Sizes of induced subgraphs of Ramsey graphs

Noga Alon*, József Balogh†, Alexandr Kostochka‡ and Wojciech Samotij§

January 27, 2009

Abstract

An n-vertex graph G is *c-Ramsey* if it contains neither a complete nor an empty induced subgraph of size greater than $c \log n$. Erdős, Faudree and Sós conjectured that every c-Ramsey graph with n vertices contains $\Omega(n^{5/2})$ induced subgraphs any two of which differ either in the number of vertices or in the number of edges, i.e. the number of distinct pairs $(|V(H)|, |E(H)|)$, as H ranges over all induced subgraphs of G, is $\Omega(n^{5/2})$. We prove an $\Omega(n^{2.3693})$ lower bound.

1 Introduction

For a graph $G = (V, E)$, call a set $W \subseteq V$ *homogenous*, if W induces a clique or an independent set. Let $\hom(G)$ denote the maximum size of a homogenous set of vertices of G. For a positive constant $c > 0$, an n-vertex graph G is called *c-Ramsey* if $\hom(G) \leq c \log n$.

Ramsey theory states that every n-vertex graph G satisfies $\hom(G) \geq (\log n)/2$, and for almost all such G, we have $\hom(G) \leq 2 \log n$. In other words, in a random graph G, the value $\hom(G)$ is of logarithmic order. Moreover, the only known examples of graphs with $\hom(G) = O(\log n)$ come from various constructions based on random graphs with edge density bounded away from 0 and 1. Therefore it is natural to ask whether c-Ramsey graphs look "random" in some sense.

This question has been an object of intense study. The first result in this area is due to Erdős and Szemerédi [11], who showed that the edge density of c-Ramsey graphs is bounded away from 0 and 1. Not much later Erdős and Hajnal [10] proved that for a fixed integer k, such graphs are *k-universal*, i.e. they contain every graph on k vertices as an induced subgraph. This was improved by Prömel and Rödl in [12], where they proved that in fact

*Tel Aviv University, Tel Aviv 69978, Israel and IAS, Princeton, NJ, 08540, USA. Research supported in part by the Israel Science Foundation, by a USA-Israel BSF grant, by NSF grant CCF 0832797 and by the Ambrose Monell Foundation. E-mail address: nogaa@tau.ac.il

†Department of Mathematics, University of Illinois, Urbana, IL 61801. E-mail address: jobal@math.uiuc.edu. This material is based upon work supported by NSF CAREER Grant DMS-0745185 and DMS-0600303, UIUC Campus Research Board Grants 06139, 07048 and 08086, and OTKA grant 049398.

‡Department of Mathematics, University of Illinois, Urbana, IL, 61801, USA and Sobolev Institute of Mathematics, Novosibirsk, Russia. E-mail address: kostochk@math.uiuc.edu. Research of this author is supported in part by NSF grant DMS-06-50784 and by grant08-01-00673 of the Russian Foundation for Basic Research.

§Department of Mathematics, University of Illinois, Urbana, IL, 61801, USA. Research supported in part by UIUC Campus Research Board Grant 08086. E-mail address: samotij2@illinois.edu.

c-Ramsey graphs are $d \log n$-universal, where the constant d depends only on c, which is asymptotically best possible. A similarly flavored result was obtained by Shelah. In [13] he proved that every c-Ramsey graph contains 2^{dn} non-isomorphic induced subgraphs, where again d is some constant depending only on c, settling a conjecture of Erdős and Rényi.

A related question was asked by Erdős and McKay (see [7], [8]), who conjectured that every c-Ramsey graph contains an induced subgraph with exactly m edges, for every $1 \leq m \leq dn^2$, where again the constant d depends only on c. The conjecture is still open, and the best currently known result is due to Alon, Krivelevich and Sudakov [3].

In this paper we tackle a similar problem, first posed by Erdős, Faudree and Sós (see [7], [8]), who stated the following conjecture:

Conjecture 1. *For every positive constant c, there is a positive constant $b = b(c)$, such that if G is a c-Ramsey graph on n vertices, then the number of distinct pairs $(|V(H)|, |E(H)|)$, as H ranges over all induced subgraphs of G, is at least $bn^{5/2}$.*

At the time the conjecture was stated, its authors knew how to prove an $\Omega(n^{3/2})$ lower bound. The same lower bound was also obtained as a corollary of a much stronger result of Bukh and Sudakov in [6]. Very recently it has been improved to $\Omega(n^2)$ by Alon and Kostochka in [2]. Here we further improve this bound to $\Omega(n^{1+\sqrt{30}/4-\epsilon}) \approx \Omega(n^{2.3693-\epsilon})$.

Theorem 2. *For every positive constants c and ϵ, there is a positive constant $b = b(c, \epsilon)$, such that if G is a c-Ramsey graph on n vertices, then the number of distinct pairs $(|V(H)|, |E(H)|)$, as H ranges over all induced subgraphs of G, is at least $bn^{1+\frac{\sqrt{30}}{4}-\epsilon} \approx bn^{2.3693-\epsilon}$.*

Remark. *In fact, as it will become clear in the proof of Theorem 2, we prove a bit stronger statement. Namely, we show that for $\Theta(n)$ different values of k, there are k-vertex induced subgraphs with $\Omega(n^{\frac{\sqrt{30}}{4}-\epsilon})$ different sizes.*

The paper is organized as follows. In Section 2 we introduce some notation and state a few older results that we will repeatedly use throughout the paper. At the end of Section 2 we formulate Theorem 12, a rather technical statement, from which we are able to quite easily derive the main result, Theorem 2. This is done in Section 3. Finally, in Section 4 we prove some technical lemmas that will later be used in the proof of Theorem 12, which we postpone till Section 5.

2 Basics

For a graph G we denote the number of vertices of G by $v(G)$, and the number of edges by $e(G)$. The (edge) density of G is $a(G) = e(G)\binom{v(G)}{2}^{-1}$. For any $v \in V(G)$ and a subset $W \subseteq V(G)$, let $d(v, W)$ be the number of neighbors of v lying in W. Similarly, if H is an induced subgraph of G and $u \in V(H)$, then $d_H(u) = d(u, V(H))$ will denote the degree of u in the subgraph H. The relation $H \leq G$ will always mean that H is an induced subgraph of G. For a subset $A \subseteq V(G)$ we denote the subgraph of G induced on A by $G[A]$. To increase clarity of presentation, if G is clear from the context, we will abbreviate $e(G[A])$ by $e(A)$. Finally, for every integer k, with $0 \leq k \leq v(G)$, we define the following quantities:

$$\begin{aligned} \psi(k, G) &= \max\{e(H) - e(H') : H, H' \leq G \text{ with } v(H) = v(H') = k\}, \\ \phi(k, G) &= |\{e(H) : H \leq G \text{ with } v(H) = k\}|. \end{aligned}$$

Note that the number of distinct pairs $(|V(H)|, |E(H)|)$ as in Conjecture 1 can be now written as

$$\left|\{(v(H), e(H)) : H \leq G\}\right| = \sum_{k=0}^{v(G)} \phi(k, G).$$

Erdős, Goldberg, Pach and Spencer ([9]; see also [5] and [2]) derived the following bound on $\psi(k, G)$ for graphs with edge density bounded away from 0 and 1.

