Induced subgraphs with many repeated degrees

Yair Caro [*]   Raphael Yuster [†]

Abstract

Erdős, Fajtlowicz and Staton asked for the least integer $f(k)$ such that every graph with more than $f(k)$ vertices has an induced regular subgraph with at least k vertices. Here we consider the following relaxed notions. Let $g(k)$ be the least integer such that every graph with more than $g(k)$ vertices has an induced subgraph with at least k repeated degrees and let $h(k)$ be the least integer such that every graph with more than $h(k)$ vertices has an induced subgraph with at least k maximum degree vertices. We obtain polynomial lower bounds for $h(k)$ and $g(k)$ and nontrivial linear upper bounds when the host graph has bounded maximum degree.

1   Introduction

We consider undirected simple and finite graphs. Erdős, Fajtlowicz and Staton (c.f. [7], page 85) asked for the order of magnitude of the least integer $f(k)$ such that every graph with more than $f(k)$ vertices has an induced regular subgraph with at least k vertices. Clearly $f(k) \leq R(k) - 1$ where $R(k)$ is the diagonal Ramsey number, but already small values of k suggest that it is smaller. They conjectured that $f(k)$ is less than exponential in k and this major open problem is still unresolved. The best lower bound for $f(k)$ is polynomial in k. Alon, Krivelevich and Sudakov [1] proved that $f(k) = \Omega(k^2/\sqrt{\log k})$ improving an earlier result of Bollobás (c.f. [7]) who proved $f(k) = \Omega(k^{2-\epsilon})$ for every $\epsilon > 0$.

In this paper we consider a version of the aforementioned question which relaxes the regularity requirement. For a graph G, let $rep(G)$ be the maximum multiplicity of a mode (i.e. most common value) of the degree sequence of G. This parameter, also called the *repetition number* has been recently studied by several researchers, see [4, 5]. Likewise, let $maxrep(G)$ be the number of vertices with maximum degree in G; this has recently been studied in [6, 9]. Let, therefore, $g(k)$ be the least integer such that every graph with more than $g(k)$ vertices has an induced subgraph with repetition number at least k and let $h(k)$ be the least integer such that every graph with more

than $h(k)$ vertices has an induced subgraph with at least k maximum degree vertices. We clearly have

$$\Theta(k) \leq g(k) \leq h(k) \leq f(k) \leq R(k) - 1$$

and small values of k show that all of these parameters are separated.

Our first main contribution in this paper is a polynomial lower bound for $g(k)$ and hence for $h(k)$.

Theorem 1. *There exists an absolute constant c such that $g(k) \geq ck^{3/2}/(\log k)^{1/2}$.*

Determining these functions even with the further restriction that the graph in question has bounded maximum degree also seems challenging. Let $g(k, d)$ denote the least integer such that every graph with maximum degree at most d and more than $g(k, d)$ vertices has an induced subgraph H with $rep(H) \geq k$. Similarly define $h(k, d)$ and $f(k, d)$. Again, we have here that for every nonnegative integer d, $g(k, d) \leq h(k, d) \leq f(k, d) \leq (k-1)(d+1)$ as in every graph with $(k-1)(d+1) + 1$ vertices and with maximum degree d there is an independent set of size k. Once again, it is not difficult to construct some small examples of distinct pairs (k, d) showing that these 2-valued functions are separated.

For very small fixed d, the values of $f(k, d)$ and $g(k, d)$ are easy to obtain. For example, it clearly holds that $g(k, 0) = h(k, 0) = f(k, 0) = k - 1$. It is also easy to show that $h(k, 1) = g(k, 1) = f(k, 1) = \lfloor 1.5(k - 1) \rfloor$. For $d = 2$ things are only slightly more involved but all parameters can be precisely determined as the host graphs in this case are just disjoint unions of cycles, paths, and isolated vertices. For example, it is not difficult to prove that $h(k, 2) = 2(k - 1)$ for all odd k and $h(k, 2) = 2k - 3$ for all even k as was shown in [3]. However, already for $d = 3$, even the asymptotic behaviors of $h(k, 3)$ as well as $g(k, 3)$ and $f(k, 3)$ seem elusive. It is not difficult to show that for every fixed d, each of $h(k, d)/k$, $g(k, d)/k$, $f(k, d)/k$ has a limit. The following theorem provides nontrivial upper and lower bounds for the limits of the first two, while Proposition 4.2 provides an upper bound for the limit of the third.

