# Induced subgraphs of prescribed size

Noga Alon[*]      Michael Krivelevich[†]      Benny Sudakov [‡]

## Abstract

A subgraph of a graph $G$ is called *trivial* if it is either a clique or an independent set. Let $q(G)$ denote the maximum number of vertices in a trivial subgraph of $G$. Motivated by an open problem of Erdős and McKay we show that every graph $G$ on $n$ vertices for which $q(G) \leq C \log n$ contains an induced subgraph with exactly $y$ edges, for every $y$ between 0 and $n^{\delta(C)}$. Our methods enable us also to show that under much weaker assumption, i.e., $q(G) \leq n/14$, $G$ still must contain an induced subgraph with exactly $y$ edges, for every $y$ between 0 and $e^{\Omega(\sqrt{\log n})}$.

## 1 Introduction

All graphs considered here are finite, undirected and simple. For a graph $G = (V, E)$, let $\alpha(G)$ denote the independence number of $G$ and let $w(G)$ denote the maximum number of vertices of a clique in $G$. Let $q(G) = max\{\alpha(G), w(G)\}$ denote the maximum number of vertices in a trivial induced subgraph of $G$. By Ramsey Theorem (see, e.g., [10]), $q(G) \geq \Omega(\log n)$ for every graph $G$ with $n$ vertices. Let $u(G)$ denote the maximum integer $u$, such that for every integer $y$ between 0 and $u$, $G$ contains an induced subgraph with precisely $y$ edges. Erdős and McKay [5] (see also [6], [7] and [4], p. 86) raised the following conjecture

**Conjecture 1.1** *For every $C > 0$ there is a $\delta = \delta(C) > 0$, such that every graph $G$ on $n$ vertices for which $q(G) \leq C \log n$ satisfies $u(G) \geq \delta n^2$.*

Very little is known about this conjecture. In [3] it is proved for random graphs. For non-random graphs, Erdős and McKay proved the following much weaker result: if $G$ has $n$ vertices and $q(G) \leq C \log n$, then $u(G) \geq \delta(C) \log^2 n$. Here we prove the following, which improves the $\Omega(\log^2 n)$ estimate considerably, but is still far from settling the conjecture.

**Theorem 1.2** *For every $C > 0$ there is a $\delta = \delta(C) > 0$, such that every graph $G$ on $n$ vertices for which $q(G) \leq C \log n$ satisfies $u(G) \geq n^\delta$.*

We suspect that $u(G)$ has to be large even if $q(G)$ is much larger than $C \log n$. In fact, we propose the following conjecture.

**Conjecture 1.3** *Every graph $G$ on $n$ vertices for which $q(G) \leq n/20$ satisfies $u(G) \geq \Omega(n)$.*

Our methods enable us to show that the following weaker statement holds for any graph $G$ satisfying the above assumption.

**Theorem 1.4** *There exists a constant $c > 0$, such that every graph $G$ on $n$ vertices for which $q(G) \leq n/14$ satisfies $u(G) \geq e^{c\sqrt{\log n}}$.*

Throughout the paper we omit all floor and ceiling signs whenever these are not crucial. All logarithms are in base 2, unless otherwise specified. We make no attempt to optimize the absolute constants in our estimates and assume, whenever needed, that the number of vertices $n$ of the graph considered is sufficiently large.

## 2    Sets with large intersection and large complements-intersection

**Lemma 2.1** *Let $\mathcal{F}$ be a family of $s$ subsets of $M = \{1, 2, \ldots, m\}$, and suppose that each $F \in \mathcal{F}$ satisfies $\epsilon m \leq |F| \leq (1 - \epsilon)m$. Suppose, further, that there are integers $a, b, t$ such that*

$$s\big(\epsilon(1 - \epsilon)\big)^t - \binom{s}{a}\delta^t > b - 1.$$

*Then there is a subset $\mathcal{G} \subset \mathcal{F}$ of $b$ members of $\mathcal{F}$ such that the intersection of every $a$ members of $\mathcal{G}$ has cardinality larger than $\delta m$, and the intersection of the complements of every $a$ members of $\mathcal{G}$ has cardinality larger than $\delta m$.*

**Proof.** We apply a modified version of an argument used in [2]. Let $A_1$ and $A_2$ be two random subsets of $M$, each obtained by picking, randomly, independently and with repetitions, $t$ members of $M$. Define $\mathcal{G}' = \{F \in \mathcal{F} : A_1 \subset F, \ F \cap A_2 = \emptyset\}$. The probability that a fixed set $F \in \mathcal{F}$ lies in $\mathcal{G}'$ is

$$\left(\frac{|F|}{m}\right)^t \left(\frac{m - |F|}{m}\right)^t \geq \big(\epsilon(1 - \epsilon)\big)^t.$$

