Disjoint induced subgraphs of the same order and size

Béla Bollobás, Teeradej Kittipassorn, Bhargav Narayanan, and Alex Scott

Abstract. For a graph G, let $f(G)$ be the largest integer k for which there exist two vertex-disjoint induced subgraphs of G each on k vertices, both inducing the same number of edges. We prove that $f(G) \geq n/2 - o(n)$ for every graph G on n vertices. This answers a question of Caro and Yuster from 2009.

1. Introduction

Given a graph G, can we guarantee that G contains two large, vertex-disjoint copies of the same graph? It follows from Ramsey's theorem that any graph on n vertices contains two vertex-disjoint isomorphic induced subgraphs on $\Omega(\log n)$ vertices; by considering a random graph on n vertices, it is easy to check that this is also best-possible up to constant factors.

What if, rather than asking for two isomorphic subgraphs, we ask for two subgraphs that are the same with respect to one or more graph parameters? Caro and Yuster [6] considered the question of finding two vertex-disjoint subgraphs of a given graph of the same order which induce the same number of edges. For a graph G, let $f(G)$ be the largest integer k such that there are two vertex-disjoint induced subgraphs of G each on k vertices, both inducing the same number of edges and let $f(n)$ be the minimum value of $f(G)$ taken over all graphs on n vertices. Trivially, $f(n) \leq \lfloor n/2 \rfloor$; also, as shown by Ben-Eliezer and Krivelevich [4], equality holds (with high probability) for the Erdős–Rényi random graphs $G(n,p)$ for all $0 \leq p \leq 1$.

There was a large gap between the best known upper and lower bounds for $f(n)$. From below, one can easily show using the pigeonhole principle that $f(n) = \Omega(n^{1/3})$. As observed by Caro and Yuster, it is possible to improve this to $f(n) = \Omega(n^{1/2})$ using a well-known result of Lovász determining the chromatic number of Kneser graphs. By considering a carefully constructed disjoint union of cliques, each on an odd number of vertices, Caro and Yuster showed that $f(n) \leq n/2 - \Omega(\log \log n)$.

As expected, one can say more about $f(G)$ when G belongs to certain special graph classes. For example, Axenovich, Martin and Ueckerdt [3] showed that $f(G) \geq \lceil n/2 \rceil - 1$

Date: 5 January 2014.

2010 *Mathematics Subject Classification.* Primary 05C35; Secondary 05C07.

when G is a forest; this is clearly best-possible. Indeed, it is possible to get quite close to the trivial upper bound of $n/2$ when we restrict our attention to sparse graphs. In their paper, Caro and Yuster showed, for any fixed $\alpha > 0$, that if G is a graph on n vertices, then $f(G) \geq n/2 - o(n)$ provided G has at most $n^{2-\alpha}$ edges (or non-edges). Axenovich, Martin and Ueckerdt [3] later showed that the same holds for graphs with at most $o(n^2/(\log n)^2)$ edges.

Our main aim in this paper is to narrow considerably the gap between the best known upper and lower bounds for $f(n)$, and thereby answer a question of Caro and Yuster [6].

Theorem 1.1. *For every $\varepsilon > 0$, there exists a natural number $N = N(\varepsilon)$ such that for any graph G on $n > N$ vertices, $f(G) \geq n/2 - \varepsilon n$. Consequently,*

$$n/2 - o(n) \leq f(n) \leq n/2 - \Omega(\log \log n).$$

We remark that much research has been done on the family of induced subgraphs of a graph. For example, call a graph *k-universal* if it contains every graph of order k as an induced subgraph. Very crudely, if G is a k-universal graph with n vertices, then

$$\binom{n}{k} \geq \frac{2^{\binom{k}{2}}}{k!},$$

so $n \geq 2^{(k-1)/2}$. As remarked in [5], almost all graphs with $k^2 2^{k/2}$ vertices are k-universal, and the Paley graphs come close to providing examples which are almost as good. Hajnal conjectured that if a graph only has a 'small' number of distinct (non-isomorphic) induced subgraphs, then it contains a trivial (complete or empty) subgraph with linearly many vertices. This was proved, shortly after the conjecture was made, by Alon and Bollobás [1], and Erdős and Hajnal [8], the latter in a stronger form. In [1] only a few parameters, like order, size and maximal degree, were used to distinguish non-isomorphic graphs.

Erdős and Hajnal [9] then went much further: they realised that forbidding a single graph as an induced subgraph severely constrains the structure of a graph. More precisely, they made the major conjecture that for every graph H, there is a positive constant $\gamma(H)$ such that if a graph of order n does not contain H as an induced subgraph, then the graph contains a trivial subgraph with at least $n^{\gamma(H)}$ vertices. In spite of all the work on this conjecture, see [7, 11, 13] for instance, we are very far from the desired bound.

Let us finally mention another interesting line of research about finding disjoint isomorphic (not necessarily induced) subgraphs. Jacobson and Schönheim (see [10, 12]) independently raised the question of finding edge-disjoint isomorphic subgraphs. Improving on results of Erdős, Pyber and Pach [10], it has been shown by Lee, Loh and Sudakov [12] that every graph on m edges contains a pair of edge-disjoint isomorphic

subgraphs with at least $\Omega((m \log m)^{2/3})$ edges and that this is also best-possible up to a multiplicative constant.

The rest of this paper is organised as follows. We give an overview of our approach in Section 3, and then fill in the details and prove Theorem 1.1 in Section 4. There are many natural questions about induced subgraphs which are close to Theorem 1.1 in spirit; we conclude in Section 5 by mentioning some of these.

2. Preliminaries

Our objective in this section is to establish some notational conveniences and collect together, for easy reference, some simple propositions that we shall make use of when proving our main result.

2.1. **Notation.** It will be convenient to establish some notation for working with sets of pairs. A *pair* $\{x, y\}$ will always mean an unordered pair with $x \neq y$, and a *collection* of pairs $\mathcal{P}$ will always mean a set of disjoint pairs; for example, $\mathcal{P} = \{\{1, 2\}, \{3, 4\}\}$ is a collection of pairs, but $\mathcal{Q} = \{\{1, 2\}, \{2, 3\}\}$ is not. For a collection of pairs denoted by $\mathcal{P}$, we shall write P for the underlying ground set of elements, i.e., $P = \bigcup_{\{x,y\}\in\mathcal{P}} \{x, y\}$; in other words, we reserve the corresponding upper case letter for the ground set. We shall say that two collections of pairs $\mathcal{P}$ and $\mathcal{Q}$ are disjoint if $P \cap Q = \varnothing$; for example, the collections $\mathcal{P}_1 = \{\{1, 2\}, \{3, 4\}\}$ and $\mathcal{Q}_1 = \{\{5, 6\}, \{7, 8\}\}$ are disjoint, while the collections $\mathcal{P}_2 = \{\{1, 2\}, \{3, 4\}\}$ and $\mathcal{Q}_2 = \{\{1, 3\}, \{2, 4\}\}$ are not.

As usual, given a graph $G = (V, E)$, we write $\mathrm{d}(v)$ and $\Gamma(v)$ respectively for the degree and for the neighbourhood of a vertex v in G. For a subset $U \subset V$, we write $G[U]$ for the subgraph induced by U, $e(U)$ for the number of edges of $G[U]$, and $\mathrm{d}(U)$ for the sum of the degrees (in G) of the vertices of U. Given two disjoint subsets $A, B \subset V$, we write $e(A, B)$ for the number of edges with one endpoint each in A and B.

We shall also use the following less common terminology and notation. For any two vertices $x, y \in V$, we write $\delta(x, y)$ for the *degree difference* between x and y, namely the quantity $|\mathrm{d}(x) - \mathrm{d}(y)|$. We say that two vertices x and y *disagree on a vertex* $v \neq x, y$ if v is adjacent to exactly one of x and y; otherwise x and y *agree* on v. For any two vertices $x, y \in V$, the *difference neighbourhood* $\Gamma(x, y)$ of x and y is the set of vertices $v \neq x, y$ on which x and y disagree; we write $\Delta(x, y)$ for the size of the difference neighbourhood, so that $\delta(x, y) \leq \Delta(x, y)$. If two vertices x and y agree on every vertex $v \neq x, y$, we say that the pair $\{x, y\}$ is a *clone pair*. When the graph G in question is not clear from the context, we shall, for example, write $\delta(x, y, G)$ to denote the degree difference between x and y in G.

We say that a graph G is *splittable* if there is a partition $V = A \cup B$ of its vertex set into two sets A and B of equal size with $e(A) = e(B)$; in this case, we call (A, B) a *splitting* of G. Note that $e(A) = e(B)$ if and only if $\mathrm{d}(A) = \mathrm{d}(B)$, since $\mathrm{d}(A) = 2e(A) + e(A, B)$.