Theorem 2. *There exists an absolute constant $c > 0$ such that for every $d \geq 2$,*

$$c\left(\frac{d}{\log d}\right)^{1/3} \leq \lim_{k \to \infty} \frac{g(k, d)}{k} \leq \lim_{k \to \infty} \frac{h(k, d)}{k} \leq \frac{d}{2} + 1 \ .$$

Regarding the first nontrivial case $d = 3$ we obtain the following more specific bounds.

Theorem 3.
$$\frac{53}{24} \leq \lim_{k \to \infty} \frac{h(k, 3)}{k} \leq \frac{5}{2} \ , \qquad \frac{13}{6} \leq \lim_{k \to \infty} \frac{g(k, 3)}{k} \leq \frac{12}{5} \ .$$

In the next section we prove the general lower bound, namely Theorem 1. Section 3 considers bounded degree graphs where we prove Theorems 2 and 3. The final section contains some concluding remarks and open problems.

2 A polynomial lower bound for repetition of induced subgraphs

Proof of Theorem 1. Our construction is probabilistic and follows the non-uniform random graph model as in [1]. Let $\overline{p} = (p_1, \ldots, p_n)$ where $p_i = (1-\epsilon)/\sqrt{2} + i(4+\sqrt{2})\epsilon/(7n)$ and let $G(n, \overline{p})$ be the probability space of graphs on vertex set $[n]$ where the pair (i, j) is an edge with probability $p_i p_j$ independently of all other pairs. In this proof we will use $\epsilon = 1 - \sqrt{2}/4$ so that $p_i = 1/4 + i/(2n)$.

Let $C > 0$ be an absolute constant to be chosen later and let $k = 3Cn^{2/3}(\ln n)^{1/3}$. We will assume that k is an integer multiple of 3 as this does not affect the asymptotic claim and also assume that n is sufficiently large to satisfy the claimed inequalities. For every $k \leq t \leq n$ we will prove that the probability that $G \sim G(n, \overline{p})$ has an induced subgraph H with t vertices having $rep(H) \geq k$ is less than $1/n$. Hence, by the union bound it will follow that with positive probability, a graph $G \sim G(n, \overline{p})$ has no subgraph with repetition number at least k and therefore $g(k) \geq n = ck^{3/2}/(\log k)^{1/2}$ for a suitable absolute constant c, hence Theorem 1 follows.

Fix some t with $k \leq t \leq n$ and fix some $T \subseteq [n]$ with $t = |T|$. Let $H = G[T]$ where $G \sim G(n, \overline{p})$. We will prove that the probability that $rep(H) \geq k$ is less than $(n\binom{n}{t})^{-1}$ and as there are only $\binom{n}{t}$ choices for T, this will prove that the probability of G having an induced subgraph H with t vertices having $rep(H) \geq k$ is less than $1/n$, as claimed.

Fix some $R \subseteq T$ with $|R| = k$. We will prove that the probability of all the vertices of R having the same degree in H is less than $(\binom{t}{k})^{-1}(n\binom{n}{t})^{-1}$ hence it will imply the probability that $rep(H) \geq k$ is less than $(n\binom{n}{t})^{-1}$, as claimed.

Recall that the vertices of R are a subset of the total order $[n]$ hence let $R = \{v_1, \ldots, v_k\}$ where $v_i < v_{i+1}$ for $i = 1, \ldots, k-1$. Let A denote the first $k/3$ vertices of R and let B denote the last $k/3$ vertices of R. Our goal is to upper bound the probability that the sum of the degrees in H of the vertices in A equals the sum of the degrees in H of the vertices in B. This probability clearly upper bounds the probability that all the vertices of R have the same degree in H.

For a pair of vertices (i, j), let $X_{i,j}$ denote the indicator random variable which equals 1 if (i, j) is an edge. Let X_A denote the sum of the degrees of A in H. Let Y_A denote the sum of the indicator random variables corresponding to pairs (i, j) such that $i \in A$ and $j \in T \setminus (A \cup B)$ and let Z_A denote the sum of the indicator random variables corresponding to pairs (i, j) such that both i, j are in A. Analogously define X_B, Y_B, Z_B. Notice that $X_B - X_A = Y_B + 2Z_B - Y_A - 2Z_A$ (indicator random variables corresponding to pairs with one endpoint in A and another in B are canceled out in the difference $X_B - X_A$). Also notice that all the $2\binom{k/3}{2} + 2(k/3)(t - 2k/3)$ indicator variables involved in forming Y_A, Y_B, Z_A, Z_B are independent.