Call a subfamily $S$ of $a$ members of $\mathcal{F}$ *bad* if either the cardinality of the intersection of all members of $S$ or the cardinality of the intersection of all the complements of these members is at most $\delta m$. If $S$ is such a bad subfamily, then the probability it lies in $\mathcal{G}'$ is at most $\delta^t$. Indeed, if the cardinality of the intersection of all members of $S$ is at most $\delta m$, then the probability that all members of $A_1$ lie in all these members is at most $\delta^t$, and if the cardinality of the intersection of the complements is at most $\delta m$, then the probability that all members of $A_2$ lie in all complements is at most $\delta^t$. By

linearity of expectation, it follows that the expected value of the random variable counting the size of $\mathcal{G}'$ minus the number of bad $a$-tuples contained in $\mathcal{G}'$ is at least

$$s\big(\epsilon(1-\epsilon)\big)^t - \binom{s}{a}\delta^t > b - 1.$$

Hence there is a particular choice of $A_1, A_2$ such that the corresponding difference is at least $b$. Let $\mathcal{G}$ be a subset of $\mathcal{G}'$ of cardinality $b$ obtained from $\mathcal{G}$ by removing at least one member from each bad $a$-tuple. This $\mathcal{G}$ clearly possesses the required properties. $\qquad\square$

We need the following two special cases of the last lemma.

**Corollary 2.2** *Let $\mathcal{F}$ be a family of subsets of $M = \{1, 2, \ldots, m\}$, and suppose that each $F \in \mathcal{F}$ satisfies $\epsilon m \le |F| \le (1 - \epsilon)m$.*
*(i) If $|\mathcal{F}| \ge (4/\epsilon)^2$ then $\mathcal{F}$ contains two sets such that the size of their intersection and the size of the intersection of their complements are both at least $(\epsilon/4)^2 m$.*
*(ii) If $|\mathcal{F}| \ge m^{3/4}$ and $\epsilon(1-\epsilon) \ge m^{-1/30}$, then $\mathcal{F}$ contains a family of at least $m^{0.6}$ sets, so that the intersection of each three of them is of size at least $m^{1/2}$, and the intersection of the complements of each three of them is of size at least $m^{1/2}$.*

**Proof.** Part (i) follows by applying the lemma with $s = (4/\epsilon)^2$, $a = b = t = 2$ and $\delta = (\epsilon/4)^2$. Part (ii) follows by applying the lemma with $s = m^{3/4}$, $a = 3$, $t = 4$, $\delta = m^{-1/2}$ and $b = m^{0.6}$. $\qquad\square$

## 3 Density and induced subgraphs

For a real $\gamma < 1/2$ and an integer $t$, call a graph $G$ $(\gamma, t)$-*balanced* if the number of edges in the induced subgraph of $G$ on any set of $r \ge t$ vertices is at least $\gamma\binom{r}{2}$ and at most $(1 - \gamma)\binom{r}{2}$ edges. We need the following simple fact.

**Lemma 3.1** *Let $G = (V, E)$ be a $(\gamma, n/3)$-balanced graph on $n$ vertices. Then there is a set $U$ of $\gamma n/6$ vertices of $G$ such that for $W = V - U$ and for each $u \in U$, $u$ has at least $\gamma|W|/6$ and at most $(1 - \gamma/6)|W|$ neighbors in $W$.*

**Proof.** If $G$ has at least $n/2$ vertices of degree at most $(n-1)/2$, let $V_1$ be a set of $n/2$ such vertices. By assumption there are at least $\gamma\binom{n/2}{2}$ edges in the induced subgraph on $V_1$, and hence it contains a vertex of degree bigger than $\gamma n/3$. Omitting this vertex from $V_1$ and applying the same reasoning to the remaining subgraph, we get another vertex of degree at least $\gamma n/3$. Continue in this manner $\gamma n/6$ steps to get a set $U$ of $\gamma n/6$ vertices. This set clearly satisfies the requirements. Indeed, by definition, every vertex in $U$ has at most $(n-1)/2 < (1 - \gamma/6)^2 n = (1 - \gamma/6)|V - U|$ neighbors. On the other hand it has at least $\gamma n/3 - \gamma n/6 = \gamma n/6$ neighbors outside $U$. If $G$ does not have at least $n/2$ vertices of degree at most $(n-1)/2$, we apply the same argument to its complement. $\qquad\square$

The following lemma is crucial in the proof of the main results.

**Lemma 3.2** Let $G = (V, E)$ be a $(6\epsilon, n^{0.2})$-balanced graph on $n$ vertices, and suppose that $\epsilon \geq n^{-0.01}$. Define $\delta = 0.5(\epsilon/4)^2$ and put $k = 0.3\frac{\log n}{\log(1/\delta)} - 1$. Then there are pairwise disjoint subsets of vertices $A_0, A_1, \ldots, A_{k+1}, B_0, B_1, \ldots, B_{k+1}$ with the following properties:

(i) $|A_i| = 2^i$ for each $i$.

