Large disjoint subgraphs with the same order and size

Y. Caro [*] R. Yuster [†]

Abstract

For a graph G let $f(G)$ be the largest integer k so that there are two vertex-disjoint subgraphs of G, each with k vertices, and that induce the *same* number of edges. Clearly $f(G) \leq \lfloor n/2 \rfloor$ but this is not always achievable.

Our main result is that for any fixed $\alpha > 0$, if G has n vertices and at most $n^{2-\alpha}$ edges then $f(G) = n/2 - o(n)$, which is asymptotically optimal. The proof also yields a polynomial time randomized algorithm.

More generally, let t be a fixed nonnegative integer and let H be a fixed graph. Let $f_H(G,t)$ be the largest integer k so that there are two k-vertex subgraphs of G having at most t vertices in common, that induce the *same* number of copies of H. We prove that if H has r vertices then $f_H(G,t) = \Omega(n^{1-(2r-1)/(2r+2t+1)})$. In particular, there are two subgraphs of the same order $\Omega(n^{1/2+1/(8r-2)})$ that induce the same number of copies of H and that have no copy of H in common.

1 Introduction

All graphs in this paper are finite, undirected and simple. We follow the notation and terminology of [3]. Many basic questions in extremal graph theory can be stated as asking for the existence of at least two large induced subgraphs that share some property, and that are "far apart". For example, Ramsey's Theorem asserts that we can always find two vertex-disjoint isomorphic subgraphs with a logarithmic number of vertices.

In this paper we address perhaps the most basic property of having the same order and size (throughout this paper *order* refers to the number of vertices while *size* refers to the number of edges), and, more generally, having the same order and the same number of induced copies of a fixed graph H.

Formally, if t is any fixed nonnegative integer and H is any fixed graph, then let $f_H(G,t)$ denote the largest integer k so that there exist two induced subgraphs of G, having the same order k, having the same number of induced copies of H, and that intersect in at most t vertices. Especially interesting are the case $H = K_2$ (having the same number of edges) that we denote by $f(G,t)$, the case $t = 0$ (the vertex-disjoint case) that we denote by $f_H(G)$ or by $f(G)$ if $H = K_2$, and the case $t = 1$ (the edge-disjoint and complement edge-disjoint case).

[*]Department of Mathematics and Physics, University of Haifa, Tivon 36006, Israel. email: yacaro@kvgeva.org.il
[†]Department of Mathematics, University Haifa, Haifa 31905, Israel. email: raphy@math.haifa.ac.il

Clearly, if G has n vertices we always have $f(G) \le \lfloor n/2 \rfloor$. However, this is not always achievable since, for example $f(K_{1,n-1}) = n/2 - 1$. In fact, it is not difficult to construct examples of graphs G where $n/2 - f(G)$ grows with n and is at least $\Omega(\log\log n)$ (see Section 4). Obtaining good lower bounds for $f(G,t)$, and, more generally, for $f_H(G,t)$, seems to be a nontrivial task, and the goal of this paper is to provide such bounds.

Before stating our results, let us first observe that we always have $f(G) \ge n^{1/3}$ by a simple pigeonhole argument. Indeed, arbitrarily select $\lfloor n^{2/3} \rfloor - 1$ vertex disjoint induced subgraphs of order $\lceil n^{1/3} \rceil$ each, and notice that the number of possible values for the number of edges in the induced subgraphs is less than the number of disjoint induced subgraphs. This trivial lower bound is also, of course, algorithmic.

However, there is a significantly better, though still simple, lower bound for $f(G)$ that follows from a difficult result of Lovász on the chromatic number of the Kneser graph.

Proposition 1.1 *If k satisfies $n - 2k + 2 > \binom{k}{2} + 1$ then $f(G) \ge k$. In particular, $f(G) > \sqrt{2n + 17/4} - 2$.*

In Section 2 we give the simple proof of this proposition in the more general setting of $f_H(G)$, in which case it yields a $\Theta(n^{1/r})$ lower bound for $f_H(G)$ where r denotes the number of vertices of H.

Our main result is that the trivial upper bound $f(G) \le n/2 - o(n)$ is asymptotically tight if G is any graph which is not too dense. In fact, we prove that if $\alpha > 0$ is any fixed constant and G has at most $n^{2-\alpha}$ edges then, indeed $f(G) = n/2 - o(n)$. More formally:

Theorem 1.2 *For every fixed $\alpha > 0$ and for every $\epsilon > 0$ there exists $N = N(\alpha, \epsilon)$ so that for all $n > N$, if G is a graph with n vertices and at most $n^{2-\alpha}$ edges then $f(G) \ge n/2 - \epsilon n$.*

The proof is based upon several probabilistic and combinatorial arguments and yields, in particular, a polynomial time randomized algorithm. Notice that since $f(G)$ is complement invariant then Theorem 1.2 also applies to graphs that have at most $n^{2-\alpha}$ non-edges.

