives us $\mathbb{P}(\mathcal{A}_i) + \mathbb{P}(\mathcal{A}_{ii}) + \mathbb{P}(\mathcal{A}_{iii}) \leq 3/4$. Therefore, by the union bound we can choose a set $W \subset S_1$ that satisfies all the conditions in (i)−(iii).

To continue the proof in this case, note that by (i) and (ii) we can apply Turán's theorem to $J[W]$ to find an independent set $U_0 \subset W$ with $|U_0| = 2k+1$. However, this means that U_0 is $\left(\delta(H)/48\right)$-diverse to $V(H)$. By (iii) the set U_0 is γ-balanced to $V(H)$, where $\gamma := \log_2 n \cdot m_\Delta/k$. Letting $\mathcal{D} := \mathcal{B}_\beta(U_0, V(H))$ denote the blended probability distribution on $[0.1, 0.9]^{V(H)}$, by applying Lemma 4.3 with $\beta^{-1} := 10\log_2 n\sqrt{m_\Delta}$ we obtain that for all distinct $u, v \in U_0$:

$$\mathrm{bad}_{\mathcal{D}}(u,v) \leq \frac{960\log_2 n\sqrt{m_\Delta}}{\delta(H)} + \frac{\delta(H)}{48}\exp\left(\frac{-4.5 \cdot m_\Delta \log_2^2 n}{2m_\Delta \log_2 n}\right).$$

By noting that $\Delta := \Delta(H) \leq 5\log_2 n \cdot \delta(H)$, this can be further reduced to:

$$\mathrm{bad}_{\mathcal{D}}(u,v) \leq \frac{12 \cdot (20\log_2 n)^2 \cdot \sqrt{m_\Delta}}{\Delta} + \frac{\delta(H)}{48n^2}.$$

Our next claim is that $\Delta^{-1}\sqrt{m_\Delta} < 28kn^{-1}\log_2 k$. Indeed, on the one hand, when $m_\Delta = 1$ then $\Delta^{-1}\sqrt{m_\Delta} \leq \Delta^{-1} \leq 10kn^{-1}\log_2 n < 28kn^{-1}\log_2 k$ by (10), as required. On the other hand, $m_\Delta \geq 1$ implies $\Delta^{-1} \leq 240kn^{-1}$ and so $\Delta^{-1}\sqrt{m_\Delta} \leq \sqrt{240k/(n\Delta)} \leq 240kn^{-1} < 28kn^{-1}\log_2 k$, which proves the claim.

Recall that $\log_2 n \leq 3\log_2 k$ by (10) and that $n \geq 8000k^2$ and $\delta(H) < n$. Therefore, we can deduce that for all distinct $u, v \in U_0$ we have:

$$\mathrm{bad}_{\mathcal{D}}(u,v) \leq 12 \cdot (60\log_2 k)^2 \cdot \frac{28k\log_2 k}{n} + \frac{1}{10^5k^2} \leq \frac{10^3(\log_2 k)^3}{k}.$$

To finish the proof in this case, choose a subset $U \subset U_0$ of size $10^{-3}k\log_2^{-2} k \geq k^{1/4}$. It follows that $\mathrm{bad}_{\mathcal{D}}(u,v) \leq 16|U|^{-1}\log_2|U|$ for all $u, v \in U$.

By summing over all pairs of distinct vertices in U, it immediately follows, as required, that:

$$\mathrm{bad}_{\mathcal{D}}(U) \leq \frac{16\log_2|U|}{|U|} \cdot \binom{|U|}{2} = 8|U|\log_2|U|.$$

CASE 2: $|S_2| \geq m/2$.

Our first step here is to find a set $W \subset S_2$ and for each vertex $w \in W$ two sets $S_w, T_w \subset V(H)$ with the following properties:

(i) $|W| \geq |S_2|/16\Delta(H)$;

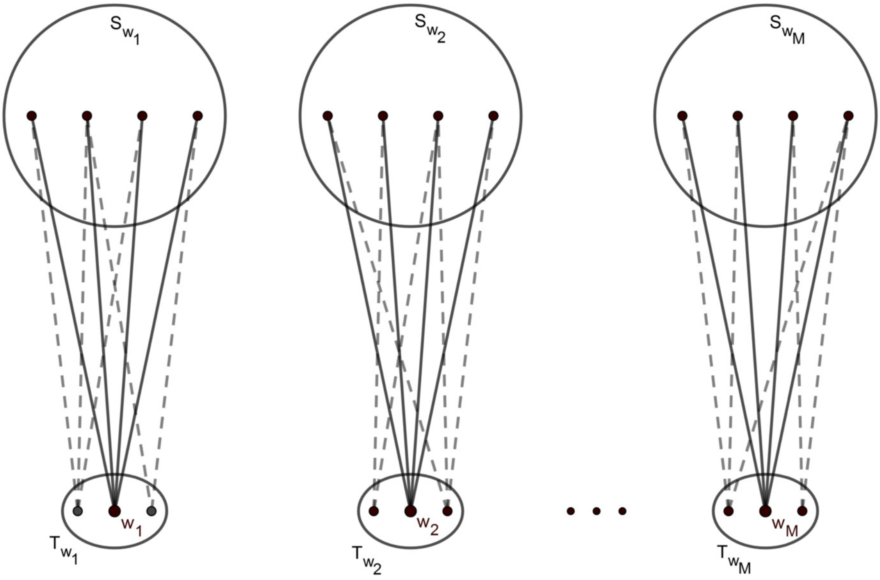

Fig. 2. The clusters formed around each $w \in W$.

(ii) $S_w \subset N_H(w)$ and $|S_w| \geq |N_H(w)|/2$ for each $w \in W$;
(iii) $S_w \cap N_H(w') = \emptyset$ for all distinct $w, w' \in W$;
(iv) $T_w \subset N_J(w)$ with $|T_w| = t := 2^{-19} \cdot 9k \log_2^{-2} k$ for all $w \in W$.

With these sets in hand, our set U will (roughly) be of the form $U = \bigcup_{w \in W} U_w$, where each U_w is a set produced by applying Theorem 5.4 to $S_w \subset N_H(w)$, while the sets T_w will be used to establish 'bad' control between vertices in distinct U_w.

As the diagram in Fig. 2 suggests, our partition is guided by the neighbourhoods of vertices in the set $W = \{w_i\}_i$. The high 'J-degree' of vertices in S_2 guarantees a strong clustering behaviour, so that each vertex w_i has a large set T_{w_i} of vertices which behave similarly. These sets can be used to obtain 'bad' control between vertices in distinct S_{w_i}.

We now proceed with the details. To begin, select a set $W_0 \subset S_2$ by including each element independently with probability $p := 1/8\Delta$, where $\Delta := \Delta(H)$. Next, for each $w \in W_0$ we define the set:

$$S_w := \{v \in V(H) : N_H(v) \cap W_0 = \{w\}\}.$$

We then let $W \subset W_0$ be the set $W := \{w \in W_0 : |S_w| \geq |N_H(w)|/2\}$. Lastly, each $w \in W$ is also an element of S_2, by definition, so we have $d_J(w) \geq m/600k > t$. We take T_w to be an arbitrary subset of size t from $N_J(w)$.

Having specified the sets, it remains to show that with positive probability properties (i)−(iv) hold for our choices. To see this, note that (ii) holds by definition of S_w and W. Property (iii) also always holds as if $v \in S_w \cap N_H(w')$ then $v \in N_H(w) \cap N_H(w')$ and

$\{w, w'\} \subset N_H(v) \cap W_0$, which by definition of S_w implies $w = w'$. Lastly (iv) immediately holds by construction.

It only remains to prove that (i) holds with positive probability. To see this, note that given $w \in S_2$ and $v \in N(w)$ we have:

$$\mathbb{P}\big(v \in S_w \big| w \in W_0\big) = (1-p)^{d_H^{S_2}(v)-1} > (1-p)^{\Delta} \geq e^{-2p\Delta} = e^{-1/4}.$$

Therefore $\mathbb{P}(v \notin S_w | w \in W_0) \leq 1 - e^{-1/4} \leq 1/4$ and this inequailty quickly implies that $\mathbb{E}\big[|N(w) \setminus S_w| \big| w \in W_0\big] \leq |N(w)|/4$. By Markov's inequality we obtain:

$$\mathbb{P}(w \notin W | w \in W_0) = \mathbb{P}\big(|N(w) \setminus S_w| \geq |N(w)|/2 \big| w \in W_0\big) \leq 1/2.$$

We can now further deduce that:

$$\mathbb{E}\big[|W_0 \setminus W|\big] = \sum_{w \in S_2} \mathbb{P}(w \notin W | w \in W_0) \cdot \mathbb{P}(w \in W_0) \leq \mathbb{E}\big[|W_0|\big]/2 = |S_2|p/2,$$

since $\mathbb{E}[|W_0|] = |S_2|p$, as we recall that $|W_0| \sim \text{Bin}(|S_2|, p)$. From here, it follows next that $\mathbb{E}[|W|] = \mathbb{E}[|W_0| - |W_0 \setminus W|] \geq |S_2|p/2 = |S_2|/16\Delta$. Thus we can fix a choice of W so that (i), and hence (i)−(iv), are satisfied.