We next estimate $\mathbb{E}[X_B - X_A]$. First observe that $\mathbb{E}[Z_B - Z_A] \geq 0$ since a pair of vertices in B have a higher chance to be an edge than a pair of vertices in A. Therefore,

$$\mathbb{E}[X_B - X_A] = \mathbb{E}[Y_B - Y_A] + 2\mathbb{E}[Z_B - Z_A] \geq \mathbb{E}[Y_B - Y_A] .$$

For every vertex $j \in T \setminus (A \cup B)$ and for every $i \in A$ and $i' \in B$ the probability of the pair i, j to be an edge is less than the probability of the pair (i', j) to be an edge by at least $k/(24n)$ since

$$\left(\frac{1}{4} + \frac{i'}{2n}\right)\left(\frac{1}{4} + \frac{j}{2n}\right) - \left(\frac{1}{4} + \frac{i}{2n}\right)\left(\frac{1}{4} + \frac{j}{2n}\right) \geq \left(\frac{i' - i}{8n}\right) > \frac{k}{24n} .$$

It therefore follows that $\mathbb{E}[Y_B - Y_A] \geq (k/3)(t - 2k/3)\frac{k}{24n}$. Consequently, since $t \geq k$ we obtain

$$\mathbb{E}[X_B - X_A] \geq \mathbb{E}[Y_B - Y_A] \geq \frac{k}{3}\left(t - \frac{2k}{3}\right)\frac{k}{24n} \geq \frac{tk^2}{216n} .$$

Let us next consider the random variable $S = \frac{1}{2}\mathbb{E}[X_B - X_A] - \frac{1}{2}(X_B - X_A)$. Then $E[S] = 0$ and S is the sum of $q = 2\binom{k/3}{2} + 2(k/3)(t - 2k/3)$ independent random variables where each of these random variables takes only two values and those two values are at most 1 apart. Then by a large deviation result of Chernoff (see [2], Theorem A.1.18),

$$\Pr[S > a] < e^{-\frac{2a^2}{q}} .$$

We therefore obtain:

$$\Pr[X_B = X_A] = \Pr\left[S = \frac{\mathbb{E}[X_B - X_A]}{2}\right]$$
$$\leq \Pr\left[S > \frac{\mathbb{E}[X_B - X_A]}{3}\right]$$
$$\leq \exp\left(-\frac{2\left(\frac{\mathbb{E}[X_B - X_A]}{3}\right)^2}{q}\right)$$
$$\leq \exp\left(-\frac{2t^2k^4}{9 \cdot 216^2 n^2 q}\right)$$
$$\leq \exp\left(-\frac{tk^3}{9 \cdot 216^2 n^2}\right) .$$

Recalling that $k = 3Cn^{2/3}(\ln n)^{1/3}$ we obtain from the last inequality that for a sufficiently large

absolute constant C,

$$\Pr[X_B = X_A] \leq \exp\left(-\frac{tk^3}{9 \cdot 216^2 n^2}\right)$$
$$\leq \exp\left(-2t \ln n\right)$$
$$= \frac{1}{n^{2t}}$$
$$< \frac{1}{n\binom{t}{k}\binom{n}{t}}$$

as required. $\qquad\square$

3   Repetition in induced subgraphs of bounded degree graphs

We first observe that for every fixed d, the sequences $g(k,d)/k$, $h(k,d)/k$, $f(k,d)/k$ have a limit. First observe that as mentioned in the introduction, all of these sequences are bounded from above by $d+1$. We next show that $g(mk,d) \geq m \cdot g(k,d)$, $h(mk,d) \geq m \cdot h(k,d)$, $f(mk,d) \geq m \cdot f(k,d)$ holds for any positive integer m implying that the claimed limits exist in all cases. Consider a graph G with maximum degree at most d and with $g(k,d)$ vertices for which every induced subgraph H has $rep(H) < k$. Take m vertex-disjoint copies of G, thereby obtaining a graph with maximum degree at most d and with $m \cdot g(k,d)$ vertices for which every induced subgraph H has $rep(H) < mk$. Thus, by its definition, $g(mk,d) \geq m \cdot g(k,d)$. An identical argument holds for $h(k,d)$ and $f(k,d)$.