Our conventions for asymptotic notation are largely standard; however, we feel obliged to point out that we write $o_{k\to\infty}(1)$ to denote a function (of k) that goes to 0 as $k \to \infty$, and that when we write, say $\Omega_k(.)$, we mean that the constant suppressed by the asymptotic notation is allowed to depend on (but is completely determined by) the parameter k. For the sake of clarity of presentation, we systematically omit floors and ceilings whenever they are not crucial.

2.2. **Preliminary observations.** We shall make use of the following simple observation repeatedly when constructing a splitting.

Proposition 2.1. *Given positive real numbers $x_1, x_2, \ldots, x_t$ in the interval $[a, b]$ with $0 \le a \le b$, we may, for every $y \in [-ta, ta]$, choose signs $\zeta_i \in \{-1, +1\}$ such that $|y + \sum \zeta_i x_i| \le b$.* □

The following first moment bound will prove useful; it is easily checked that the bound is the best-possible.

Proposition 2.2. *Let X be a random variable such that $X \le N$ and $\mathbb{E}[X] \ge Np$. Then*

$$\mathbb{P}\left(X \ge \frac{\mathbb{E}[X]}{2}\right) \ge \frac{p}{2-p}.$$ □

We will also need the following two easy propositions.

Proposition 2.3. *Given $x_1, x_2, \ldots, x_t$ in the interval $[0, a]$, a positive real b and a natural number N, we may find $\lfloor t/N \rfloor - \lceil a/b \rceil$ disjoint subsets of $\{x_1, x_2, \ldots, x_t\}$, each of size N, such that $|x_i - x_j| \le b$ for any x_i and x_j belonging to the same subset.*

Proof. Suppose that $x_1 \le x_2 \le \cdots \le x_t$. Let $i_0 = 1$ and define i_j to be the smallest index such that $x_{i_j} > x_{i_{j-1}} + b$ and consider the sets $S_j = \{x_{i_j}, x_{i_j+1}, \ldots, x_{i_{j+1}-1}\}$. Since $x_1 \ge 0$ and $x_t \le a$, there are at most $\lceil a/b \rceil$ such sets. Now, by discarding at most N numbers from each S_j if necessary, we can assume that N divides $|S_j|$ for each j. We now partition each S_j into subsets of size N. Clearly, $|x_i - x_j| \le b$ for any x_i and x_j belonging to the same subset. The number of elements we have discarded is at most $N\lceil a/b \rceil$. So the number of subsets of size N we are left with is at least $\lfloor t/N \rfloor - \lceil a/b \rceil$. □

Remark. We shall often apply Proposition 2.3 to the degrees of a subset of vertices of a graph; we consequently obtain disjoint groups of vertices such that the degree difference of any two vertices in the same group is suitably bounded.

Proposition 2.4. *Let x, y and z be three vertices and U some subset of vertices of a graph G. Then some two of the vertices x, y and z disagree on at most two thirds of the vertices of U.*

Proof. Any vertex $v \in U$ belongs to at most two of the three difference neighbourhoods $\Gamma(x,y)$, $\Gamma(y,z)$ and $\Gamma(z,x)$. The claim follows by averaging. □

2.3. **Binomial random variables.** We will need some easily proven statements about binomial random variables. We collect these here. As usual, for a random variable with distribution $\mathrm{Bin}(N,p)$, we write $\mu(=Np)$ for its mean and $\sigma^2(=Np(1-p))$ for its variance.

The first proposition we shall require is an easy consequence of the fact that $e^{-2x} \leq 1-x \leq e^{-x}$ for all $0 \leq x \leq 1/2$.

Proposition 2.5. *Let X be a random variable with distribution $\mathrm{Bin}(N,p)$, with $p \leq 1/2$. Then for every $1 \leq k \leq N$,*

$$\exp(-2\mu)(\mu/k)^k \leq \mathbb{P}(X=k) \leq \exp(-\mu)(2e\mu/k)^k.$$

Also, $\exp(-2\mu) \leq \mathbb{P}(X=0) \leq \exp(-\mu)$. □

We shall make use of the following standard concentration result which first appeared in a paper of Bernstein and was later rediscovered by Chernoff and Hoeffding; see [2] for example.

Proposition 2.6. *Let X be a random variable with distribution $\mathrm{Bin}(N,p)$. Then*

$$\mathbb{P}(|X-Np| > t) \leq 2\exp\left(\frac{-2t^2}{N}\right).$$ □

Proposition 2.7. *Let X be a random variable with distribution $\mathrm{Bin}(N,p)$. Then*

$$\mathbb{P}(X \text{ is even}) = \frac{1}{2}(1+(1-2p)^N).$$ □

Proposition 2.8. *Let X_1 and X_2 be two independent random variables both with distribution $\mathrm{Bin}(N,p)$. Then*

$$\mathbb{P}(X_1 = X_2) = o_{\sigma\to\infty}(1).$$

In particular, when $p \leq 1/2$, $\mathbb{P}(X_1 = X_2) = o_{\mu\to\infty}(1)$. □

Proposition 2.9. *Let X_1 and X_2 be two independent random variables with distributions $\mathrm{Bin}(N_1,p)$ and $\mathrm{Bin}(N_2,p)$ respectively, with $p \geq 1/2$. Then*

$$\mathbb{P}(|X_1 - X_2| < |N_1 - N_2|^{1/3}) = o_{|N_1-N_2|\to\infty}(1).$$ □

Proposition 2.10. *Let X_1 and X_2 be two independent random variables with distributions* $\mathrm{Bin}(N_1, p)$ *and* $\mathrm{Bin}(N_2, p)$ *respectively, with $p \geq 1/2$. Suppose $N_1 \leq N$, $N_2 \leq N$ and $|N_1 - N_2| \leq cN^{1/2}$ for some absolute constant c. Then*

$$\mathbb{P}(|X_1 - X_2| > N^{2/3}) = O\left(\exp\left(\frac{-N^{1/3}}{5}\right)\right). \qquad \square$$

3. Overview of our strategy

To prove Theorem 1.1, we need to show that if $\varepsilon > 0$ and n is sufficiently large, then any graph G on n vertices contains two disjoint subsets of vertices of the same size, each of cardinality at least $(1/2 - \varepsilon)n$, which induce the same number of edges. Equivalently, we need to show that it is possible to transform G into a splittable graph by deleting at most $2\varepsilon n$ vertices from G. Recall that a graph is splittable if and only if there is a partition of its vertex set into two sets of equal size such that the sums of the degrees of the vertices in the two sets are equal.

We shall show that there is a probability $0 < p \leq \varepsilon$ (depending on G) such that if we delete vertices from G with probability p, then the resulting graph H is splittable with positive probability.

To show that this random subgraph H is splittable, we shall exhibit a large collection of 'gadgets' in H. Given $0 \leq a \leq b$, by an $[a, b]$-*gadget*, we mean a pair of vertices $\{x, y\}$ such that $a \leq \delta(x, y) \leq b$; a gadget, in other words, is just a pair of vertices whose degree difference we can control.

Once we have found sufficiently many suitable gadgets in H, we construct a splitting of H as follows: we use Proposition 2.1 to decide, one-by-one for each gadget, which way round to assign the vertices of the gadget to the sides of the splitting. The following lemma makes this idea precise.

Lemma 3.1. *Let H be a graph on an even number of vertices and suppose that we can partition $V(H)$ into disjoint collections of pairs $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_k$ such that the pairs in $\mathcal{P}_i$ are $[a_i, b_i]$-gadgets, where $0 \leq a_1 \leq b_1$ and $0 < a_i \leq b_i$ for $2 \leq i \leq k$. If $b_{i-1} \leq a_i|\mathcal{P}_i|$ for each $2 \leq i \leq k$, then $V(H)$ can be partitioned into two sets A, B of the same size such that $|\,\mathrm{d}(A) - \mathrm{d}(B)| \leq b_k$. In particular, if $b_k = 1$, then H is splittable.*

Proof. We show by induction on i that it is possible to partition the vertices of the gadgets in $\mathcal{P}_1, \ldots, \mathcal{P}_i$ into two sets A_i and B_i of equal size such that $|\,\mathrm{d}(A_i) - \mathrm{d}(B_i)| \leq b_i$. The lemma follows by taking $A = A_k$ and $B = B_k$.

We set $b_0 = 0$ and $A_0 = B_0 = \varnothing$, so the claim is trivially true when $i = 0$. So suppose that $i \geq 1$ and that we have constructed A_{i-1} and B_{i-1}. Denote the $[a_i, b_i]$-gadgets in $\mathcal{P}_i$ by (x_j, y_j), where $\mathrm{d}(x_j) \geq \mathrm{d}(y_j)$ for $1 \leq j \leq |\mathcal{P}_i|$. Using the fact that $b_{i-1} \leq a_i|\mathcal{P}_i|$, it follows from Proposition 2.1 that there is a choice of signs $\zeta_j \in \{-1, +1\}$ for $1 \leq j \leq |\mathcal{P}_i|$

such that

$$\left|(\mathrm{d}(A_{i-1}) - \mathrm{d}(B_{i-1})) + \sum_j \zeta_j \delta(x_j, y_j)\right| \leq b_i.$$

Given ζ_j as above, we construct A_i and B_i from A_{i-1} and B_{i-1} as follows: for each $1 \leq j \leq |\mathcal{P}_i|$, we add x_j to A_{i-1} and y_j to B_{i-1} if $\zeta_j = 1$, and y_j to A_{i-1} and x_j to B_{i-1} if $\zeta_j = -1$. The claim follows.