(ii) $|B_0| = 3$ and $|B_i| = 2$ for each $i \geq 1$.

(iii) $B_0$ is an independent set in $G$.

(iv) Each vertex of $B_i$ is connected to each vertex of $A_i$, and is not connected to any vertex of $A_j$ for $j > i$.

(v) There are no edges connecting vertices of two distinct sets $B_i$.

(vi) The induced subgraph of $G$ on $A_i$ contains at least $6\epsilon\binom{|A_i|}{2}$ edges.

**Proof.** By Lemma 3.1 there is a set $U_0$ of $\epsilon n$ vertices of $G$, so that for $W_0 = V - U_0$, $|W_0| = m_0$, each vertex of $U$ has at least $\epsilon m_0$ and at most $(1 - \epsilon)m_0$ neighbors in $W_0$. Therefore, by Corollary 2.2, part (ii), there is a set $S$ of at least $m_0^{0.6} > n^{0.5}$ vertices in $U_0$, so that any three of them have at least $m_0^{0.5} \geq n^{0.5}/2$ common neighbors and at least $m_0^{0.5} \geq n^{0.5}/2$ common non-neighbors in $W_0$. Since $G$ is $(6\epsilon, n^{0.2})$-balanced, its induced subgraph on $S$ contains an independent set of size three (simply by taking, repeatedly, a vertex of minimum degree in this subgraph, and omitting all its neighbors). Let $A_0$ be a set consisting of an arbitrarily chosen common neighbor of these three vertices, and let $B_0$ be the set of these three vertices. Define also $C_1$ to be the set of all common non-neighbors of the vertices in $B_0$ inside $W_0$. Therefore $|C_1| \geq n^{0.5}/2$. By Lemma 3.1 applied to the induced subgraph on $C_1$, it contains a set $U_1$ of $\epsilon|C_1| \geq (4/\epsilon)^2$ vertices such that for $W_1 = C_1 - U_1$, $|W_1| = m_1$, each vertex of $U_1$ has at least $\epsilon m_1$ and at most $(1 - \epsilon)m_1$ neighbors in $C_1$. By Corollary 2.2, part (i), there is a set $B_1$ of two vertices of $U_1$, having at least $(\epsilon/4)^2 m_1 \geq 0.5(\epsilon/4)^2|C_1| = \delta|C_1| \geq 0.5\delta n^{0.5}$ common non-neighbors and at least $0.5\delta n^{0.5}$ common neighbors. Let $A_1$ be a set of two of these common neighbors that contains the maximum number of edges among all possible choices for the set $A_1$ (one edge in this particular case), and let $C_2$ be the set of all common non-neighbors. Note that by averaging, and as $G$ is $(6\epsilon, n^{0.2})$-balanced, the number of edges on $A_1$ is at least $6\epsilon\binom{|A_1|}{2}$. Note also that there are no edges between $B_0$ and $A_1$.

Applying the same argument to the induced subgraph on $C_2$ we find in it pairwise disjoint sets $B_2$ of two vertices, $A_2$ of four vertices, and $C_3$ of size at least $0.5\delta^2 n^{0.5}$, so that each member of $B_2$ is connected to each member of $A_2$ but to no member of $C_3$, and $A_2$ contains at least $6\epsilon\binom{|A_2|}{2}$ edges. We can clearly continue this process as long as the resulting sets $C_i$ are of size that exceeds $n^{0.2}$, thus obtaining all sets $A_i, B_i$. The construction easily implies that these sets satisfy all the required conditions (i)-(vi). $\square$

**Corollary 3.3** Let $G, \epsilon, n, \delta$ and $k$ be as in Lemma 3.2. Then $u(G) \geq 6\epsilon\binom{2^k}{2}$, that is, for every integer $y$ between $0$ and $6\epsilon\binom{2^k}{2}$, $G$ contains an induced subgraph with precisely $y$ edges.

**Proof.** Let $a_1, a_2, a_3, \ldots, a_p$ be an ordering of all vertices of $\cup_i A_i$ starting with the unique vertex of $A_0$, followed by those in $A_1$, then by those in $A_2$, etc. Given an integer $y$ in the range above, let

$j$ be the largest integer such that the induced subgraph of $G$ on $A = \{a_1, a_2, \ldots, a_j\}$ has at most $y$ edges. Let $z$ denote the number of edges in this induced subgraph. Notice that as $y \leq 6\epsilon\binom{2^k}{2}$, $j < p$, as the sets $A_i$ satisfy the conclusion of Lemma 3.2, (vi). By the maximality of $j$, the number of neighbors of $a_{j+1}$ in $A$ is bigger than $y - z$, and hence $|A| > y - z$. To complete the proof we show that one can append to $A$ appropriate vertices from the sets $B_i$ to get an induced subgraph with the required number of edges. Note that the set $A$ consists of the union of all sets $A_i$ for $i$ between 0 and some $d \leq k$, together with some vertices of $A_{d+1}$. Therefore we have $|A| \leq \sum_{i=0}^{d+1} 2^i$. Observe also that for each $1 \leq i \leq d$ and $b \in B_i$,

$$2^i = |A_i| = d(b, A_i) \leq d(b, A) \leq \sum_{j=0}^{i} |A_j| = 1 + \ldots + 2^i < 2^{i+1}. \tag{1}$$

The proof will easily follow from the proposition below.

