INDUCED SUBGRAPHS WITH MANY DISTINCT DEGREES

BHARGAV NARAYANAN AND ISTVÁN TOMON

ABSTRACT. Let $\hom(G)$ denote the size of the largest clique or independent set of a graph G. In 2007, Bukh and Sudakov proved that every n-vertex graph G with $\hom(G) = O(\log n)$ contains an induced subgraph with $\Omega(n^{1/2})$ distinct degrees, and raised the question of deciding whether an analogous result holds for every n-vertex graph G with $\hom(G) = O(n^\varepsilon)$, where $\varepsilon > 0$ is a fixed constant. Here, we answer their question in the affirmative and show that every graph G on n vertices contains an induced subgraph with $\Omega((n/\hom(G))^{1/2})$ distinct degrees. We also prove a stronger result for graphs with large cliques or independent sets and show, for any fixed $k \in \mathbb{N}$, that if an n-vertex graph G contains no induced subgraph with k distinct degrees, then $\hom(G) \geq n/(k-1) - o(n)$; this bound is essentially best-possible.

1. INTRODUCTION

A subset of the vertices of a graph is called *homogeneous* if it induces either a clique or an independent set. What can one say about the structure of graphs with no large homogeneous sets? This is a central question in graph Ramsey theory and has received considerable attention over the last sixty years. Let $\hom(G)$ denote the size of the largest homogeneous set of a graph G. Erdős and Szekeres [8] proved that $\hom(G) \geq (\log n)/2$ for any graph G on n vertices and subsequently, Erdős [7] used probabilistic arguments to demonstrate the existence of an n-vertex graph G with $\hom(G) \leq 2 \log n$; here, and throughout the paper, all logarithms are base 2. Despite considerable effort, see [11, 2, 6, 5] for instance, we know of no deterministic constructions of graphs with no large homogeneous sets; this suggests that such graphs should perhaps 'look like' random graphs. This belief is supported by many results which show that graphs with no large homogeneous sets possess many of the same properties as random graphs.

For a constant $C > 0$, we say that an n-vertex graph G is *C-Ramsey* if $\hom(G) \leq C \log n$. There are a number of results which show that Ramsey graphs share various properties with dense random graphs. For example, Erdős and Szemerédi [10] proved that Ramsey graphs must have edge densities bounded away from 0 and 1. Prömel and Rödl [12] later showed that every C-Ramsey graph on n vertices is $(\delta \log n)$-universal for some positive constant $\delta = \delta(C)$; here, a graph is *k-universal* if it contains an induced copy of every graph on at most k vertices. As a final example, let us mention that Shelah [13] proved that Ramsey graphs contain exponentially many non-isomorphic induced subgraphs.

The results of this paper are motivated by a line of questioning proposed by Bukh and Sudakov. Bukh and Sudakov [3] proved, settling a conjecture of Erdős, Faudree and Sós [9], that every C-Ramsey graph on n vertices contains an induced subgraph with at least $\delta n^{1/2}$ distinct degrees for some constant $\delta(C) > 0$. They then raised the possibility of a similar result holding for graphs with much larger homogeneous sets, and suggested in particular that for any $0 < \varepsilon < 1/2$, every n-vertex graph G with $\hom(G) < n^{\varepsilon}$ should contain an induced subgraph with $\Omega(n^{1/2-\varepsilon})$ distinct degrees; by building on their work, we establish this fact as a special case of Theorem 1.1 below.

For a graph G, let $f(G)$ denote the largest integer k for which G contains an induced subgraph with k distinct degrees. Our first result is the following.

Theorem 1.1. *For every graph G on n vertices, we have*

$$f(G) \geq \frac{1}{250} \left(\frac{n}{\hom(G)} \right)^{1/2}.$$

We believe that this result is far from sharp however; for example, we are unable to construct an n-vertex graph G either with $\hom(G) < n^{1/2}$ and $f(G) = o(n^{1/2})$, or with $\hom(G) \geq n^{1/2}$ and $f(G) = o(n/\hom(G))$.

Our next result sharpens Theorem 1.1 for graphs with very large homogeneous sets. For a positive integer $k \in \mathbb{N}$, the disjoint union of $n/(k-1)$ cliques each of size $k-1$ gives us an example of a graph G on n vertices with $f(G) = k-1$ and $\hom(G) = \max\{n/(k-1), k-1\}$. We prove that if k is fixed and n is sufficiently large, then such a construction is essentially best-possible.

Theorem 1.2. *Fix $k \in \mathbb{N}$ and $\varepsilon > 0$. If n is sufficiently large, then $f(G) \geq k$ for every n-vertex graph G with $\hom(G) < n/(k-1+\varepsilon)$.*

We remark that with ε fixed, the minimal n for which we are able to verify Theorem 1.2 is exponential in k. We further believe that it should be possible to prove Theorem 1.2 without an ε-dependent error term; however, we are unable to do this at present.

This paper is organised as follows. In the next section, we establish some notation and collect together a few basic tools. In Section 3, we introduce the main ideas used in this paper and prove Theorem 1.1. In Section 4, we refine the ideas used to prove Theorem 1.1 and prove Theorem 1.2. We conclude with a discussion of some open problems in Section 5.

For the sake of clarity of presentation, we systematically omit floor and ceiling signs whenever they are not crucial. We also make no attempt to optimise the absolute constants in our results.

2. Preliminaries

In this short section, we introduce some notation and collect together some facts that we shall rely on repeatedly in the sequel.