If $b_k = 1$, notice that we have a partition of $V(H)$ into two sets A and B of equal size such that $|\mathrm{d}(A) - \mathrm{d}(B)| \leq 1$. As $\mathrm{d}(A) + \mathrm{d}(B)$ is the sum of all the vertex degrees, we conclude that $\mathrm{d}(A) = \mathrm{d}(B)$ since $\mathrm{d}(A) - \mathrm{d}(B)$ must be even. □

Lemma 3.1 tells us that a graph is splittable if we can find the right gadgets in the graph. The majority of the work in proving Theorem 1.1 is in showing that it is possible to find a good collection of gadgets.

4. Proof of the main result

We now try and make the intuition presented in Section 3 precise. We shall show that if $\varepsilon > 0$ and n is sufficiently large, it is possible to transform any graph G on n vertices into a splittable graph by deleting at most $2\varepsilon n$ vertices from G. Before we begin, we remark that the various constants suppressed by the asymptotic notation throughout the proof are allowed to depend on ε. We shall use $c_1, c_2, \ldots$ to represent small constants depending on ε and $C_1, C_2, \ldots$ for large constants depending on ε. All our estimates will hold when n is sufficiently large.

Proof of Theorem 1.1. Let $\varepsilon > 0$ be fixed. By deleting an arbitrary vertex of G if necessary, assume that $n = |V(G)|$ is even. Let $\beta = \beta(\varepsilon)$ be a small constant whose value we shall fix at the end of the argument in Case 1.

Call a pair of vertices $\{x, y\}$ a 'large' pair if $\delta(x, y) \in [n^{1/3}, \beta n]$. Let $c_1 = \varepsilon/2$. We distinguish two cases depending on how many disjoint large pairs we can find in G. We first deal with the case when G contains many disjoint large pairs.

Case 1: G contains $c_1 n$ disjoint large pairs of vertices. In this case, we shall show that G has an induced subgraph H of even order on at least $(1 - 2\varepsilon)n$ vertices that contains

(1) a collection $\mathcal{S}_H$ of $[1, 1]$-gadgets of size $\Omega(n/\log n)$,
(2) a collection $\mathcal{M}_H$ of $[1, n^{2/3}]$-gadgets of size at least $2\beta n$, and
(3) a collection $\mathcal{L}_H$ of $[n^{1/9}, 2\beta n]$-gadgets of size $\Omega(n)$

such that the collections $\mathcal{S}_H$, $\mathcal{M}_H$, and $\mathcal{L}_H$ are disjoint. It is straightforward to check that such a graph H is splittable using Lemma 3.1. Indeed, pair up the vertices $V(H) \setminus (L_H \cup M_H \cup S_H)$ arbitrarily; any such pair is a $[0, n]$-gadget, so we have a partition of $V(H)$ into disjoint collections of $[0, n]$-gadgets, $[n^{1/9}, 2\beta n]$-gadgets, $[1, n^{2/3}]$-gadgets

and $[1,1]$-gadgets. The sizes of these collections satisfy the conditions of Lemma 3.1 if n is sufficiently large and it follows that H is splittable.

We shall now show that G does indeed contain such an induced subgraph H. We shall construct H by deleting vertices from G at random.

To avoid notational clutter, in the rest of the argument in Case 1, we shall write *large-gadget* for an $[n^{1/9}, 2\beta n]$-gadget, *medium-gadget* for a $[1, n^{2/3}]$-gadget and *one-gadget* for a $[1,1]$-gadget.

Let $\mathcal{L}$ be a collection of $c_1 n$ large pairs of vertices of G. The pairs in $\mathcal{L}$ will be the candidates for the large-gadgets we hope to find in H. Our next task is to find a large collection $\mathcal{M}$ of 'medium' pairs and a reasonably large collection $\mathcal{S}$ of 'small' pairs; the collections $\mathcal{M}$ and $\mathcal{S}$ will provide the candidate pairs for the medium-gadgets and one-gadgets that we would like to find in H.

Now, $|V \setminus L| = (1-2c_1)n$; recall that in our notation, L denotes the underlying ground set of $\mathcal{L}$. If we find more than $(1/2-\varepsilon)n$ disjoint clone pairs $\{x,y\}$ in $G[V \setminus L]$, we are done. Indeed, we can delete all the other $(\leq 2\varepsilon n)$ vertices not in any of these clone pairs to get a splittable graph: we split this graph by assigning different vertices of each clone pair to different halves of the partition. So we may assume that we can find a set $V' \subset V \setminus L$ of vertices of G such that any two vertices of V' disagree on some vertex of $V \setminus L$ and $|V'| \geq (2\varepsilon - 2c_1)n \geq \varepsilon n$.

Let $C_1 = 4/\varepsilon$ and let $c_2 = \varepsilon/12$. We now apply Proposition 2.3 to the degrees of the vertices of V'; by our choice of C_1 and c_2, we see that we can find $c_2 n$ disjoint groups of three vertices from V' such that $\delta(x,y) \leq C_1$ for any two vertices x and y in the same group. By Proposition 2.4, from each of these triples, we may choose a pair of vertices $\{x,y\}$ such that $\Delta(x,y) \leq 2n/3$. Write $\mathcal{P}$ for this collection of $c_2 n$ pairs.

For $0 \leq i \leq \log n - 1$, let $\mathcal{P}_i$ be the collection of those pairs $\{x,y\}$ in $\mathcal{P}$ such that $\Delta(x,y) \in [2^i, 2^{i+1})$. There are two possibilities that we need to consider. It might be that no collection $\mathcal{P}_i$ contains too many pairs; we deal with this case next. The case where one of these collections contains many pairs is easier; we deal with this scenario later with a modification of the argument that follows.

Let $C_2 \geq 4$ be a (large) constant depending on ε; we shall fix the value of C_2 later in the proof at the end of Case 1A. Also, let $c_3 = c_2/3C_2 \leq c_2/12$.

Case 1A: None of the collections $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_{\log n - 1}$ contains $c_3 n$ pairs. It is clear that at least one of the collections $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_{\log n - 1}$ contains at least $c_2 n/\log n$ pairs. Let k be the smallest index such that $|\mathcal{P}_k| \geq c_3 n/\log n$ and let us define our collection of small pairs $\mathcal{S}$ by setting $\mathcal{S} = \mathcal{P}_k$. We now define our collection of medium pairs $\mathcal{M}$ by setting

$$\mathcal{M} = \mathcal{P}_{k+C_2} \cup \cdots \cup \mathcal{P}_{\log n - 1}.$$

Since k is minimal and $c_3 \leq c_2/12$, we see that $|\mathcal{M}| \geq c_2 n/2$.

We shall now restrict our attention to the collections $\mathcal{S}$, $\mathcal{M}$ and $\mathcal{L}$; note that they are disjoint. We shall make use of the following facts about these collections.

(1) $\mathcal{S}$ contains $c_3 n/\log n$ pairs of vertices $\{x, y\}$ with $\delta(x, y) \leq C_1$, $\Delta(x, y) \in [2^k, 2^{k+1})$, and $\Delta(x, y) \leq 2n/3$.
(2) $\mathcal{M}$ contains $c_2 n/2$ pairs of vertices $\{x, y\}$ such that $\delta(x, y) \leq C_1$, and $\Delta(x, y) \geq 2^{k+C_2}$.
(3) $\mathcal{L}$ contains $c_1 n$ pairs of vertices $\{x, y\}$ with $\delta(x, y) \in [n^{1/3}, \beta n]$.
(4) For any pair of vertices $\{x, y\}$ in $\mathcal{S}$ or $\mathcal{M}$, there exists at least once vertex in $V \setminus L$ on which x and y disagree.

We are now in a position to describe how we intend to construct a splittable graph from G. We shall delete vertices from G independently with a fixed probability. We shall show that with positive probability, many of the small pairs from $\mathcal{S}$ form one-gadgets in the resulting graph, many of the medium pairs from $\mathcal{M}$ form medium-gadgets, and many of the large pairs from $\mathcal{L}$ form large-gadgets in the resulting graph.

Fix $p = \min\{\varepsilon, 2^{-k}\}$. We now delete vertices from G independently with probability p. Let H be the resulting graph. We shall show that with probability $\Omega(1)$, the graph H is splittable and contains at least $(1 - 2\varepsilon)n$ vertices; this clearly implies the result we are trying to prove.