Theorem 3. *For any positive $0 < \epsilon < 1/2$ and k and n satisfying $5/\epsilon < k < n/2$, and for any graph G on n vertices with density satisfying $\epsilon < a(G) < 1 - \epsilon$, we have $\psi(k, G) \geq 10^{-4} k^{3/2} \epsilon^{1/2}$.*

Suppose that each vertex $v \in V(G)$ is given a nonnegative weight $\omega(v)$. For a subgraph G' of G let its *weight* be defined as $\omega(G') = e(G') + \sum_{v \in V(G')} \omega(v)$. Generalizing the above definitions to weighted graphs we introduce a new parameter

$$\phi_\omega(k, G) \quad = \quad |\{\omega(G') : G' \leq G \text{ with } v(G') = k\}|.$$

Also, for a vertex v, let $d^\omega(v, W) = d(v, W) + \omega(v)$ and similarly $d^\omega_H(u) = d_H(u) + \omega(u)$. We will refer to these values as *weighted degrees.*

In the sequel we will repeatedly use the following results of Alon and Kostochka [2]. Although Theorem 4 does not appear there in the form in which it is stated below, it can be inferred from the proof of the main result of [2] (see concluding remarks in [2]).

Theorem 4. *For every $0 < \epsilon < 1/2$ there is an $n_0 = n_0(\epsilon)$ so that the following holds. Let $n \geq n_0$ and let G be an n-vertex graph with $\epsilon < a(G) < 1 - \epsilon$. Assume that $k \leq \frac{\epsilon n}{3}$ and every vertex $v \in V(G)$ is given a weight $\omega(v) \in [0, x \cdot \psi(k, G)/k]$, where $x \geq 1$. Then*

$$\phi_\omega(k, G) \geq 10^{-8} \frac{k}{x}.$$

Moreover one can find $10^{-8} k/x$ distinct sizes of induced k-vertex subgraphs of G, such that the difference between consecutive weights is at least mx, where $m = 500\psi(k, G)/k$.

Definition 5. *Let G be a graph on n vertices and let $0 \leq k \leq n$. Define*

$$m = m(k, G) = 500 \frac{\psi(k, G)}{k}.$$

For a k-element subset W of $V(G)$, call a vertex $v \in V(G)$ W-typical if

$$\left|d(v, W) - a(G)(k - 1)\right| \leq m + 1.$$

Theorem 6. *Let G be a graph on n vertices and let W be a k-element subset of $V(G)$, with $20 < k \leq n/3$. Then, all but at most $|W|/5$ vertices inside W are W-typical, and all but at most $|W|/5$ vertices outside W are W-typical.*

It is also good to keep in mind the following simple observation:

Observation 7. *Let G be a graph, and W a k-element subset of $V(G)$. If each vertex of G is given a nonnegative weight ω, then every W-typical vertex v satisfies*

$$\left|d^{\omega}(v,W) - a(G)(k-1)\right| \leq m + \omega(v) + 1.$$

In particular, if the weights are in the range $[0, x \cdot \psi(k,G)/k]$ for some $x \geq 1$, then all typical vertices satisfy

$$\left|d^{\omega}(v,W) - a(G)(k-1)|\right| \leq 2mx.$$

The two main definitions we are about to state are motivated by the following result of Erdős and Szemerédi from [11]:

Theorem 8. *For every positive constant c, there is some $\epsilon = \epsilon(c) > 0$, such that if G is an n-vertex c-Ramsey graph, then $\epsilon < a(G) < 1 - \epsilon$.*

Assume that G is a graph on n-vertices and H is an induced subgraph of G of order n^{δ}. It is clear that $\hom(H) \leq \hom(G)$. Therefore, if G is c-Ramsey, then H is c/δ-Ramsey, so in particular $a(H)$ is bounded away from 0 and 1. Informally, all large subgraphs of a c-Ramsey graph have edge density bouded away from 0 and 1. It makes sense to define a similar property for an arbitrary graph.

Definition 9. *For $0 < \epsilon < 1/2$ and $0 < \delta \leq 1$, let $\mathcal{D}(\epsilon, \delta)$ denote the family of graphs G, such that all induced subgraphs $H \leq G$ with $v(H) \geq v(G)^{\delta}$ have density $\epsilon < a(H) < 1 - \epsilon$.*

Having defined the class $\mathcal{D}(\epsilon, \delta)$, it is immediate to derive the following corollary of Theorem 8:

Corollary 10. *Let c and δ be positive constants. There are constants $0 < \epsilon < 1/2$ and n_0 (depending on c and δ), such that every n-vertex c-Ramsey graph with $n \geq n_0$ belongs to $\mathcal{D}(\epsilon, \delta)$.*

Keeping in mind the statement of Corollary 10, from now on we can concentrate our attention on graphs in classes $\mathcal{D}(\epsilon, \delta)$. Our aim will be to show that large enough graphs in $\mathcal{D}(\epsilon, \delta)$ have many induced subgraphs that differ either by number of vertices or weight (for a reasonably chosen weight function). The following definition should make this a little more precise.

Definition 11. *For every $0 < \epsilon < 1/2$ and $0 < \delta \leq 1$, let $\mathcal{P}(\epsilon, \delta)$ be the set of pairs (α, β), such that for some positive constants C, D, F and n_0 the following holds:*
All $G \in \mathcal{D}(\epsilon, \delta)$ with $n \geq n_0$ vertices satisfy

$$\phi_{\omega}(k,G) \quad \geq \quad C\frac{k}{x^F \log^D n} \min\left\{k^{\alpha}, \left(\frac{\psi(k,G)}{k}\right)^{\beta}\right\} \tag{1}$$

for all $k \in [\frac{\epsilon n}{100}, \frac{\epsilon n}{3}]$ and weight functions $0 \leq \omega \leq x \cdot \psi(k,G)/k$, where $x \geq 1$.

We will be working only with graphs whose edge density is bounded away from 0 and 1, and for all such G, Theorem 3 guarantees that $(\psi(k,G)/k)^{\beta} \geq \Omega(k^{\beta/2})$ for all $k \in [\frac{\epsilon n}{100}, \frac{\epsilon n}{3}]$. Therefore if $\alpha < \beta/2$, the minimum in (1) is equal to k^{α}, and can change only by a constant multiplicative factor when we decrease β to 2α. Since we do not care about the constants, but only the order of magnitude of $\phi_\omega(k,G)$, we can always assume that whenever $(\alpha,\beta) \in \mathcal{P}(\epsilon,\delta)$, we have $\alpha \geq \beta/2$.

Also, since $\psi(k,G) \leq k(k-1)/2$, trivially $(\psi(k,G)/k)^\beta \leq k^\beta$. Therefore if $\alpha > \beta$, the minimum in (1) is equal to $(\psi(k,G)/k)^\beta$, and will not change when we decrease α to β. Therefore we can also assume that all pairs $(\alpha,\beta) \in \mathcal{P}(\epsilon,\delta)$ satisfy $\alpha \leq \beta$.

Finally, we are able to state the main theorem, from which the main result, Theorem 2, will be derived as a simple corollary.

Theorem 12. *Suppose that* $(\alpha,\beta) \in \mathcal{P}(\epsilon,\delta)$. *Then* $\left(\frac{\beta+2}{\beta+5}, \frac{\alpha+1}{2}\right) \in \mathcal{P}(\epsilon,\delta/10)$.

We postpone the proof of Theorem 12 till Section 5. Instead we will now show how it implies the main result of this paper.