Since, by definition, $f_H(G,t) \ge f_H(G,t-1)$, Proposition 1.1 also gives a $\Theta(\sqrt{n})$ lower bound for $f(G,1)$, as well as a $\Theta(n^{1/r})$ for $f_{K_r}(G,1)$. But can we do better if we allow the subgraphs to intersect in a single vertex? The answer is yes, but the proof becomes more complicated. the following theorem supplies a general lower bound for $f_H(G,t)$ which, already for $t = 1$, is far better than the one given by Proposition 1.1.

Theorem 1.3 *If H has r vertices then $f_H(G,t) = \Omega(n^{1-(2r-1)/(2r+2t+1)})$.*

The proof of theorem 1.3 is based upon probabilistic arguments, and the use of a generalized Erdős-Ko-Rado Theorem of Wilson. One can immediately see that Theorem 1.3 asserts that, for example, $f(G,1) = \Omega(n^{4/7})$, and $f_{K_3}(G,1) = \Omega(n^{4/9})$.

Another interesting case is $t = r - 1$, as two subgraphs that intersect in at most $r - 1$ vertices cannot share a copy of H. We therefore have the following corollary:

Corollary 1.4 *Let H be a fixed graph with r vertices. Then, a graph G with n vertices has two subgraphs of the same order $\Omega(n^{1/2+1/(8r-2)})$ that induce the same number of copies of H and that have no copy of H in common.*

The rest of this paper is organized as follows. In Section 2 we focus on vertex-disjoint subgraph with the same order and prove Theorem 1.2 and Proposition 1.1. Section 3 focuses on almost disjoint subgraphs with the same order and contains the proof of Theorem 1.3. The final section contains some concluding remarks and open problems.

2 Large disjoint subgraphs with the same order

Before proving Theorem 1.2 we need to establish a few lemmas. Suppose G is any graph on the vertex set $V(G) = \{v_1, \ldots, v_n\}$ where n is even and where $d(v_i) \leq d(v_{i+1})$ for $i = 1, \ldots, n-1$ (hence the vertices are sorted according to their degrees). We say that (A, B) is a *paired partition* of $V(G)$ if $|A \cap \{v_{2i-1}, v_{2i}\}| = 1$ for $i = 1, \ldots, n/2$. Notice that G has precisely $2^{n/2}$ (ordered) paired partitions. For a paired partition (A, B) let $e(A)$ (resp. $e(B)$) denote the number of edges induced by A (resp. B). Let $e(A, B)$ denote the number of edges with one endpoint in A.

Lemma 2.1 *For any paired partition (A, B) we have $|e(A) - e(B)| \leq \Delta(G)/2 = d(v_n)/2$. Furthermore, if $S \subset V(G)$ then, in a random paired partition both A and B contain at least $|S|/3$ elements of S with probability at least $1 - 2e^{-|S|/108}$.*

Proof: Clearly $2e(A) + e(A, B) = \sum_{v \in A} d(v)$ and also $2e(B) + e(A, B) = \sum_{v \in B} d(v)$. Hence

$$2(e(A) - e(B)) = \sum_{v \in A} d(v) - \sum_{v \in B} d(v)\,.$$

Clearly $\sum_{v \in A} d(v) - \sum_{v \in B} d(v)$ is maximal when $A = \{v_2, v_4, \ldots, v_n\}$ but even then the difference is at most $d(v_n) - d(v_1)$.

For the second part of the lemma, let $s = |S|$, let j be the number of pairs $\{v_{2i-1}, v_{2i}\}$ that contain one element of S, and let $S' \subset S$ be those corresponding j elements. Thus, precisely $(s-j)/2$ pairs have both of their elements in S. In a random paired partition (A, B) there are precisely $(s-j)/2$ elements of $S \setminus S'$ in A and the same holds for B. Thus, if $j \leq s/3$ we are done. Otherwise, each element of S' is chosen to A independently with probability $1/2$. Notice that the expectation of $|S' \cap A|$ is $j/2$. By a standard large deviation Chernoff estimate (cf. [2] Theorem A.1.13) we have that

$$\Pr[|S' \cap A| < j/3] < e^{-(j/6)^2/j} = e^{-j/36} < e^{-s/108}.$$

Thus, with probability at least $1 - e^{-s/108}$ we have $|S \cap A| \geq j/3 + (s-j)/2 = s/2 - j/6 \geq s/3$ and the same holds for B. ■

The proof of Theorem 1.2 can be deduced from the proof of the following, seemingly more restricted theorem.