Note that for a graph to be splittable, it must necessarily contain an even number of vertices. With this in mind, let $\mathcal{E}$ be the event that an even number of vertices have been deleted, in other words, $\mathcal{E}$ is the event that $|V(H)|$ is even. By Proposition 2.7, we see that $\mathbb{P}(\mathcal{E}) \geq 1/2$. We now analyse what happens to the degree differences of the pairs in $\mathcal{S}$, $\mathcal{M}$ and $\mathcal{L}$ in the graph H.

One-gadgets. We first show that many of the pairs in $\mathcal{S}$ form one-gadgets in H.

Lemma 4.1. *For any pair $\{x, y\} \in \mathcal{S}$,*

$$\mathbb{P}(\{x, y\} \text{ is a one-gadget in } H \mid \mathcal{E}) \geq f(\varepsilon, C_1) > 0.$$

The crucial fact about Lemma 4.1 is that the lower bound on the probability is independent of C_2.

Proof of Lemma 4.1. Let $A = \Gamma(x) \setminus (\Gamma(y) \cup \{y\})$ and $B = \Gamma(y) \setminus (\Gamma(x) \cup \{x\})$. Thus, $\delta(x, y) = ||A| - |B||$ and $\Delta(x, y) = |A| + |B|$. Note that since x and y disagree on at least one vertex of $V \setminus L$, it cannot be the case that both A and B are empty. Suppose without loss of generality that $|A| \geq |B|$ and that in particular, $A \neq \varnothing$.

Let E_1 be the event that both x and y are not deleted, E_2 the event that no vertices are deleted from B, E_3 the event that exactly $|\delta(x, y) - 1|$ vertices are deleted from A, and E_4 the event that the number of vertices deleted from $V \setminus (A \cup B \cup \{x, y\})$ has the same parity as $|\delta(x, y) - 1|$. It is obvious that the family $\{E_1, E_2, E_3, E_4\}$ is independent

since these events correspond to disjoint sets of vertices, and it is easy to check that

$$\mathbb{P}(\{x,y\} \text{ is a 1-gadget in } H \,|\, \mathcal{E}) \geq \prod_{i=1}^{4} \mathbb{P}(E_i).$$

To complete our proof of the claim, we shall bound the factors on the right one by one. Clearly, $\mathbb{P}(E_1) \geq (1-\varepsilon)^2$.

We trivially have $|A|, |B| \leq 2^{k+1}$. Furthermore $|A|, |B| \geq 2^{k-1} - C_1/2$, since $0 \leq \delta(x,y) \leq C_1$. Also, we know that $\varepsilon 2^{-k} \leq p \leq 2^{-k}$. To bound $\mathbb{P}(E_2)$, first note that $p|B| \leq 2$. Now, $\mathbb{P}(E_2) = \mathbb{P}(\mathrm{Bin}(|B|,p) = 0)$, so by Proposition 2.5, $\mathbb{P}(E_2) \geq \exp(-4)$.

We now bound $\mathbb{P}(E_3)$. Clearly, $p|A| \leq 2$. If $2^k \geq 2C_1$, then $|A| \geq 2^{k-2}$, so $p|A| \geq \varepsilon/4$. If $2^k \leq 2C_1$, then $p \geq \varepsilon 2^{-k} \geq \varepsilon/2C_1$, so $p|A| \geq \varepsilon/2C_1$ since $|A| \geq 1$. Consequently,

$$\min\{\varepsilon/4, \varepsilon/2C_1\} \leq p|A| \leq 2.$$

Now, $\mathbb{P}(E_3) = \mathbb{P}(\mathrm{Bin}(|A|,p) = |\delta(x,y) - 1|)$. Using the above estimates for $p|A|$ and the fact that $0 \leq \delta(x,y) \leq C_1$ in Proposition 2.5, we see that $\mathbb{P}(E_3) = \Omega_{\varepsilon,C_1}(1)$.

Finally, since $\Delta(x,y) \leq 2n/3$, it follows that $|V \setminus (A \cup B)| \geq n/3$, so by Proposition 2.7, $\mathbb{P}(E_4) \geq 1/6$ for all sufficiently large n. The claim follows. □

From Lemma 4.1 and Proposition 2.2 we see that, conditional on $\mathcal{E}$, the number of one-gadgets in H from $\mathcal{S}$ is $\Omega(n/\log n)$ with probability at least $f(\varepsilon, C_1)/2$; furthermore, and crucially, we note that this lower bound on the probability is independent of C_2.

Medium-gadgets. We next shift our attention to the pairs in $\mathcal{M}$.

Lemma 4.2. *For any pair $\{x,y\} \in \mathcal{M}$,*

$$\mathbb{P}(1 \leq \delta(x,y,H) \leq n^{2/3} \,|\, x,y \in V(H)) = 1 - o_{C_2 \to \infty}(1) - o(1).$$

Proof. Let $N_1 = |\Gamma(x) \setminus (\Gamma(y) \cup \{y\})|$ and let $N_2 = |\Gamma(y) \setminus (\Gamma(x) \cup \{x\})|$ and suppose without loss of generality that $N_1 \geq N_2$. Note that $\delta(x,y) = |N_1 - N_2| \leq C_1$. Let X_1 and X_2 be independent random variables with distributions $\mathrm{Bin}(N_1, 1-p)$ and $\mathrm{Bin}(N_2, 1-p)$ respectively. Observe that $\delta(x,y,H)$ has the same distribution as $|X_1 - X_2|$.

We condition on $x, y \in V(H)$. Let E_1 be the event that $\delta(x,y,H) = 0$. Clearly, $\mathbb{P}(E_1) = \mathbb{P}(X_1 = X_2)$. Let E_2 denote the event that $\delta(x,y,H) \geq n^{2/3}$. It is enough to show that $\mathbb{P}(E_1 \cup E_2) = o_{C_2 \to \infty}(1) + o(1)$.

For any fixed values of p and N_2, it is not hard to check that $\mathbb{P}(X_1 = X_2)$ attains its maximum when $N_1 = N_2$; clearly, $\mathbb{P}(X_1 = X_2) = \sum_{i=0}^{N_2} \mathbb{P}(X_1 = i)\mathbb{P}(X_2 = i)$ and the required conclusion follows from Cauchy–Schwarz inequality. Thus $\mathbb{P}(E_1)$ is bounded above by the probability of two independent random variables with the distribution $\mathrm{Bin}(N_2, 1-p)$, or equivalently $\mathrm{Bin}(N_2, p)$, being equal. Now, $N_2 \geq 2^{k+C_2-1} - C_1/2$ and $p \geq \varepsilon 2^{-k}$. So, $pN_2 \geq \varepsilon 2^{C_2-1} - 2^{-k+1}$ which, since $k \geq 0$, means that $pN_2 \geq \varepsilon 2^{C_2-1} - 2$.

As ε is fixed, we note that pN_2 can be made arbitrarily large by choosing C_2 large enough. Since $p \leq 1/2$, by Proposition 2.8, we see that $\mathbb{P}(E_1) = o_{C_2 \to \infty}(1)$.

Clearly, $\mathbb{P}(E_2) = \mathbb{P}(|X_1 - X_2| \geq n^{2/3})$. Applying Proposition 2.10 to X_1 and X_2, we conclude that $\mathbb{P}(E_2) = O(\exp(-n^{1/3}/5))$. □

Let $\mathcal{M}'$ be the collection of those pairs $\{x, y\} \in \mathcal{M}$ such that both x and y survive in H. Since the family of events $\{x, y \in V(H)\}$ is a family of mutually independent events for different pairs $\{x, y\} \in \mathcal{M}$ and since $\mathbb{P}(x, y \in V(H)) \geq (1-\varepsilon)^2$, it follows from Proposition 2.6 that $\mathbb{P}(|\mathcal{M}'| < (1-\varepsilon)^2|\mathcal{M}|/2) = \exp(-\Omega(n))$.

Consequently, from Lemma 4.2, it follows that for any pair $\{x, y\} \in \mathcal{M}$,

$$\mathbb{P}\left(1 \leq \delta(x, y, H) \leq n^{2/3} \,\middle|\, \{x, y\} \in \mathcal{M}',\, |\mathcal{M}'| > \frac{(1-\varepsilon)^2|\mathcal{M}|}{2}\right) = 1 - o_{C_2 \to \infty}(1) - o(1).$$

Thus by Proposition 2.2, conditional on $|\mathcal{M}'| > (1-\varepsilon)^2|\mathcal{M}|/2$, the number of medium-gadgets from $\mathcal{M}'$ in H is at least $|\mathcal{M}'|/3$ with probability $1 - o_{C_2 \to \infty}(1) - o(1)$. Thus, the number of medium-gadgets in H is at least $(1-\varepsilon)^2|\mathcal{M}|/6$ with probability $(1 - \exp(-\Omega(n)))(1 - o_{C_2 \to \infty}(1) - o(1))$, which is still $1 - o_{C_2 \to \infty}(1) - o(1)$.