The proofs of Theorems 2 and 3 follow directly from the proofs of the following Lemmas and proposition.

Proposition 3.1. *There exists an absolute constant $c > 0$ such that for all $d \geq 2$,*

$$c\left(\frac{d}{\log d}\right)^{1/3} \leq \lim_{k\to\infty} \frac{g(k,d)}{k} \ .$$

Proof. Let $d \geq 2$. By Theorem 1, there exists an absolute constant $C > 0$ such that there are graphs with at most d vertices and for which every induced subgraph has repetition less than $Cd^{2/3}(\log d)^{1/3}$. Fixing d, let G_d be such a graph. Now take $k/(Cd^{2/3}(\log d)^{1/3})$ pairwise vertex-disjoint copies of G_d. The resulting graph, which clearly has maximum degree less than d, has $kd^{1/3}/(C(\log d)^{1/3})$ vertices and every induced subgraph has repetition less than k. It follows that $g(k,d)/k \geq d^{1/3}/(C(\log d)^{1/3})$ and the proposition follows. $\qquad\square$

Lemma 3.2. *For infinitely many k it holds that $h(k,3) \geq (53/24)(k-1)$.*

Proof. Consider the following set of graphs $\{K_1, H_1, H_2, H_3\}$ where K_1 is the isolated vertex and H_1, H_2, H_3 are the graphs shown in Figure 1. Observe that each of them has maximum degree at

5

most 3. Now, suppose $k-1$ is a multiple of 24. Construct a graph G_k as follows. Take $(7/24)(k-1)$ disjoint copies of K_1, $(1/24)(k-1)$ disjoint copies of H_1, $(1/24)(k-1)$ disjoint copies of H_2 and $(1/6)(k-1)$ disjoint copies of H_3. The graph G_k has

$$1 \cdot \frac{7}{24}(k-1) + 6 \cdot \frac{1}{24}(k-1) + 8 \cdot \frac{1}{24}(k-1) + 8 \cdot \frac{1}{6}(k-1) = \frac{53}{24}(k-1)$$

vertices. It is not hard to check that for an induced subgraph H of G_k we have $maxrep(H) \leq k-1$ as follows. A subgraph with maximum degree 0 can be obtained from of all copies of K_1, two vertices from a copy of H_1, three vertices from a copy of H_2 and three vertices from a copy of H_3, yielding a total of $k-1$. A subgraph with maximum degree 1 can be obtained from 4 vertices from each copy of H_1, H_2, H_3, yielding a total of $k-1$. A subgraph with maximum degree 2 can be obtained from $3, 5, 4$ vertices from each copy of H_1, H_2, H_3 respectively, yielding a total of $k-1$. A subgraph with maximum degree 3 can be obtained from $2, 2, 5$ vertices from each copy of H_1, H_2, H_3 respectively, yielding a total of $k-1$. □
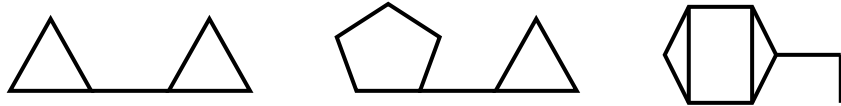

Figure 1: The graphs H_1, H_2, H_3 of Lemma 3.2 shown from left to right.

The reasoning behind the construction of Lemma 3.2 is as follows. A close to optimal construction for $h(k, 3)$ is a disjoint union of bounded size components since we have observed in the beginning of this section that a lower bound for $h(2k, 3)$ is obtained by taking two copies of an optimal construction for $h(k, 3)$ and that the sequence divided by k converges. So, given a positive integer constant r, we would like to optimize over all graphs G consisting of components of size at most r each and maximum degree at most 3. Denote by $D(r, 3)$ the set of all connected graphs with at most r vertices and maximum degree at most 3. If we can generate $D(r, 3)$, we can use linear programming to decide upon the optimal ratio of components of each possible element belonging to $D(r, 3)$. Construct a linear programming instance as follows. For each $H \in D(r, 3)$ define a nonnegative variable x_H. Now maximize the sum $|H|x_H$ under the following constraints, one constraint for each $p = 0, 1, 2, 3$. Let $c(H, p)$ denote the maximum number of vertices of maximum degree p in an induced subgraph of H. Then we require that the sum $c(H, p)x_H$ is at most $k - 1$ (or one can normalize dividing by $k - 1$). Now, if r is too large, we do not have enough computing power to determine $D(r, 3)$. We have, however, determined $D(10, 3)$, and have run the aforementioned linear

