Combinatorics, Probability and Computing (2013) 22, 874–884. © Cambridge University Press 2013
doi:10.1017/S0963548313000369

On Independent Sets in Graphs with Given Minimum Degree

HIU-FAI LAW[1] and COLIN McDIARMID[2]

[1]Mathematisches Seminar der Universität Hamburg, Germany
(e-mail: hiufai.law@gmail.com)
[2]Department of Statistics, Oxford University, UK
(e-mail: cmcd@stats.ox.ac.uk)

Received 2 October 2012; revised 5 August 2013; first published online 16 September 2013

We consider numbers and sizes of independent sets in graphs with minimum degree at least d. In particular, we investigate which of these graphs yield the maximum numbers of independent sets of different sizes, and which yield the largest random independent sets. We establish a strengthened form of a conjecture of Galvin concerning the first of these topics.

2010 *Mathematics subject classification*: Primary 05C35
Secondary 05C69, 05C07

Given a finite graph G, let $\mathcal{I}(G)$ be the set of independent sets and let $i(G) = |\mathcal{I}(G)|$; and for $k \geqslant 0$ let $\mathcal{I}_k(G)$ be the set of independent sets of order k and let $i_k(G) = |\mathcal{I}_k(G)|$. Thus $i(G) = \sum_{k \geqslant 0} i_k(G)$.

There are many extremal results on $i(G)$ and $i_k(G)$, where G ranges over a certain family of graphs, for example, trees or regular graphs (see [2]–[5], [7]–[10], [12]). Here we investigate graphs with a given lower bound on the vertex degrees. For $d \geqslant 0$, let $\mathcal{G}_n(d)$ be the set of graphs of order n with minimum degree at least d (n, k and d will always be integers). We are interested in which of these graphs yield the maximum numbers of independent sets of different sizes, and which yield the largest random independent sets. Let us discuss numbers first.

Recall that the *independence number* $\alpha(G)$ is the maximum size of an independent set. Clearly $\alpha(G) \leqslant n - d$ for each $G \in \mathcal{G}_n(d)$. Recently, Galvin [5] proved that, for n suitably larger than d, we have $i(G) < i(K_{d,n-d})$ for any $G \in \mathcal{G}_n(d)$ that is not (isomorphic to) $K_{d,n-d}$. Moreover, he conjectured essentially that for any $d \geqslant 1$, there exist integers $N(d)$ and $C(d)$ such that for each $n \geqslant N(d)$, $K_{d,n-d}$ maximizes i_k over all graphs in $\mathcal{G}_n(d)$ for each k satisfying $C(d) \leqslant k \leqslant n - d$, and he proved such a result in the case when $d = 1$. This conjecture has been proved for bipartite graphs with $N(d) = 2d$ and $C(d) = 3$ by Alexander, Cutler and Mink [1]. Further evidence is provided by Engbers and Galvin [4].

We shall see that this conjecture holds with $N(d) = O(d^3)$ and $C(d) = 3$. Observe that we need $C(d) \geqslant 3$ here, since each n-vertex graph has $i_0(G) = 1$ and $i_1(G) = n$. Also $i_2(G) = \binom{n}{2} - e(G)$, where $e(G)$ is the number of edges, and graphs $G \in \mathcal{G}_n(d)$ can have $i_2(G) > i_2(K_{d,n-d})$. (For example, if d is fixed and n is large and even, $K_{d,n-d}$ has $d(n-d) \sim dn$ edges, whereas a d-regular graph has $dn/2$ edges.) We do not know if $N(d)$ can be significantly reduced.

Theorem 1. *Let $1 \leqslant d \leqslant n^{1/3}/2$. For each graph $G \in \mathcal{G}_n(d)$ and each $k \geqslant 3$ we have $i_k(G) \leqslant i_k(K_{d,n-d})$, and if G is not $K_{d,n-d}$ then $i_2(G) + i_4(G) < i_2(K_{d,n-d}) + i_4(K_{d,n-d})$, and so $i(G) < i(K_{d,n-d})$.*

A graph $G \in \mathcal{G}_n(d)$ with $\alpha(G) = n-d$ has the form $G = H + I_{n-d}$ for a graph H of order d and the empty graph I_{n-d} on $n-d$ vertices. (Recall that for graphs G, G' with disjoint vertex sets, the sum $G + G'$ denotes the graph obtained by adding all edges between them.) Let $K^*_{a,b}$ denote the graph $K_a + I_b$.