Theorem 2.2 *For every positive integer $r \geq 4$ and for every $\epsilon_1 > 0$ there exists $N_1 = N_1(r, \epsilon_1)$ so that for all $n > N_1$, if G is a graph with n vertices and maximum degree at most $n^{1-3/r}$ then $f(G) \geq n/2 - \epsilon_1 n$.*

Before proving Theorem 2.2 let us first see how to obtain Theorem 1.2 from it.

Proof of Theorem 1.2 given Theorem 2.2: Let α and ϵ be as in Theorem 1.2. We can assume $\alpha < 2$ otherwise the theorem is trivial. Choose r to be the smallest positive integer so that so that $3/r < \alpha/2$. Choose $\epsilon_1 = \epsilon/2$. Let $N_1 = N_1(r, \epsilon_1)$ be the constant from theorem 2.2. Choose

$$N = \max\{(4/\epsilon)^{r/3}\ ,\ 2N_1\}\,.$$

Now, suppose G has $n > N$ vertices and at most $n^{2-\alpha}$ edges. Clearly, by deleting from G the vertices with degree greater than $0.5n^{1-3/r}$ (if there are any) we remain with a subgraph G' having $n' > n - 4n^{1-3/r}$ vertices and maximum degree at most $0.5n^{1-3/r} < (n')^{1-3/r}$. Since $n' > N_1(r, \epsilon_1)$ we have that

$$f(G) \geq f(G') \geq \frac{n'}{2} - \epsilon_1 n' \geq \frac{n}{2} - n\left(\epsilon_1 + \frac{2}{n^{3/r}} - \frac{4\epsilon_1}{n^{3/r}}\right) \geq \frac{n}{2} - \epsilon n.$$

■

Proof of Theorem 2.2: Throughout the proof we ignore floors and ceilings of fractional powers of n as these have no effect on the asymptotic nature of our results. The proof proceeds by induction on r starting with the basic case of $r = 4$. For this basic case we need to prove that for any $\epsilon' > 0$ there exists $N'(\epsilon')$ so that for all $n > N$, if G is a graph with $n > N'$ vertices and maximum degree at most $n^{1/4}$ then $f(G) \geq n/2 - \epsilon' n$.

We start by greedily finding an independent set of G of size $n^{1/4}$ and denote it by T_0. Denote the set of neighbors of all vertices of T_0 by X_0 and notice that since $\Delta(G) \leq n^{1/4}$ we have $|X_0| \leq n^{1/2}$. Delete $X_0 \cup T_0$ from the graph G. In the remaining graph, find $0.5n^{1/4}$ independent edges in the sense that any two of these edges only induce a matching. Denote this set of edges by T_1 and observe that T_1 can be constructed greedily by picking an edge, and deleting both of its endpoints and their at most $2n^{1/4} - 2$ neighbors from the graph, and continuing in the same way. If we cannot find $0.5n^{1/4}$ edges for T_1 then we are left with $n - n^{1/4} - n^{1/2} - n^{1/2}$ isolated vertices which means that there are two vertex disjoint independent sets of order larger than $n/2 - 2n^{1/2}$, and in particular $f(G) \geq n/2 - \epsilon' n$ as required (we assume whenever necessary that N' is sufficiently large to satisfy the inequalities).

Otherwise, by deleting at most one additional vertex we are now left with a graph with $s > n - 3n^{1/2}$ vertices where s is even. By Lemma 2.1 an arbitrary paired partition (A, B) of this remaining graph has $|e(A) - e(B)| \leq 0.5n^{1/4}$. Suppose $e(A) - e(B) = t \geq 0$. Add to A precisely $2t$ isolated vertices from T_0. Add to B precisely t independent edges from T_1. The expanded sets are still vertex-disjoint, have precisely $s/2 + 2t > n/2 - \epsilon' n$ vertices each, and induce precisely $e(A)$ edges each.

We may now assume that the theorem has been proved for $r - 1$ and for all ϵ. We need to prove it for r and any given $\epsilon > 0$. By the induction hypothesis we know that for every $\epsilon' > 0$ there exists $N' = N'(\epsilon')$ so that for all $n > N'$, if G is a graph with n vertices and maximum degree at most $n^{1-3/(r-1)}$ then $f(G) \geq n/2 - \epsilon' n$. We wish to use this fact in order to prove that for every $\epsilon > 0$, there exists $N = N(\epsilon)$ so that for all $n > N$, if G is a graph with n vertices and maximum degree at most $n^{1-3/r}$ then $f(G) \geq n/2 - \epsilon n$.