3 Proof of Theorem 2

First note that by Theorem 4, for all $0 < \epsilon' < 1/2$, the pair $(0,0)$ is in $\mathcal{P}(\epsilon',1)$. Define

$$i(\alpha,\beta) = \left(\frac{\beta+2}{\beta+5}, \frac{\alpha+1}{2}\right), \qquad \text{and note that} \qquad i^2(\alpha,\beta) = \left(\frac{\alpha+5}{\alpha+11}, \frac{2\beta+7}{2\beta+10}\right).$$

Now it is easy to see that both coordinates of the sequence $i^{2n}(0,0)$ are increasing and bounded, and hence the sequence converges to a pair (α,β) satisfying

$$\alpha = \frac{\alpha+5}{\alpha+11} \quad \text{and} \quad \beta = \frac{2\beta+7}{2\beta+10},$$

namely

$$(\alpha,\beta) := (\sqrt{30}-5, \frac{\sqrt{30}}{2}-2) \approx (0.4772, 0.7386).$$

By iteratively applying Theorem 12, we get that for every positive constant ϵ there is some δ, such that $(\alpha-\epsilon,\beta-\epsilon) \in \mathcal{P}(\epsilon',\delta)$ for every $\epsilon' > 0$. Let G be a c-Ramsey graph with $n = v(G)$ large enough. By Corollary 10, $G \in \mathcal{D}(\epsilon',\delta)$ for sufficiently small ϵ'. By the definition of $\mathcal{P}(\epsilon',\delta)$, if we set $x = 1$ and $\omega(v) = 0$ for all $v \in V(G)$, then for all $k \in [\frac{\epsilon' n}{100}, \frac{\epsilon' n}{3}]$

$$\begin{aligned}
\phi(k,G) &\geq C\frac{k}{\log^D n} \min\left\{k^{\alpha-\epsilon}, \left(\frac{\psi(k,G)}{k}\right)^{\beta-\epsilon}\right\} \\
&\geq \Omega\left(\frac{k \cdot \min\{k^{\alpha-\epsilon}, k^{(\beta-\epsilon)/2}\}}{\log^D n}\right) \\
&\geq \Omega\left(\min\{k^{1+\alpha-2\epsilon}, k^{1+\beta/2-\epsilon}\}\right) \\
&\geq \Omega\left(k^{\frac{\sqrt{30}}{4}-\epsilon}\right),
\end{aligned}$$

where the first inequality follows from Theorem 3, the second inequality holds because $k = \Theta(n)$ and hence $\log^D n = o(k^\epsilon)$, and the last one is due to $\beta/2 < \alpha$. Hence the number of distinct pairs $(v(H), e(H))$ can be bounded as follows:

$$\sum_{k=0}^{n} \phi(k,G) \geq \sum_{k \in [\frac{\epsilon' n}{100}, \frac{\epsilon' n}{3}]} \phi(k,G) \geq \Omega(n^{1+\frac{\sqrt{30}}{4}-\epsilon}).$$

4 Technical lemmas

Theorem 13. *Let M_j denote the family of all $n(n-1)\cdot\ldots\cdot(n-j+1)$ ordered subsets $A=\{a_1,\ldots,a_j\}$ of $[n]$ of cardinality j. Let $F: M_j \to \mathbb{R}$ be a real function, and suppose that if for A and $B=\{b_1,\ldots,b_j\}\in M_j$, the number of indices i for which $a_i \neq b_i$ is at most 2 then $|F(A)-F(B)|\leq 1$. Let $\mu = E(F)$ denote the expected value of $F(T)$, where T is chosen randomly and uniformly in M_j. Then, for every $\lambda > 0$,*

$$\Pr[|F(T)-\mu| \geq \lambda\sqrt{j}] \leq 2e^{-\lambda^2/2}.$$

Proof. We apply the method in [1], Lemma 2.2, which is based on known arguments, see, for example, Chapter 7 in [4]. Define a martingale $X_0, X_1, \ldots, X_j$ on the members T of M_j, where $X_i(T)$ is the expected value of $F(T')$ as T' ranges over all ordered subsets T of size j satisfying $t_1 = t'_1, \ldots, t_i = t'_i$. Thus $X_0 = \mu$ is a constant and $X_j(T) = F(T)$. This is clearly a Doob martingale. We claim that if two ordered sets A and B agree on their first i elements and differ in element number $i+1$, then $|X_{i+1}(A) - X_{i+1}(B)| \leq 1$. Indeed, there is a one to one correspondence π between all ordered sets $T \in M_j$ that agree with A on their first $i+1$ elements and all those that agree with B on their first $i+1$ elements, so that the symmetric difference between T and $\pi(T)$ is at most 2 for all T. (In this correspondence one simply swaps b_{i+1} and a_{i+1}). Thus the two averages $X_{i+1}(A)$ and $X_{i+1}(B)$ differ by at most 1. This easily implies that $|X_{i+1}(T) - X_i(T)| \leq 1$ for all i, as $X_i(T)$ is the average of numbers of the form $X_{i+1}(T')$ any pair of which differ by at most 1. The result now follows from Azuma's inequality (see, e.g., Theorem 7.2.1 in [4]). □

From the above it is easy to get the following:

Lemma 14. *Let s be a fixed integer. Let G be a graph on n vertices and let $N_1, \ldots, N_{n^s} \subseteq V(G)$, with N_i having size $0 \leq n_i \leq n$. Then there is an ordering $(v_1, \ldots, v_n)$ of the vertices in $V(G)$ such that, if we let $S_j = \{v_1, \ldots, v_j\}$, we have*

(i) for each $1 \leq j \leq n$ and every $1 \leq i \leq n^s$, $|S_j \cap N_i|$ differs from the expectation, $\frac{j}{n}\cdot n_i$, by at most $2j^{1/2}\sqrt{2(s+1)\log(2n)}$,

(ii) for each $1 \leq j \leq n$, the number of edges in $G[S_j]$ differs from the expectation, $\binom{j}{2}\cdot e(G)\binom{n}{2}^{-1}$, by at most $2j^{3/2}\sqrt{2\log(2n)}$.

Proof. Take a random ordering of the vertices of $V(G)$. For every fixed j, the set S_j is a uniform random subset of size j of the set of vertices of G. Fix $1 \leq i \leq n^s$. By applying Theorem 13 to the function $F(T) = |T \cap N_i|/2$, we get that

$$\Pr\left[\left||S_j \cap N_i| - \frac{j}{n}\cdot n_i\right| > 2j^{1/2}\sqrt{2(s+1)\log(2n)}\right] \leq \frac{1}{(2n)^{s+1}}.$$

Using the union bound we show that the probability that our ordering does not satisfy (i) is at most $n^s \cdot n \cdot 1/(2n)^{s+1} = 1/2^{s+1} < 1/2$.

Similarly, applying Theorem 13 to the function $F(T) = e(G[T])/(2|T|)$ yields

$$\Pr\left[\left|e(G[S_j]) - e(G)\cdot\binom{j}{2}\binom{n}{2}^{-1}\right| > 2j^{3/2}\sqrt{\log(2n)}\right] \leq \frac{1}{2n}.$$

Again the union bound implies that the probability that our ordering does not satisfy (ii) is at most $n \cdot 1/(2n) = 1/2$. Hence the probability that a random ordering of $V(G)$ satisfies both (i) and (ii) is greater than zero. □

Finally, we need a folklore lemma, whose proof we present for the sake of completeness.

Lemma 15. *Given $I_1, \ldots, I_n$ open bounded intervals, there exists a set $J \subseteq [n]$, such that $(I_j)_{j \in J}$ are disjoint and*

$$\sum_{j \in J} l(I_j) = l(\bigcup_{j \in J} I_j) \geq \frac{1}{2} l(\bigcup_{j=1}^{n} I_j),$$

where $l(I)$ denotes the length of I.

Proof. First let us delete all "redundant" intervals, i.e. every time some $I_i \subseteq \bigcup_{j \neq i} I_j$, we remove I_i. Since the union of all intervals does not change after any such deletion, without loss of generality we can assume that the family $I_1, \ldots, I_n$ contains no redundant intervals. We may also assume that the left ends of our intervals form a non-decreasing sequence. It easily follows that also the right ends form a non-decreasing sequence, or otherwise some $I_{i+1} \subseteq I_i$.