Our current aim is to find distinct expected degrees in each subgraph $G[S_w]$ with $w \in W$ by appealing to Theorem 5.4 and to use the control sets $\{T_w\}_{w \in W}$ that ensure we can control the degrees between the different sets, so that we can find our required set U in $\bigcup_{w \in W} S_w$.

To proceed with this, first observe that the sets $\{S_w\}_{w \in W}$ are pairwise disjoint, since for distinct $w, w' \in W$ we have $S_w \cap S_{w'} \subset S_w \cap N_H(w') = \emptyset$ by (ii) and (iii).

Next, notice that the sets $\{T_w\}_{w \in W}$ are also pairwise disjoint. Indeed, suppose there is some $v \in T_{w_1} \cap T_{w_2}$ for some distinct $w_1, w_2 \in W$ and assume $|N(w_1)| \leq |N(w_2)|$. Let $S_v := S_{w_2} \cap N(v)$ and $\overline{S_v} := S_{w_2} \setminus N(v)$. As $v \sim w_1$ in J, $N_H(w_1) \cap S_{w_2} = \emptyset$ and $S_{w_2} \subset N_H(w_2)$, we deduce that $|S_v| \leq \varepsilon |N(w_1)| \leq \varepsilon |N(w_2)|$. However $v \sim w_2$ in J, thus $|\overline{S_v}| \leq \varepsilon |N(w_2)|$. Therefore $|N(w_2)|/2 \leq |S_{w_2}| = |S_v| + |\overline{S_v}| \leq 2\varepsilon |N(w_2)| = |N(w_2)|/24$, which is a contradiction. It follows that $T_{w_1} \cap T_{w_2} = \emptyset$ for any $w_1 \neq w_2$ in W.

We want to ensure that vertices of S_w have high degree in T_w, whereas their degree in $T_{w'}$ with $w' \neq w$ is low. Given $w \in W$ we define:

$$R_w := \big\{v \in \bigcup_{w' \neq w} S_{w'} : d_H^{T_w}(v) \geq t/3\big\}, \qquad \text{and} \qquad L_w := \big\{v \in S_w : d_H^{T_w}(v) \leq 2t/3\big\}.$$

If we count the non-edges between S_w and T_w we see there are at least $(t/3)|L_w|$ of them, whereas their number is at most $t(\varepsilon|N(w)|)$ since each vertex of T_w is connected to w in J. It follows that $|L_w| \leq 3\varepsilon|N(w)| \leq 3\varepsilon(2|S_w|) \leq |S_w|/8$, using (ii) above and that $\varepsilon = 1/48$. Similarly, by double counting the edges between $\bigcup_{w' \neq w} S_{w'}$ and T_w we obtain $|R_w| \leq |S_w|/8$.

We now set $S'_w := S_w \setminus \bigcup_{v\in W}(L_v \cup R_v \cup T_v) \subset S_w$ for each $w \in W$. Discarding elements if necessary, we may assume that $|S'_w| > 1$ for all $w \in W$. As the sets $\{S_w\}_{w\in W}$ are pairwise disjoint, this also holds for $\{S'_w\}_{w\in W}$. From our bounds above we find that:

$$\begin{aligned}|\bigsqcup_{w\in W} S'_w| &\geq |\bigsqcup_{w\in W} S_w| - |\bigcup_{v\in W}(L_v \cup R_v \cup T_v)| - |W| \\ &\geq \sum_{w\in W}\big(|S_w| - |L_w| - |R_w| - |T_w| - 1\big) \\ &\geq \sum_{w\in W}\big(|S_w| - |S_w|/4 - t - 1\big) \\ &\geq \sum_{w\in W}\frac{|S_w|}{2} \geq |W|\frac{\delta(G)}{4}.\end{aligned}$$

The second inequality here comes from $|L_w|, |R_w| \leq |S_w|/8$, whereas the third one uses that:

$$1 + t := 1 + \frac{9k}{2^{19}\log_2^2 k} \overset{(10)}{\leq} \frac{k}{400\log_2^2 n} \leq \frac{\Delta(H)}{40\log_2 n} \leq \frac{\delta(H)}{8} \leq \frac{|S_w|}{4}.$$

The final inequality above comes from (ii). Continuing with the previous expression, using that $\delta(H) \geq \Delta/(5\log_2 n)$ and that, by property (i), $|W| \geq |S_2|/16\Delta \geq m/32\Delta$, we obtain:

$$\sum_{w\in W}|S'_w| \geq \frac{|W|\delta(G)}{4} \geq \left(\frac{|S_2|}{16\Delta}\right)\left(\frac{\Delta}{20\log_2 n}\right) = \frac{m}{640\log_2 n} \geq \frac{n}{2^{15}\log_2^2 n}. \qquad (11)$$

We are now in good position to find the desired set U. To do this, we want to apply Theorem 5.4 to each graph $G[S'_w]$. With this in mind, for each $w \in W$ let $k_w := |S'_w|k/n$, and note that $|S'_w|/k_w = n/k$. Also recall that $|S'_w| \leq |N_H(w)| \leq \Delta \leq 4nk^{-1/3}$, hence:

$$\frac{|S'_w|}{k_w^{5/2}} = \frac{|S'_w|}{(|S'_w|k/n)^{5/2}} = \frac{n^{5/2}}{k^{5/2}|S'_w|^{3/2}} \geq \frac{n^{5/2}}{k^{5/2}\cdot\Delta^{3/2}} \geq \frac{n^{5/2}}{k^{5/2}(4nk^{-1/3})^{3/2}} = \frac{n}{8k^2} \geq 1000.$$

We also have $\hom(G[S'_w]) \leq \hom(G) \leq n/12k = |S'_w|/12k_w$. Thus for each $w \in W$, provided $k_w \geq 1$, we can apply Theorem 5.4 to $G[S'_w]$ to obtain a set $U_w \subset S'_w$ with $|U_w| \geq k_w + 1$ and a probability distribution $\mathcal{D}_w$ on $[0.1, 0.9]^{S'_w}$ such that:

$$\mathrm{bad}^{S'_w}_{\mathcal{D}_w}(U_w) \leq |U_w| \cdot f(|U_w|), \quad \text{where } f(x) := 8\log_2 x. \qquad (12)$$

As in the proof of Theorem 5.4, if $k_w \leq 1$ then any set $U_w \subset S'_w$ of size $2 \geq k_i + 1$ trivially satisfies (12), thus the above computations all make sense.

We now set $U := \bigcup_{w\in W} U_w$ and $S' := \bigcup_{w\in W} S'_w$. Our distribution $\mathcal{D}$ will again be a product distribution, with $\mathcal{D}_w$ the forming factors. For each $w \in W$ we also take

$\mathcal{E}_w$ to denote the uniformly constant distribution on T_w given by $\mathcal{E}_w := \mathcal{U}_{T_w}$ and set $T := \cup_{w\in W} T_w$. We note that given distinct $w, w' \in W$ and $u \in U_w$, $u' \in U_{w'}$ we have $d_H^{T_w}(u) \geq 2t/3 \geq d_H^{T_w}(u') + t/3$. Therefore, by the choice of $\mathcal{E}_w$ and from Lemma 4.2 we find that:

$$\mathrm{bad}_{\mathcal{E}_w}^{T_w}(u, u') \leq 9/t. \tag{13}$$

We also let $\mathcal{T}_R$ denote the trivial R-induced distribution, where $R := V(G) \setminus (S' \cup T)$. Let $\mathcal{D}$ be the product distribution on $[0.1, 0.9]^{S'} \times [0.1, 0.9]^T \times [0.1, 0.9]^R = [0.1, 0.9]^{V(G)}$ below:

$$\mathcal{D} := \Big(\prod_{w\in W} \mathcal{D}_w\Big) \times \Big(\prod_{w\in W} \mathcal{E}_w\Big) \times \mathcal{T}_R.$$

To complete the proof, we are only left to lower bound $|U|$ and upper bound $\mathrm{bad}_{\mathcal{D}}(U)$. For the lower bound, using (11) and that $\log_2 n \leq 2\sqrt{2}\log_2 k$ from (10), we obtain:

$$|U| = \sum_{w\in W} |U_w| \geq \sum_{w\in W} k_w \geq \sum_{w\in W} |S'_w| \cdot \frac{k}{n} = \frac{k}{n}\Big(\big|\bigcup_{w\in W} S'_w\big|\Big) \geq \frac{k}{2^{19}\log^2 k} = \frac{t}{9}.$$