**Proposition 3.4** *If $2^{i+1} \leq y - z < 2^{i+2}$ for $1 \leq i \leq k$, then one can add some vertices from $B_i$ to $A$ to get $y - z' < 2^{i+1}$, where $z'$ is the number of edges in the induced graph on the augmented set.*

**Proof of Proposition 3.4.** Let $B_i = \{b_{i1}, b_{i2}\}$. First we add $b_{i1}$ to $A$. Recalling (1), we get $y - z < 2^{i+2} - 2^i = 3 \cdot 2^i$ (where $z$ is the number of edges in the new set $A$). If $y - z < 2^{i+1}$, then we are done, otherwise $y - z \geq 2^{i+1}$, while the degree of $b_{i2}$ to the new set $A$ is less than $2^{i+1} + 1$, i.e. at most $2^{i+1}$, again due to (1). Adding $b_{i2}$ decreases the difference $y - z$ by at least $2^i$, thus making it less than $3 \cdot 2^i - 2^i = 2^{i+1}$. □

To prove the lemma, recall that initially $y - z < |A| \leq \sum_{i=0}^{d+1} |A_i| < 2^{d+2}$. As long as $y - z \geq 2^2$, we find an index $1 \leq i \leq d + 1$ such that $2^{i+1} \leq y - z < 2^{i+2}$ and apply Proposition 3.4 to reduce the difference $y - z$ below $2^{i+1}$, using vertices from $B_i$. Once we reach $y - z < 2^2 = 4$, we add $y - z$ vertices from $B_0$, obtaining the required number of edges. □

Finally to prove Theorem 1.2 we need also the following lemma of Erdős and Szemerédi [8].

**Lemma 3.5** *Let $G$ be a graph of order $n$ with at most $n^2/s$ edges. Then $G$ contains a trivial subgraph on at least $\Omega\left(\frac{s}{\log s} \log n\right)$ vertices.*

**Proof of Theorem 1.2.** Let $G = (V, E)$ be a graph with $n$ vertices satisfying $q(G) \leq C \log n$. Then, from Lemma 3.5, it follows easily that there is an $\beta = \beta(C) > 0$, so that $G$ is $(\beta, n^{0.2})$-balanced. By Lemma 3.2 this implies that $u(G) \geq n^\delta$ for some constant $\delta = \delta(C) > 0$. □

## 4 Graphs with large trivial subgraphs

In this section we present the proof of Theorem 1.4. First we need to obtain a lower bound on $u(H)$ for a bipartite graph $H$ with positive degrees. This is done in the following simple lemma, which may be of independent interest.

**Lemma 4.1** *Let $H$ be a bipartite graph with classes of vertices $A$ and $B$ such that every vertex of $A$ has a positive degree. Then $u(H) \geq |A|$.*

**Proof.** Let $B' \subset B$ be a subset of minimum cardinality of $B$ such that each vertex of $A$ has at least one neighbor in $B'$. Put $|A| = n$, and let $d_1 \leq d_2 \leq \ldots \leq d_n$ be the degrees of the vertices of $A$ in the induced subgraph $H'$ of $H$ on $A \cup B'$. We assume that the vertices of $A$ are $1, \ldots, n$. Also for every $j \in A$, let $N_j$ be the set of neighbors of this vertex in $B'$. By the minimality of $B'$, $d_1 = 1$. Similarly, by the minimality of $B'$ for each $i > 1$ we have $d_i \leq d_1 + d_2 + \ldots + d_{i-1}$. Otherwise we can delete an arbitrary vertex of $B'$ not in $\bigcup_{j=1}^{i-1} N_j$, keeping all degrees in $A$ positive and contradicting the minimality. Therefore it is easy to see that every integer up to $\sum_{i=1}^{n} d_i \geq n$ can be written as a sum of a subset of the set $\{d_1, \ldots, d_n\}$, and is thus equal to the number of edges in the corresponding induced subgraph of $H'$ (and hence of $H$). $\qquad\square$

**Remark.** This result is clearly tight, as shown by a star (that is, by the trivial case $|B| = 1$) or by a complete bipartite graph on $A$ and $B$ with $|B| \leq |A|$, in case $|A| + 1$ is a prime.

Next we need the following easy lemma which deals with the possible sizes of induced subgraphs in a graph which is either a disjoint union of cycles or a long path.