2.1. **Notation.** For us, a pair $\{x, y\}$ will always mean an unordered pair with $x \neq y$. For a set X, we write $X^{(2)}$ for the family of all pairs on the ground set X.

Let $G = (V, E)$ be a graph. We write $v(G)$ and $e(G)$ respectively for the number of vertices and edges of G. The complement of G is denoted by $\overline{G}$. The neighbourhood of a vertex x is denoted by $\Gamma(x)$, and the non-neighbourhood of x is denoted by $\overline{\Gamma}(x)$. Also, let $\mathrm{d}(x) = |\Gamma(x)|$ denote the degree of x in G, and let $\overline{\mathrm{d}}(x) = |\overline{\Gamma}(x)|$ be the degree of x in $\overline{G}$.

For a subset $U \subset V$, we write $G[U]$ for the subgraph of G induced by U. For a vertex $x \in V$, we write $\Gamma_U(x)$ for the set $\Gamma(x) \cap U$ and $\overline{\Gamma}_U(x)$ for the set $\overline{\Gamma}(x) \cap U$; we also set $\mathrm{d}_U(x) = |\Gamma_U(x)|$ and $\overline{\mathrm{d}}_U(x) = |\overline{\Gamma}_U(x)|$.

Finally, we define the *neighbourhood-distance* between two vertices $x, y \in V$ by

$$\delta(x, y) = |(\Gamma(x) \setminus \{y\}) \triangle (\Gamma(y) \setminus \{x\})|.$$

It is not hard to check that this distance satisfies the triangle inequality, i.e., $\delta(x, y) + \delta(y, z) \geq \delta(x, z)$ for any three vertices $x, y, z \in V$.

2.2. **Graph theoretic estimates.** We need the following simple fact about the neighbourhood-distance.

3

Proposition 2.1. *For each $K \in \mathbb{N}$, there exists a $\Delta \in \mathbb{N}$ such that the following holds. If G is a graph with $\delta(x, y) \leq K$ for all $x, y \in V(G)$, then either G or $\overline{G}$ has maximum degree at most Δ.*

Proof. We prove the claim with $\Delta = 4K$. The proposition is trivial if $v(G) \leq 4K$, so we may assume that $v(G) \geq 4K + 1$. Fix a vertex $v \in V(G)$ and let s and t be the number of edges and non-edges between $\Gamma(v)$ and $\overline{\Gamma}(v)$ respectively.

If $x \in \Gamma(v)$, then $\mathrm{d}_{\overline{\Gamma}(v)}(x) \leq \delta(v, x) \leq K$. Hence, $s \leq K|\Gamma(v)|$ and analogously, $t \leq K|\overline{\Gamma}(v)|$. Since $s + t = |\Gamma(v)||\overline{\Gamma}(v)|$, it follows that

$$|\Gamma(v)||\overline{\Gamma}(v)| \leq K(|\Gamma(v)| + |\overline{\Gamma}(v)|) = K(v(G) - 1).$$

Since $v(G) \geq 4K + 1$, this is only possible if one of $|\Gamma(v)|$ or $|\overline{\Gamma}(v)|$ is at most $2K$. If $|\Gamma(v)| \leq 2K$, then it is not hard to see that every vertex has degree at most $3K$; indeed, this follows from the trivial observation that $\mathrm{d}(x) \leq \delta(x, v) + \mathrm{d}(v)$ for any $x \in V(G)$. If $|\overline{\Gamma}(v)| \leq 2K$ on the other hand, then it is clear that the maximum degree of $\overline{G}$ is at most $3K$. $\square$

We shall also require the following fact.

Proposition 2.2. *For any $k, \Delta \in \mathbb{N}$, there exists an $L \in \mathbb{N}$ such that the following holds. If G is a graph with $f(G) < k$ and maximum degree at most Δ, then G contains at most L vertices of degree at least $k - 1$.*

Proof. We shall prove the claim with $L = (\Delta^2 + 1)k$. We say that two vertices $x, y \in V(G)$ are *independent* if $(\Gamma(x) \cup \{x\}) \cap (\Gamma(y) \cup \{y\}) = \varnothing$. As the maximum degree of G is at most Δ, a vertex of G is dependent on at most $\Delta + \Delta(\Delta - 1) = \Delta^2$ other vertices. Let $S \subset V(G)$ be the set of vertices of degree at least $k - 1$ and suppose for the sake of contradiction that $|S| > L$. Since $|S|/(\Delta^2 + 1) > k$, we can select k pairwise independent vertices from S; let $x_1, x_2, \ldots, x_k$ be these vertices. For $1 \leq i \leq k$, let X_i be an arbitrary $(i - 1)$-element subset of $\Gamma(x_i)$ and let

$$X = \{x_1, x_2, \ldots, x_k\} \cup X_1 \cup X_2 \cup \cdots \cup X_k.$$

Since the vertices $x_1, x_2, \ldots, x_k$ are pairwise independent, we see that $\mathrm{d}_X(x_i) = i - 1$ for each $1 \leq i \leq k$. It follows that $x_1, x_2, \ldots, x_k$ have different degrees in $G[X]$; this contradicts the fact that $f(G) < k$. $\square$

Finally, we need the Caro–Wei theorem [4, 14] which refines Turán's theorem.