Let, therefore, $\epsilon > 0$ be given. Throughout the rest of the proof we will always pick N to be sufficiently large so as to guarantee the inequalities and we shall use the induction hypotheses with a value ϵ' sufficiently small (but still only a function of ϵ and r).

We start by finding in G an independent set T_0 of size $3n^{2.9/r}$. We delete T_0 and all of its neighbors from G to obtain a graph G_0. Notice that G_0 has at least $n - O(n^{1-0.1/r})$ vertices. Similarly, we find in G_0 a set of $n^{2.9/r}$ independent edges, denote their $2n^{2.9/r}$ endpoints by T_1, and delete T_1 and all of their neighbors. As in the case $r = 4$, if we cannot find such a T_1 then there is a huge independent set in G_0 and we are done. Otherwise, notice that the remaining graph, denoted G_1, has at least $n - O(n^{1-0.1/r})$ vertices.

Starting with the graph G_1, we now construct a sequence of graphs $G_2, G_3, \ldots, G_{r-2}$ as described in the following process. Each G_{i+1} will be a subgraph of G_i obtained by deleting several vertices from G_i. Each G_i will contain at least $n - O(n^{1-0.1/r})$ vertices (and recall that this initially holds for G_1).

While constructing G_{i+1} from G_i we will *color* some vertices with a color i (vertices may be colored by more than one color; for example a vertex may have color 3 as well as color 7 which means that it was colored while creating G_4 from G_3 and also colored wile creating G_8 from G_7). Uncolored vertices are those that have yet to receive any color. Initially, G_1 has no colored vertices at all before we construct G_2 from it.

We denote by $d_i(v)$ the degree of a vertex $v \in G_i$ in G_i. We denote by $c_i(v)$ the number of neighbors of v in G_i that have color i or less. Set $p_i = 0.5n^{-1+3/r+i/(r+1)}$ to denote a probability that will be used later. A property that we shall maintain is the following. If $d_i(v) > 0.5n^{1-3/(r-1)}$ then

$$\frac{1}{2}p_i d_i(v) \le c_i(v) \le 2p_i d_i(v).$$

We now describe how to create G_{i+1} from G_i. Randomly and independently color each vertex of G_i with the color i with probability p_i. Consider any vertex $v \in G_i$. The expected number of i-colored neighbors is precisely $d_i(v)p_i$. Now, if $d_i(v) \ge 0.5n^{1-3/(r-1)} \ge 0.5n^{1/4}$ we can use the standard Chernoff large deviation bounds to obtain that with exponentially small probability, the number of i-colored neighbors deviates from its mean $d_i(v)p_i$ by a factor of at most 1.5 (or any $1+\delta$ for that matter). Similarly, if C_i denotes the number of vertices colored by i then its expectation is $E[C_i] = |G_i|p_i$. Again, we have that the probability that C_i deviates from its mean by more than a factor of 2 is exponentially small. Hence, we can *fix* an i-coloring of some of the vertices of G_i so that the following holds:

$$C_i < 2|G_i|p_i \le n^{3/r+i/(r+1)} \; , \tag{1}$$

and also the following holds for all $v \in G_i$ with $d_i(v) \ge 0.5n^{1-3/(r-1)}$:

$$\frac{1}{2}p_i d_i(v) \le c_i(v) \le 1.5p_i d_i(v) + \sum_{j=1}^{i-1} 2p_j d_i(v) < 2p_i d_i(v) \; . \tag{2}$$

We construct an independent set T_{i+1} of vertices of G_i greedily as follows. As long as there is an uncolored vertex $v \in G_i$ independent of all previous vertices that were selected to T_{i+1} and so that $d_i(v) \geq 0.5n^{1-3/(r-1)}$, we add v to T_{i+1}. We halt when either no such vertex can be found anymore, or once the following inequality holds for the first time:

$$\sum_{v \in T_{i+1}} c_i(v) \geq 3n^{(i+1)/(r+1)}. \tag{3}$$

Let B_{i+1} denote the set of uncolored neighbors of the vertices of T_{i+1}. We define G_{i+1} to be the graph obtained from G_i by removing the vertices $T_{i+1} \cup B_{i+1}$.