For the upper bound on $\mathrm{bad}_{\mathcal{D}}(U)$, we have:

$$\begin{aligned}
\mathrm{bad}_{\mathcal{D}}(U) &= \sum_{w\in W} \mathrm{bad}_{\mathcal{D}}(U_w) + \sum_{\{w,w'\}\subset W} \mathrm{bad}_{\mathcal{D}}(U_w, U_{w'}) \\
&\leq \sum_{w\in W} \mathrm{bad}_{\mathcal{D}_w}^{S'_w}(U_w) + \sum_{\{w,w'\}\subset W} \mathrm{bad}_{\mathcal{E}_w}^{T_w}(U_w, U_{w'}) \\
&\leq \sum_{w\in W} \mathrm{bad}_{\mathcal{D}_w}^{S'_w}(U_w) + \sum_{\{w,w'\}\subset W} |U_w| \cdot |U_{w'}| \cdot \max_{(u,u')\in U_w\times U_{w'}} \mathrm{bad}_{\mathcal{E}_w}^{T_w}(u_w, u_{w'}) \\
&\leq \sum_{w\in W} |U_w| \cdot f(|U_w|) + \sum_{\{w,w'\}\subset W} |U_w| \cdot |U_{w'}| \cdot \left(\frac{9}{t}\right).
\end{aligned}$$

The first inequality here follows Lemma 4.1, the second is immediate from the definition of $\mathrm{bad}_{\mathcal{D}}^{S}(U, V)$, whereas the third one holds by (12) and (13).

Choose a smallest subset $W' := \{w_1, \ldots, w_M\}$ of W so that $|\bigsqcup_{w\in W'} U_w| \geq t/9$. If $W' = \{w'\}$ for some $w' \in W$ then we are done by simply taking $U = U_{w'}$ since $\mathrm{bad}_{\mathcal{D}}(U_{w'}) \leq |U_{w'}|f(|U_{w'}|)$. Otherwise we can assume that the sequence $U_i := U_{w_i}$ is non-increasing in size with i, i.e. that $|U_1| \geq |U_2| \geq \ldots \geq |U_M|$. Setting $U_{<i} := \bigcup_{j<i} U_i$, we immediately see from our choice of W' that $|U_{<i}| \leq t/9$. Our bound on $\mathrm{bad}_{\mathcal{D}}(U)$ from above thus gives:

$$\mathrm{bad}_{\mathcal{D}}(U) \leq \sum_{i\in[M]} |U_i| \cdot f\big(|U_i|\big) + \sum_{i\in[M]} \left(\frac{9|U_{<i}|}{t}\right)|U_i| \leq \sum_{i\in[M]} |U_i| \cdot f\big(|U_i|\big) + \sum_{i\in[2,M]} |U_i|.$$

For each $i \geq 2$ we have $|U_i| \leq |U_{<i}|$ from the ordering and so by applying Lemma 2.5 we get $|U_{<i}| \cdot f(|U_{<i}|) + |U_i| \leq |U_{<i+1}| \cdot f(|U_{<i+1}|)$. Repeatedly applying this as i increases gives us $\mathrm{bad}_{\mathcal{D}}(U) \leq |U| \cdot f(|U|)$, noting that $U_{<m+1} = U$. This completes the proof. □

Let us remark that combining the two cases in the proof above yealds the lower bound $|U| \geq 2^{-25} k \log_2^{-2}(k+1)$ for our desired set U.

We are finally able to prove Theorem 5.1. The proof proceeds in a very similar way to that of Theorem 5.4 but, as the details are involved, for completeness we will go through it with care.

Proof of Theorem 5.1. We will prove a slightly more convenient statement, namely that given the hypothesis there is a set $U \subset V(G)$ and a distribution $\mathcal{D}$ on $[0.1, 0.9]^{V(G)}$ such that $|U| \geq 2^{-26}(k + k^{3/4}) \log_2^{-2}(k+1)$ and $\mathrm{bad}_{\mathcal{D}}(U) \leq 8|U| \log_2 |U|$. We will prove this by induction on $|V(G)|$. Note that the theorem trivially holds in the first case where the hypothesis applies, when $n = 20000$ and $k = 1$ (taking U to be any sets of size 1 and $\mathcal{D}$ the trivial distribution). Also, as in Theorem 5.9, when k is small there is nothing to prove, so we can assume $k \geq 2^{25}$.

Let V_0 be a largest vertex set of G such that $|\operatorname{div}(u, v)| \geq 2k^{3/2}$ for all $u, v \in V_0$. If $|V_0| \geq k+1$ then we are done by Lemma 5.2, otherwise assume that $V_0 = \{v_1, v_2, \dots, v_L\}$ for some $L \leq k$ and for each $i \in [L]$ define the set $V_i := \{v \in V(G) : |\operatorname{div}(v, v_i)| < 2k^{3/2}\}$. Due to the maximality of S_0 we get $V(G) = \bigcup_{i=1}^{L} V_i$. The proof splits into the two already familiar cases:

Case I: Every $j \in [L]$ with $d_G(v_j) \in [nk^{-1/3}, n - 1 - nk^{-1/3}]$ satisfies $|V_j| \leq 3k$.

We have seen that at most $3k^2$ vertices of G do not lie in a set V_j of size at least $3k$. Moreover, $d_G(v_i) - 2k^{3/2} < d_G(v) < d_G(v_i) + 2k^{3/2}$ for all $i \in [L]$ and $v \in V_i$. Hence $d_G(v) \notin [nk^{-1/3} + 2k^{3/2}, n - 1 - nk^{-1/3} - 2k^{3/2}]$ for at least $n - 3k^2$ of the vertices $v \in V(G)$. Therefore for all such vertices we have $d_G(v) \leq nk^{-1/3} + 2k^{3/2}$ or $d_G(v) \geq n - 1 - nk^{-1/3} - 2k^{3/2}$. We will assume that at least half of these vertices satisfy the first condition, as otherwise we can follow an identical argument working with $\overline{G}$ instead. Consequently we find a set $V \subset V(G)$ with $|V| \geq (n - 3k^2)/2 \geq 12n/25$ and $\Delta(G[V]) \leq nk^{-1/3} + 2k^{3/2} < 4|V|k^{-1/3}$. Moreover, $|V| \geq 8000k^2$ and $\hom(G[V]) \leq \hom(G) \leq n/25k \leq |V|/12k$, hence we can apply Theorem 5.9 to $G[V]$ to obtain a distribution $\mathcal{D}_1$ on $[0.1, 0.9]^V$ and a vertex set $U \subset V$ with $\mathrm{bad}_{\mathcal{D}}(U) \leq 8|U| \log_2 |U|$ and $|U| \geq 2^{-25} k \log_2^{-2}(k+1) > 2^{-26}(k + k^{3/4}) \log_2^{-2}(k+1)$. We also set $\mathcal{D}_0 := \mathcal{T}_{V(G) \setminus V}$, i.e. the trivial distribution induced by $V(G) \setminus V$, and let $\mathcal{D} := \mathcal{D}_0 \times \mathcal{D}_1$ denote the product distribution on $[0.1, 0.9]^{V(G)}$. By making use of Lemma 4.1 we can, once again, obtain $\mathrm{bad}_{\mathcal{D}}(U) \leq \mathrm{bad}^V_{\mathcal{D}_1}(U) \leq 8|U| \log_2 |U|$, as required.

Case II: There is $j \in [L]$ such that $d_G(v_j) \in [nk^{-1/3}, n - 1 - nk^{-1/3}]$ and $|V_j| \geq 3k$.

As we did in Theorem 5.4, pick a subset V of V_j of size $3k$ such that $v_j \in V$ and set $X_1 := N(v_j) \setminus V$ and $X_2 =: V(G) \setminus (V \cup N(v_j))$. Then both $|X_1|, |X_2| \geq nk^{-1/3} - 3k$.

The same double counting argument from Theorem 5.4 works here to give us the sets $Y_1 = \{u \in X_1 : d_G^V(u) \geq 2k\}$ and $Y_2 = \{u \in X_2 : d_G^V(u) \leq k\}$, both of size at least $nk^{-1/3} - 3k - 6k^{3/2} > 4096k^{3/2} > 2^{48}$, such that $|X_i \setminus Y_i| \leq 6k^{3/2}$ for each $i \in \{1, 2\}$. Since $V(G) = V \cup X_1 \cup X_2$ is a partition, this shows that $Z := V(G) \setminus (Y_1 \cup Y_2)$ satisfies $|Z| \leq 2 \cdot 6k^{3/2} + 3k \leq 15k^{3/2}$.