4

Proposition 2.3. *Every graph G contains an independent set of size at least*

$$\sum_{v \in V(G)} \frac{1}{\mathrm{d}(v) + 1} \geq \frac{v(G)^2}{2e(G) + v(G)}.$$

Proof. Order the vertices of G uniformly at random and consider the set I of those vertices that precede all their neighbours in this ordering. Clearly, I is independent; the proposition follows since

$$\mathbb{E}[|I|] = \sum_{v \in V(G)} \frac{1}{\mathrm{d}(v) + 1}. \qquad \square$$

2.3. **Probabilistic inequalities.** We need the following two well-known inequalities; the proofs of these claims may be found in [1]. We shall require Markov's inequality.

Proposition 2.4. *Let X be a non-negative real-valued random variable. For any $t \geq 0$, we have*

$$\mathbb{P}(X > t) < \frac{\mathbb{E}[X]}{t}. \qquad \square$$

We shall also require Hoeffding's inequality.

Proposition 2.5. *Let $X = \sum_{i=1}^{n} X_i$ where $X_1, X_2, \ldots, X_n$ are independent real-valued random variables with $0 \leq X_i \leq 1$ for each $1 \leq i \leq n$. For any $t \geq 0$, we have*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq 2 \exp\left(\frac{-2t^2}{n}\right). \qquad \square$$

3. Small homogeneous sets

This section is devoted to the proof of Theorem 1.1. Before we prove the result, let us give an overview of the proof. To show that a graph G contains an induced subgraph with many distinct degrees, we pick a random subset U of the vertices and consider the subgraph of G induced by U. We first show that if the neighbourhood-distance between a pair of vertices $x, y \in U$ is large, then the probability of $\mathrm{d}_U(x)$ and $\mathrm{d}_U(y)$ being equal is small. Next, to show that there are many such pairs of vertices with large neighbourhood-distances, we bound the number of pairs of vertices with small neighbourhood-distances in terms of $\hom(G)$. We now make this sketch precise.

Proof of Theorem 1.1. The result is trivial if our graph has fewer than 250 vertices, so we may assume that $G = (V, E)$ is a graph on $n \geq 250$ vertices. Let U be a random subset of V obtained by selecting each vertex of G with probability $1/2$, independently of the other vertices. Define an auxiliary *degree graph* D on U where two vertices $x, y \in U$ are joined by an edge if they have the same degree in $G[U]$; note that D is the union of vertex disjoint cliques, one for each vertex degree in $G[U]$. For any pair of distinct vertices $x, y \in V$, the probability that xy is an edge of D is precisely $\mathbb{P}(\mathrm{d}_U(x) = \mathrm{d}_U(y))/4$. Hence, the expected number of edges of D is given by

$$\mathbb{E}[e(D)] = \frac{1}{4} \sum_{\{x,y\} \in V^{(2)}} \mathbb{P}(\mathrm{d}_U(x) = \mathrm{d}_U(y)).$$

We shall bound the probability of the event $\mathrm{d}_U(x) = \mathrm{d}_U(y)$ by the neighbourhood-distance between x and y; a similar result appears in [3].

Lemma 3.1. *For any pair of distinct vertices* $x, y \in V$, *we have*

$$\mathbb{P}(\mathrm{d}_U(x) = \mathrm{d}_U(y)) < \frac{20}{\sqrt{\delta(x, y) + 1}}.$$

Proof. Let $s = |\Gamma(x) \setminus \Gamma(y)|$ and $t = |\Gamma(y) \setminus \Gamma(x)|$. Observe that $s + t$ is equal to $\delta(x, y) + 2$ if $xy \in E$ and $\delta(x, y)$ otherwise. Without loss of generality, we may suppose that $s \geq t$. Therefore, $s \geq \delta(x, y)/2$ and it follows that

$$\begin{aligned}
\mathbb{P}(\mathrm{d}_U(x) = \mathrm{d}_U(y)) &= \frac{1}{2^{s+t}} \sum_{i=0}^{t} \binom{s}{i} \binom{t}{i} \\
&\leq \frac{1}{2^s} \max_{0 \leq i \leq s} \binom{s}{i} \left(\frac{1}{2^t} \sum_{i=0}^{t} \binom{t}{i} \right) \\
&= \frac{1}{2^s} \binom{s}{\lfloor s/2 \rfloor} < \frac{10}{\sqrt{s+1}} \leq \frac{20}{\sqrt{\delta(x, y) + 1}}. \qquad \square
\end{aligned}$$

As an immediate consequence of this lemma, we have

$$\mathbb{E}[e(D)] < \sum_{\{x,y\} \in V^{(2)}} \frac{5}{\sqrt{\delta(x, y) + 1}}. \tag{1}$$

We now bound the right hand side of (1) in terms of $\hom(G)$. Fix a vertex $x \in V$ and consider the sets $S = \Gamma(x)$ and $T = \overline{\Gamma}(x)$. Note that $\delta(x, y) \geq \mathrm{d}_T(y)$ for any

6

$y \in T$. By applying Proposition 2.3 to the graph $G[T]$, we get

$$\hom(G) \geq \sum_{y \in T} \frac{1}{d_T(y) + 1} \geq \sum_{y \in T} \frac{1}{\delta(x, y) + 1}.$$

Similarly, if $y \in S$, then $\delta(x, y) \geq \overline{d}_S(y)$. Applying Proposition 2.3 to the complement of $G[S]$, we similarly get

$$\hom(G) \geq \sum_{y \in S} \frac{1}{\overline{d}_S(y) + 1} \geq \sum_{y \in S} \frac{1}{\delta(x, y) + 1}.$$