We now consider the two cases that caused the procedure for creating T_{i+1} to halt. Suppose (3) still does not hold, but we cannot find another uncolored vertex to add to T_{i+1}. Each uncolored vertex u of G_{i+1} has degree $d_{i+1}(u) < 0.5n^{1-3/(r-1)}$. Observe that if $v \in T_{i+1}$ then $d_i(v) \geq 0.5n^{1-3/(r-1)}$ and hence, by (2), $c_i(v) \geq p_i d_i(v)/2$. Now, since (3) did not yet occur we have $\sum_{v \in T_{i+1}} c_i(v) < 3n^{(i+1)/(r+1)}$. Therefore

$$\sum_{v \in T_{i+1}} d_i(v) < \frac{2}{p_i} 3n^{(i+1)/(r+1)} = 12n^{1-3/r-i/(r+1)+(i+1)/(r+1)} = 12n^{1-3/r+1/(r+1)}.$$

Thus, $|T_{i+1} \cup B_{i+1}| \leq O(n^{1-3/r+1/(r+1)})$. Since, by (1) the number vertices colored by i or less is at most $2rnp_i \leq O(n^{3/r+(r-3)/(r+1)}) = O(n^{1-1/r+4/r(r+1)})$ we obtain that the graph G' consisting of the uncolored vertices of G_{i+1} has at least

$$|G'| \geq |G_i| - O(n^{1-3/r+1/(r+1)}) - O(n^{1-1/r+4/r(r+1)}) = n - O(n^{1-1/r+4/r(r+1)})$$

vertices and, furthermore, the maximum degree of G' is at most $0.5n^{1-3/(r-1)} < |G'|^{1-3/(r-1)}$. We may therefore apply the induction hypotheses to G' and obtain that

$$f(G) \geq f(G') \geq |G'|/2 - \epsilon'|G'| \geq n/2 - \epsilon n.$$

Assume, therefore that T_{i+1} has been created and the last vertex added to it caused (3) to hold for the first time. In this case we go to step $i+1$ of the algorithm using the constructed graph G_{i+1}. Notice that G_{i+1} has at least

$$|G_{i+1}| = |G_i| - |T_{i+1} \cup B_{i+1}| = n - O(n^{1-0.1/r}) - O(n^{1-3/r+1/(r+1)}) \geq n - O(n^{1-0.1/r})$$

vertices. We will also need to bound $|T_{i+1}|$ from above. Before inserting the last vertex to T_{i+1} we know that (3) still does not hold. It follows that

$$\sum_{v \in T_{i+1}} d_i(v) \leq \sum_{v \in T_{i+1}} \frac{2}{p_i} c_i(v) \leq n^{1-3/r} + 12n^{1-3/r-i/(r+1)+(i+1)/(r+1)} \leq 13n^{1-3/r+1/(r+1)}.$$

Since the degree if each vertex in T_{i+1} is at least $0.5n^{1-3/(r-1)}$ we have

$$|T_{i+1}| \leq O\left(\frac{n^{1-3/r+1/(r+1)}}{n^{1-3/(r-1)}}\right) = O\left(n^{-3/r+1/(r+1)+3/(r-1)}\right) \ll \frac{1}{r}(n^{2.9/r}). \tag{4}$$

Consider the final graph G_{r-2} and the sets of vertices $T_0, T_1, \ldots, T_{r-2}$ that have been constructed during the process. Notice that $T_0 \cup \cdots \cup T_{r-2}$ is an independent set. We know that G_{r-2} has at least $n - O(n^{1-0.1/r})$ vertices, and has maximum degree $n^{1-3/r}$. By deleting from G_{r-2} at most one vertex we can assume that G_{r-2} has an even number of vertices. Let (A, B) be a random paired partition of G_{r-2}. By Lemma 2.1 we know that $e(A) - e(B) = k < n^{1-3/r}$ (assuming w.l.o.g. that $e(A) \geq e(B)$).

For $j = 1, \ldots r-3$ and for each vertex $v \in T_{j+1}$, the set S_v of neighbors of v in G_{r-2} has cardinality $|S_v| = c_j(v)$. By (2) we know that $c_j(v) \geq 0.5 p_j d_j(v) \geq 0.25 p_j n^{1-3/(r-1)} \geq O(n^{1/r})$. Hence, by Lemma 2.1 there exists a paired partition (A, B) so that for all $j = 1, \ldots r-3$ and for each vertex $v \in T_{j+1}$, v has at least $c_j(v)/3$ neighbors in B.

As A induces k more edges than B, we will attempt to correct this gap by carefully adding to B vertices of $T_1 \cup \cdots \cup T_{r-2}$ and by adding the same amount of isolated vertices to A from T_0. By (4) we recall that $|T_0| > |T_1| + |T_2| + \cdots + |T_{r-2}|$ and thus we always have enough isolated vertices to add to A. Hence, we just need to show that it is possible to add to B vertices of $T_1 \cup \cdots \cup T_{r-2}$ so that the sum of the number of neighbors in B of these added vertices is *precisely* k. For $i = 1, \ldots, r-2$ we will gradually add vertices of T_{r-1-i} to B so that k slowly decreases until it becomes zero. We will denote the gap after step i by k_i. Hence, initially $k_0 = k$ and we need to show that $k_{r-2} = 0$. We will also make sure that $k_i < n^{1-(3+i)/(r+1)}$. We denote by B_i the extension of B after step i.