**Lemma 4.2** *Let $k$ be an integer and let $H$ be a graph which is either (i) a disjoint union of $k$ cycles or (ii) a path of length $3k - 1$. Then for every integers $m \leq k$ and $t \leq m - 2$ there exists an induced subgraph of $H$ with exactly $m$ vertices and $t$ edges.*

**Proof.** Since we can always consider a subgraph of $H$ induced by the union of the first $m$ cycles in case (i) or a path of length $3m$ in case (ii), it is enough to prove this statement only for $m = k$.
**(i).** Denote by $c_1, \ldots, c_k$ the lengths of the cycles forming $H$. Given an integer $t \leq k - 2$, let $j$ be the index such that $c_1 + \ldots + c_{j-1} \leq t < c_1 + \ldots + c_j$. If $c_1 + \ldots + c_j \geq t + 2$ then we can delete a few consecutive vertices from the $j$-th cycle to obtain a graph with exactly $t$ edges. It is easy to see that the number of vertices of this graph is $t + 1 < k$. Otherwise $c_1 + \ldots + c_j = t + 1$. Then we can delete one vertex from the $j$-th cycle and add any two vertices which form an edge from the next cycle. In this case we obtain a graph with $t$ edges and $t + 2 \leq k$ vertices. Note that in both cases we constructed an induced subgraph of $H$ with exactly $t$ edges and at most $k$ vertices. Since the total number of disjoint cycles is $k$ we can now add to our graph one by one isolated vertices from the remaining cycles until we obtain a graph with exactly $k$ vertices.
**(ii).** To prove the assertion of the lemma in this case just pick the first $t + 1$ vertices of the path and add an independent set of size $k - (t + 1)$ which is a subset of the last $2k - 2$ vertices of the path. $\qquad\square$

A number is *triangular* if it is of the form $\binom{a}{2}$ for some positive integer $a$. We need the following well known result proved by Gauss (see, e.g., [9], p 179).

**Proposition 4.3** *Every positive integer is a sum of at most three triangular numbers.*

Having finished all the necessary preparations we are now ready to complete the proof of our second theorem.

**Proof of Theorem 1.4.** Let $G = (V, E)$ be a graph of order $n$ such that $q(G) \leq n/14$ and let $I$ be a largest independent set in $G$. Denote by $G'$ the subgraph of $G$ induced by the set $V' = V - I$ and let $I'$ be the maximum independent set in $G'$. By the definition of $I$, every vertex of $I'$ has at least one neighbor in $I$. Therefore the set $I \cup I'$ induces a bipartite subgraph $H$ of $G$ which satisfies the condition of Lemma 4.1. This implies that $u(G) \geq u(H) \geq |I'|$. Thus, if $|I'| \geq e^{0.2\sqrt{\log n}}$ then we are done. Otherwise we have $\alpha(G') \leq e^{0.2\sqrt{\log n}}$.

Next suppose that there is a subset $X \subseteq V'$ of size at least $n^{1/2}$ such that the induced subgraph $G'[X]$ contains no clique of order at least $e^{0.2\sqrt{\log n}}$. Then, by the above discussion, $q(G'[X]) \leq e^{0.2\sqrt{\log n}}$ and it follows easily from Lemma 3.5 that $G'[X]$ is $(e^{-0.2\sqrt{\log n}}, n^{0.2})$-balanced. Now Lemma 3.2 and Corollary 3.3 imply that $u(G) \geq u(G'[X]) \geq e^{\Omega(\sqrt{\log n})}$ and we are done again. Denote by $m$ the number of vertices in $G'$. Then $m \geq 13n/14$ and $q(G') \leq q(G) \leq n/14 \leq m/13$. In addition, we now may assume that every subset of vertices of $G'$ of order at least $2m^{1/2} > n^{1/2}$ contains a clique of size larger than $e^{0.2\sqrt{\log n}} > e^{0.2\sqrt{\log m}}$. Since $u(G) \geq u(G')$ it is enough to bound $u(G')$.

Now we start with $G'$ and delete repeatedly maximal sized cliques till we are left with less than $2m^{1/2}$ vertices. Let $W_1, \ldots, W_k$ be the deleted cliques and $w_1 \geq \ldots \geq w_k$ be their corresponding sizes. According to the above discussion $w_1 \leq m/13$ and $w_k \geq e^{0.2\sqrt{\log m}}$. Also, $\sum_{i=1}^{k} w_i \geq m - 2m^{1/2}$. If for all $1 \leq i \leq k - 5$ we have $w_{i+5}/w_i < 3/5$, then $w_j < (3/5)^i w_1$ for all $j \geq 5i$, and thus $\sum_{i=1}^{k} w_i \leq 5 \sum_{i=1}^{k/5} w_{5i} < 5 \sum_{i=1}^{k/5} (3/5)^i w_1 < 25w_1/2 \leq 25m/26$ - a contradiction. Hence we conclude that there is an $i_0$, $1 \leq i_0 \leq k - 5$, such that $w_j \geq (3/5)w_{i_0}$ for $i_0 + 1 \leq j \leq i_0 + 5$.