Thus, conditional on the event $\mathcal{E}$, the number of medium-gadgets in H from $\mathcal{M}$ is $\Omega(n)$ with probability $1 - o_{C_2 \to \infty}(1) - o(1)$.

Large-gadgets. We finally consider the pairs of vertices in $\mathcal{L}$. Recall that every pair $\{x, y\} \in \mathcal{L}$ is such that $\delta(x, y) \in [n^{1/3}, \beta n]$, where β is a (small) constant whose value we have yet to fix. (Indeed, the value of β has so far played no role in our calculations.)

Lemma 4.3. *For any pair $\{x, y\} \in \mathcal{L}$,*

$$\mathbb{P}(n^{1/9} \leq \delta(x, y, H) \leq 2\beta n \,|\, x, y \in V(H)) = 1 - o(1).$$

Proof. We condition on $x, y \in V(H)$. Let E_1 be the event that $\delta(x, y, H) < n^{1/9}$. Since $\delta(x, y) \geq n^{1/3}$, it follows immediately from Proposition 2.9 that $\mathbb{P}(E_1) = o(1)$.

Let E_2 be the event that $\delta(x, y, H) > 2\beta n$. Let $A = \Gamma(x) \setminus (\Gamma(y) \cup \{y\})$ and $B = \Gamma(y) \setminus (\Gamma(x) \cup \{x\})$, and let X_1 and X_2 be random variables that denote the the number of vertices from A and B respectively which survive in H. Clearly, the distributions of X_1 and X_2 are $\mathrm{Bin}(|A|, 1-p)$ and $\mathrm{Bin}(|B|, 1-p)$ respectively.

If E_2 were to occur, i.e., it were the case that $|X_1 - X_2| > 2\beta n$, then this would imply that either $|X_1 - (1-p)|A|| \geq \beta n/2$ or $|X_2 - (1-p)|B|| \geq \beta n/2$, since $(1-p)||A| - |B|| \leq \delta(x, y) \leq \beta n$. It follows that $\mathbb{P}(E_2) = o(1)$ since the probability of either of the above two possibilities is $\exp(-\Omega(n))$ by Proposition 2.6. □

Arguing as in the case of medium-gadgets, we see from Lemma 4.3 that conditional on the event $\mathcal{E}$, the number of large-gadgets in H from $\mathcal{L}$ is $\Omega(n)$ with probability $1 - o(1)$.

Constructing a splitting. We now have a reasonably clear picture of what the degree differences in H of the pairs of vertices in $\mathcal{S}$, $\mathcal{M}$ and $\mathcal{L}$ look like. In summary, conditional on $\mathcal{E}$, we have demonstrated that in H, we can find

(1) a collection $\mathcal{S}_H$ of $\Omega(n/\log n)$ one-gadgets with probability $f(\varepsilon, C_1)/2$,
(2) a collection $\mathcal{M}_H$ of $\Omega(n)$ medium-gadgets with probability $1 - o_{C_2 \to \infty}(1) - o(1)$, and
(3) a collection $\mathcal{L}_H$ of $\Omega(n)$ large-gadgets with probability $1 - o(1)$

such that the collections $\mathcal{S}_H, \mathcal{M}_H$ and $\mathcal{L}_H$ are disjoint.

Thus by choosing C_2 to be a sufficiently large constant depending on ε, by the union bound, we find all of the above with probability $\Omega(1)$ conditional on $\mathcal{E}$, provided n is sufficiently large. Also, the expected number of vertices deleted is at most εn, so by Proposition 2.6, the probability that we have deleted more than $2\varepsilon n$ vertices is $\exp(-\Omega(n))$.

Consequently, we see that H, with probability $\Omega(1)$, has the aforementioned collections of gadgets, and furthermore, also has an even number of vertices and at least $(1 - 2\varepsilon)n$ vertices. We are done if we can guarantee that $2\beta n \leq |\mathcal{M}_H|$; this is possible if we choose $\beta = \beta(\varepsilon)$ to be a suitably small constant because $|\mathcal{M}_H| = \Omega(n)$.

We now consider the case where one of the sets $\mathcal{P}_i$ contains many pairs.

Case 1B: One of the sets $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_{\log n - 1}$ contains $c_3 n$ pairs. This case is easier to deal with than the previous one. We shall argue exactly as before; however we shall have no need of medium-gadgets and it will suffice to consider one-gadgets and large-gadgets alone.

Let k be any index such that $|\mathcal{P}_k| \geq c_3 n$ (while we chose k to be minimal previously, any index k such that $|\mathcal{P}_k| \geq c_3 n$ will do in this case). As before, we set $p = \min\{\varepsilon, 2^{-k}\}$ and $\mathcal{S} = \mathcal{P}_k$. We now delete vertices from G independently with probability p. Let H be the resulting graph; as before, we condition on deleting an even number of vertices. We claim that H is splittable with probability $\Omega(1)$.

It is not hard to check that Lemma 4.1 and Lemma 4.3 hold in this case as well. We conclude that we can delete an even number of vertices from G to obtain a graph H with $|V(H)| \geq (1 - 2\varepsilon)n$ in such a way that in H, we can find

(1) a collection $\mathcal{S}_H$ of $\Omega(n)$ one-gadgets, and
(2) a collection $\mathcal{L}_H$ of $\Omega(n)$ large-gadgets

such that $\mathcal{S}_H$ and $\mathcal{L}_H$ are disjoint. As before, it follows from Lemma 3.1 that H is splittable when n is sufficiently large provided $2\beta n \leq |\mathcal{S}_H|$; this is possible if we choose $\beta = \beta(\varepsilon)$ to be a suitably small constant because $|\mathcal{S}_H| = \Omega(n)$.

Thus, for all sufficiently small β (so as to satisfy the conditions from both Case 1A and 1B), we see that we are done if G contains many disjoint large pairs. Note that we have now fixed the value of β. We now deal with the case G does not contain many disjoint large pairs.

Case 2: G does not contain $c_1 n$ disjoint large pairs. In this case, we shall show that G has an induced subgraph H of even order on at least $(1-2\varepsilon)n$ vertices such that $V(H)$ may be partitioned into

(1) a collection $\mathcal{S}_H$ of $[1,1]$-gadgets of size $\Omega(n/\log n)$, and
(2) a collection $\mathcal{M}_H$ of $[0, n^{2/3}]$-gadgets.

In the rest of the argument in Case 2, we shall, as before, call $[1,1]$-gadgets *one-gadgets* and we call $[0, n^{2/3}]$-gadgets (as opposed to $[1, n^{2/3}]$-gadgets as we did earlier) *medium-gadgets*.

It is easily seen from Lemma 3.1 that H is splittable if n is sufficiently large. We construct our splitting by starting with the pairs in $\mathcal{M}_H$ - we can use these pairs to construct a partition such that sums of the degrees of the vertices of the two halves of the partition differ by at most $n^{2/3}$. We then use the the pairs in $\mathcal{S}_H$ to reduce the difference to at most one; we are done by parity considerations.

We now show how to find such a subgraph H. We start by describing how to find pairs of vertices which will be the candidates for the medium-gadgets we hope to find in H.

Let $\mathcal{L}$ be a maximal collection of large pairs in G. Note that since $\mathcal{L}$ is maximal, we have either $\delta(x,y) < n^{1/3}$ or $\delta(x,y) > \beta n$ for any two vertices $x, y \in V \setminus L$. As $\beta n > 2n^{1/3}$ for all sufficiently large n, there is a partition $V \setminus L = K_1 \cup K_2 \cup \cdots \cup K_m$ into 'clumps' K_i with $m \leq 1/\beta$ in such a way that $\delta(x,y) < n^{1/3}$ for any $x, y \in K_i$ and $\delta(x,y) > \beta n$ if $x \in K_i$ and $y \in K_j$ with $i \neq j$.

We ignore the way in which vertices are originally paired in $\mathcal{L}$ and focus on the ground set L. By Proposition 2.3, we can find from L, at least $|L|/2 - n^{1/2}$ disjoint pairs $\{x, y\}$ such that $\delta(x,y) \leq n^{1/2}$; call this collection of pairs $\mathcal{Q}$.

Let F be the graph obtained from G as follows. Delete every vertex of $L \setminus Q$. Delete one vertex from every clump K which contains an odd number of vertices. Having done this, delete a clump K (i.e., delete all the vertices of K) if $|K| \leq n^{1/2}$.

Note that the vertex set of F consists of the surviving clumps, each of which has even size and cardinality at least $n^{1/2}$, and the (possibly empty) set of pairs $\mathcal{Q}$. Since we had at most $1/\beta$ clumps initially, we have deleted $O(n^{1/2})$ vertices in total from G to obtain

F. Hence, for any two vertices $x, y \in V(F)$, $|\delta(x, y, F) - \delta(x, y, G)| = O(n^{1/2})$. Hence, if either x and y both belong to the same (surviving) clump or if the pair $\{x, y\}$ is in $\mathcal{Q}$, then $\delta(x, y, F) = O(n^{1/2})$. Let us say that two vertices $x, y \in V(F)$ are *proximate* if either both x and y belong to the same clump in F or if $\{x, y\} \in \mathcal{Q}$; these proximate pairs of vertices will be our candidates for medium-gadgets in H.