Now observe that whenever $j > i+1$, the interval I_j lies to the right of I_i (so in particular they are disjoint), or otherwise $I_{i+1} \subseteq I_i \cup I_j$. Hence all the intervals with even indices are pairwise disjoint, and similarly all the intervals with odd indices are pairwise disjoint. Obviously one of those families has to cover at least half of the entire union. □

5 Proof of Theorem 12

Fix some $\delta, \epsilon > 0$ and any pair $(\alpha, \beta) \in \mathcal{P}(\epsilon, \delta)$. Let $\gamma = (\beta+2)/(\beta+5)$ and C, D, F be as in Definition 11. Recall that by the remark following this definition, we can assume that $\beta/2 \leq \alpha \leq \beta$. Furthermore let $G \in \mathcal{D}(\epsilon, \delta/10)$ with n vertices and fix any $k \in [\frac{\epsilon n}{100}, \frac{\epsilon n}{3}]$. For each $v \in V(G)$, let $\omega(v)$ be its weight, satisfying $0 \leq \omega(v) \leq x \cdot \psi(k, G)/k$ for some $x \geq 1$. To simplify the notation we let $m = 500\psi(k, G)/k$. Throughout the proof we will assume that n is big enough. We will also omit all ceiling and floor signs, as they are not crucial. Finally, in order to avoid tedious constant computations, $C', D', F', C'', D'', F'' \ldots$ will denote some constants that depend only on δ, ϵ, α and β, and not on k, n, ω or G. In order to limit the number of different symbols in the proof, these constants will be often "recycled". We hope this does not cause too much confusion. Similarly, $C_1, C_2, \ldots$ will denote some constants depending only on δ, ϵ, α and β, but their values will remain fixed throughout the entire proof.

Theorem 4 guarantees the existence of a sequence $H_1, \ldots, H_{10^{-8}k/x}$ of k-vertex induced subgraphs of G such that $\omega(H_{i+1}) - \omega(H_i) \geq mx$ for every $1 \leq i < 10^{-8}k/x$.

Before we start, let us outline our general strategy. First, for each i in the above range, we will find an interval I_i centered at $\omega(H_i)$ that contains some N_i different weights of k-vertex induced subgraphs of G. Then, using Lemma 15 we will find a large family of pairwise disjont I_i's (thus making sure that they all contain different weights) and add up the corresponding N_i's. The sum we obtain will surely be a lower bound on the number of distinct weights of k-vertex induced subgraphs of G. In order to prove the promised lower bound, we will make sure that for all i, the ratio of N_i and the length $l(I_i)$ of the interval I_i will satisfy

$$\frac{N_i}{l(I_i)} \quad \geq \quad \frac{C'}{mx^{F'} \log^{D'} n} \min\{k^{\gamma}, m^{\frac{1+\alpha}{2}}\}, \tag{2}$$

and the total length of this disjoint family of I_i's will be of order $\Omega(k \cdot m)$.

Fix some i. By Theorem 6, at least $0.8k$ vertices in $V(H_i)$ are $V(H_i)$-typical. Hence we can find either a sequence $u_1, \dots, u_{0.5k^\gamma}$ of typical vertices with different weighted degrees $d^\omega_{H_i}$ or a set $B_i \subseteq V(H_i)$ of typical vertices with the same value of $d^\omega_{H_i}$, say d'_i, of size at least $k^{1-\gamma}$. Similarly, there are either typical vertices $v_1, \dots, v_{0.5k^\gamma} \in V(G) - V(H_i)$ with different weighted degrees $d^\omega(v_j, V(H_i))$ or a set $A_i \subseteq V(G) - V(H_i)$ of typical vertices with the same value of $d^\omega(-, V(H_i))$, say d_i, of size at least $k^{1-\gamma}$.

Assume first that we have found a sequence $u_1, \dots, u_{0.5k^\gamma} \in V(H_i)$ of typical vertices with different weighted degrees $d^\omega_{H_i}$. Let v be an arbitrary $V(H_i)$-typical vertex from $V(G) - V(H_i)$. Either at least $0.25k^\gamma$ vertices in the sequence (u_j) are adjacent to v, or at least $0.25k^\gamma$ vertices in that sequence are non-neighbors of v. Without loss of generality we can assume that the former holds and $u_1, \dots, u_{0.25k^\gamma} \in N_G(v)$. Consider graphs $H_{i,j} = G[V(H_i) + v - u_j]$. Then for a fixed i

$$\omega(H_{i,j}) = \omega(H_i) + d^\omega(v, H_i) - 1 - d^\omega_{H_i}(u_j)$$

are all distinct as j ranges from 1 to $0.25k^\gamma$. Moreover, since both u_j and v are $V(H_i)$-typical, by our assumption on ω and Observation 7,

$$|\omega(H_{i,j}) - \omega(H_i)| \leq |d^\omega(v, H_i) - d^\omega_{H_i}(u_j)| + 1 < 5mx.$$

Hence if we set $I_i = \omega(H_i) + (-5mx, 5mx)$, all the weights $\omega(H_{i,j})$ will belong to I_i. Therefore $N_i \geq C'k^\gamma$, and so (2) is satisfied.

We deal with the case when we can find a sequence $v_1, \dots, v_{0.5k^\gamma}$ in a similar fashion, exchanging v_j's in turn with some fixed $V(H_i)$-typical vertex $u \in V(H_i)$, again obtaining at least $0.25k^\gamma$ different weights in the interval $I_i = \omega(H_i) + (-5mx, 5mx)$. Hence for the remainder of the proof we assume that there are sets A_i, B_i and numbers d_i, d'_i, as described above.

Let $t = k^{2(1-\gamma)/3}$. Fix an arbitrary subset B'_i of B_i of size t. We can find a t-element subset $A'_i \subseteq A_i$ such that the (non-weighted) degrees $d(-, B'_i)$ of every pair of vertices of A'_i differ by at most $\sqrt{t}$. It is possible since

$$|A_i| \quad \geq \quad t^{3/2} = t \cdot \frac{|B'_i|}{\sqrt{t}}.$$

Let d^*_i be the edge density between A'_i and B'_i, that is

$$d^*_i \quad = \quad \sum_{v \in A'_i} \frac{d(v, B'_i)}{t^2}.$$

Then by the choice of A'_i, we have $|d(v, B'_i) - td^*_i| \leq \sqrt{t}$ for all $v \in A'_i$. Applying Lemma 14 to the graph $G[B'_i]$ and the neighborhoods of vertices from A'_i, or rather their traces on B'_i, one gets the following statement.

Claim 1. *We can enumerate $B'_i = \{b_1, \dots, b_t\}$, such that for all $1 \leq z \leq t$:*

(i) $|e(S_z) - a'_i\binom{z}{2}| \leq 2z^{3/2}\sqrt{2\log k}$,

(ii) $|d(v, S_z) - zd^*_i| \leq 2z^{1/2}\sqrt{5\log k}$ *for all* $v \in A'_i$,

where S_z abbreviates $\{b_1, \dots, b_z\}$ and $a'_i = e(G[B'_i])/\binom{t}{2}$.

Proof. Let $A'_i = \{v_1, \dots, v_t\}$ and then define N_j to be the set of neighbors of v_j in the set B'_i. By the remark preceeding the statement of this Lemma, we have

$$n_j = |N_j| = d(v_j, B'_i) \in [td^*_i - \sqrt{t}, td^*_i + \sqrt{t}]. \tag{3}$$

Lemma 14 applied to the graph $G[B'_i]$ and sets $N_1, \dots, N_t$ yields the desired enumeration $B'_i = \{b_1, \dots, b_t\}$. To see that (i) holds, it just suffices to note that $t \ll k$, so $\log(2t) \leq \log k$. For (ii), Lemma 14 (i) guarantees that

$$|d(v_j, S_z) - \frac{z}{t} n_j| \leq 2z^{1/2}\sqrt{4\log(2t)}, \tag{4}$$

and combining (3) with (4) gives the desired bound. $\square$

What we would like to do now is to obtain many k-vertex induced subgraphs of G with different weights by exchanging the set of vertices $S_z \subseteq B'_i$ with many subsets of A'_i, possibly for many values of z.