For $i = 1$, let us first use vertices of T_{r-2} to add to B in order to decrease k. Each vertex $v \in T_{r-2}$ has at least $c_{r-3}(v)/3$ neighbors in G_{r-2} and at most $c_{r-3}(v)$ neighbors. Since, by (2) $c_{r-3}(v) < 2p_{r-3} n^{1-3/r} = n^{1-4/(r+1)}$ each addition of a vertex from T_{r-2} does not decrease the gap by more than $n^{1-4/(r+1)}$. Hence, by adding sufficiently many vertices from T_{r-2} we can make the gap smaller than $n^{1-4/(r+1)}$. But how can we make sure that we do not exhaust T_{r-2} before getting this smaller gap? We therefore need to show that the sum of the number of neighbors of the vertices of T_{r-2} in B is greater than k, or, equivalently, that the sum of the number of neighbors of the vertices of T_{r-2} in G_{r-2} is greater than $3k$. But, by (3) this latter sum is at least $3n^{(r-2)/(r+1)} > 3n^{1-3/r} > 3k$, as required. We have therefore proved that $e(A) - e(B_1) = k_1 < n^{1-4/(r+1)}$.

Let us now consider a general step $i = 2, \ldots, r-3$ (the last step $i = r-2$ will be handled separately). Each addition of a vertex from from T_{r-1-i} to B_{i-1} does not decrease k_{i-1} by more than $c_{r-2-i}(v) < 2p_{r-2-i} n^{1-3/r} = n^{1-(3+i)/(r+1)}$. Thus we can make sure that $k_i < n^{1-(3+i)/(r+1)}$. Again, we must make sure that we do not exhaust T_{r-1-i} before getting this smaller gap. We therefore need to show that the sum of the number of neighbors of the vertices of T_{r-1-i} in B (which is precisely the same as in B_{i-1} because the union of all of the T_j is an independent set) is greater than k_{i-1}, or, equivalently, that the sum of the number of neighbors of the vertices of T_{r-1-i} in G_{r-2} is greater than $3k_{i-1}$. But, by (3) this latter sum is at least $3n^{(r-1-i)/(r+1)} = 3n^{1-(2+i)/(r+1)} > 3k_{i-1}$, as required. We have therefore proved that $e(A) - e(B_i) = k_i < n^{1-(3+i)/(r+1)}$.

Consider now the final step $i = r-2$. Prior to this step we have $e(A) - e(B_{r-3}) = k_{r-3} < n^{1-r/(r+1)} = n^{1/(r+1)}$. But T_1 has $n^{2.9/r}$ independent edges which is far more than what we need in order to close the gap and make $e(A) - e(B_{r-2}) = 0$. We have therefore proved that $f(G) \geq |G_{r-2}|/2 \geq n/2 - O(n^{1-0.1/r})$, as required. ■

The *Kneser graph* $KG(n,k)$ has as its vertex set all k-subsets of $\{1,\ldots,n\}$ and two vertices of $KG(n,k)$ are adjacent if the corresponding k-subsets are disjoint. Kneser conjectured in [5] that the chromatic graph of $KG(n,k)$ is $n-2k+2$. This conjecture was solved in a seminal paper of Lovász [6]. Using this result one can easily derive the following proposition, generalizing Proposition 1.1.

Proposition 2.3 *Suppose that P is a graph-theoretic parameter and let $g_P(k)$ denote the number of possible values that P can attain in the family of k-vertex graphs. Let k be the largest integer for which $n-2k+2 > g_P(k)$. Then any n-vertex graph has two induced vertex-disjoint subgraphs of order k for which the value of P is the same.*

proof: Consider the set of all induced k-vertex subgraphs of G. If H is such a subgraph then color H with the color $P(H)$. This corresponds to a coloring of the vertices of $KG(n,k)$ with $g_P(k)$ colors. The coloring cannot be proper since $\chi(KG(n,k)) = n-2k+2 > g_P(k)$. Hence two disjoint k-vertex subgraphs receive the same color. ■

If P is the property "number of edges" then $g_P(k) = \binom{k}{2}+1$. If P is the property "number of induced copies of H" then $g_P(k) \le \binom{k}{r}+1$ where r is the number of vertices of H. Thus, by Proposition 2.3, $f_H(G) = \Omega(n^{1/r})$.