We now show how to find pairs of vertices which will be the candidates for the one-gadgets we hope to find in H. We shall henceforth work with F as opposed to G. We shall write V for $V(F)$ and all degrees and degree differences, unless specified otherwise, will be with respect to F.

Since $|\mathcal{L}| \leq c_1 n = \varepsilon n/2$ and since we have only deleted $O(n^{1/2})$ vertices so far, note that $|V \setminus Q| \geq (1 - 3\varepsilon/2)n$ for n sufficiently large.

If we find at least $(1/2 - \varepsilon)n$ disjoint clone pairs $\{x, y\}$ in $F[V \setminus Q]$, we are done. So we may assume that we can find a set $V' \subset V \setminus Q$ of vertices of F with $|V'| \geq (2\varepsilon - 3\varepsilon/2)n = \varepsilon n/2$ such that any two vertices of V' disagree on some vertex in $V \setminus Q$.

We claim that if C_3 is sufficiently large (as a function of β), then we can find from any subset of C_3 vertices of V', two vertices x and y such that for each clump K, the number of vertices of K on which x and y disagree is at most $2|K|/3$. To see this, suppose that we have found C_3 vertices such that any two of them x and y disagree on more than two thirds of some clump $K_{x,y}$. Applying Ramsey's theorem (with $1/\beta$ colours) to the complete graph on these C_3 vertices where the edge between x and y is labelled by the clump $K_{x,y}$, we see that we can find a monochromatic triangle provided C_3 is large enough. But by Proposition 2.4, out of any three vertices, at least two disagree on at most two thirds of the vertices of K. We have a contradiction.

Choose C_3 as described above and set $C_4 = 4C_3/\varepsilon$ and $c_4 = \beta/2C_4$. By Proposition 2.3, we can find from V', at least n/C_4 disjoint groups of size C_3 such that that $\delta(x, y) \leq C_4$ for any two vertices x and y in the same group. From each of these n/C_4 groups of size C_3, choose a pair of vertices $\{x, y\}$ such that x and y disagree on at most two thirds of every clump. Choose a clump K^* such that at least a β fraction of these pairs $\{x, y\}$ are such that x and y disagree on at least one vertex in K^*; this is possible because any two vertices of V' disagree on $V(F) \setminus Q$ and consequently, on at least one clump and furthermore, there are at most $1/\beta$ clumps. Let $\mathcal{P}$ be this collection of pairs which all disagree on at least one vertex in K^*; clearly, $|\mathcal{P}| \geq \beta n/C_4 = 2c_4 n$.

We shall proceed as in Case 1 by pigeonholing the pairs in $\mathcal{P}$ into different boxes based on the size of their difference neighbourhoods, but with one important difference. Note that while *any two* vertices in the same clump have a small ($O(n^{1/2})$) degree difference, we can only guarantee that two vertices of Q have small ($O(n^{1/2})$) degree difference if the pair belongs to $\mathcal{Q}$. Consequently, when we later delete vertices at random, we shall either delete both vertices of a pair in $\mathcal{Q}$ or retain both; hence, we

shall treat a pair of vertices in $\mathcal{Q}$ as a single vertex when it comes to pigeonholing the pairs in $\mathcal{P}$. This is made precise below.

Let $F_{\mathcal{Q}}$ be the multigraph without loops obtained from F by contracting every pair $\{x, y\}$ in $\mathcal{Q}$ (we ignore the loops that might arise). Note that there are at most two parallel edges between any two vertices of $F_{\mathcal{Q}}$ unless both vertices correspond to contracted pairs from $\mathcal{Q}$, in which case there are at most four parallel edges between them. In $F_{\mathcal{Q}}$, we say that two vertices x and y disagree on a vertex $v \neq x, y$ if the number of edges between v and x is not equal to the number of edges between v and y. For $0 \leq i \leq \log n - 1$, let $\mathcal{P}_i$ be the collection of those pairs $\{x, y\}$ in $\mathcal{P}$ such that $\Delta(x, y, F_{\mathcal{Q}}) \in [2^i, 2^{i+1})$, where $\Delta(x, y, F_{\mathcal{Q}})$ is the number of vertices of $F_{\mathcal{Q}}$ on which x and y disagree.

As before, let k be any index such that $|\mathcal{P}_k| \geq 2c_4 n/\log n$; take $\mathcal{S} = \mathcal{P}_k$ and set $p = \min\{\varepsilon, 2^{-k}\}$.

In summary, $\mathcal{S}$ consists of pairs $\{x, y\}$ such that

(1) x and y disagree on at most two thirds of every clump,
(2) x and y disagree on at least one vertex of K^*,
(3) $\delta(x, y) \leq C_4$, and
(4) $\Delta(x, y, F_{\mathcal{Q}}) \in [2^k, 2^{k+1})$.

Furthermore, since $\delta(x, y) \leq C_4 = o(n^{1/2})$ for any $\{x, y\} \in \mathcal{S}$, both members of any pair in $\mathcal{S}$ must belong to the same clump.

Consider the partition $\mathcal{S} = \mathcal{S}_o \cup \mathcal{S}_e$ where $\mathcal{S}_o$ is the set of those pairs $\{x, y\} \in \mathcal{S}$ such that $\delta(x, y)$ is odd. Recall that $|\mathcal{S}| \geq 2c_4 n/\log n$, so one of $\mathcal{S}_o$ or $\mathcal{S}_e$ contains more than $c_4 n/\log n$ pairs. At this point, we need slightly different arguments depending on whether we have more pairs with odd degree difference or even degree difference in $\mathcal{S}$.

Case 2A: $\mathcal{S}$ contains many odd pairs. We first consider the case where $|\mathcal{S}_o| \geq c_4 n/\log n$. We shall delete vertices from F as follows. We pick vertices of $F_{\mathcal{Q}}$ independently with probability p. For every vertex of $F_{\mathcal{Q}}$ that we pick, we delete (as appropriate) either the corresponding vertex or both vertices of the corresponding pair of vertices from $\mathcal{Q}$ in $F_{\mathcal{Q}}$. Let H be the resulting graph. Our aim is to show that H is splittable with probability $\Omega(1)$.

Earlier, we conditioned on deleting an even number of vertices from G. In this case, we need a little more. Let $\mathcal{E}^*$ be the event that an even number of vertices were deleted from each clump. By Proposition 2.7, we see that $\mathbb{P}(\mathcal{E}^*) \geq (1/2)^{1/\beta}$. Note that a consequence of $\mathcal{E}^*$ is that $|V(H)|$ is even.

ONE-GADGETS. First, we shall show that many of the pairs in $\mathcal{S}_o$ become one-gadgets in H.

Lemma 4.4. *For any pair* $\{x, y\} \in \mathcal{S}_o$,

$$\mathbb{P}(\{x, y\} \text{ is a one-gadget in } H \mid \mathcal{E}^*) = \Omega(1).$$

Proof. In $F_{\mathcal{Q}}$, let A be the set of those vertices $v \neq x, y$ such that number of edges between v and x is more than the number of edges between v and y and let B be defined analogously by interchanging x and y. Let $A = A_1 \cup A_2$ where A_1 and A_2 are respectively those vertices v in A such that the number of edges between v and x is one, respectively two, more than the number of edges from v to y; define B_1 and B_2 analogously.

The proof follows that of Lemma 4.1. Clearly,

$$2^k \leq |A_1| + |A_2| + |B_1| + |B_2| < 2^{k+1}.$$

Furthermore, $\delta(x, y) = ||A_1| + 2|A_2| - |B_1| - 2|B_2||$, so

$$-C_4 \leq |A_1| + 2|A_2| - |B_1| - 2|B_2| \leq C_4.$$

Using the above two inequalities, it is not hard to check that

$$\max\{|A_1|, |A_2|\}, \max\{|B_1|, |B_2|\} \geq 2^{k-3} - C_4/4.$$

Since $\delta(x, y)$ is odd, suppose without loss of generality that $\mathrm{d}(x) > \mathrm{d}(y)$. Let E_1 be the event that both x and y are not picked to be deleted, E_2 the event that no vertices are picked from B, E_3 the event that $X_1 + 2X_2 = \delta(x, y) - 1$ where X_1 and X_2 are the number of vertices picked from A_1 and A_2 respectively, and E_4 the event that the number of vertices picked from $K \setminus (A \cup B \cup \{x, y\})$ has the same parity as the number of vertices picked from $K \cap (A \cup B \cup \{x, y\})$ for every clump K. The collection of events $\{E_1, E_2, E_3\}$ is clearly independent, and it is easy to check that

$$\mathbb{P}(\{x, y\} \text{ is a one-gadget} \mid \mathcal{E}^*) \geq \mathbb{P}(E_1)\mathbb{P}(E_2)\mathbb{P}(E_3)\mathbb{P}(E_4|E_1, E_2, E_3).$$

Clearly, $\mathbb{P}(E_1) \geq (1-\varepsilon)^2$. As in Lemma 4.1, note that $p|B| \leq 2$, so by Proposition 2.5, $\mathbb{P}(E_2) \geq \exp(-4)$.