To get started and see how this idea works in practice, let T_z be some set of z vertices from A'_i, and let $H'_i(z) = G[V(H_i) \cup T_z - S_z]$. We compute the weight of this graph.

$$\begin{aligned}
\omega(H'_i(z)) &= \omega(H_i) - \sum_{j=1}^{z}(\omega(b_j) + d_{H_i}(b_j)) + e(S_z) + \sum_{v \in T_z}(\omega(v) + d(v, H_i - S_z)) + e(T_z) \\
&= \omega(H_i) - d'_i z + e(S_z) + \sum_{v \in T_z}(d_i - d(v, S_z)) + e(T_z) \\
&= \omega(H_i) + \Delta_i z + e(S_z) + e(T_z) - \sum_{v \in T_z} d(v, S_z), \qquad (5)
\end{aligned}$$

where $\Delta_i = d_i - d'_i \in [-4mx, 4mx]$. If all the degrees $d(v, S_z)$, where v ranges over $A'_i \supseteq T_z$, were equal, for fixed i and z the weight of $H'_i(z)$ would depend only on $e(T_z)$. Even though it does not have to be the case (our last claim only guarantees that $d(v, S_z)$ are all "close" to $d^*_i z$), this will not be a big issue for us, since, as we will later see, by assigning carefully chosen weights to vertices in A'_i, we can compensate for the possibly uneven distribution of the degrees.

Fix some $z \geq n^{1/10}$. Let $d^{\max}_i(z) = \max_{v \in A'_i} d(v, S_z)$ and for each $v \in A'_i$ set $\omega'(v) = d^{\max}_i(z) - d(v, S_z)$. If we again let T_z be some z-subset of A'_i, then

$$\omega'(T_z) = e(T_z) + \sum_{v \in T_z} \omega'(v) = e(T_z) - \sum_{v \in T_z} d(v, S_z) + d^{\max}_i(z) \cdot z. \tag{6}$$

Hence if we combine (5) with (6), the weight of $H'_i(z) = G[H_i \cup T_z - S_z]$ can be written in the form

$$\omega(H'_i(z)) = \omega(H_i) + \Delta_i z + e(S_z) - d^{\max}_i(z) \cdot z + \omega'(T_z),$$

where only the last term depends on the choice of T_z as a particular z-subset of A'_i.

Claim 2. *There are positive constants C_1 and D_1, such that for all $n^{1/10} \leq z \leq t' = \epsilon t/3$, we have*

$$\phi_{\omega'}(z, G[A'_i]) \geq \frac{C_1 z}{\log^{D_1} n} \min\left\{ z^{\alpha}, \left(\frac{\psi(z, A'_i)}{z}\right)^{\beta} \right\}. \tag{7}$$

Proof. Let A_i'' be any $(3z/\epsilon)$-element subset of A_i' that contains some z vertices spanning the most edges among all z-vertex subsets of A_i' and some z vertices spanning the least edges among all z-vertex subsets of A_i'. By construction, $\psi(z, A_i'') = \psi(z, A_i')$. Since we assumed that z is big enough, $G \in \mathcal{D}(\epsilon, \delta/10)$ implies that $G[A_i''] \in \mathcal{D}(\epsilon, \delta)$. By our assumption $(\alpha, \beta) \in \mathcal{P}(\epsilon, \delta)$, we have

$$\phi_{\omega'}(z, G[A_i'']) \geq \frac{C'z}{\log^D z \cdot \log^F n} \min\left\{ z^\alpha, \left(\frac{\psi(z, A_i'')}{z}\right)^\beta \right\},$$

since from Claim 1 (ii) and Theorem 3 it follows that (provided n is large enough) $0 \leq \omega' \leq \log n \cdot \psi(z, A_i'')/z$. Finally, (7) follows because $A_i'' \subseteq A_i'$ and therefore $\phi_{\omega'}(z, G[A_i']) \geq \phi_{\omega'}(z, G[A_i''])$. $\square$

Let us again rewrite formula (5).

$$\begin{aligned} \omega(H_i'(z)) &= \omega(H_i) + \Delta_i z + a_i'\binom{z}{2} + (e(S_z) - a_i'\binom{z}{2}) \\ &\quad + e(T_z) - \sum_{v \in T_z} (d(v, S_z) - d_i^* z) - d_i^* z^2. \end{aligned}$$

Now let $a_i = a(G[A_i'])$ and recall that $z \geq n^{1/10}$. Since:

- $|e(S_z) - a_i'\binom{z}{2}| \leq 2z^{3/2}\sqrt{2\log k}$ by Claim 1 (i),
- $|d(v, S_z) - d_i^* z| \leq 2z^{1/2}\sqrt{5 \log k}$ by Claim 1 (ii),
- $|e(T_z) - a_i \cdot \binom{z}{2}| \leq \psi(z, A_i')$ by the definition of ψ and
- $\psi(z, A_i') \geq 10^{-4}\epsilon^{1/2}z^{3/2}$ by Theorem 3 and the assumptions on G and z,

we conclude that $\omega(H_i'(z)) - \omega(H_i)$ lands in the interval

$$I_i(z) \quad = \quad \Delta_i z + (a_i' + a_i)\binom{z}{2} - d_i^* z^2 + C_2 \log k \cdot \big(- \psi(z, A_i'), \psi(z, A_i')\big), \tag{8}$$

where C_2 is some constant depending only on ϵ. In particular, the following is true.

Claim 3. *We can find k-vertex induced subgraphs of G with at least $\phi_{\omega'}(z, G[A_i'])$ different weights in the interval $\omega(H_i) + I_i(z)$.*

In the remainder of the proof we will carefully estimate the number of different weights in all these intervals. Recall that $n^{1/10} \leq z \leq t' = \frac{\epsilon t}{3}$ is the number of vertices we want to exchange. For the sake of brevity let $|I_i(z)|$ denote $\max\{|\inf I_i(z)|, |\sup I_i(z)|\}$, i.e. how much the weight of $H_i'(z)$ can possibly differ from the weight of H_i, and

$$c(I_i(z)) \quad = \quad \Delta_i z + (a_i' + a_i)\binom{z}{2} - d_i^* z^2 \tag{9}$$

will denote the center of the interval $I_i(z)$.

Before we proceed with the counting, first let us prove a technical lemma.

Claim 4. *For all $z \geq n^{1/10}$, the centers $c(I_i(z+1))$ and $c(I_i(z))$ satisfy*

$$|c(I_i(z+1)) - c(I_i(z)) - \Delta_i| < 5z.$$

In particular,

$$|\Delta_i| - 5z < |c(I_i(z+1)) - c(I_i(z))| < |\Delta_i| + 5z.$$

Proof. Looking at the definition of $c(I_i(z))$ in (9), it is easy to see that the difference $\delta = c(I_i(z+1)) - c(I_i(z))$ can be computed as follows:

$$\begin{aligned}\delta &= \Delta_i + (a_i' + a_i) \cdot (\binom{z+1}{2} - \binom{z}{2}) + d_i^*(z^2 - (z+1)^2) \\ &= \Delta_i + (a_i' + a_i)z - d_i^*(2z+1).\end{aligned}$$

Hence,

$$\left|\delta - \Delta_i\right| \leq (a_i' + a_i)z + d_i^*(2z+1) \leq 4z + 1,$$

where the last inequality holds because $a_i, a_i', d_i^* \in [0,1]$. □

We will now analyze the function $z \mapsto |I_i(z)|$ and split the proof into several cases. First, recall that $m = 500\psi(k,G)/k$ and z is in the range $n^{1/10} \leq z \leq t' = \epsilon t/3$.