For a vertex $v \in W_j$ denote by $N_{W_i}(v)$ the set of neighbors of $v$ in $W_i, i \neq j$. First consider the case when for some $v$, both $N_{W_i}(v)$ and $W_i - N_{W_i}(v)$ have size at least $e^{0.01\sqrt{\log m}}$. Then for every two integers $0 \leq a \leq e^{0.01\sqrt{\log m}}$ and $0 \leq b \leq a - 1$ if we pick any $b$ vertices from $N_{W_i}(v)$ and $a - b$ vertices from $W_i - N_{W_i}(v)$, then together with $v$ we obtain a set which spans exactly $\binom{a}{2} + b$ edges. This immediately implies that $u(G) \geq e^{0.01\sqrt{\log m}} = e^{\Omega(\sqrt{\log n})}$. Hence we can assume that for all $i \neq j$, the degree $d_{W_i}(v)$ of every vertex $v \in W_j$ is either less than $e^{0.01\sqrt{\log m}}$ or larger than $|W_i| - e^{0.01\sqrt{\log m}}$.

Denote by $e(W_i, W_j)$ the number of edges between $W_i$ and $W_j$. Let $X \subseteq W_j$ be the set of all vertices with degree in $W_i$ less than $e^{0.01\sqrt{\log m}}$ and suppose that $|X| \geq e^{0.01\sqrt{\log m}}$. Let $X'$ be a subset of $X$ of size $e^{0.01\sqrt{\log m}}$ and let $Y'$ be a set of all neighbors of vertices from $X'$ in $W_i$. Clearly $|Y'| \leq e^{0.01\sqrt{\log m}}|X'| = e^{0.02\sqrt{\log m}}$. Consider a vertex $u \in W_i - Y'$. This vertex is not adjacent to all vertices of $X'$ and hence has at least $e^{0.01\sqrt{\log m}}$ non-neighbors in $W_j$. Thus, by the above discussion, we know that $d_{W_j}(u) \leq e^{0.01\sqrt{\log m}}$. Using the fact that $|W_i|, |W_j| \geq e^{0.2\sqrt{\log m}}$ we can conclude that

$$e(W_i, W_j) \leq e^{0.01\sqrt{\log m}}|W_i - Y'| + |Y'||W_j| \leq e^{0.01\sqrt{\log m}}|W_i| + e^{0.02\sqrt{\log m}}|W_j| \leq e^{-0.1\sqrt{\log m}}|W_i||W_j|.$$

On the other hand, if $|X| \leq e^{0.01\sqrt{\log m}}$ then all the vertices in $W_j - X$ have at least $|W_i| - e^{0.01\sqrt{\log m}}$ neighbors in $|W_i|$. This, together with the fact that $|W_i|, |W_j| \geq e^{0.2\sqrt{\log m}}$, implies

$$e(W_i, W_j) \geq \left(|W_j| - e^{0.01\sqrt{\log m}}\right)\left(|W_i| - e^{0.01\sqrt{\log m}}\right) \geq \left(1 - e^{-0.1\sqrt{\log m}}\right)|W_i||W_j|.$$

Now we can assume that for every $i_0 \leq j_1 < j_2 \leq i_0 + 5$ either

$$\frac{e(W_{j_1}, W_{j_2})}{|W_{j_1}||W_{j_2}|} \leq e^{-0.1\sqrt{\log m}} \qquad \text{or} \qquad \frac{e(W_{j_1}, W_{j_2})}{|W_{j_1}||W_{j_2}|} \geq 1 - e^{-0.1\sqrt{\log m}}.$$

Then, by the well known, simple fact that the diagonal Ramsey number $R(3,3)$ is 6 the above set of cliques $W_{i_0}, \ldots, W_{i_0+5}$ contains a triple $W_{j_1}, W_{j_2}, W_{j_3}, i_0 \leq j_1 < j_2 < j_3 \leq i_0 + 5$, in which either all the pairs satisfy the first inequality or all the pairs satisfy the second inequality. To finish the proof of the theorem it is enough to consider the following two cases.