To complete the proof we will apply the induction hypothesis to both Y_1 and Y_2. Let $t_i := |Y_i|$ for $i \in \{1, 2\}$ and set $k_i := k(t_i/n) \leq k$. This gives us $\hom(G[Y_i]) \leq \hom(G) \leq n/25k = t_i/25k_i$. We also have $t_i k_i^{-2} \geq k_i^{-1}(t_i/k_i) = k_i^{-1}(n/k) \geq nk^{-2} \geq 20000$. Therefore, for each $i = 1, 2$ we can apply the induction hypothesis to $G[Y_i]$ to find a probability distribution $\mathcal{D}_i$ on $[0.1, 0.9]^{Y_i}$ and a set $U_i \subset Y_i$ which satisfies $|U_i| \geq 2^{-26}(k_i + k_i^{3/4}) \log_2^{-2}(k_i + 1)$ and:

$$\mathrm{bad}_{\mathcal{D}_i}(U_i) \leq |U_i| \cdot f(|U_i|), \quad \text{where } f(x) := 8 \log_2 x. \tag{14}$$

Let us remark that if $k_i < 2^{33}$ then any set $U_i \subset Y_i$ of size $2 \geq 2^{-25} k_i \log_2^{-2}(k_i + 1)$ trivially satisfies (14), as already noted many times before, thus the above computations all make sense.

We can also assume that $\max\{|U_1|, |U_2|\} \leq k$, as otherwise the theorem follows immediately by just taking U to equal one of these sets.

Next, let $\mathcal{D}_0 := \mathcal{U}_V$ denote the uniformly constant distribution on $[0.1, 0.9]^V$ and pick $\mathcal{D}_3 := \mathcal{T}_Z$, the trivial Z-induced distribution. Setting $U := U_1 \cup U_2$, let $\mathcal{D}$ denote the product distribution $\prod_{i \in [0,3]} \mathcal{D}_i$ on $[0.1, 0.9]^V \times \prod_{i \in [2]} [0.1, 0.9]^{Y_i} \times [0.1, 0.9]^Z = [0.1, 0.9]^{V(G)}$.

Note that $d_G^V(u) \geq 2k \geq d_G^V(v) + k$ for all $u \in Y_1$ and $v \in Y_2$, by definition of Y_1 and Y_2. It then follows from Lemma 4.2 that for all such vertices we have:

$$\mathrm{bad}_{\mathcal{D}_0}^V(u, v) \leq \frac{3}{k}. \tag{15}$$

As $V(G) = Y_1 \cup Y_2 \cup Z$ is a partition and $|Z| \leq 15k^{3/2}$, we get $t_1 + t_2 \geq n - 15k^{3/2}$, therefore we obtain $k_1 + k_2 \geq k - 15n^{-1}k^{5/2} \geq k - \sqrt{k}/300$. Moreover, by recalling that $t_i \geq nk^{-1/3} - 3k - 6k^{3/2} > nk^{-1/3}/2$, we deduce that $k_i = k(t_i/n) \geq k^{2/3}/2$ for $i \in \{1, 2\}$. By Lemma 2.4 we get that:

$$k_1^{3/4} + k_2^{3/4} \geq \frac{\sqrt{k}}{\sqrt[4]{8}} + \left(k - \frac{\sqrt{k}}{300} - \frac{\sqrt[3]{k^2}}{2}\right)^{3/4} > \frac{\sqrt{k}}{2} + k^{3/4} \cdot \left(1 - \frac{2k^{-1/3}}{3}\right)^{3/4}.$$

Using the inequalities $1 - t \geq \exp(-2t)$ and $\exp(-t) \geq 1 - t$, which hold for any $t \in [0, 0.5]$ and in particular for $t = \Theta(k^{-1/3})$, we can further deduce that:

$$k_1^{3/4} + k_2^{3/4} \geq \frac{\sqrt{k}}{2} + k^{3/4} \cdot \exp\left(-k^{-1/3}\right) > \frac{\sqrt{k}}{2} + k^{3/4} - k^{-5/12} > k^{3/4} + \frac{\sqrt{k}}{300}. \tag{16}$$

We are now in a position to lower bound the size of U:

$$\begin{aligned}|U| = |U_1| + |U_2| &\geq \frac{1}{2^{26}} \cdot \left(\frac{k_1 + k_1^{3/4}}{\log_2^2(k_1+1)} + \frac{k_2 + k_2^{3/4}}{\log_2^2(k_2+1)} \right) \geq \\ &\geq \frac{1}{2^{26}\log_2^2(k+1)} \cdot \left(k_1 + k_1^{3/4} + k_2 + k_2^{3/4}\right) \overset{(16)}{\geq} \\ &\geq \frac{1}{2^{26}\log_2^2(k+1)} \cdot \left(k_1 + k_2 + \frac{\sqrt{k}}{300} + k^{3/4}\right) \geq \frac{k + k^{3/4}}{2^{26}\log_2^2(k+1)}.\end{aligned}$$

Finally, we are able to estimate $\mathrm{bad}_{\mathcal{D}}(U)$ as follows:

$$\begin{aligned}\mathrm{bad}_{\mathcal{D}}(U) &= \sum_{\{u,v\}\subset U_1} \mathrm{bad}_{\mathcal{D}}(u,v) + \sum_{\{u,v\}\subset U_2} \mathrm{bad}_{\mathcal{D}}(u,v) + \sum_{(u,v)\in U_1\times U_2} \mathrm{bad}_{\mathcal{D}}(u,v) \\ &= \mathrm{bad}_{\mathcal{D}}(U_1) + \mathrm{bad}_{\mathcal{D}}(U_2) + |U_1||U_2| \cdot \max_{(u,v)\in U_1\times U_2} \left\{ \mathrm{bad}_{\mathcal{D}}(u,v) \right\} \\ &\leq \mathrm{bad}_{\mathcal{D}_1}(U_1) + \mathrm{bad}_{\mathcal{D}_2}(U_2) + |U_1||U_2| \cdot \max_{(u,v)\in U_1\times U_2} \left\{ \mathrm{bad}^V_{\mathcal{D}_0}(u,v) \right\} \\ &\leq |U_1| \cdot f(|U_1|) + |U_2| \cdot f(|U_2|) + \frac{3}{k} \cdot |U_1||U_2| \leq |U| \cdot f(|U|).\end{aligned}$$

The final three inequalities here respectively follow from Lemma 4.1, then from (14) and (15), and lastly using that $\max\{|U_1|, |U_2|\} \leq k$ and Lemma 2.5. This completes the proof. □

6. Distinct degrees in random graphs

In this section we will study $f(G(n,p))$, the number of distinct degrees which can be found in an induced subgraph of the Erdős–Rényi random graph $G(n,p)$. Our results extend the estimates for the case of constant p due to Bukh and Sudakov [9] and to Conlon, Morris, Samotij and Saxton [10]. We restate Theorem 1.3 for the reader's convenience.

Theorem 1.3. *Let $n \in \mathbb{N}$ and let $p := p(n) \in [0, 1/2]$. Then whp the random graph $G(n,p)$ satisfies the following:*

(i) $f\big(G(n,p)\big) = \Theta\left(\sqrt[3]{pn^2}\right)$ *for $p \in [n^{-1/2}, 1/2]$;*
(ii) $f\big(G(n,p)\big) = \Theta\big(\Delta(G(n,p))\big)$ *for $p \leq n^{-1/2}$.*

Although the estimation of $f(G(n,p))$ is quite natural in itself, we believe, as discussed in the concluding remarks, that the behaviour for $p \in [n^{-1/2}, 1/2]$ essentially determines the extremal relationship between $\hom(G)$ and $f(G)$ beyond the range of the Narayanan–Tomon conjecture, when $\hom(G) < n^{1/2}$. As a result, our calculations will focus on the case (i) of Theorem 1.3. The next subsection contains the proof the upper bound on $f(G(n,p))$ in this case, whereas the second subsection contains the more difficult lower bound. In the final subsection we briefly indicate how to approach the case when $p \leq n^{-1/2}$.

6.1. Upper bound on $f(G(n,p))$

In this subsection we prove the upper bound on $f(G(n,p))$. Our approach closely follows that of Bukh and Sudakov (see Proposition 2.4 in [9]), but we include the complete details, as the estimates are more involved in the sparse case.

Proposition 6.1. *When $n \in \mathbb{N}$ and $p \in [n^{-1/2}, 1/2]$, then $f\big(G(n,p)\big) = O\big(\sqrt[3]{pn^2}\big)$ whp.*

Proof. Suppose $G \sim G(n,p)$ has a subset $A \subset V(G)$ of size a such that $G[A]$ has $8b$ distinct degrees, where $b = 16\sqrt[3]{pn^2}$. As at most $6b-1$ of our distinct degrees can lie in the interval $(pa-3b, pa+3b)$, either there are at least b vertices of A that have degree at least $pa+3b$ or at least b vertices that have degree at most $pa-3b$.