program [1], the result of which is the construction of Lemma 3.2 which yielded, after normalization $x_{K_1} = 7/24$, $x_{H_1} = 1/24$, $x_{H_2} = 1/24$ $x_{H_3} = 1/6$. The remaining variable are zero. In particular, this means that in any improved construction, if exists, one would need to use components of size at least 11.

Observe that the construction of Lemma 3.2 fails as a lower bound construction for $g(k, 3)$ since there is an induced subgraph (of maximum degree 3) where the degree 2 appears more than $k-1$ times.

Lemma 3.3. *For infinitely many k it holds that $g(k, 3) \geq (13/6)(k - 1)$.*

Proof. Consider the following set of graphs $\{K_1, K_2, P_4, Q\}$ where P_4 is the path on four vertices and Q is the complement of C_6. Observe that each of them has maximum degree at most 3. Now, suppose $k - 1$ is a multiple of 6. Construct a graph G_k as follows. Take $(k - 1)/6$ disjoint copies of each of K_1, K_2, P_4, Q. The graph G_k has

$$1 \cdot \frac{1}{6}(k - 1) + 2 \cdot \frac{1}{6}(k - 1) + 4 \cdot \frac{1}{6}(k - 1) + 6 \cdot \frac{1}{6}(k - 1) = \frac{13}{6}(k - 1)$$

vertices. It is not hard to check that for an induced subgraph H of G_k we have $rep(H) \leq k - 1$ as follows. The degree 0 can be repeated at most $1, 1, 2, 2$ times in an induced subgraph of K_1, K_2, P_4, Q respectively. The degree 1 can be repeated at most $0, 2, 2, 2$ times in an induced subgraph of K_1, K_2, P_4, Q respectively. The degree 2 can be repeated at most $0, 0, 2, 4$ times in an induced subgraph of K_1, K_2, P_4, Q respectively. The degree 3 can be repeated at most $0, 0, 0, 6$ times in an induced subgraph of K_1, K_2, P_4, Q respectively. In all four cases, the yielded total is $k - 1$. $\square$

We now turn to the upper bounds. Our next lemma implies, in particular, $\lim_{k\to\infty} \frac{h(k,3)}{k} \leq \frac{5}{2}$.

Lemma 3.4. $h(k, d) \leq (k - 1)(d/2 + 1)$.

Proof. We prove the lemma by induction on d where the base case $d = 0$ is trivial. Suppose G is a graph with $n > (k-1)(d/2+1)$ vertices and with maximum degree at most d. We must show that there is an induced subgraph H with $maxrep(H) \geq k$. If G has k or more vertices of degree d, we are done as we can take $H = G$. So, assume otherwise. This means that the sum of the degrees of the vertices of G is at most $(d - 1)n + k - 1$ so G has at most $n(d - 1)/2 + (k - 1)/2$ edges.

We perform the following process which has $d - 1$ stages. At the beginning of stage t, we have an induced subgraph of G denoted by G_{d+1-t} with maximum degree at most d. For $t = 1$ set $G_d = G$. Stage t consists of repeatedly removing vertices of degree $d + 1 - t$ form G_{d+1-t} until we remain with a subgraph with maximum degree at most $d - t$ denoted by G_{d-t}. Let x_{d+1-t} be such that the number of vertices removed in stage t is $x_{d+1-t}(k - 1)$ and observe that we may assume

[1]Soure code of our program is available at `https://www.dropbox.com/s/27acszyntajyh6i/Induced%20subgraphs%20with%20many%20max%20degree%20vertices.zip?dl=0`

that $0 \le x_i \le 1$ for $i = 2, \ldots, d$ as otherwise we are trivially done. When the process ends after stage $d - 1$, the remaining graph G_1 has maximum degree at most 1.