**Case 1.** For every $1 \leq i_1 < i_2 \leq 3$, $\frac{e(W_{j_{i_1}}, W_{j_{i_2}})}{|W_{j_{i_1}}||W_{j_{i_2}}|} \leq e^{-0.1\sqrt{\log m}}$. Then an easy counting shows that there are at least $|W_{j_1}|/3$ vertices in $W_{j_1}$ with at most $3e^{-0.1\sqrt{\log m}}|W_{j_i}|$ neighbors in $W_{j_i}, i = 2, 3$. Let $X_1$ be any set of such vertices of size $e^{0.05\sqrt{\log m}}$ and let $W'_{j_2}, W'_{j_3}$ be the sets of all neighbors of vertices from $X_1$ in $W_{j_2}$ and $W_{j_3}$, respectively. Then by definition, $|W'_{j_i}| \leq 3e^{-0.1\sqrt{\log m}}|W_{j_i}||X_1| = 3e^{-0.05\sqrt{\log m}}|W_{j_i}|$. Also, since $e(W_{j_2}, W_{j_3}) \leq e^{-0.1\sqrt{\log m}}|W_{j_2}||W_{j_3}|$, there exist at least $2|W_{j_2}|/3$ vertices in $W_{j_2}$ with at most $3e^{-0.1\sqrt{\log m}}|W_{j_3}|$ neighbors in $W_{j_3}$. Let $X_2$ be a set of such vertices of size $e^{0.05\sqrt{\log m}}$ which is disjoint from $W'_{j_2}$. Note that the existence of $X_2$ follows from the fact that $2|W_{j_2}|/3 \gg |W'_{j_2}|$. Let $W''_{j_3}$ be the set of all neighbors of vertices from $X_2$ in $W_{j_3}$. Then $|W''_{j_3}| \leq 3e^{-0.05\sqrt{\log m}}|W_{j_3}|$ and hence $|W_{j_3}| \gg |W'_{j_3}| + |W''_{j_3}|$. Finally let $X_3$ be any subset of $W_{j_3}$ of size $e^{0.05\sqrt{\log m}}$ which is disjoint from $W'_{j_3} \cup W''_{j_3}$. Then $\bigcup_i X_i$ induces subgraph $H$ of $G$, which is disjoint union of cliques of order $e^{0.05\sqrt{\log m}}$ with no edges between them. Clearly, every number of the form $\binom{x}{2} + \binom{y}{2} + \binom{z}{2}, 0 \leq x, y, z \leq e^{0.05\sqrt{\log m}}$ can be obtained as a number of edges in the induced subgraph of $H$. Therefore, by Proposition 4.3, $u(G) \geq u(H) \geq e^{0.05\sqrt{\log m}} = e^{\Omega(\sqrt{\log n})}$.

**Case 2.** For every $1 \leq i_1 < i_2 \leq 3$, $\frac{e(W_{j_{i_1}}, W_{j_{i_2}})}{|W_{j_{i_1}}||W_{j_{i_2}}|} \geq 1 - e^{-0.1\sqrt{\log m}}$. Denote by $H$ the subgraph of $G'$ induced by the union of sets $W_{j_1}, W_{j_2}, W_{j_3}$. Let $\overline{H}$ be the complement of $H$ and let $l$ be the number of vertices in $\overline{H}$. Note that by our construction the set $W_{j_1}$ corresponds to a largest independent set in $\overline{H}$ and that $|W_{j_2}|, |W_{j_3}| \geq (3/5)|W_{j_1}|$. Therefore

$$\alpha(\overline{H}) \leq \frac{|W_{j_1}|}{\sum_{i=1}^{3} |W_{j_i}|} l \leq \frac{5}{11} l.$$

We also have that the number of edges in $\overline{H}$ is bounded by $e^{-0.1\sqrt{\log m}}\binom{l}{2}$. This implies that there are at least $21l/22$ vertices in $\overline{H}$ with degree at most $22le^{-0.1\sqrt{\log m}}$. Denote this set by $X_1$ and consider the following process. Let $C_1$ be a shortest cycle in the induced subgraph $\overline{H}[X_1]$. Note that such cycle must span no other edges of $\overline{H}$. The existence of $C_1$ follows from the fact that if $X_1$ spans an acyclic graph in $\overline{H}$ then it should contain an independent set of size at least $|X_1|/2 \geq 21l/44 > 5l/11 \geq \alpha(\overline{H})$, contradiction. If the length of $C_1$ is at least $3e^{0.01\sqrt{\log m}}$, then in particular $\overline{H}$ contains an induced path of length $3e^{0.01\sqrt{\log m}} - 1$ and we stop. Otherwise, $|C_1| \leq 3e^{0.01\sqrt{\log m}}$. Let $X_2$ be the set of vertices of $X_1$ not adjacent to any vertex of $C_1$. Since all vertices in $X_1$ have degree at most $22le^{-0.1\sqrt{\log m}}$ and $|X_1| \geq 21l/22$ we obtain that

$$|X_2| \geq |X_1| - 22le^{-0.1\sqrt{\log m}}|C_1| \geq |X_1| - 22le^{-0.09\sqrt{\log m}} \geq \left(1 - e^{-0.05\sqrt{\log m}}\right)|X_1|.$$