5.1 Case 1. $\max_z |I_i(z)| < 3mx$.

In particular $I_i(t') \subseteq (-3mx, 3mx)$. We set $I_i = \omega(H_i) + (-3mx, 3mx)$ and note that by Claims 2 and 3,

$$\begin{aligned}N_i &\geq \phi_{\omega'}(t', G[A_i']) \geq \frac{C_1 t'}{\log^{D_1} n} \min\left\{(t')^\alpha, \left(\frac{\psi(t', A_i')}{t'}\right)^\beta\right\} \\ &\geq \frac{C'}{\log^{D'} n} \min\{k^{\frac{2}{3}(1-\gamma)(1+\alpha)}, k^{\frac{2}{3}(1-\gamma)(1+\frac{\beta}{2})}\} = \frac{C'}{\log^{D'} n} k^\gamma, \qquad (10)\end{aligned}$$

since $\alpha \geq \beta/2$ and $\frac{2}{3}(1-\gamma)(1+\frac{\beta}{2}) = \gamma$. Finally, note that $l(I_i) = 6mx$ and therefore inequality (2) is satisfied. This completes the proof in Case 1.

5.2 Case 2. $\max_z |I_i(z)| \geq 3mx$.

Let z_0 be the minimal z such that $|I_i(z)| \geq 3mx$. First we show that $|I_i(z_0)|$ is in fact not much larger than $3mx$. To make it precise, let us prove the following claim.

Claim 5. *If $z_0 > n^{1/10}$, then there is a constant C_3 depending only on ϵ, such that*

$$|I_i(z_0)| < C_3 mx.$$

Proof. Minimality of z_0 implies that $|c(I_i(z_0 - 1))|$ and $C_2 \log k \cdot \psi(z_0 - 1, A_i')$ are at most $3mx$. By Claim 4,

$$|c(I_i(z_0))| \leq 3mx + |\Delta_i| + 5z_0 < C'mx,$$

where the second inequality holds since, by Theorem 3, we have $3mx \geq \psi(z_0 - 1, A_i') = \Omega(z_0^{3/2})$ and, by Observation 7, we have $|\Delta_i| \leq 4mx$ (recall that we work only with typical vertices). Finally, note that $\psi(z_0, A_i')$ differs from $\psi(z_0 - 1, A_i')$ by at most z_0 and therefore

$$|I_i(z_0)| = |c(I_i(z_0))| + C_2 \log k \cdot \psi(z_0, A_i') < C'mx + C'' \log k \cdot \psi(z_0 - 1, A_i') < C_3 mx.$$

□

From (8) it easily follows that

$$|I_i(z)| \;\leq\; |\Delta_i| z + |(a_i' + a_i)\binom{z}{2} - d_i^* z^2| + C_2 \log k \cdot \psi(z, A_i'), \tag{11}$$

and therefore we can split the proof into further subcases, depending on which of the three terms on the right-hand side of (11) is the "dominant" term.

5.2.1 Case 2a. $C_2 \log k \cdot \psi(z_0, A_i') \geq mx$.

First note that $z_0 = \Omega(\sqrt{\frac{mx}{\log k}})$, simply because $\psi(z, -) \leq \binom{z}{2}$. Claim 5 allows us to set $I_i = \omega(H_i) + (-C_3 mx, C_3 mx) \supseteq \omega(H_i) + I_i(z_0)$. Finally by Claims 2 and 3,

$$\begin{aligned} N_i &\geq \phi_{\omega'}(z_0, G[A_i']) \geq \frac{C_1 z_0}{\log^{D_1} n} \min\left\{(z_0)^{\alpha}, \left(\frac{\psi(z_0, A_i')}{z_0}\right)^{\beta}\right\} \\ &= \frac{C_1}{\log^{D_1} n} \min\{z_0^{(1+\alpha)}, \psi(z_0, A_i')^{\beta} z_0^{1-\beta}\} \\ &\geq \frac{C'}{\log^{D'} n} \min\{(mx)^{\frac{1+\alpha}{2}}, (mx)^{\beta}(mx)^{\frac{1-\beta}{2}}\} = \frac{C'}{\log^{D'} n}(mx)^{\frac{1+\alpha}{2}}, \end{aligned} \tag{12}$$

since $\alpha \leq \beta$. The length of I_i is $l(I_i) = 2C_3 mx$, hence the inequality (2) is satisfied. This completes the proof in Case 2a.

5.2.2 Case 2b. $z_0 \geq \sqrt{mx/3}$ (takes care of $|(a_i' + a_i)\binom{z_0}{2} - d_i^* z_0^2| \geq mx$).

Note that $|\Delta_i| < \sqrt{mx \log k}$ or else the center of $I_i(z_1)$, where $z_1 = \sqrt{mx}/\log^{1/4} k < z_0$ would be at distance at least

$$|\Delta_i| z_1 - |(a_i' - a_i(z_1))\binom{z_1}{2} - d_i^* z_1^2| \geq mx \log^{1/4} k - O(\frac{mx}{\sqrt{\log k}}) > 3mx$$

from 0, contradicting the minimality of z_0. From (11) and the above bound on $|\Delta_i|$ it follows that $|I_i(\sqrt{mx}/\log k)| \leq C' \frac{mx}{\sqrt{\log k}} \leq 0.1mx$. In the sequel we will combine this simple observation with the following claim.

Claim 6. *For $\sqrt{mx}/\log k \leq z < z_0$, the intervals $I_i(z)$ and $I_i(z+1)$ intersect.*

Proof. By Claim 4, the distance δ between the centers of these two intervals is at most $|\Delta_i| + 5z$, and $\psi(z, A_i') \geq C' z^{3/2} \gg (mx)^{1/2+1/6} \gg |\Delta_i|$. Now we are done, since $l(I_i(z)) = 2C_2 \log k \cdot \psi(z, A_i')$. □

The above observation, together with Claim 6 show that the family $\{I_i(z) : \sqrt{mx}/\log k \leq z \leq z_0\}$ covers an interval of length at least $2.9mx$ (either $[0.1mx, 3m]$ or $[-3m, -0.1mx]$). Also, Claim 5 shows that $I_i(z_0)$ (and by the choice of z_0 also all the other $I_i(z)$'s in question) is entirely contained in $-\omega(H_i) + I_i = (-C_3mx, C_3mx)$.

Again, by Claims 2 and 3, in each of the $\omega(H_i) + I_i(z)$'s we can find at least

$$\begin{aligned}\phi_{\omega'}(z, G[A_i']) &\geq \frac{C_1}{\log^{D_1} n}\min\{z^{1+\alpha}, \psi(z, A_i')^{\beta} z^{1-\beta}\} \\ &\geq \frac{C'}{\log^{D'} n}\min\{m^{\frac{1+\alpha}{2}}, l(I_i(z))^{\beta} m^{\frac{1-\beta}{2}}\}\end{aligned}$$

weights. Lemma 15 ensures we can find a collection of disjoint $I_i(z)$'s, indexed by $z \in Z$, of total length at least $1.45m$. Hence

$$\begin{aligned}N_i &\geq \sum_{z\in Z}\frac{C'}{\log^{D'} n}\min\{m^{\frac{1+\alpha}{2}}, l(I_i(z))^{\beta} m^{\frac{1-\beta}{2}}\} \\ &\geq \frac{C'}{\log^{D'} n}\min\{m^{\frac{1+\alpha}{2}}, \sum_{z\in Z} l(I_i(z))^{\beta} m^{\frac{1-\beta}{2}}\} \\ &\geq \frac{C'}{\log^{D'} n}\min\{m^{\frac{1+\alpha}{2}}, \left(\sum_{z\in Z} l(I_i(z))\right)^{\beta} m^{\frac{1-\beta}{2}}\} \\ &\geq \frac{C''}{\log^{D''} n}\min\{m^{\frac{1+\alpha}{2}}, m^{\beta} m^{\frac{1-\beta}{2}}\} \\ &= \frac{C''}{\log^{D''} n} m^{\frac{1+\alpha}{2}},\end{aligned} \tag{13}$$

where the third inequality holds because $0 \leq \beta \leq 1$ and therefore $y \mapsto y^{\beta}$ is concave. Once again, $l(I_i) = 2C_3mx$ and therefore inequality (2) is satisfied. This completes the proof in Case 2b.