The number of vertices of G_1 is

$$n - (k-1) \sum_{i=2}^{d} x_i .$$

The number of edges of G_1 is

$$|E(G)| - (k-1) \sum_{i=2}^{d} i x_i \le n(d-1)/2 + (k-1)/2 - (k-1) \sum_{i=2}^{d} i x_i .$$

Now suppose that for some $2 \le i \le d$ we have $\sum_{j=i}^{d} x_j < n/(k-1) - (i+1)/2$. In this case we have that the number of vertices of G_{i-1} is

$$n - (k-1) \sum_{j=i}^{d} x_j > (k-1) \left(\frac{i-1}{2} + 1 \right)$$

which implies by the induction hypothesis that G_{i-1} has an induced subgraph H with $maxrep(H) \ge k$. So, we may assume that for all $2 \le i \le d$ it holds that $\sum_{j=i}^{d} x_j \ge n/(k-1) - (i+1)/2$. In particular, this means that

$$\sum_{i=2}^{d} (i-1) x_i \ge \frac{n(d-1)}{k-1} - (d+2)(d+1)/4 + 1.5. \tag{1}$$

Returning now to G_1, it has an independent set of size its number of vertices minus its number of edges, namely, using (1) of size at least

$$\left(n - (k-1) \sum_{i=2}^{d} x_i \right) - \left(n(d-1)/2 + (k-1)/2 - (k-1) \sum_{i=2}^{d} i x_i \right) \tag{2}$$

$$= n \left(1 - \frac{d-1}{2} \right) - \frac{k-1}{2} + (k-1) \sum_{i=2}^{d} (i-1) x_i$$

$$\ge n \left(1 - \frac{d-1}{2} \right) - \frac{k-1}{2} + (k-1) \left(\frac{n(d-1)}{k-1} - \frac{(d+2)(d+1)}{4} + 1.5 \right)$$

$$= n \left(1 + \frac{d-1}{2} \right) + (k-1) \left(1 - \frac{(d+2)(d+1)}{4} \right)$$

$$> (k-1) \left[\left(\frac{d}{2} + 1 \right) \left(1 + \frac{d-1}{2} \right) + 1 - \frac{(d+2)(d+1)}{4} \right]$$

$$= k - 1 .$$

So, there is an independent set of size at least k in G which is, in particular an induced subgraph H with $maxrep(H) \geq k$. □

We can slightly modify the proof of Lemma 3.4 to obtain an upper bound for $g(k,d)$. This upper bound gives, in particular, that $\lim_{k\to\infty} \frac{g(k,3)}{k} \leq \frac{12}{5}$.

Lemma 3.5. $g(k,d) \leq (k-1)(2d+6)/5$ for $d \geq 2$ and $g(k,1) \leq 1.5(k-1)$.

Proof. We prove the lemma by induction on d where the base case $d=1$ is an easy exercise. Assume now that $d \geq 2$ and suppose G is a graph with $n > (k-1)(2d+6)/5$ vertices and with maximum degree at most d. We must show that there is an induced subgraph H with $rep(H) \geq k$. If G has k or more vertices of the same degree we are done as we can take $H = G$. So, assume otherwise. This means that there are at most $k-1$ vertices of any given degree. Let $r \geq 2$ be any integer such that $r(k-1) \leq n$ then the sum of the degrees of the vertices of G is at most

$$dn - (k-1) - 2(k-1) - \cdots - (r-1)(k-1) - r(n - r(k-1)) = (d-r)n - (k-1)\binom{r}{2} + r^2(k-1) \ .$$

So the number of edges of G is at most

$$\frac{(d-r)n}{2} + (k-1)\left(\frac{r^2}{2} - \frac{r(r-1)}{4}\right) \ .$$

We perform the exact same process with $d-1$ stages as in Lemma 3.4 and using the same notations of G_i and x_i as in that lemma. When the process ends after stage $d-1$, the remaining graph G_1 has maximum degree at most 1. The number of vertices of G_1 is $n - (k-1)\sum_{i=2}^{d} x_i$ and the number of edges of G_1 is

$$|E(G)| - (k-1)\sum_{i=2}^{d} ix_i \leq \frac{(d-r)n}{2} + (k-1)\left(\frac{r^2}{2} - \frac{r(r-1)}{4}\right) - (k-1)\sum_{i=2}^{d} ix_i \ .$$