It follows that

$$2 \hom(G) \geq \sum_{y \in V \setminus \{x\}} \frac{1}{\delta(x, y) + 1}.$$

Finally, by summing this inequality over all the vertices of G and applying the Cauchy–Schwarz inequality, we get

$$n \hom(G) \geq \sum_{\{x,y\} \in V^{(2)}} \frac{1}{\delta(x, y) + 1} \geq \binom{n}{2}^{-1} \left(\sum_{\{x,y\} \in V^{(2)}} \frac{1}{\sqrt{\delta(x, y) + 1}} \right)^2. \qquad (2)$$

From (1) and (2), we conclude that

$$\mathbb{E}[e(D)] < 4\sqrt{n^3 \hom(G)}.$$

It follows from Markov's inequality that

$$\mathbb{P}\left(e(D) > 12\sqrt{n^3 \hom(G)} \right) < \frac{1}{3}.$$

Next, since $|U|$ is the sum of $n \geq 250$ independent indicator random variables and $\mathbb{E}[|U|] = n/2$, it follows from Hoeffding's inequality that

$$\mathbb{P}(|U| < n/3) < \frac{1}{3}.$$

Hence, with positive probability, the degree graph D satisfies $v(D) > n/3$ and $e(D) < 12(n^3 \hom(G))^{1/2}$. Applying Proposition 2.3 to D, we see that this graph contains an independent set of size at least

$$\frac{v(D)^2}{2e(D) + v(D)} \geq \frac{1}{250} \sqrt{\frac{n}{\hom(G)}}.$$

The vertices of this independent set have different degrees in $G[U]$ by the definition of D. Therefore, the subgraph induced by U has at least $(1/250)(n/\hom(G))^{1/2}$ distinct degrees with positive probability; the result follows. $\qquad \square$

4. Large homogeneous sets

In this section, we consider graphs with large homogeneous sets and prove Theorem 1.2. Before we turn to the proof of Theorem 1.2, let us identify the inefficiencies in the proof of Theorem 1.1; to this end, we consider some examples of graphs with large homogeneous sets that contain no induced subgraphs with many distinct degrees. Let $b, k, n \in \mathbb{N}$ be positive integers satisfying $b \leq k \leq (\log n)/2$ and consider

  (1) a disjoint union of n/k cliques each of size k,
  (2) a disjoint union of k cliques each of size n/k, and
  (3) a disjoint union of k/b copies of $\overline{H}$, where H is the disjoint union of n/k cliques each of size b.

It is not hard to see that if G is the graph in either of the above three examples, then $\hom(G) = n/k$ and $f(G) = k$. Suppose that G is one of the graphs described above; let U be a $1/2$-random subset of $V(G)$ and define the degree graph D on U as in the proof of Theorem 1.1.

If G is the graph in example (1), then with high probability, $G[U]$ has about

$$\frac{n}{2^k} \binom{k-1}{i}$$

vertices of degree i for each $0 \leq i \leq k - 1$. This means that while $G[U]$ has k distinct degrees, the k cliques in the auxiliary graph D have very different sizes. Hence, the final application of the Caro–Wei bound in the proof of Theorem 1.1 is too crude for this graph.

If G is the graph in example (2), then $G[U]$ again has k distinct degrees with high probability. However, notice that $\delta(x, y)$ is either 0 or $2n/k - 2$ for any pair of vertices $x, y \in V(G)$ with the former being the case for about $1/k$ of the pairs. Our application of the Cauchy–Schwartz inequality in the proof of Theorem 1.1 is therefore suboptimal for this graph.

Our third example is a generalisation of the first two examples. The graph defined in example (3) is the complement of the graph in our first example if $b = k$, and is the graph in our second example when $b = 1$. When $1 < b < k$, this example illustrates both the aforementioned inefficiencies in the proof of Theorem 1.1.

Nevertheless, notice that in all of our examples, a random induced subgraph has at least k distinct degrees with high probability; our strategy for proving

Theorem 1.2 will be informed by this fact. Let us now sketch our strategy. First, we show that we may partition the vertex set of our graph G into groups in such a way that vertices within a group are all close together in neighbourhood-distance while vertices from different groups are far apart. We then select a random subset U of the vertices as before. We use the bound on $\hom(G)$ to show that in $G[U]$, the degrees of the vertices within a group are distributed like the vertex degrees in a random induced subgraph of the graph in example (1). We then argue that most pairs of vertices from different groups have different degrees in $G[U]$. We finally deduce from these facts that the number of distinct degrees in $G[U]$ is large.

We are ready to prove Theorem 1.2; while following the proof, the reader is encouraged to keep example (3) in mind.

Proof of Theorem 1.2. The result is trivial when $k = 1$, so fix $k \geq 2$. Also, we may assume without loss of generality that $0 < \varepsilon < 1/2$. With the benefit of hindsight, we define the constants $\beta = \varepsilon/10k$, $\eta = \varepsilon\beta/10^5 k^2$ and $J = 10^4 k^{12} 2^{4k}/\eta^4$.

Assume that n is sufficiently large and suppose for the sake of contradiction that $G = (V, E)$ is a graph on n vertices with $\hom(G) < n/(k - 1 + \varepsilon)$ and $f(G) < k$.

Let U be a random subset of V obtained by selecting each vertex of V with probability $1/2$, independently of the other vertices. As before, define an auxiliary degree graph D on U where two vertices $x, y \in U$ are joined by an edge if they have the same degree in $G[U]$.

For a set of vertices $W \subset V$, we define

$$\hat{\delta}(W) = \sum_{\{x,y\} \in W^{(2)}} \frac{5}{\sqrt{\delta(x, y) + 1}}.$$

We first prove the following.