We will assume first that we are in the former case, as this is the more intricate one. Let $B \subset A$ be a set of b vertices which all have degree at least $pa+3b$. Let us now look at the number $e(A,B)$ of edges with one endpoint in A and one in B (those with both their endpoints in B will be counted twice since $B \subset A$). We then have $pab + 3b^2 \leq e(A,B) = 2e(B) + e(A \setminus B, B)$. As $|B| = b$ implies $e(B) < b^2$, we find that:

$$e(A \setminus B, B) \geq pab + b^2 \geq p(a-b)b + b^2. \tag{17}$$

Letting F denote the event that there are sets A and B which satisfy (17), it suffices to show that $\mathbb{P}(F) = o(1)$. To see this, first suppose that $16p(a-b) \geq b$. As $\mathbb{E}[e(A \setminus B, B)] = p(a-b)b$, by using Chernoff's Inequality with $\delta = b/16p(a-b) \leq 1$ we get that:

$$\begin{aligned}\mathbb{P}\left(e(A \setminus B, B) \geq pb(a-b) + b^2\right) &\leq \mathbb{P}\left(e(A \setminus B, B) \geq pb(a-b) + 2^{-4}b^2\right) \\ &\leq \exp\left(\frac{-b^3}{2^{10}p(a-b)}\right) \leq \exp\left(\frac{-b^3}{2^{10}pn}\right),\end{aligned}$$

where the final inequality uses that $a-b \leq a \leq n$. Therefore, the union bound implies that event F can happen with probability at most:

$$\mathbb{P}(F) \leq 2^n \cdot \binom{n}{b} \cdot \exp\left(\frac{-b^3}{4 \cdot pn}\right) \leq 2^{2n} \cdot \exp\left(\frac{-b^3}{2^{10} \cdot pn}\right).$$

This tends to zero as $n \to \infty$, as $b \geq 16\sqrt[3]{pn^2}$.

Now suppose instead that $16p(a-b) < b$. As $e(A \setminus B, B)$ has binomial distribution, we have:

$$\mathbb{P}\left(e(A \setminus B, B) \geq pb(a-b) + b^2\right) \leq \binom{b(a-b)}{b^2} \cdot p^{b^2} \leq \left(\frac{2ep(a-b)}{b}\right)^{b^2} \leq 2^{-b^2}.$$

Recalling that $p \geq n^{-1/2}$ and that $b \geq 16\sqrt[3]{pn^2} \geq 16\sqrt[3]{n^{-1/2}n^2} \geq 16\sqrt{n}$, using the union bound we find that the event F occurs with probability at most:

$$\mathbb{P}(F) \leq 2^n \binom{n}{b} 2^{-b^2} \leq 2^{2n} \cdot 2^{-(16\sqrt{n})^2}.$$

Hence, it follows again that $\mathbb{P}(F) = o(1)$ in this second case.

Finally, if there is a set $B \subset A$ of b vertices that all have degree at most $pa - 3b$ in $G[A]$, then $e(A \setminus B, B) \leq e(A, B) \leq b(pa - 3b) \leq pb(a-b) - b^2$. If $p(a-b) < b$ then it is clear that such a set B exists with 0 probability since we cannot have a negative number of edges. Otherwise we just apply Chernoff's Inequality for $\delta = \dfrac{b}{p(a-b)} \leq 1$ to get:

$$\mathbb{P}\left(e(A \setminus B, B) \leq pb(a-b) - b^2\right) \leq \exp\left(-\frac{b^3}{2p(a-b)}\right) \leq \exp\left(-\frac{b^3}{2pn}\right),$$

where the last inequality follows as $a - b \leq a \leq n$. We have seen before that the union bound gives us probability of at most $2^n \cdot \binom{n}{b} \cdot \exp\left(-b^3/2pn\right)$ for such a set B to exist and we have shown in the previous case that this probability tends to 0 as $n \to \infty$. □

6.2. Lower bound on $f(G(n,p))$

We now focus on proving our sharp lower bound for $f(G(n,p))$. Before we start, we will present a few results that will help us along the way.

Given $D > 0$ and a graph G, we call a set $U \subset V(G)$ *D-diverse* if it is D-diverse to $V(G)$. We say that the graph G is *D-diverse* if $V(G)$ is D-diverse (see Section 4 before Lemma 4.3).

Proposition 6.2. *If $p \gg \log n/n$ then all vertices of $G(n,p)$ have degree asymptotic to np whp. In particular, whp they all have degrees less than $2np$.*

Proof. Let u be a vertex of $G(n,p)$. Then $d_{G(n,p)}(u) \sim Bin(n-1,p)$, so we can apply Chernoff's Inequality for $\delta = 3\sqrt{\dfrac{\log n}{np}}$ to get $\mathbb{P}\left(|d(u) - np| \geq 3\sqrt{np\log n}\right) \leq 2n^{-9/2}$. The result now follows by the union bound. The last part is a consequence of the fact that $\delta \leq 1$. □

Lemma 6.3. *If $p \gg \log n/n$ and $p \leq 1/2$ then whp $G(n,p)$ is $p(n-1)$-diverse.*

Proof. Let us first notice that $|\text{div}(u,v)| \sim Bin(n-2,q)$ for all distinct $u, v \in V(G)$, where $q := 2p(1-p)$. This holds as every vertex of $\text{div}(u,v)$ has to be in $N(u) \setminus N(v)$

or in $N(v) \setminus N(u)$ and these two possibilities represent disjoint events that happen with probability $p(1-p)$.

Note that $q \geq p \gg \log n/n$ and so we can apply Chernoff's Inequality for $\delta = 3\sqrt{\dfrac{\log n}{nq}}$ to obtain $\mathbb{P}\left(|\mathrm{div}(u,v)| \leq nq - 3\sqrt{nq \log n}\right) \leq n^{-9/2}$. By the union bound, we therefore deduce that $|\mathrm{div}(u,v)| > nq(1-\delta)$ for any $u, v \in V(G)$ whp. Since $q \geq p$ and $\delta \to 0$, we get that $G(n,p)$ is $p(n-1)$-diverse whp in this case. □

Lemma 6.4. *Given $n \in \mathbb{N}$ and $p \in [n^{-1/2}, 1/2]$, let $G \sim G(n,p)$ and let $V(G) := U \sqcup S$ be a vertex partition such that $\sqrt{n} \leq 4|U| \leq pn$. Then, with high probability, there is a subset $W \subset S$ such that U is $pn/3$-diverse to W and $d_G^U(w) \leq 10p|U|$ for all $w \in W$.*

Proof. Define the set $S_B := \{v \in V : d_G^U(v) \geq 10p|U|\}$. For any $v \in S$ the random variable $d_G^U(v)$ has distribution $Bin(|U|, p)$, hence, as $p|U| \geq 2.5$, we deduce that:

$$\tilde{p} := \mathbb{P}(v \in S_B) \leq (e/10)^{10p|U|} < 3^{-2.5} < 1/10.$$

Now for each subset $W \subset S$ we see that $|W \cap S_B|$ has distribution $Bin(|W|, \tilde{p})$. Therefore, for any $u_1, u_2 \in U$ we can deduce by using Theorem 2.7 and Lemma 6.3 that:

$$\mathbb{P}\left(|\mathrm{div}(u_1,u_2) \cap S_B| > \frac{|\mathrm{div}(u_1,u_2)|}{3}\right) \leq (3e\tilde{p})^{|\mathrm{div}(u_1,u_2)|/3} \leq \left(\frac{9}{10}\right)^{\sqrt{n}/4}.$$

Call a pair $\{u_1, u_2\} \subset U$ of vertices *big* if $|\mathrm{div}(u_1,u_2) \cap S_B| > |\mathrm{div}(u_1,u_2)|/3$. By using the union bound we immediately deduce that U contains a *bad* pair of vertices with probability at most $p^2n^2/8 \cdot (9/10)^{\sqrt{n}/4} \to 0$ as $n \to \infty$. Set now $W := S \setminus S_B$ and note that whp we have $|\mathrm{div}(u_1,u_2) \setminus S_B| > 2|\mathrm{div}(u_1,u_2)|/3 \geq 2p(n-1)/3$ for all distinct vertices $u_1, u_2 \in U$. We then get:

$$\left|N_G^W(u_1) \triangle N_G^W(u_2)\right| \geq \left|\mathrm{div}(u_1,u_2) \setminus S_B\right| - |U| \geq \frac{2p(n-1)}{3} - \frac{\sqrt{n}}{4} > \frac{pn}{3}.$$

The second property follows directly from the definition of W, proving our result. □

Definition 6.5. Let G be a n-vertex graph and let $0 < p \leq 1/2$. We call a set U of vertices of G *p-convenient* if $d_G(u) \leq 2pn$ for all $u \in U$ and there is a set $W \subset V(G) \setminus U$ such that U is $pn/3$-diverse to W and $d_G^U(w) \leq 10p|U|$ for all $w \in W$.

We now expose the randomness in $G(n,p)$ and obtain a fixed graph G. According to Proposition 6.2 and Lemma 6.4, we may assume that G contains a p-convenient set U of size $\sqrt[3]{pn^2}/4$.