Now suppose that for some $3 \leq i \leq d$ we have $\sum_{j=i}^{d} x_j < n/(k-1) - (2i+4)/5$. In this case we have that the number of vertices of G_{i-1} is

$$n - (k-1)\sum_{j=i}^{d} x_j > (k-1)\left(\frac{2(i-1)+6}{5}\right)$$

which implies by the induction hypothesis that G_{i-1} has an induced subgraph H with $rep(H) \geq k$. So, we may assume that for all $3 \leq i \leq d$ it holds that $\sum_{j=i}^{d} x_j \geq n/(k-1) - (2i+4)/5$. In the

case $i = 2$, if we have $\sum_{j=2}^{d} x_j < n/(k-1) - 1.5$ then the number of vertices of G_1 is

$$n - (k-1) \sum_{j=2}^{d} x_j > 1.5(k-1)$$

which implies that G_1 has an induced subgraph H with $rep(H) \geq k$. So, we may assume that $\sum_{j=2}^{d} x_j \geq n/(k-1) - 1.5$. In particular, this means that

$$\sum_{i=2}^{d}(i-1)x_i \geq \frac{n(d-1)}{k-1} - \frac{4}{5}(d-1) - \frac{d(d+1)}{5} + 0.5.$$

Returning now to G_1, it has an independent set of size its number of vertices minus its number of edges, namely of size at least

$$\left(n - (k-1) \sum_{i=2}^{d} x_i \right) - \left(\frac{(d-r)n}{2} + (k-1)\left(\frac{r^2}{2} - \frac{r(r-1)}{4} \right) - (k-1) \sum_{i=2}^{d} ix_i \right)$$

$$\geq n\left(\frac{d+r}{2} \right) + (k-1)\left(\frac{r(r-1)}{4} - \frac{r^2}{2} + \frac{1}{2} - \frac{4}{5}(d-1) - \frac{d(d+1)}{5} \right)$$

$$> (k-1)\left[\left(\frac{2d+6}{5} \right)\left(\frac{d+r}{2} \right) + \frac{r(r-1)}{4} - \frac{r^2}{2} + \frac{1}{2} - \frac{4}{5}(d-1) - \frac{d(d+1)}{5} \right].$$

Now, the last inequality is at least $k-1$ if $r \geq 2$ and equals $k-1$ if $r = 2$. So, there is an independent set of size at least k in G which is, in particular an induced subgraph H with $rep(H) \geq k$. $\qquad \square$

4 Concluding remarks and open problems

Recall the important notion of quasi-random graphs defined by Chung, Graham, and Wilson [8]. We say that a graph G is quasi-random if for any subset of vertices $U \subseteq V(G)$ we have $e(U) = \frac{1}{2}\binom{|U|}{2} \pm o(n^2)$. In other words, it resembles the behavior of the random graph $G(n, \frac{1}{2})$ with respect to edge distributions. It is well-known that quasi-random graphs possess many (though not all) properties of a typical random graph. Observe that in our proof of Theorem 1 we have used $\epsilon = 1 - \sqrt{2}/4$ but the proof works equally well with any fixed $\epsilon > 0$ (this only affects the value of the constant c) or even when ϵ is a function of n tending slowly to zero (in which case we will still have $g(k) \geq ck^{3/2}/\log k$). But then observe that the distribution $G(n, \bar{p})$ in the proof is a quasi-random graph with very high probability, since each pair is an edge with probability $\frac{1}{2} \pm o_n(1)$. Consequently, this shows that there are quasi-random graphs with n vertices for which every induced subgraph has repetition at most $n^{2/3} \log n$. This should be compared to the result proved by Krivelevich, Sudakov, and Wormald [10] showing that with high probability a largest

induced *regular* subgraph of $G(n, \frac{1}{2})$ has about $n^{2/3}$ vertices. So, a random graph usually has a regular subgraph with about $n^{2/3}$ vertices. On the other hand, there are quasi-random graphs which do not contain an induced subgraph with repetition number much larger than $n^{2/3}$. Noticing this, it seems interesting to determine the typical maximum repetition of an induced subgraph of a random graph.

The functions $f(k, d)$, $g(k, d)$, $h(k, d)$ have the following sub-additive property.