We now bound $\mathbb{P}(E_3)$. First suppose that $2^{k-3} - C_4/4 > C_4$. Recall that $\delta(x, y)$ is odd. If $|A_2| \geq |A_1|$, we consider the event that $(\delta(x, y) - 1)/2$ vertices are picked from A_2 and no vertices are picked from A_1 in $F_{\mathcal{Q}}$; as in Lemma 4.1, we see that $p|A_2| = \Theta(1)$, so this event occurs with probability $\Omega(1)$. Hence, E_3 occurs with probability $\Omega(1)$. If $|A_1| > |A_2|$, we consider the event that $\delta(x, y) - 1$ vertices are picked from $|A_1|$ and no vertices are picked from A_2 and note that this event occurs with probability $\Omega(1)$; hence, E_3 occurs with probability $\Omega(1)$.

If, on the other hand, $2^{k-3} - C_4/4 \leq C_4$, then clearly $k = \Theta(1)$l hence, $p, |A_1|, |A_2|$ are all $\Theta(1)$. In this case, we consider the event that $t = \min\{(\delta(x, y) - 1)/2, |A_2|\}$ vertices are picked from A_2 and $\delta(x, y) - 1 - 2t$ vertices are picked from A_1. Now, $|A_1| + 2|A_2| \geq \delta(x, y)$ since we assumed that $\mathrm{d}(x) > \mathrm{d}(y)$, so $|A_1| \geq \delta(x, y) - 1 - 2t$.

Also, as noted above, $p, |A_1|, |A_2|$ are all $\Theta(1)$. So this event occurs with probability $\Omega(1)$. Hence, the event E_3 occurs with probability $\Omega(1)$.

Since x and y disagree on at most two thirds of every clump and since every clump has size at least $n^{1/2}$, it follows from Proposition 2.7 that $\mathbb{P}(E_4|E_1, E_2, E_3) \geq (1/6)^{1/\beta}$ for all sufficiently large n. □

Let $\mathcal{S}_H$ be the set of pairs from $\mathcal{S}_o$ that form one-gadgets in H. From Lemma 4.4 and Proposition 2.2, we see that conditional on $\mathcal{E}^*$, $|\mathcal{S}_H| \geq \mathbb{E}[|\mathcal{S}_H|]/2 = \Omega(n/\log n)$ with probability $\Omega(1)$.

Medium-gadgets. We now show that the degree difference of any pair of vertices which are proximate in F cannot become too large in H.

Lemma 4.5. *Conditional on $\mathcal{E}^*$ and $|\mathcal{S}_H| \geq \mathbb{E}[|\mathcal{S}_H|]/2$, the probability that there exist $x, y \in V(H)$ which are proximate in F and satisfy $\delta(x, y, H) > n^{2/3}$ is $o(1)$.*

Proof. Recall that for any two vertices x and y which are proximate in F, $\delta(x, y) = O(n^{1/2})$. For such a pair of vertices x and y, note by Proposition 2.10 that

$$\mathbb{P}(\delta(x, y, H) > n^{2/3} \,|\, x, y \in V(H)) = O(\exp(-n^{1/3}/5)).$$

Consequently, since we have conditioned on an event with probability $\Omega(1)$, the probability that there exist some vertices $x, y \in V(H)$ such that x and y are proximate and $\delta(x, y, H) > n^{2/3}$ is $O(n^2 \exp(-n^{1/3}/5)) = o(1)$. □

Constructing a splitting. We now describe how to construct a splitting of H. Let $\mathcal{Q}_H$ be the set of pairs from $\mathcal{Q}$ that survive in H. For a clump K in F, let K_H denote the set $(K \setminus S_H) \cap V(H)$. Clearly, $V(H)$ is the disjoint union of S_H, Q_H and the clumps K_H. Note that conditional on $\mathcal{E}^*$, the size of K_H is even for every clump K since both members of any pair in $\mathcal{S}_H$ must necessarily belong to the same clump. Since each K_H has even cardinality, we may group the vertices of each K_H into pairs. Pair up the vertices in each K_H arbitrarily; let $\mathcal{M}_H$ be the collection consisting of these pairs and the pairs in $\mathcal{Q}_H$. Clearly, every pair of vertices in $\mathcal{M}_H$ are proximate in F and by Lemma 4.5, the probability that there exists some pair $\{x, y\} \in \mathcal{M}_H$ with $\delta(x, y, H) > n^{2/3}$ is $o(1)$.

The expected number of vertices deleted from F is at most εn and the number of vertices deleted from G to obtain F is $O(n^{1/2})$. Hence, by Proposition 2.6, the probability that we have deleted more than $2\varepsilon n$ vertices from G is $\exp(-\Omega(n))$.

We conclude that there exists an induced subgraph H of G such that $|V(H)| \geq (1 - 2\varepsilon)n$, and with the further property that $V(H)$ may be partitioned into

(1) a collection $\mathcal{S}_H$ of one-gadgets of size $\Omega(n/\log n)$, and
(2) a collection $\mathcal{M}_H$ of medium-gadgets.

It follows from Lemma 3.1 that H is splittable and we are done.

Case 2B: $\mathcal{S}$ contains many even pairs. Now we consider the case where $|\mathcal{S}_e| \geq c_4 n/\log n$.

Note that since we intend to delete either both vertices of a pair in $\mathcal{Q}$ or neither, it might be the case that it is impossible to make the parity of the degree difference of a pair in $\mathcal{S}_e$ odd in H. Consequently, in this case, we will need to work with $[2,2]$-gadgets, or *two-gadgets* for short, in addition to one-gadgets. With the exception of this slight change of tack to account for parity considerations, the argument is quite similar to the one in the previous case, and we only sketch it.

Let c_5 be a (small) constant depending on ε; the value of c_5 will be chosen later, following the statement of Lemma 4.6.

Recall that every pair of vertices in $\mathcal{S}_e$ disagree on some vertex in the clump K^*. Suppose there exists a vertex $v \in K^*$ such that $c_5 n/\log n$ pairs from $\mathcal{S}_e$ all disagree on v. In this case, we may complete the proof as follows. Let $\mathcal{S}_v \subset \mathcal{S}_e$ be the collection of pairs in $\mathcal{S}_e$ that disagree on v. We shall delete vertices from F as follows. We first delete v and then delete one other vertex uniformly at random from K^*. Following this, we proceed as before by picking vertices of $F_{\mathcal{Q}}$ independently with probability p and then deleting the corresponding vertices or pairs of vertices from $\mathcal{Q}$ in F. Let H be the resulting graph. Note that when we delete v, the degree difference of every pair in $\mathcal{S}_v$ changes parity and becomes odd. When we then delete another vertex uniformly at random from K^*, the parity of the degree difference of a pair in $\mathcal{S}_v$ is unaltered with probability at least $1/3$ since every pair in $\mathcal{S}$ disagree on at most two thirds of any clump. Arguing as in Lemma 4.4, for any pair in $\mathcal{S}_v$, we see that the probability that this pair forms a one-gadget in H, conditional on deleting an even number of vertices from every clump, is $\Omega(1)$ (albeit with a smaller constant than in Case 2A). Since $|\mathcal{S}_v| \geq c_5 n/\log n$, we can conclude the proof exactly as in the case where $\mathcal{S}$ contains many odd pairs.

Hence, we may assume that for every vertex $v \in K^*$, the number of pairs in $\mathcal{S}_e$ that disagree on v is at most $c_5 n/\log n$. We delete vertices from F as before by picking vertices of $F_{\mathcal{Q}}$ independently with probability p and then deleting the corresponding vertices or pairs of vertices from $\mathcal{Q}$ in G. Let H be the resulting graph.

As before, let $\mathcal{E}^*$ be the event that an even number of vertices were deleted from each clump. The proof of Lemma 4.4, with minor modifications for the change in parity, yields a proof of the following lemma.

Lemma 4.6. *For any $\{x,y\} \in \mathcal{S}_e$, $\mathbb{P}(\{x,y\}$ is a two-gadget in $H \mid \mathcal{E}^*) = \Omega(1)$.* □

Let $\mathcal{S}_H$ be the collection of pairs from $\mathcal{S}_e$ that form two-gadgets in H. From Lemma 4.6 and Proposition 2.2, we see that there exists a small positive constant c_6 such that, conditional on $\mathcal{E}^*$, $|\mathcal{S}_H| \geq c_6 n/\log n$ with probability $\Omega(1)$. Let us now fix $c_5 = c_6/4$.