Claim 4.1. *For each $W \subset V$, we have*

$$\hat{\delta}(W) > \frac{|W|^2 - 3k|W|}{54k}.$$

Proof. The statement is trivial if $|W| \leq 3k$, so suppose that $|W| \geq 3k + 1 \geq 7$. Note that by Lemma 3.1, the quantity $\hat{\delta}(W)$ is an upper bound for the expected number of edges of D spanned by $W \cap U$; in other words,

$$\mathbb{E}[e(D[W \cap U])] \leq \hat{\delta}(W).$$

9

Applying Markov's inequality, we get

$$\mathbb{P}(e(D[W \cap U]) > 3\hat{\delta}(W)) < 1/3.$$

Also, $|W \cap U|$ is the sum of $|W| \geq 7$ independent indicator random variables, so by Hoeffding's inequality,

$$\mathbb{P}(|W \cap U| < |W|/3) \leq 2\exp(-14/9) < 1/2.$$

Hence, with positive probability, we have both $e(D[W \cap U]) \leq 3\hat{\delta}(W)$ and $|W \cap U| \geq |W|/3$. By Proposition 2.3, with positive probability, the graph $D[W \cap U]$ contains an independent set of size at least

$$\frac{|W \cap U|^2}{2e(D[W \cap U]) + |W \cap U|} \geq \frac{|W|^2}{54\hat{\delta}(W) + 3|W|}.$$

However, since $f(G) < k$, we know that D cannot contain an independent set of size k. It follows that

$$k > \frac{|W|^2}{54\hat{\delta}(W) + 3|W|},$$

or equivalently,

$$\hat{\delta}(W) > \frac{|W|^2 - 3k|W|}{54k}. \qquad \square$$

This tells us that every large subset of V must have many pairs of vertices whose neighbourhood-distance is small. This suggests that we should be able to group the vertices of G into large groups in such a way that the neighbourhood-distance between any two vertices within a group is small. We prove such a statement this next; recall that $\beta = \varepsilon/10k$ and $J = 10^{24}k^{20}2^{4k}/(\varepsilon\beta)^4$.

Claim 4.2. *There exists a $K = K(k, \varepsilon) > 0$ such that the following holds. There are pairwise disjoint sets $A_1, A_2, \ldots, A_m \subset V$ such that*

(1) $|A_1 \cup A_2 \cup \cdots \cup A_m| > (1 - \beta)n$,
(2) $|A_i| > \beta n/10^4 k$ *for each* $1 \leq i \leq m$,
(3) *if* $x, y \in A_i$, *then* $\delta(x, y) < K$, *and*
(4) *if* $x \in A_i$ *and* $y \in A_j$ *with* $i \neq j$, *then* $\delta(x, y) > J$.

Proof. We prove the claim with

$$K = 2\left(10^6 k^2 + \frac{J\log(10^4 k)}{\log(1 + \beta/2)}\right).$$

We shall construct a sequence of pairwise disjoint sets $A_1, A_2, \ldots$ and another nested sequence of sets $B_0 \subset B_1 \subset \ldots$ such that for each $l \in \mathbb{N}$,

10

(i) B_l is disjoint from $A_1 \cup A_2 \cup \cdots \cup A_l$,

(ii) $|A_l| > \beta n/10^4 k$,

(iii) $|B_l| < (\beta/2) \sum_{i=1}^{l} |A_i|$,

(iv) if $x, y \in A_l$, then $\delta(x, y) < K$, and

(v) if $x \in A_l$ and $y \in V \setminus (A_l \cup B_l)$, then $\delta(x, y) > J$.

We set $B_0 = \varnothing$. Let $l \geq 0$ and suppose that the sets $A_1, A_2, \ldots, A_l$ and B_l have already been constructed satisfying the above properties. The claim is proved if $\sum_{i=1}^{l} |A_i| > (1 - \beta)n$. Suppose otherwise; we construct A_{l+1} and B_{l+1} as follows.

Define $W = V \setminus (B_l \cup A_1 \cup A_2 \cup \cdots \cup A_l)$ and note that $|W| > \beta n/2$. Let p be the number of pairs $\{x, y\} \in W^{(2)}$ such that $\delta(x, y) < 10^6 k^2$. Note that we have

$$\hat{\delta}(W) = \sum_{\{x,y\} \in W^{(2)}} \frac{5}{\sqrt{\delta(x, y) + 1}} < 5p + \frac{|W|^2}{10^3 k}.$$

On the other hand, if n is sufficiently large, then by Claim 4.1, we have

$$\hat{\delta}(W) > \frac{|W|^2 - 3k|W|}{54k} > \frac{|W|^2}{100k}.$$

It follows that $p > |W|^2/10^3 k$. Thus, there is a vertex $w \in W$ and a subset $S \subset W$ with $|S| > |W|/10^3 k$ where every $x \in S$ satisfies $\delta(w, x) < 10^6 k^2$.

We set $C_1 = \{w\} \cup S$ and iteratively construct an increasing sequence of sets $C_1 \subset C_2 \subset \ldots$ as follows. Having constructed C_i, consider the set T of vertices $x \in W \setminus C_i$ for which there exists a vertex $y \in C_i$ such that $\delta(x, y) \leq J$. If $|T| \geq \beta |C_i|/2$, then set $C_{i+1} = C_i \cup T$. Otherwise, stop and define $A_{l+1} = C_i$ and $B_{l+1} = B_l \cup T$.