Lemma 4.1. *Let $z \in \{f, g, h\}$, then $z(k, d_1) + z(k, d_2) \geq z(k, d_1 + d_2 + 1)$.*

Proof. By a result of Lovász [11], the vertices of a graph G with maximum degree $d_1 + d_2 + 1$ can be partitioned into two parts A_1 and A_2 where the maximum degree of $G[A_i]$ is at most d_i for $i = 1, 2$. Now, if G has more than $h(k, d_1) + h(k, d_2)$ vertices, there is some i for which $G[A_i]$ has more than $h(k, d_i)$ vertices, so by definition $G[A_i]$ has an induced subgraph H with $maxrep(H) \geq k$. The same argument holds for g and f. $\qquad\square$

While Lemma 4.1 is not strong enough to improve upon Lemmas 3.4 and 3.5, it is useful for obtaining the following upper bound:

Proposition 4.2.
$$\lim_{k \to \infty} \frac{f(k, d)}{k} \leq \frac{11d}{15} \ .$$

Proof. We prove that $f(k, 2) \leq 11(k-1)/5$. Let G be a graph with $n > 11(k-1)/5$ vertices and of maximum degree at most 2. Let t denote the number of vertices on all the cycles of G. If $t \geq k$ we are done. Otherwise, let G' denote the induced subgraph on the path components which consists of $n - t$ vertices. Let t_0 denote the size of an independent set of G' and let t_1 denote the number of vertices in a maximum induced matching of G'. It is straightforward to verify that $t_0 + t_1 \geq n - t$. Now, in each cycle of length r, the number of independent vertices plus the number of vertices in an induced matching is at least $4r/5$ where the worst case occurs for C_5. So, if s_0 denotes the size of an independent set of G and s_1 denotes the number of vertices in a maximum induced matching of G we have that $s_0 + s_1 \geq n - t + 4t/5$. So there is an induced regular subgraph of G of order at least $n/2 - t/10$. But
$$\frac{n}{2} - \frac{t}{10} > \frac{11(k-1)}{10} - \frac{k-1}{10} = k - 1 \ .$$
It is worth noting that the bound $11(k-1)/5$ is tight as for any k such that $k - 1$ is a multiple of 10 we can take $(k-1)/5$ copies of C_5 and $3(k-1)/10$ copies of P_4 and obtain a graph with $11(k-1)/5$ vertices and no induced regular subgraph on k vertices.

Now, as we have proved $f(k, 2) \leq 11(k-1)/5$, we can apply the sub-additive inequality of Lemma 4.1 and obtain (with room to spare) that $f(k, d) \leq (11/5)d(k-1)$ hence $f(k, d)/k \leq 11d/15$. As we have already shown in Section 3 that the limit exists, the upper bound $11d/15$ serves as an upper bound for the limit. $\qquad\square$

Finally, it would be interesting to obtain a sub-linear upper bound for the limit of $h(k,d)/k$ or improve the polynomial lower bound for $g(k)$ which would imply, by Proposition 3.1, an improved lower bound for the limit of $g(k,d)/k$.

Acknowledgment

References

[1] N. Alon, M. Krivelevich, and B. Sudakov. Large nearly regular induced subgraphs. *SIAM Journal on Discrete Mathematics*, 22(4):1325–1337, 2008.

[2] N. Alon and J. Spencer. *The probabilistic method.* John Wiley & Sons, 2004.

[3] Y. Caro, J. Lauri, and C. Zarb. Equating two maximum degrees. *Australasian Journal of Combinatorics*, 72(2):206–225, 2018.

[4] Y. Caro, A. Shapira, and R. Yuster. Forcing k-repetitions in degree sequences. *The Electronic Journal of Combinatorics*, 21(1), 2014. P1.24.

[5] Y. Caro and D. B. West. Repetition number of graphs. *The electronic journal of combinatorics*, 16(1), 2009. R7.

[6] Y. Caro and R. Yuster. Large induced subgraphs with equated maximum degree. *Discrete Mathematics*, 310(4):742–747, 2010.

[7] F. R. K. Chung and R. L. Graham. *Erdős on graphs: his legacy of unsolved problems.* A K Peters, 1998.

[8] F. R. K. Chung, R. L. Graham, and R. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[9] A. Girao and K. Popielarz. Large induced subgraphs with k vertices of almost maximum degree. *SIAM Journal on Discrete Mathematics*, 32(1):450–454, 2018.

[10] M. Krivelevich, B. Sudakov, and N. Wormald. Regular induced subgraphs of a random graph. *Random Structures & Algorithms*, 38(3):235–250, 2011.

[11] L. Lovász. On decomposition of graphs. *Studia Scientiarum Mathematicarum Hungarica*, 1(273):238, 1966.