At this point, the reader might have already noticed that the p-convenient conditions fit in very well with those from Lemma 4.3. Indeed, the set U is $pn/3$-diverse to W and $10p$-balanced, so the hypothesis of the lemma is satisfied. The most natural thing to do would now be to apply the lemma with the blended distribution $\mathcal{B}_\beta(U, W)$. However, in order to obtain a set of $\Theta(|U|)$ degrees, we would like the first term in the RHS of (5) to be of order $|U|^{-1}$, which forces $\beta := \Theta\left(|U|/pn\right)$. This would then make the second term in the RHS of (5) to be a constant, so it seems that we cannot get the desired 'bad' control. One can obtain weaker bounds on $f(G(n,p))$ by altering the parameters here, but there is an unavoidable loss as things stand.

There is though a way around this issue. In Lemma 4.3 we solve the problem of coordinates lying outside $[0.1, 0.9]$ by dealing with each pair of vertices $\{u_1, u_2\} \subset U$ individually. However, in certain situations it is possible to show that many vertices $u_1 \in U$ are *simultaneously* good for all pairs $\{u_1, u_2\} \subset U$. The crucial twist here is that the diversity term $D := pn/3$ satisfies $D = \Omega(\Delta(G))$. This allows us to guarantee that a fixed vertex $u_1 \in U$ whp is likely to have no neighbours in $\operatorname{div}(u_1, u_2)$ whose coordinates are 'outliers', and this happens *for all* $u_2 \in U$. The approach here builds upon that of Jenssen, Keevash, Long and Yepremyan [19].

A slight change in our notation will be convenient below. Given a graph G with vertex partition $V(G) = U \sqcup W$ and a probability vector $\underline{\mathbf{p}} = (p_w)_{w\in W} \in [0,1]^W$, we write $G(\underline{\mathbf{p}})$ to denote the probability space on the set of induced subgraphs of G *that contain* U, where for each vertex set $S \subset W$, the induced subgraph $G[U \cup S]$ is selected with probability $\prod_{v\in S} p_v \prod_{v\in W\setminus S}(1 - p_v)$.

Proposition 6.6. *Given $n \in \mathbb{N}$ and $p \in [n^{-1/2}, 1/2]$, let G be an n-vertex graph with a p-convenient set $U \subset V(G)$ of size $\sqrt[3]{pn^2}/4$. Then there is a vector $\underline{\mathbf{p}} \in [0.1, 0.9]^{V(G)\setminus U}$ and a set $U' \subset U$ with $|U'| \geq |U|/500$ so that $\left|\mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_1)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_2)]\right| \geq 1$ for all distinct vertices $u_1, u_2 \in U'$.*

Proof. We may assume that n is large enough so that all asymptotic bounds hold. First set $\beta := |U|/5pn < 0.1$ and let $S \subset V(G) \setminus U$ be such that U is $pn/3$-diverse to S and $d_G^U(v) \leq 10p|U|$ for all $v \in S$. As in the proof of Lemma 4.3, for each $u \in U$ define the random vector $\underline{\mathbf{q}}^u$ on $\mathbb{R}^S$ by $\underline{\mathbf{q}}^u := \underline{\mathbf{p}}' - \alpha_u \cdot \operatorname{proj}_S(\mathbf{u})$, where $\underline{\mathbf{p}}'$ is the vector from before the truncation in the definition of the blended distribution $\mathcal{B}_\beta(U, S)$. Recall we do this so that $\underline{\mathbf{q}}^u$ is independent of α_u.

We call a vertex $u \in U$ *good* if there are at most $d_G^S(u)/25$ coordinates $v \in S \cap N(u)$ so that $\underline{\mathbf{q}}_v^u \notin [0.2, 0.8]$. Let $U^g \subset U$ denote the set of *good* vertices.

We claim that $\mathbb{P}(|U^g| \geq |U|/2) > 1/2$. To prove this, take $u \in U$ and note that $\underline{\mathbf{q}}_v^u$ is a sum of at most $10p|U|$ uniform independent random variables, thus by applying Hoeffding's inequality we obtain:

$$\mathbb{P}\left(\underline{\mathbf{q}}_v^u \notin [0.2, 0.8]\right) = \mathbb{P}\left(|\underline{\mathbf{q}}_v^u - 1/2| > 0.3\right) \leq 2\exp\left(\frac{-50 \cdot 0.09 \cdot p^2 n^2}{10p|U|^3}\right) = 2e^{-7.2} < \frac{1}{100}.$$

We deduce that the expected number of coordinates $v \in V \cap N(u)$ with $\underline{\mathbf{q}}_v^u \notin [0.2, 0.8]$ is at most $d_G^V(u)/100$. By Markov we get that the vertex u is not good with probability less than 1/4. Therefore, the expected number of vertices $u \in U$ that are not good is at most $|U|/4$ and the claim follows from a simple application of Markov's Inequality.

We now set $T := V(G) \setminus (S \cup U)$ and let $\mathcal{T}_T$ denote the trivial T-induced distribution. Define $\mathcal{D}$ to be the product distribution $B_\beta(U, S) \times \mathcal{T}_T$ on $[0.1, 0.9]^{S \cup T}$ and for distinct vertices $u_1, u_2 \in U$ let E_{u_1,u_2} denote the event that $\left|\mathbb{E}_{\underline{\mathbf{p}} \sim \mathcal{D}}\left[d_{G(\underline{\mathbf{p}})}(u_1) - d_{G(\underline{\mathbf{p}})}(u_2)\right]\right| \leq 1$. Moreover, since U is $pn/3$-diverse to S, we know that either $|N^S_G(u_1) \setminus N^S_G(u_2)| \geq pn/6$ or $|N^S_G(u_2) \setminus N^S_G(u_1)| \geq pn/6$. We set $m_S(u_1, u_2) := u_1$ in the first case and $m_S(u_1, u_2) := u_2$ in the second one.

Our next claim is that $\mathbb{P}\big(E_{u,u'} \mid m_S(u, u') \in U^g\big) \leq 120|U|^{-1}$ for all $u \neq u'$ in U. To prove it, we can assume that $u = m_S(u, u')$. As $u \in U^g$, at most $2pn/25$ vertices in $N_G(u)$ represent coordinates v such that $\underline{\mathbf{q}}^u_v \notin [0.2, 0.8]$. Therefore, we can find a subset $Y \subset N^S_G(u) \setminus N^S_G(u')$ of size $pn/12$ such that $\underline{\mathbf{q}}^u_v \in [0.2, 0.8]$ for all $v \in Y$. Since $\underline{\mathbf{p}}'_v = \underline{\mathbf{q}}^u_v + \alpha_u \mathbf{u}_v$ and $|\alpha_u| < 0.1$, we deduce that no Y-coordinate of $\underline{\mathbf{p}}'$ gets truncated when creating $\underline{\mathbf{p}} \sim B_\beta(U, S)$. Condition now on any choice of $\boldsymbol{\alpha} := (\alpha_w)_{w \neq u}$ such that $u \in U^g$ and note that α_u is independent of it.

By looking at the following expression (when $\underline{\mathbf{p}} \sim \mathcal{D}$) as a function of α_u:

$$\begin{aligned}\mathbb{E}[d_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(u')] &= \text{ constant } + \mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d^S_{G(\underline{\mathbf{p}})}(u')] \\ &= \text{ constant } + (\mathrm{proj}_S(\mathbf{u}) - \mathrm{proj}_S(\mathbf{v})) \cdot \mathrm{proj}_S(\underline{\mathbf{p}})\end{aligned}$$

we observe that $E_{u,u'}$ holds provided that, conditioned on $\boldsymbol{\alpha}$, this difference lies in an interval of length 2. The same argument as in Lemma 4.3 gives us that $E_{u,u'}|\boldsymbol{\alpha}$ happens with probability at most $24(pn\beta)^{-1} = 120|U|^{-1}$. The claim follows from the law of total probability.