5.2.3 Case 2c. $|\Delta_i|z_0 \geq mx$ and $\psi(z_0, A_i') \geq |\Delta_i|$.

We can easily assume that neither of the previous subcases holds, so in particular $C_2 \log k \cdot \psi(z_0, A_i') < mx$ and $z_0 < \sqrt{mx/3}$, which implies $|\Delta_i| > \sqrt{3mx}$. It is not hard to see that $|\Delta_i|z_0$ cannot be larger than $8mx$. If it was the case, i.e. $|\Delta_i|z_0 > 8mx$, then by Claim 4,

$$|c(I_i(z_0 - 1))| > 8mx - |\Delta_i| - 5z \geq 8mx - 4mx - mx = 3mx,$$

contradicting the minimality of z_0. Moreover, z_0 is not too small either, since

$$z_0^2 > \psi(z_0, A_i') \geq |\Delta_i| > \sqrt{3mx} = \sqrt{1500x \cdot \psi(k, G)/k} \geq C'k^{1/4}. \tag{14}$$

Before we proceed, we need the following claim.

Claim 7. *For all $z_0/30 \leq z \leq z_0$,*

$$10^{-3}\psi(z_0, A_i') \leq \psi(z, A_i') \leq 48\psi(z_0, A_i').$$

Proof. The first inequality follows from a simple averaging argument (see Observation 4 in [2]), which implies that

$$\psi(z, A'_i) \geq \psi(z_0, A'_i) \cdot \binom{z}{2} / \binom{z_0}{2}.$$

Finally, Lemma 6 in [2] yields that for every n-vertex graph G and all $0 < s < k < n/3$, we have $\psi(s, G) \leq 48\psi(k, G)$. This implies the second inequality. □

Claim 7 implies that $0 \leq \omega' \leq C' z^{1/2}\sqrt{\log k} \leq \log n \cdot \psi(z, A'_i)/z$ for all z in the range $[z_0/30, z_0]$. Moreover, recall that (14) implies that $z_0/30 \geq n^{1/10}$, and hence by Claims 2 and 3, in each interval $\omega(H_i) + I_i(z)$ we find at least

$$\begin{aligned}
\phi_{\omega'}(z, G[A'_i]) &\geq \frac{C_1 z}{\log^{D_1} n} \min\left\{ z^{\alpha}, \left(\frac{\psi(z, A'_i)}{z}\right)^{\beta} \right\} \\
&= \frac{C_1}{\log^{D_1} n} \min\{z^{1+\alpha}, \psi(z, A'_i)^{\beta} z^{1-\beta}\} \\
&\geq \frac{C'}{\log^{D'} n} \min\{l(I_i(z))^{\frac{1+\alpha}{2}}, l(I_i(z))^{\beta} l(I_i(z))^{\frac{1-\beta}{2}}\} \\
&= \frac{C'}{\log^{D''}} l(I_i(z))^{\frac{1+\alpha}{2}}
\end{aligned}$$

k-vertex induced subgraphs with different weights. Now recall that $|\Delta_i| z_0 \leq 8mx$ and therefore

$$|c(I_i(z_0/30))| \leq |\Delta_i| \frac{z_0}{30} + 3\left(\frac{z_0}{30}\right)^2 \leq \frac{4mx}{15} + \frac{mx}{900} \leq \frac{mx}{2}.$$

Moreover, by Claims 4 and 7,

$$|c(I_i(z+1)) - c(I_i(z))| < |\Delta_i| + 5z \leq 2\psi(z_0, A'_i) \leq C_2 \log k \cdot \psi(z, A'_i) = l(I_i(z))/2,$$

so the intervals $I_i(z)$ and $I_i(z+1)$ intersect and hence the family $\{I'_i(z) : z_0/30 \leq z \leq z_0\}$ will cover an interval of length at least $2.5mx$. Lemma 15 ensures we can find a collection of disjoint $I_i(z)$'s, indexed by $z \in Z$, of total length not smaller than $1.25m$. By (11),

$$|I_i(z_0)| \;\leq\; |\Delta_i| z_0 + z^2 + C_2 \log k \cdot \psi(z_0, A'_i) < 8mx + mx + mx,$$

and therefore all $I_i(z)$'s in question are entirely contained in the interval $(-10mx, 10mx)$. Hence, if we set $I_i = \omega(H_i) + (-10mx, 10mx)$, we will have

$$N_i \;\geq\; \frac{C'}{\log^{D'} n} \sum_{z\in Z} l(I_i(z))^{\frac{1+\alpha}{2}} \geq \frac{C'}{\log^{D'} n} \left(\sum_{z\in Z} l(I_i(z))\right)^{\frac{1+\alpha}{2}} \geq \frac{C'}{\log^{D'} n} m^{\frac{1+\alpha}{2}}, \tag{15}$$

where the second inequality follows by concavity of $y \mapsto y^{\frac{1+\alpha}{2}}$, since $0 \leq \alpha \leq 1$. Finally, let us note that inequality (2) is satisfied. This completes the proof in Case 2c.

5.2.4 Case 2d. $|\Delta_i|z_0 \geq mx$ and $\psi(z_0, A_i') < |\Delta_i|$.

Recall that $t' = \varepsilon t/3$. This time we have to let z be a little larger, i.e. we define

$$z_1 = \min\{t', \min\{z : \psi(z, A_i') \geq |\Delta_i|\}\}.$$

Note that there are two distinct cases to consider, depending on which value in the above minimum is smaller.

Case 2d-A. $z_1 = t'$ and $\psi(z, A_i') < |\Delta_i|$ for all $z \leq z_1$.

First note that $z_1 \ll \psi(z_1, A_i') < |\Delta_i|$, so

$$c(I_i(z_1)) \geq |\Delta_i|z_1 - z_1^2 \geq 0.5|\Delta_i|z_1,$$

and

$$c(I_i(z_1/30)) \leq |\Delta_i|\frac{z_1}{30} + \left(\frac{z_1}{30}\right)^2 \leq 0.1|\Delta_i|z_1.$$

Claim 8. *There are at least $C'z_1/\log k$ pairwise disjoint intervals among $\{I_i(z) : z_1/30 \leq z \leq z_1\}$.*