It is clear that $B_{l+1} \cap (A_1 \cup A_2 \cup \cdots \cup A_{l+1}) = \varnothing$ and that $\delta(x, y) > J$ for all $x \in A_{l+1}$ and $y \in V \setminus (B_{l+1} \cup A_{l+1})$. It is also clear that

$$|A_{l+1}| \geq |C_1| > \frac{|W|}{10^3 k} > \frac{\beta n}{10^4 k}.$$

Finally, note that we must have $|T| < \beta |C_i|/2$ when we define A_{l+1} and B_{l+1}, so we inductively have

$$|B_{l+1}| < \frac{\beta}{2} \sum_{i=1}^{l+1} |A_i|.$$

Next, observe that for each $i \geq 1$, we have

$$|C_i| \geq \left(1 + \frac{\beta}{2}\right)^{i-1} |C_1| \geq \left(1 + \frac{\beta}{2}\right)^{i-1} \frac{|W|}{10^3 k}.$$

11

As each C_i is a subset of W, it is clear that the sets A_{l+1} and B_{l+1} get defined after at most $\log(10^4 k)/\log(1 + \beta/2)$ iterations. The neighbourhood-distance satisfies the triangle inequality; consequently,

$$\delta(w, x) < 10^6 k^2 + (i - 1)J$$

for all $x \in C_i$ and therefore,

$$\delta(x, y) < 2(10^6 k^2 + (i - 1)J)$$

for all $x, y \in C_i$. Thus, for all $x, y \in A_{l+1}$, we have

$$\delta(x, y) < 2\left(10^6 k^2 + \frac{J \log(10^4 k)}{\log(1 + \beta/2)}\right) = K.$$

Therefore, the sets $A_1, A_2, \ldots, A_{l+1}$ and B_{l+1} also satisfy the properties described above.

To finish the proof, note that since the sets $A_1, A_2, \ldots$ are pairwise disjoint, the above described construction procedure must terminate; indeed, if $m > 10^4 k/\beta$, then $\sum_{i=1}^{m} |A_i| > (1 - \beta)n$. $\square$

Let K and $A_1, A_2, \ldots, A_m$ be as promised by Claim 4.2. Note that $m \le 10^4 k/\beta$ since $|A_i| > \beta n/10^4 k$ for each $1 \le i \le m$.

To proceed, we need to introduce the following notion of independence. If $S \subset V$ and $x, y \in S$, we say that x and y are S-independent in G if

$$(\Gamma_S(x) \cup \{x\}) \cap (\Gamma_S(y) \cup \{y\}) = \varnothing;$$

analogously, we say that x and y are S-independent in $\overline{G}$ if

$$(\overline{\Gamma}_S(x) \cup \{x\}) \cap (\overline{\Gamma}_S(y) \cup \{y\}) = \varnothing.$$

Recall that U is a $1/2$-random subset of the vertices; therefore, if x and y are S-independent, then $\mathrm{d}_{U \cap S}(x)$ and $\mathrm{d}_{U \cap S}(y)$ are independent random variables.

For $1 \le i \le m$, let r_i be the unique non-negative integer such that

$$\frac{r_i n}{k - 1 + \varepsilon} < |A_i| - \eta n \le \frac{(r_i + 1)n}{k - 1 + \varepsilon}.$$

Our next step is to show that each A_i has a large subset of pairwise A_i-independent vertices; recall that $\eta = \varepsilon \beta/10^5 k^2$.

Claim 4.3. *For each $1 \le i \le m$, there is a subset $B_i \subset A_i$ with $|B_i| > \eta n/2k^2$ such that either*

(1) $r_i \leq d_{A_i}(x) \leq k - 2$ *for every* $x \in B_i$, *and*

(2) *any pair of distinct vertices* $x, y \in B_i$ *are* A_i-*independent in* G,

or

(1) $r_i \leq \overline{d}_{A_i}(x) \leq k - 2$ *for every* $x \in B_i$, *and*

(2) *any pair of distinct vertices* $x, y \in B_i$ *are* A_i-*independent in* $\overline{G}$.

It is clear from the definition of r_i that $r_i \leq k - 1$; the above claim implicitly tells us that our assumptions about G actually imply that $r_i \leq k - 2$ for each $1 \leq i \leq m$.

Proof of Claim 4.3. Fix $1 \leq i \leq m$ and let $r = r_i$, $A = A_i$ and $F = G[A_i]$. To avoid confusion, we write d_F and δ_F to denote the degrees and neighbourhood-distances in the graph F.

Note that $\delta_F(x, y) \leq \delta(x, y) < K$ for all $x, y \in A$. Hence, by applying Proposition 2.1 to F, we see that there exists a $\Delta = \Delta(k, \varepsilon)$ such that either $d_F(x) \leq \Delta$ for all $x \in A$ or $\overline{d}_F(x) \leq \Delta$ for all $x \in A$. We suppose that the former holds; the other case may be handled analogously.

We now apply Proposition 2.2 to F and conclude that all but at most $L = L(k, \varepsilon)$ vertices $x \in A$ satisfy $d_F(x) \leq k - 2$. Let q be the number of vertices $x \in A$ with $d_F(x) \leq r - 1$. Then, by Proposition 2.3, F contains an independent set of size at least q/r. From our assumption about $\hom(G)$, it follows that

$$q \leq r \hom(F) \leq r \hom(G) < \frac{rn}{k - 1 + \varepsilon};$$

therefore, there are at least ηn vertices $x \in A$ with $d_F(x) \geq r$.