Constructing a splitting. As before, let $\mathcal{Q}_H$ be the collection of pairs from $\mathcal{Q}$ that survive in H and for each clump K in F, let us write K_H for the set $(K \setminus S_H) \cap V(H)$.

We have shown that with probability $\Omega(1)$, the graph H is such that

(1) $|K_H|$ is even for every clump K, and
(2) $|\mathcal{S}_H| \geq c_6 n/\log n$.

Consider any pair $\{x, y\} \in \mathcal{S}_H$ and note that in F, x and y disagree on at most two thirds of any clump; in particular, x and y agree on at least a third of K^*. Consequently, the probability that x and y disagree on every vertex of K_H^* is $\exp(-\Omega(n^{1/2}))$. Hence, with probability $1 - o(1)$, for every $\{x, y\} \in \mathcal{S}_H$, there exists some vertex in K_H^* on which x and y agree.

Next, it follows from Lemma 4.5 that with probability $1 - o(1)$, any two vertices $x, y \in V(H)$ which are proximate satisfy $\delta(x, y, H) \leq n^{2/3}$. Finally, the probability that we have deleted more that $2\varepsilon n - 2$ vertices of total from G is, by Proposition 2.6, $\exp(-\Omega(n))$. It follows that with probability $\Omega(1)$, the graph H, in addition to possessing the aforementioned properties, also has the following properties.

(3) For every $\{x, y\} \in \mathcal{S}_H$, there exists some vertex in K_H^* on which x and y agree.
(4) For any $x, y \in V(H)$ such that x and y are proximate in F, $\delta(x, y, H) \leq n^{2/3}$.
(5) $|V(H)| \geq (1 - 2\varepsilon)n + 2$.

With a view of making the graph H splittable, we alter H as follows. Fix a pair $(x^*, y^*) \in \mathcal{S}_H$ and a vertex $v \in K^*$ on which x^* and y^* disagree. We know that there is a vertex $u \in K_H^*$ on which x^* and y^* agree. Delete u from H. If $v \in V(H)$, delete v from H and if $v \notin V(H)$, add v back. After these alterations, note that H still has an even number of vertices. Note also that now, $|V(H)| \geq (1 - 2\varepsilon)n$ and $\delta(x^*, y^*, H) \in \{1, 3\}$.

Before we altered H, at most $c_5 n/\log n$ pairs in $\mathcal{S}_H$ disagreed on any vertex in K^*; the alterations above change the degree differences of at most $2c_5 n/\log n = c_6 n/2 \log n$ pairs in $\mathcal{S}_H$. Hence, H contains a collection $\mathcal{S}_H$ of least $c_6 n/2 \log n - 1$ pairs of vertices $\{x, y\}$ such that $\delta(x, y, H) = 2$ and a pair (x^*, y^*) such that $\delta(x^*, y^*, H) \in \{1, 3\}$. Furthermore, all the vertices of $V(H) \setminus (S_H \cup \{x^*, y^*\})$ may be grouped into pairs $\{x, y\}$ such that $\delta(x, y, H) \leq n^{2/3} + 2$; let $\mathcal{M}_H$ denote this collection of pairs.

It is now easy to check that H is splittable using the argument used to prove Lemma 3.1. Indeed, we can use pairs in $\mathcal{M}_H$ to construct a partition such that sums of the degrees of the vertices of the two halves of the partition differ by at most $n^{2/3} + 2$. For n sufficiently large, we can then reduce the difference to at most two by using all but one of the pairs in $\mathcal{S}_H$. Finally, using the one remaining pair in $\mathcal{S}_H$ and the pair (x^*, y^*), we can reduce the difference to at most one; we are done constructing a splitting of H by parity considerations. This completes the proof of Theorem 1.1. $\square$

5. Conclusion

We have shown that $f(n) \geq n/2 - o(n)$. In fact, it should be possible to read out a bound of $f(n) \geq n/2 - n/(\log\log n)^c$ from our proof for some absolute constant $c > 0$; we chose not to include a proof of this fact to keep the presentation simple, and because we do not believe that such a bound is close to the truth. While we have managed to pin down f up to its first order term, there is still a large gap between the upper and lower bounds for $n/2 - f(n)$.

Problem 5.1. *What is the asymptotic behaviour of* $n/2 - f(n)$?

We know that $n/2 - f(n) = \Omega(\log\log n)$ and $n/2 - f(n) = o(n)$; we suspect that the truth lies closer to the lower bound and that in particular, $n/2 - f(n) = o(n^\varepsilon)$ for every $\varepsilon > 0$. Indeed, it is not inconceivable that $n/2 - f(n) = \Theta(\log n)$.

It is natural to generalise the problem to the case where we have more than one type of edge, or ask for more than two disjoint subgraphs. For any $r, l \in \mathbb{N}$, given an edge colouring Δ of the complete graph on n vertices with r colours, let $g(\Delta)$ be the largest integer k for which we can find l disjoint subsets $V_1, V_2, \ldots, V_l$ of $[n]$, each of cardinality k, such that for each $1 \leq i \leq r$, the number of edges induced by V_j of colour i is the same for every $1 \leq j \leq l$. Let $g(n, r, l)$ be the minimum value of $g(\Delta)$ taken over all edge colourings of the complete graph on n vertices. In particular, note that $g(n, 2, 2) = f(n)$. We conjecture that $g(n, r, 2) = n/2 - o(n)$ and more generally, ask the following question.

Problem 5.2. *For* $r, l \in \mathbb{N}$, *what is the asymptotic behaviour of* $g(n, r, l)$?

Finally, we mention a question about digraphs that we find particularly appealing. Given a digraph D on n vertices, let $h(D)$ denote the largest integer k for which there exist disjoint subsets $A, B \subset V$ such that $|A| = |B| = k$ and the number of directed edges from A to B is equal to the number of directed edges from B to A. Let $h(n)$ be the minimum value of $h(D)$ taken over all digraphs on n vertices.

Problem 5.3. *Determine* $h(n)$.

Acknowledgements

The first author was partially supported by NSF grant DMS-1301614 and EU MULTIPLEX grant 317532.

The majority of the research in this paper was carried out while the authors were visitors at Microsoft Research, Redmond. We are grateful to Yuval Peres and the other members of the Theory Group at Microsoft Research for their hospitality.

References

1. N. Alon and B. Bollobás, *Graphs with a small number of distinct induced subgraphs*, Discrete Math. **75** (1989), 23–30. 2
2. 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. 5
3. M. Axenovich, R. R. Martin, and T. Ueckerdt, *Twins in graphs*, European J. Combin. **39** (2014), 188–197. 1, 2
4. I. Ben-Eliezer and M. Krivelevich, *Perfectly balanced partitions of smoothed graphs*, Electron. J. Combin. **16** (2009), Note 14. 1
5. B. Bollobás and A. Thomason, *Graphs which contain all small graphs*, European J. Combin. **2** (1981), 13–15. 2
6. Y. Caro and R. Yuster, *Large disjoint subgraphs with the same order and size*, European J. Combin. **30** (2009), 813–821. 1, 2
7. M. Chudnovsky and S. Safra, *The Erdős-Hajnal conjecture for bull-free graphs*, J. Combin. Theory Ser. B **98** (2008), 1301–1310. 2
8. P. Erdős and A. Hajnal, *On the number of distinct induced subgraphs of a graph*, Discrete Math. **75** (1989), 145–154. 2
9. ———, *Ramsey-type theorems*, Discrete Appl. Math. **25** (1989), 37–52. 2
10. P. Erdős, J. Pach, and L. Pyber, *Isomorphic subgraphs in a graph*, Combinatorics (Eger, 1987), Colloq. Math. Soc. János Bolyai, vol. 52, North-Holland, Amsterdam, 1988, pp. 553–556. 2
11. J. Fox and B. Sudakov, *Induced Ramsey-type theorems*, Adv. Math. **219** (2008), 1771–1800. 2
12. C. Lee, P. Loh, and B. Sudakov, *Self-similarity of graphs*, SIAM J. Discrete Math. **27** (2013), 959–972. 2
13. H. J. Prömel and V. Rödl, *Non-Ramsey graphs are $c \log n$-universal*, J. Combin. Theory Ser. A **88** (1999), 379–384. 2

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK, *and* Department of Mathematical Sciences, University of Memphis, Memphis TN 38152, USA, *and* London Institute for Mathematical Sciences, 35a South St., Mayfair, London W1K 2XF, UK

Email address: b.bollobas@dpmms.cam.ac.uk

Department of Mathematical Sciences, University of Memphis, Memphis TN 38152, USA

Email address: t.kittipassorn@memphis.edu

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

Email address: b.p.narayanan@dpmms.cam.ac.uk

Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK

Email address: scott@maths.ox.ac.uk