Proof. Since $|\Delta_i|z$ "dominates" both z and $l(I_i(z))$, intuitively it is clear that whenever $z_2 - z_1$ is big enough, $I_i(z_1)$ and $I_i(z_2)$ are disjoint. Formally, by Claim 4,

$$\begin{aligned} |c(I_i(z_2)) - c(I_i(z_1))| &\geq (z_2 - z_1) \cdot |\Delta_i| - 5z_2(z_2 - z_1) \\ &= (z_2 - z_1) \cdot (|\Delta_i| - 5z_2) \geq \frac{(z_2 - z_1)|\Delta_i|}{2}, \end{aligned}$$

and therefore whenever

$$z_2 - z_1 \geq 4C_2 \log k \geq \frac{l(I_i(z_1)) + l(I_i(z_2))}{|\Delta_i|},$$

the intervals $I_i(z_1)$ and $I_i(z_2)$ are disjoint. $\square$

Note that for each z,

$$|I_i(z)| \leq |\Delta_i|z + z^2 + C_2 \log k \cdot \psi(z, A_i') < 2|\Delta_i|z,$$

so $(-2|\Delta_i|t', 2|\Delta_i|t')$ contains all the intervals $I_i(z)$, with $z_1/30 \leq z \leq z_1$. Finally, $z_1/30 \geq n^{1/10}$, and so by Claims 2, 3 and 8, if we set $I_i = \omega(H_i) + (-2|\Delta_i|t', 2|\Delta_i|t')$, we will get

$$\begin{aligned} N_i &\geq \frac{C'}{\log k}t' \cdot \frac{C_1 t'}{\log^{D_1} n} \min\left\{(t')^\alpha, \left(\frac{\psi(t', A_i')}{t'}\right)^\beta\right\} \\ &\geq t' \cdot \frac{C''}{\log^{D'} n} \min\{k^{\frac{2}{3}(1-\gamma)(1+\alpha)}, k^{\frac{2}{3}(1-\gamma)(1+\frac{\beta}{2})}\} = t' \cdot \frac{C''}{\log^{D'} n} k^\gamma. \end{aligned} \tag{16}$$

Recall that we are exchanging only $V(H_i)$-typical vertices and therefore $|\Delta_i| \leq 4mx$. Hence $l(I_i) \leq 16mx \cdot t'$ and therefore inequality (2) is satisfied. That completes the proof in Case 2d-A.

Case 2d-B. $\psi(z_1, A_i') \geq |\Delta_i|$.

We can simply rewrite the proof of Case 2c here, replacing z_0 with z_1. The only change is that $I_i = \omega(H_i) + (-C'Mx, C'Mx)$, where $M = |\Delta_i|z_1$ and $|c(I_i'(\frac{z_1}{30}))| \leq 0.5Mx$, and in (15), $m^{\frac{1+\alpha}{2}}$ will be replaced by $M^{\frac{1+\alpha}{2}}$. Hence we consider Case 2d-B resolved.

To finish the proof, note that each time (see: (10), (12), (13), (15), (16)) we were able to construct at least N_i graphs with different weights in the interval I_i, such that the aforementioned inequality (2) holds:

$$\frac{N_i}{l(I_i)} \quad \geq \quad \frac{C'}{mx^{F'} \log^{D'} n} \min\{k^\gamma, m^{\frac{1+\alpha}{2}}\}. \tag{2}$$

Moreover, the intervals I_i, which are centered at $\omega(H_i)$, all have length at least mx. Therefore these intervals cover the (disjoint!) family $\{\omega(H_i) + [-0.5mx, 0.5mx] : 1 \leq i \leq 10^{-8}k/x\}$. Hence

$$l(\bigcup_{i=1}^{10^{-8}k/x} I_i) \geq 0.5 \cdot 10^{-8}k \cdot m.$$

By Lemma 15, we can find a subfamily of pairwise disjoint I_i's of joint length $C'km$. That gives us at least

$$C'km \cdot \min \frac{N_i}{l(I_i)} \geq \frac{C'k}{x^{F'} \log^{D'} n} \min\{k^\gamma, m^{\frac{1+\alpha}{2}}\} = \frac{C''k}{x^{F'} \log^{D'} n} \min\left\{k^{\frac{\beta+2}{\beta+5}}, \left(\frac{\psi(k,G)}{k}\right)^{\frac{1+\alpha}{2}}\right\}$$

different weights, for some absolute constants C'', D', F'. This completes the proof.

6 Concluding remarks

It seems that the main reason why our argument fails to prove an $\Omega(n^{5/2})$ lower bound is the lack of deeper understanding of the behavior of the function $z \mapsto \psi(z, G)$. The only estimates for $\psi(z, G)$ we are using in the proof, namely $\Omega(z^{3/2}) \leq \psi(z, G) \leq O(z^2)$, do not exploit the dependence of $\psi(z, G)$ and $\psi(z', G)$ for different values of z and z' (except for when z and z' are of the same order of magnitude, see Claim 7). Note that in a random graph $G(n, p)$, where $p \in (0, 1)$ is fixed (independent of n), with high probability we have $\psi(z, G) = \Theta(z^{3/2})$ for all $z = n^{\Omega(1)}$.

Suppose we assume that there is some $\rho = \rho(G) \in [1/2, 1]$, such that $\psi(z, G) \approx z^{1+\rho}$ for all $z = n^{\Omega(1)}$. A simple (but lengthy and tedious) analysis of the proof shows that under that additional assumption (which we will not try to make much more precise), Theorem 12 could be improved to (note that since the order of $\psi(k, G)$ is known, the parameter β is now obsolete)

$$(\alpha, *) \in \mathcal{P}(\epsilon, \delta) \implies \left(\frac{2+2\alpha}{5+2\alpha}, *\right) \in \mathcal{P}(\epsilon, \delta/10).$$

This in turn would imply an $\Omega(n^{5/2-\epsilon})$ lower bound for the number of distinct sizes of induced subgraphs. Further analysis shows that even a much more modest assumption of the form $\Omega(z^{3/2+\rho_1}) \leq \psi(z, G) \leq \Omega(z^{2-\rho_2})$, where at least one of ρ_1, ρ_2 is positive, would further improve the current lower bound.

7 Acknowledgment

The authors thank the anonymous referee for valuable comments and suggestions.

References

[1] N. Alon and G. Gutin, *Properly colored Hamilton cycles in edges colored complete graphs*, Random Structures and Algorithms **11** (1997), 179–186.

[2] N. Alon and A. Kostochka, *Induced subgraphs with distinct sizes*, Random Structures and Algorithms **34** (2009), 45–53.

[3] N. Alon, M. Krivelevich, and B. Sudakov, *Induced subgraphs of prescribed size*, Journal of Graph Theory **43** (2003), 239–251.

[4] N. Alon and J. Spencer, *The Probabilistic Method, Second Edition*, Wiley, 2000.

[5] B. Bollobás and A.D. Scott, *Discrepancy in graphs and hypergraphs*, Bolyai Soc. Math. Studies **15** (2006), 33–56.

[6] B. Bukh and B. Sudakov, *Induced subgraphs of Ramsey graphs with many distinct degrees*, Journal of Combinatorial Theory B **97** (2007), 612–619.

[7] P. Erdős, *Some of my favorite problems in various branches of combinatorics*, Matematiche (Catania) **47** (1992), 231–240.

[8] ———, *Some recent problems and results in graph theory. The Second Kraków Conference on Graph Theory (Zgorzelisko, 1994)*, Discrete Math. **164** (1997), 81–85.

[9] P. Erdős, M. Goldberg, J. Pach, and J. Spencer, *Cutting a graph into two dissimilar halves*, J. Graph Theory **2** (1988), 121–131.

[10] P. Erdős and A. Hajnal, *On spanned subgraphs of graphs*, Contributions to Graph Theory and its Applications (Internat. Colloq., Oberhof, 1977), 1977, pp. 80–96.

[11] P. Erdős and E. Szemerédi, *On a Ramsey type theorem. Collection of articles dedicated to the memory of A. Rényi*, I. Period. Math. Hungar. **2** (1972), 295–299.

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

[13] S. Shelah, *Erdős and Rényi conjecture*, Journal of Combinatorial Theory A **82** (1998), 179–185.