To complete the proof, consider the graph J on the vertex set U^g where $u_1u_2 \in E(J)$ if E_{u_1,u_2} holds. By the second claim we get $\mathbb{E}[e(J)] \leq 120|U|^{-1} \cdot |U^g|(|U^g| - 1)/2 < 60|U|$, thus by Markov $\mathbb{P}\big(e(J) > 120|U|\big) < 1/2$. It follows that $\mathbb{P}\big(e(J) \leq 120|U|\big) > 1/2$ and recall that $\mathbb{P}(|U^g| \geq |U|/2) > 1/2$. Therefore, with positive probability, we can choose $\underline{\mathbf{p}} \sim \mathcal{D}$ such that $|U^g| \geq |U|/2$ and $e(J) \leq 120|U|$. For such a choice, the average degree of the resulting graph J is $4e(J)/2|U_g| \leq 4e(J)/|U| \leq 480$. Thus, by Turán's Theorem J has an independent set of size at least $|U|/500$. This independent set in J is precisely what we required. □

Proposition 6.7. *Let G be a n-vertex graph and let $p \in [n^{-1/2}, 1/2]$. Suppose that there is a p-convenient set $U \subset V(G)$ in G, a vector $\underline{\mathbf{p}} \in [0.1, 0.9]^{V(G) \setminus U}$ and a vertex subset $U' \subset U$ so that $\left|\mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_1)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_2)]\right| \geq 1$ for all distinct vertices $u_1, u_2 \in U'$. Then $f(G) = \Omega(|U'|)$.*

Proof. We can assume n is sufficiently large. Let H be a random induced subgraph selected according to $G(\underline{\mathbf{p}})$ and define for it the following sets:

$$B = \{u \in U' : \left|d_H(u) - \mathbb{E}[d_{G(\underline{\mathbf{p}})}]\right| \leq \sqrt{2pn}\},$$

$$P = \{\{u, u'\} \subset U' : \left|\mathbb{E}[d_{G(\underline{\mathbf{p}})}(u)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(u')]\right| \leq 2\sqrt{2pn}\},$$
$$J = \{\{u, u'\} \in P : d_H(u) = d_H(u')\}.$$

Our first claim is that $\mathbb{P}(|B| \geq |U'|/2) \geq 1/2$. To prove it, we start by estimating $|B|$. For any $u \in U'$ we have $\mathbb{V}\mathrm{ar}\big(d_{G(\underline{\mathbf{p}})}(u)\big) = \sum_{v\sim u} p_v(1-p_v) \leq pn/2$, thus Chebyshev's Inequality implies that $\mathbb{P}(u \notin B) \leq 1/4$. Therefore $\mathbb{E}[|U' \setminus B|] \leq |U'|/4$, so by Markov's inequality we get that $\mathbb{P}(|U' \setminus B| \geq |U'|/2) \leq 1/2$, which is equivalent to our claim.

We now want to estimate $|J|$. First note that the separation in expected degree for U' implies that $|P| \leq 2|U'|\sqrt{2pn}$. Each $\{u, u'\}$ belongs to J with probability equal to $\mathbb{P}\big(d_H(u) - d_H(u') = 0\big)$, which we claim is $O\left(1/\sqrt{pn}\right)$. This holds true because $d_H(u) - d_H(u') = \sum \xi_v X_v$, where the sum is taken over all $v \in \mathrm{div}(u, u') \setminus U$, $\xi_v \in \{-1, 1\}$ and $X_v \sim Be(p_v)$ measures whether $v \in V$ is being picked as a vertex of H or not. As U is $pn/3$-diverse to some subset $S \subset V(G) \setminus U$, we are able to deduce that $|\mathrm{div}(u, u') \setminus U| \geq |N_G^S(u) \triangle N_G^S(u')| \geq pn/3$, so we can apply Theorem 2.3 to prove the previous claim. Therefore $\mathbb{E}[|J|] \leq |P| \cdot \max_{\{u,u'\}\in P} \mathbb{P}\big(d_H(u) = d_H(u')\big) = O(|U'|)$.

It follows that $\mathbb{P}\big(|J| = O(|U'|)\big) > 1/2$ by Markov, so together with the first claim, we are able to deduce that both $|J| = O(|U'|)$ and $|B| \geq |U'|/2$ happen with positive probability. The end of the proof follows the same idea as before: make a choice of H for which this happens and by Turán's Theorem the graph $J[B]$ obtained by building edges between the vertices of B which have equal degree in H has an independent set of size $\Omega(|U'|)$. This set must consist of vertices with distinct degrees in H, as if $u, u' \in B$ and $d_H(u) = d_H(u')$ then $\{u, u'\} \in P$ and so $\{u, u'\} \in J$, representing an edge in $J[B]$. □

With all these ingredients, we are finally able to prove the following:

Theorem 6.8. *Given $n \in \mathbb{N}$ and $p \in [n^{-1/2}, 1/2]$, one has $f\big(G(n, p)\big) = \Omega\big(\sqrt[3]{pn^2}\big)$ whp.*

Proof. We expose the randomness in $G(n, p)$ and thus move to a fixed graph G. According to Proposition 6.2 and Lemma 6.4, we can find a p-convenient set U in G of size $\sqrt[3]{pn^2}/4$. We then apply Proposition 6.6 to find a vector $\underline{\mathbf{p}} \in [0.1, 0.9]^{V(G)\setminus U}$ and a subset $U' \subset U$ of size $\Omega\left(\sqrt[3]{pn^2}\right)$ so that $\left|\mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_1)] - \mathbb{E}[d_{G(\underline{\mathbf{p}})}(u_2)]\right| \geq 1$ for all distinct $u_1, u_2 \in U'$. Lastly, Proposition 6.7 allows us to convert a constant proportion of the distinct expected degrees in U' to genuine distinct degrees, thus completing the proof. □

6.3. $f(G(n,p))$ when $p \ll n^{-1/2}$

For completeness, in this subsection we briefly analyse $f(G(n, p))$ when $p \ll n^{1/2}$. First, to see that there is a change in behaviour here over the range $p \in [n^{-1/2}, 1/2]$, note that if $G \sim G(n, p)$ with $\log n/n \ll p \ll n^{-1/2}$ then a simple concentration argument combined with the union bound shows that whp $d_G(u) = O(pn)$ for all vertices $u \in V(G)$, implying that $f(G) = O(pn) = o(\sqrt[3]{pn^2})$ in this case.

As indicated in Theorem 1.3 (ii), in this regime the maximum degree of $G(n,p)$ is a key parameter. The following simple proposition is useful here.

Proposition 6.9. *Let G be an n-vertex graph and let $U \subset V(G)$ with $|U| = k$ such that $|N_G(u) \setminus \big(U \cup \cup_{u' \in U \setminus \{u\}} N(u')\big)| \geq k$ for all $u \in U$. Then $f(G) \geq k$.*

Proof. Let $U := \{u_1, u_2 \ldots, u_k\}$ such that $|N_G(u_i) \cap U|$ is non-decreasing with i. For each $i \in [k]$ take $S_i \subset N_G(u_i) \setminus \big(U \cup \cup_{u' \in U \setminus \{u\}} N(u')\big)$ with $|S_i| = i$ – by the hypothesis such sets exist. It is now easy to see that the degrees of the vertices $u_1, u_2, \ldots, u_k$ are strictly increasing in the induced subgraph $G\big[U \cup (\cup_{i \in [k]} S_i)\big]$, giving $f(G) \geq k$, as required. □

The following observations show that $f(G(n,p)) = \Theta\big(\Delta(G(n,p))\big)$ in this regime.

(i) If $\log n/n \ll p \leq n^{-1/2}$ then whp $d_G(u) \in [pn/2, 2pn]$ for all $u \in V(G)$, therefore we obtain that $\Delta(G(n,p)) = \Theta(np)$;

(ii) If $\log n/n \ll p \leq n^{1/2}$ then given any fixed set $U \subset V(G(n,p))$ with $|U| = pn/8$ and $u \in U$ we have $\mathbb{E}\big[N_G(u) \setminus (U \cup \cup_{u' \in U \setminus \{u\}} N(u'))\big] = p(n - |U|)(1-p)^{|U|} \geq pn/4$. Chernoff's Inequality then implies that $|N_G(u) \setminus (U \cup \cup_{u' \in U \setminus \{u\}} N(u'))| \geq pn/8 = |U|$ for all $u \in U$ whp. Then Proposition 6.9 together with observation (i), combine to help us deduce that $f(G) \geq |U| = \Theta(\Delta(G(n,p)))$ for $\log n/n \ll p \leq n^{-1/2}$ whp.

(iii) If $0 \leq p \leq O(\log n/n)$ then the random graph $G(n,p)$ has $\Omega(\Delta(G(n,p)))$ vertices of degree $\Omega(\Delta(G(n,p)))$ whp (e.g. see Theorem 3.1 in [6]). It is therefore possible to find a set U of $c \cdot \Delta(G(n,p))$ vertices with degree at least $5|U|$, provided that $c > 0$ is sufficiently small.

(iv) It is also true that if $p \leq n^{-3/4}$ then whp $|N(u) \cap N(u')| \leq 3$ for all pairs of distinct vertices $u, u' \in V(G(n,p))$. With U chosen as in observation (iii), it follows that $|N(u) \setminus (U \cup \cup_{u' \in U \setminus \{u'\}} N(u'))| \geq |N(u)| - |U| - 3|U| \geq |U|$. We therefore get $f(G(n,p)) = \Theta(\Delta(n,p))$ for $p = O(\log n/n)$.

7. Concluding remarks

Theorem 1.1 proves an essentially sharp dependence between $\hom(G)$ and $f(G)$ for n-vertex graphs with $\hom(G) \geq n^{1/2}$, which asymptotically resolves a conjecture of Narayanan and Tomon from [26]. It would be appealing to further remove the logarithmic terms here.