3 Large almost-disjoint subgraphs with the same order

In this section we prove Theorem 1.3. We need the following result of Wilson [7] who generalized the Erdős-Ko-Rado Theorem for t-intersecting families.

Lemma 3.1 *If $n \ge (t+2)(k-t-2)$ then any family of more than $\binom{n-t-1}{k-t-1}$ k-subsets of n contains two subsets that intersect in at most t elements.*

Proof of Theorem 1.3: We fix a graph H with r vertices and an integer $t \ge 1$. Throughout the proof we assume that n is sufficiently large to satisfy the various inequalities. We set $k = 0.5n^{1-(2r-1)/(2r+2t+1)}$ and wish to prove that if G is a graph with n vertices then there are two k-vertex subgraphs of G that intersect in at most t vertices and that induce precisely the same number of copies of H. It will be convenient to denote the number of induced H-subgraphs of G by $m = \alpha\binom{n}{r}$, where $0 \le \alpha \le 1$.

Let R be a random subset of k vertices of G, chosen uniformly from all possible $\binom{n}{k}$ subsets. Let $G[R]$ be the subgraph induced by R and let X be the random variable corresponding to the number of induced H-subgraphs or R. As each H-subgraph of G is also a subgraph of R with probability $\binom{n-r}{k-r}/\binom{n}{k}$. we have that the expectation of X is $E[X] = \alpha\binom{k}{r}$. In fact, if $\mathcal{H}$ is the set of induced H-subgraphs of G then X is just the sum of the indicator random variables X_J for $J \in \mathcal{H}$, where $X_J = 1$ if J is a subgraph of $G[R]$. Clearly, $\Pr[X_J = 1] = \binom{n-r}{k-r}/\binom{n}{k}$.

Let us now estimate the variance of X. We recall (see [2], Page 42) that

$$Var[X] \le E[X] + \sum_{J \neq J'} Cov[X_J, X_{J'}] \ .$$

Now, to estimate $Cov[X_J, X_{J'}]$ we observe that if J and J' share no vertex then

$$Cov[X_J, X_{J'}] = \frac{k(k-1)\cdots(k-2r+1)}{n(n-1)\cdots(n-2r+1)} - \frac{k^2(k-1)^2\cdots(k-r+1)^2}{n^2(n-1)^2\cdots(n-r+1)^2} < 0 \ .$$

If J and J' share s vertices where $1 \le s \le r-1$ then

$$Cov[X_J, X_{J'}] = \frac{k(k-1)\cdots(k-2r+s+1)}{n(n-1)\cdots(n-2r+s+1)} - \frac{k^2(k-1)^2\cdots(k-r+1)^2}{n^2(n-1)^2\cdots(n-r+1)^2} < \left(\frac{k}{n}\right)^{2r-s} \ .$$

As there are less than n^{2r-s} ordered pairs (J, J') that share s vertices we obtain that

$$Var[X] < E[X] + (r-1)k^{2r-1} < rk^{2r-1}.$$

From Chebyschev's Inequality we have that for any $a > 0$

$$\Pr[|X - E[X]| \ge a] \le \frac{Var[X]}{a^2} \ .$$

We will choose $a^2 = 2Var[X]$ and obtain that with probability at least 0.5, X receives one of $1 + 2(2Var[X])^{1/2} < \sqrt{8r}k^{r-1/2}$ possible values. Since there are $\binom{n}{k}$ distinct k-subsets it follows that at least

$$\frac{1}{2}\frac{\binom{n}{k}}{\sqrt{8r}k^{r-1/2}}$$

k-subsets R have the same number of induced copies of H. It remains to show that not all of them have more than t vertices in common. By Lemma 3.1 it suffices to prove that

$$\binom{n-t-1}{k-t-1} < \frac{1}{2}\frac{\binom{n}{k}}{\sqrt{8r}k^{r-1/2}} \ .$$