We continue this process for $k = e^{0.01\sqrt{\log m}}$ steps. At step $i$, let $C_i$ be a shortest cycle in the induced subgraph $\overline{H}[X_i]$ and let $X_{i+1}$ be the set of all the vertices in $X_i$ not adjacent to any vertex in $C_i$. We assume that $|C_i| \le 3e^{0.01\sqrt{\log m}}$, otherwise we found an induced path of length $3e^{0.01\sqrt{\log m}} - 1$ and we can stop. Similarly as above one can show that for every $i$,

$$|X_i| \ge \left(1 - e^{-0.05\sqrt{\log m}}\right)|X_{i-1}| \ge \left(1 - e^{-0.05\sqrt{\log m}}\right)^i |X_1| = (1 + o(1))|X_1| > 2\alpha(\overline{H}).$$

Therefore the same argument as in the case of $C_1$ shows that a cycle $C_i$ has to exist. In the end of the process we either constructed an induced path of length $3k - 1$ or a disjoint union of induced $k$ cycles with no edges between them. In both cases this graph satisfies the assertion of Lemma 4.2. Therefore for any integers $0 \le r \le k = e^{0.01\sqrt{\log m}}$ and $0 \le t \le r - 2$, $\overline{H}$ contains an induced subgraph on $r$ vertices with exactly $t$ edges. This implies that the same set of vertices spans $\binom{r}{2} - t$ edges in $H$. Since any number $0 \le y \le e^{0.01\sqrt{\log m}}$ can be written in this form we conclude that $u(G) \ge u(H) \ge e^{0.01\sqrt{\log m}} = e^{\Omega(\sqrt{\log n})}$. This completes the proof of the theorem. $\qquad\square$

# 5 Concluding remarks

- There are several known results that show that graphs with relatively small trivial induced subgraphs have many distinct induced subgraphs of a certain type. In [11] it is shown that for every positive $c_1$ there is a positive $c_2$ such that every graph $G$ on $n$ vertices for which $q(G) \le c_1 \log n$ contains every graph on $c_2 \log n$ vertices as an induced subgraph. In [1] it is shown that for every small $\epsilon > 0$, every graph $G$ on $n$ vertices for which $q(G) \le (1 - 4\epsilon)n$ has at least $\epsilon n^2$ distinct induced subgraphs, thus verifying a conjecture of Erdős and Hajnal. In [12] it is proved that for every positive $c_1$ there is a positive $c_2$ such that every graph $G$ on $n$ vertices for which $q(G) \le c_1 \log n$ has at least $2^{c_2 n}$ distinct induced subgraphs, thus verifying a conjecture of Erdős and Rényi. The assertions of conjectures 1.1 and 1.3, as well as our results here, have a similar flavour: if $q(G)$ is small, then $G$ has induced subgraphs with all possible number of edges in a certain range.

- The following common strengthening of conjectures 1.1 and 1.3 may be true:

  **Conjecture 5.1** *Every graph $G = (V, E)$ on $n$ vertices for which $q(G) \le n/4$ satisfies $u(G) \ge \Omega(|E|)$.*

  By Lemma 3.5 the assertion of this conjecture implies that of Conjecture 1.1, and clearly it implies that of Conjecture 1.3.

- It may be interesting to have more information on the set of all possible triples $(n, q, u)$ such that there exists a graph $G$ on $n$ vertices with $q(G) = q$ and $u(G) = u$. Our results here show that if $q$ is relatively small, than $u$ must be large. Note that the union of two vertex disjoint cliques of size $n/3$ each, and $n/3$ isolated vertices show that the triple $(n, n/3 + O(1), 4)$ is

possible. It may be interesting to decide how large $u(G)$ must be for any graph $G$ on $n$ vertices satisfying $q(G) < n/(3 + \epsilon)$.

# References

[1] N. Alon and B. Bollobás, Graphs with a small number of distinct induced subgraphs, *Discrete Math.* 75 (1989), 23-30.

[2] N. Alon, M. Krivelevich and B. Sudakov, Turán numbers of bipartite graphs and related Ramsey-type questions, submitted.

[3] N. Calkin, A. Frieze and B. McKay, On subgraph sizes in random graphs, *Combinatorics, Probability and Computing* 1 (1992), 123–134.

[4] F. Chung and R. Graham, **Erdős on graphs. His legacy of unsolved problems**, A K Peters Ltd., Wellesley, MA, 1998.

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

[6] P. Erdős, Some of my favorite problems in number theory, combinatorics, and geometry, *Resenhas* 2 (1995), 165–186.

[7] P. Erdős, Some recent problems and results in graph theory, *Discrete Math.* 164 (1997), 81–85.

[8] P. Erdős and E. Szemerédi, On a Ramsey type theorem. *Periodica Math. Hungar.* 2 (1972), 295–299.

[9] D. Flath, **Introduction to number theory**, Wiley, New York, 1989.

[10] R. Graham, B. Rothschild and J. Spencer, **Ramsey theory**, 2nd edition, Wiley, New York, 1990.

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

[12] S. Shelah, Erdős and Rényi conjecture, *J. Combinatorial Theory, Ser. A* 82 (1998), 179-185.