Let S_1 be the set of vertices $x \in A$ with $d_F(x) \geq r$. Let S_2 be the set of those $x \in S_1$ which are A-dependent on some vertex $y \in A$ with $d_F(y) \geq k - 1$; the number of such y is at most L, so it follows that $|S_2| \leq L(1 + \Delta + \Delta(\Delta - 1))$. If n is sufficiently large, then the set $S_3 = S_1 \setminus S_2$ contains at least $2\eta n/3$ vertices. For every vertex $x \in S_3$, there are at most $(k-2) + (k-2)(k-3) \leq k^2 - 1$ other vertices $y \in A$ such that x and y are A-dependent. Hence, we can greedily select a set $B \subset S_3$ of pairwise A-independent vertices of size at least $|S_3|/k^2 > \eta n/2k^2$. $\qquad \square$

For $1 \leq i \leq m$, let $B_i \subset A_i$ be as guaranteed by Claim 4.3 and let $\mathcal{A}_i$ be the event that there exists an integer $d_i \geq 0$ and $r_i + 1$ pairwise disjoint sets $B_{i,0}, B_{i,1}, \ldots, B_{i,r_i} \subset B_i \cap U$ such that for all $0 \leq s, t \leq r_i$,

13

(1) $|B_{i,s}| \geq \eta n/2^{k+1}k^3$,

(2) $d_{U \cap A_i}(x) = d_{U \cap A_i}(y)$ for all $x, y \in B_{i,s}$,

(3) $d_{U \cap A_i}(x) \neq d_{U \cap A_i}(y)$ for all $x \in B_{i,s}$ and $y \in B_{i,t}$ with $s \neq t$, and

(4) $d_{U \cap (V \setminus A_i)}(x) = d_i$ for all $x \in B_{i,0} \cup B_{i,1} \cup \cdots \cup B_{i,r_i}$.

We next prove the following bound.

Claim 4.4. *For each $1 \leq i \leq m$,*

$$\mathbb{P}(\mathcal{A}_i) > 1 - \frac{1}{3m}.$$

Proof. Fix $1 \leq i \leq m$ and suppose that all pairs of distinct vertices from B_i are A_i-independent in G; the other case may be handled similarly by working with vertex degrees in $\overline{G}$ instead of G.

For each $W \subset V \setminus A_i$, we shall prove that

$$\mathbb{P}(\mathcal{A}_i \,|\, U \cap (V \setminus A_i) = W) > 1 - \frac{1}{3m}.$$

We fix $W \subset V \setminus A_i$ and condition on the event $U \cap (V \setminus A_i) = W$; note that any event that depends only on the vertices in A_i is independent of this event. If the collection of degrees $\{d_W(x)\}_{x \in B_i}$ contains k distinct integers, then the subgraph induced by $B_i \cup W$ has k different degrees, contradicting our assumption that $f(G) < k$. Hence, there exists an integer $d_i \geq 0$ for which there are least $|B_i|/k$ vertices $x \in B_i$ with $d_W(x) = d_i$; let $C_i \subset B_i$ be the set of these vertices.

For a vertex $x \in C_i$ and $0 \leq s \leq r_i$, let $I_i(x, s)$ be the indicator of the event $\{x \in U\} \cap \{d_{U \cap A_i}(x) = s\}$ and define

$$I_i(s) = \sum_{x \in C_i} I_i(x, s).$$

Observe that $\mathbb{E}[I_i(x, s)] \geq 1/2^{k-1}$ since $r_i \leq d_{A_i}(x) \leq k-2$ and consequently,

$$\mathbb{E}[I_i(s)] \geq \frac{|C_i|}{2^{k-1}} \geq \frac{\eta n}{2^k k^3}.$$

Also, the indicator random variables $\{I_i(x, s)\}_{x \in C_i}$ are independent, so by Hoeffding's inequality, we have

$$\mathbb{P}\left(I_i(s) < \frac{\eta n}{2^{k+1}k^3}\right) < \frac{1}{3mk}$$

for all sufficiently large n since $m \leq 10^4 k/\beta$. If $I_i(s) \geq \eta n/2^{k+1}k^3$ for each $0 \leq s \leq r_i$, then $\mathcal{A}_i$ clearly holds. Since $r_i \leq k-2$, the claim follows by the union bound. $\qquad\square$

Define the event $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2 \cap \cdots \cap \mathcal{A}_m$; it is clear from the previous claim that $\mathbb{P}(\mathcal{A}) > 2/3$.

Let us remind the reader that D is the graph on U where $x, y \in U$ are joined by an edge if $\mathrm{d}_U(x) = \mathrm{d}_U(y)$. Let b be the number of edges xy of D with $x \in A_i$ and $y \in A_j$ for some $1 \leq i < j \leq m$. As $\delta(x, y) > J$ for such a pair of vertices x and y, it follows from Claim 3.1 that

$$\mathbb{P}(xy \in E(D)) < \frac{5}{\sqrt{J+1}}$$

and hence, $\mathbb{E}[b] < 5n^2/J^{1/2}$.

Let $\mathcal{B}$ be the event that $b \leq 15n^2/J^{1/2}$. By Markov's inequality, $\mathbb{P}(\mathcal{B}) > 2/3$. We finish the proof of Theorem 1.2 by proving the following.