Another perhaps more compelling problem is to understand the relationship between these parameters when $\hom(G) < n^{1/2}$. Recall that Theorem 1.3 gives:

$$f\big(G(n,p)\big) = \begin{cases} \Theta\left(\sqrt[3]{pn^2}\right) & \text{for } p \in [n^{-1/2}, 1/2]; \\ \Theta\big(\Delta(G(n,p))\big) & \text{for } p \in [0, n^{-1/2}]. \end{cases}$$

It is well known that $\hom(G(n,p)) \sim -\log n/\log(1-p)$ (see e.g. [8]) when $0 < p \le 1/2$ is a fixed constant. For a general $p := p(n) \le 1/2$, the probability of having a set of size k which is homogeneous in $G(n,p)$ is at most:

$$\binom{n}{k}\left(p^{\binom{k}{2}} + (1-p)^{\binom{k}{2}}\right) \le 2n^k(1-p)^{\binom{k}{2}} \le 2n^k e^{-p\binom{k}{2}} = 2\left(ne^{-p(k-1)/2}\right)^k.$$

In particular, $\hom(G(n,p)) \le 4p^{-1}\log n$ whp. Combined with the bounds for $f(G(n,p))$ from Theorem 1.3 we find that for $p \in [n^{-1/2}, 1/2]$ we have:

$$f(G(n,p)) = \widetilde{\Omega}\left(\sqrt[3]{\frac{n^2}{\hom(G(n,p))}}\right) \quad \text{whp.}$$

We believe that a similar bound holds for any n-vertex graph G with $\hom(G) < n^{1/2}$.

Conjecture 7.1. *If G is an n-vertex graph then:*

$$f(G) \ge \min\left(\sqrt[3]{\frac{n^2}{\hom(G)}}, \frac{n}{\hom(G)}\right) n^{-o(1)}.$$

Observe that the minimum above changes exactly when $\hom(G) = n^{1/2}$, value after which the Narayanan–Tomon conjecture begins to apply. Theorem 1.1 proves it for $\hom(G) \ge n^{1/2}$. Theorem 1.3 shows that this behaviour is essentially tight for $G(n,p)$ when $p = n^{-1/2}$, when $\hom(G(n,p)) = n^{1/2+o(1)}$. At the opposite extreme, n-vertex graphs with $\hom(G)$ as small as possible (Ramsey graphs) were proven by Jenssen et al. in [19] to have $f(G) = \Omega(n^{2/3})$, and so the conjecture is true at both ends of the interval $\hom(G) \in [\Omega(\log n), n^{1/2}]$.

Data availability

No data was used for the research described in the article.

Acknowledgment

We would like to thank the referees for their careful reading of the paper, and for a number of useful suggestions.

References

[1] N. Alon, S. Friedland, G. Kalai, Regular subgraphs of almost regular graphs, J. Comb. Theory, Ser. B (ISSN 0095-8956) 37 (1) (1984) 79–91, https://doi.org/10.1016/0095-8956(84)90047-9, https://www.sciencedirect.com/science/article/pii/0095895684900479.

[2] Noga Alon, Michael Krivelevich, Benny Sudakov, Induced subgraphs of prescribed size, J. Graph Theory 43 (4) (2003) 239–251, https://doi.org/10.1002/jgt.10117, eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jgt.10117.

[3] Noga Alon, Michael Krivelevich, Benny Sudakov, Large nearly regular induced subgraphs, English (US), SIAM J. Discrete Math. (ISSN 0895-4801) 22 (4) (2008) 1325–1337, https://doi.org/10.1137/070704927.

[4] Alon Noga, Joel H. Spencer, The Probabilistic Method, second ed., Wiley, New York, ISBN 0471370460, 2004.

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

[6] Béla Bollobás, Random Graphs, 2nd ed., Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2001.

[7] B. Bollobá, Extremal Graph Theory, Dover Publications, USA, ISBN 0486435962, 2004.

[8] Béla Bollobás, Paul Erdös, Cliques in random graphs, Math. Proc. Camb. Philos. Soc. 80 (3) (1976) 419–427, https://doi.org/10.1017/S0305004100053056.

[9] Boris Bukh, Benny Sudakov, Induced subgraphs of Ramsey graphs with many distinct degrees, J. Comb. Theory, Ser. B (ISSN 0095-8956) 97 (4) (2007) 612–619, https://doi.org/10.1016/j.jctb.2006.09.006, https://www.sciencedirect.com/science/article/pii/S0095895606001080.

[10] D. Conlon, R. Morris, W. Samotij, D. Saxton, The number of distinct degrees in an induced subgraph of a random graph, unpublished.

[11] David Conlon, Jacob Fox, Benny Sudakov, Recent developments in graph Ramsey theory, in: Artur Czumaj, Agelos Georgakopoulos, Daniel Král, Vadim Lozin, Oleg Pikhurko (Eds.), Surveys in Combinatorics 2015, in: London Mathematical Society Lecture Note Series, Cambridge University Press, 2015, pp. 49–118.

[12] P. Erdös, On a lemma of Littlewood and Offord, Bull. Am. Math. Soc. 51 (12) (1945) 898–902.

[13] P. Erdös, Some of my favourite problems in various branches of combinatorics, Matematiche 47 (2) (1992) 231–240, https://lematematiche.dmi.unict.it/index.php/lematematiche/article/view/587.

[14] P. Erdös, Some remarks on the theory of graphs, Bull. Am. Math. Soc. 53 (4) (1947) 292–294.

[15] P. Erdös, G. Szekeres, A combinatorial problem in geometry, in: Ira Gessel, Gian-Carlo Rota (Eds.), Classic Papers in Combinatorics, Birkhäuser Boston, Boston, MA, ISBN 978-0-8176-4842-8, 1987, pp. 49–56.

[16] P. Erdös, A. Szemerédi, On a Ramsey-type theorem, Period. Math. Hung. 2 (1) (1972) 295–299, https://doi.org/10.1007/BF02018669, https://link.springer.com/article/10.1007%5C%2FBF02018669#citeas.

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

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

[19] Matthew Jenssen, Eoin Long, Peter Keevash, Liana Yepremyan, Distinct degrees in induced subgraphs, Proc. Am. Math. Soc. (ISSN 0002-9939) 148 (9) (Sept. 2020) 3835–3846, https://doi.org/10.1090/proc/15060.

[20] Ross Kang, Eoin Long, Viresh Patel, Guus Regts, On a Ramsey-type problem of Erdös and Pach, Bull. Lond. Math. Soc. (ISSN 0024-6093) 49 (6) (Dec. 2017) 991–999, https://doi.org/10.1112/blms.12094.

[21] Matthew Kwan, Benny Sudakov, Proof of a conjecture on induced subgraphs of Ramsey graphs, Trans. Am. Math. Soc. 372 (Dec. 2017), https://doi.org/10.1090/tran/7729.

[22] Matthew Kwan, Benny Sudakov, Ramsey graphs induce subgraphs of quadratically many sizes, Int. Math. Res. Not. 2020 (Nov. 2017), https://doi.org/10.1093/imrn/rny064.

[23] 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, Association for Computing Machinery, Montreal, Canada, ISBN 9781450345286, 2017, pp. 1144–1156.

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

[25] Bhargav Narayanan, Julian Sahasrabudhe, István Tomon, Ramsey graphs induce subgraphs of many different sizes, Combinatorica 39 (Sept. 2016), https://doi.org/10.1007/s00493-017-3755-0.

[26] Bhargav Narayanan, István Tomon, Induced subgraphs with many distinct degrees, Comb. Probab. Comput. 27 (1) (2018) 110–123, https://doi.org/10.1017/S0963548317000256.

[27] H.J. Prömel, V. Rödl, Non-Ramsey graphs are clogn-universal, J. Comb. Theory, Ser. A (ISSN 0097-3165) 88 (2) (1999) 379–384, https://doi.org/10.1006/jcta.1999.2972, https://www.sciencedirect.com/science/article/pii/S0097316599929722.

[28] L. Pyber, V. Rödl, E. Szemerédi, Dense graphs without 3-regular subgraphs, J. Comb. Theory, Ser. B (ISSN 0095-8956) 63 (1) (1995) 41–54, https://doi.org/10.1006/jctb.1995.1004, https://www.sciencedirect.com/science/article/pii/S0095895685710040.

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

[30] F.P. Ramsey, On a problem of formal logic, Proc. Lond. Math. Soc. s2–30 (1) (1930) 264–286, https://doi.org/10.1112/plms/s2-30.1.264, https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/plms/s2-30.1.264.

[31] A.D. Scott, Large induced subgraphs with all degrees odd, Comb. Probab. Comput. 1 (4) (1992) 335–349, https://doi.org/10.1017/S0963548300000389.

[32] S. Shelah, Erdös and Rényi conjecture, J. Comb. Theory, Ser. A 82 (1998) 179–185.