In particular, it suffices to show that

$$\sqrt{32r}k^{r-1/2} < \left(\frac{n}{k}\right)^{t+1} \ .$$

The latter follows immediately from $k = 0.5n^{1-(2r-1)/(2r+2t+1)}$. ■

4 Concluding remarks and open problems

- It seems that extending Theorem 1.2 to the set of all graphs is a difficult task. We do suspect, however that $f(G) = n/2 - o(n)$ for all graphs. The following construction shows that one cannot hope to replace $o(n)$ with a constant. Consider the sequence of positive integers $\{a_k\}$ defined as follows. $a_1 = 3$ and a_k is the smallest odd number so that $\sum_{i=1}^{k-1}\binom{a_i}{2} < a_k/4$. Thus, $a_2 = 13$, $a_3 = 325$ and so on. Now, clearly, if $n = a_1 + \cdots + a_k$ then $k = \Theta(\log\log n)$ (each element is of the order of a square of its predecessor). Consider, therefore, the graph G

with n vertices obtained by taking vertex disjoint cliques of sizes $a_1, \ldots, a_k$. We claim that $f(G) \leq n/2 - k/4$. Assume that A, B are disjoint sets of vertices realizing $f(G)$. Let j be the largest index for which the clique K_{a_j} of G does not contribute the same number of vertices to A and B. If $j \leq k/2$ then the fact that the cliques are odd implies that each K_{a_i} with $i > j$ has at least one vertex not in $A \cup B$ and hence $f(G) \leq (n - k/2)/2$. If $j > k/2$ there are two cases. Either K_{a_j} has at least $a_j/2 - 1$ vertices not in $A \cup B$ in which case clearly $f(G) \leq n/2 - (a_j/2 - 1)/2 > n/2 - k/4$. Otherwise, K_{a_j} has $x > a_j/2 + 1$ vertices in $A \cup B$. Without loss of generality it has y vertices in A and $x - y$ vertices in B where $y > x - y$. Since $\binom{y}{2} - \binom{x-y}{2} > a_j/4$, we have that K_{a_j} contributes to A more than $a_j/4$ edges than what it contributes to B. Hence, even if all the K_{a_i} for $i < j$ are completely with B this cannot make A and B induce the same number of edges.

- A different approach for proving a lower bound for $f(G, t)$ is through the chromatic number of *generalized* Kneser graphs. The t-generalized Kneser graph $KG(n, k, t)$ has as its vertex set all k-subsets of $\{1, \ldots, n\}$ and two vertices of $KG(n, k, t)$ are adjacent if the corresponding k-subsets are have at most t elements in common. clearly, as in Proposition 1.1, if k satisfies $\chi(KG(n, k, 1)) < \binom{k}{2} + 1$ then $f(G, 1) \geq k$. Unfortunately, the known values (and lower bounds) for $\chi(KG(n, k, 1))$ are of the order n^2/k only for values of n that are exponential in k [1, 4]. If $k = n^\alpha$ then there are no nontrivial lower bound for $\chi(KG(n, k, 1))$ and the trivial ones yield results that are inferior to those of Theorem 1.3.

- It is possible to prove an analogue of Theorem 1.2 for the parameter $f_H(G)$. In other words, for n-vertex graphs G that are not too dense, $f_H(G) = n/2 - o(n)$. The proof, however, becomes even more complicated than the present proof of Theorem 1.2 since there is no analogue of Lemma 2.1. We thus omit it from the present paper.

- It seems interesting to characterize the graph parameters for which the bound in Proposition 2.3 is far from tight. For example, the graph parameter "maximum matching number" has this property. For this parameter we have $g_P(k) = \lfloor k/2 + 1 \rfloor$. Thus, Proposition 2.3 only guarantees two vertex-disjoint subgraph of order roughly $0.4n$ having the same maximum matching number. But clearly, we can always find two vertex disjoint subgraphs of the same order greater than $n/2 - 2$ having the same maximum matching number.

- As can be seen from the proof of Theorem 1.2, all of its ingredients are algorithmic (the greedy selection of independent sets and independent edges, the randomized colorings, and the construction of the graph sequence). In fact, it is not difficult to see that, with high probability, two vertex-disjoint subsets of the same order $n/2 - o(n)$ and the same size can be constructed in $O(n^2)$ time.

References

[1] N. Alon, P. Frankl, and L. Lovász, *The chromatic number of Kneser hypergraphs*, Transactions of the AMS 298 (1986), 359–370.

[2] N. Alon and J. H. Spencer, **The Probabilistic Method**, *Second Edition*, Wiley, New York, 2000.

[3] B. Bollobás, **Extremal Graph Theory**, Academic Press, London, 1978.

[4] P. Frankl, *On the chromatic number of the general Kneser-graph*, Journal of Graph Theory 9 (1985), 217–220.

[5] M. Kneser, *Aufgabe 360*, Jahresbericht der Deutschen Mathematiker-Vereinigung 58 (1955).

[6] L. Lovász, *Kneser's Conjecture, Chromatic Numbers and Homotopy*, Journal of Combinatorial Theory, Series A 25 (1978), 319–324.

[7] R.M. Wilson, *The exact bound in the Erdős-Ko-Rado Theorem*, Combinatorica 4 (1984), 247–257.