Claim 4.5. *If both $\mathcal{A}$ and $\mathcal{B}$ hold, then $G[U]$ has at least k distinct degrees.*

Proof. We know that D is the union of vertex disjoint cliques, one for each vertex degree in $G[U]$. If $\mathcal{A}$ holds, then for each $1 \leq i \leq m$, the graph $D[A_i \cap U]$ contains at least $r_i + 1$ disjoint cliques $D_{i,0}, D_{i,1}, \ldots, D_{i,r_i}$ (which correspond to the degrees in $G[U]$ of the vertices in $B_{i,0}, B_{i,1}, \ldots, B_{i,r_i}$) each of size at least $\eta n/2^{k+1}k^3$. Additionally, if $\mathcal{B}$ holds, then there are no edges in D between $D_{i,s}$ and $D_{j,t}$ for any $1 \leq i < j \leq m$, $0 \leq s \leq r_i$ and $0 \leq t \leq r_j$. Indeed, if not, then $D_{i,s} \cup D_{j,t}$ induces a clique in D and we arrive at a contradiction since we would then have

$$b \geq |D_{i,s}||D_{j,t}| \geq \frac{\eta^2 n^2}{2^{2k+2}k^6} > \frac{15n^2}{\sqrt{J}};$$

here, the last inequality holds since $J = 10^4 2^{4k} k^{12}/\eta^4$. Hence, for all $1 \leq i \leq m$ and $0 \leq s \leq r_i$, the cliques $D_{i,s}$ all correspond to distinct degrees in $G[U]$. Thus, $G[U]$ has at least $\sum_{i=1}^m (r_i + 1)$ different degrees. Since

$$\frac{(r_i + 1)n}{k - 1 + \varepsilon} \geq |A_i| - \eta n,$$

it follows that

$$\sum_{i=1}^m (r_i + 1) \geq (1 - \beta - m\eta)(k - 1 + \varepsilon) > k - 1;$$

here, the last inequality holds since $\beta = \varepsilon/10k$, $\eta = \varepsilon\beta/10^5 k$ and $m \leq 10^4 k/\beta$. Since $\sum_{i=1}^m (r_i + 1)$ is an integer, this sum is at least k. We conclude that if both $\mathcal{A}$ and $\mathcal{B}$ hold, then $G[U]$ has at least k distinct degrees. $\qquad\square$

15

We know by the union bound that $\mathbb{P}(\mathcal{A} \cap \mathcal{B}) > 0$. Therefore, $G[U]$ has k distinct degrees with positive probability, contradicting our assumption that $f(G) < k$; the result follows. $\qquad\square$

5. Conclusion

We conclude this note by discussing two of the questions we alluded to in the introduction. First, as we mentioned earlier, we suspect that the following strengthening of Theorem 1.2 is true.

Conjecture 5.1. *Fix a positive integer $k \geq 2$. If n is sufficiently large, then $f(G) \geq k$ for every n-vertex graph G with $\mathrm{hom}(G) < n/(k-1)$.*

It may be read out of the proof of Theorem 1.2 that if G is an n-vertex graph with $\mathrm{hom}(G) > n/(100 \log n)$, then $f(G) = \Omega(n/\mathrm{hom}(G))$; for such graphs, this is a significant improvement over Theorem 1.1. We believe that it should be possible to prove a similar result for graphs with much smaller homogeneous sets.

Conjecture 5.2. *If G is an n-vertex graph with $\mathrm{hom}(G) > n^{1/2}$, then $f(G) = \Omega(n/\mathrm{hom}(G))$.*

References

1. N. Alon and J. H. Spencer, *The probabilistic method*, 3rd ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2008.
2. B. Barak, A. Rao, T. Shaltiel, and A. Wigderson, *2-source dispersers for $n^{o(1)}$ entropy, and Ramsey graphs beating the Frankl-Wilson construction*, Ann. of Math. **176** (2012), 1483–1543.
3. B. Bukh and B. Sudakov, *Induced subgraphs of Ramsey graphs with many distinct degrees*, J. Combin. Theory Ser. B **97** (2007), 612–619.
4. Y. Caro, *New results on the independence number*, Technical Report, Tel Aviv University (1979).
5. E. Chattopadhyay and D. Zuckerman, *Explicit two-source extractors and resilient functions*, Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing, ACM, New York, NY, 2016, pp. 670–683.
6. G. Cohen, *Two-source dispersers for polylogarithmic entropy and improved ramsey graphs*, Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing, ACM, New York, NY, 2016, pp. 278–284.

7. P. Erdős, *Some remarks on the theory of graphs*, Bull. Amer. Math. Soc. **53** (1947), 292–294.

8. P. Erdős and G. Szekeres, *A combinatorial problem in geometry*, Compositio Math. **2** (1935), 463–470.

9. P. Erdős, *Some of my favourite problems in various branches of combinatorics*, Matematiche (Catania) **47** (1992), 231–240.

10. P. Erdős and A. Szemerédi, *On a Ramsey type theorem*, Period. Math. Hungar. **2** (1972), 295–299.

11. V. Grolmusz, *Low rank co-diagonal matrices and Ramsey graphs*, Electron. J. Combin. **7** (2000), Research Paper 15.

12. H. J. Prömel and V. Rödl, *Non-Ramsey graphs are $c \log n$-universal*, J. Combin. Theory Ser. A **88** (1999), 379–384.

13. S. Shelah, *Erdős and Rényi conjecture*, J. Combin. Theory Ser. A **82** (1998), 179–185.

14. V. K. Wei, *A lower bound on the stability number of a simple graph*, Technical Memorandum TM 81-11217-9, Bell Laboratories (1981).

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK

E-mail address: `b.p.narayanan@dpmms.cam.ac.uk`

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK

E-mail address: `i.tomon@dpmms.cam.ac.uk`