Denote by $X(G)$ the size of an independent set chosen uniformly at random from $\mathcal{I}(G)$. Recall that X is *stochastically dominated* by Y, denoted by $X \leqslant_s Y$, if $\mathbb{P}(X \leqslant t) \geqslant \mathbb{P}(Y \leqslant t)$ for each t.

If $G \in \mathcal{G}_n(d)$ satisfies $\alpha(G) = n-d$ and G is not $K^*_{d,n-d}$, then G is (isomorphic to) a proper subgraph of $K^*_{d,n-d}$, and so $i(G) > i(K^*_{d,n-d})$; and it follows that

$$\mathbb{P}(X(G) \leqslant t) < \mathbb{P}(X(K^*_{d,n-d}) \leqslant t)$$

for $t = 0$ and $t = 1$. Hence it is *not* the case that $X(G) \leqslant_s X(K^*_{d,n-d})$. Nevertheless, our second theorem shows that, if we ignore independent sets of size at most 1, then of all graphs in $\mathcal{G}_n(d)$, the graph $K^*_{d,n-d}$ is the unique graph yielding the largest random independent sets.

Theorem 2. *Let $1 \leqslant d \leqslant n^{1/3}/2$. Then, for each graph $G \in \mathcal{G}_n(d)$ other than $K^*_{d,n-d}$, we have*

$$\mathbb{P}(X(G) \geqslant t) < \mathbb{P}(X(K^*_{d,n-d}) \geqslant t) \quad \textit{for each } t = 3, \ldots, n-d,$$

and if $\alpha(G) < n-d$ then this inequality holds also for $t = 1$ and 2.

This yields directly the following corollary.

Corollary 1. *For $1 \leqslant d \leqslant n^{1/3}/2$, each graph $G \in \mathcal{G}_n(d)$ satisfies*

$$X(G) \leqslant_s \max\{2, X(K^*_{d,n-d})\}, \tag{1}$$

and

$$\textit{if } \alpha(G) < n-d \quad \textit{then} \quad X(G) \leqslant_s X(K^*_{d,n-d}). \tag{2}$$

Also, since $\mathbb{E}(X) = \sum_{t \geqslant 1} \mathbb{P}(X \geqslant t)$, we may obtain almost directly another corollary.

https://doi.org/10.1017/S0963548313000369 Published online by Cambridge University Press

Corollary 2. *For $1 \leqslant d \leqslant n^{1/3}/2$, each graph $G \in \mathcal{G}_n(d)$ other than $K^*_{d,n-d}$ satisfies*

$$\mathbb{E}(X(G)) < \mathbb{E}(X(K^*_{d,n-d})) < (n-d)/2.$$

In order to prove these results, it turns out that the 'growth rates' α_k of the numbers of independent sets are crucial quantities. For a graph G and positive integer $k \leqslant \alpha(G)$, let

$$\alpha_k(G) := \frac{i_k(G)}{i_{k-1}(G)}.$$

Thus $\alpha_k(G)$ is $1/k$ times the average number of extensions of an independent $(k-1)$-set to an independent k-set in G; or (roughly) the 'average number of extensions per vertex' at size k.

To prove Theorem 1 we use two lemmas, one on growth rates $\alpha_k(G)$ and one on the 'base case' $i_3(G)$. To prove Theorem 2 we need one further lemma, a general result on growth rates and stochastic domination.

We adopt the following notation. For a graph G and integer d let $A = A(G,d) = \{v \in V(G) : \deg(v) > d\}$ and $B = V(G) \setminus A$; and let $a = |A|$, $b = |B|$. Also recall the standard notation that, if U is a set of vertices in G, then the neighbourhood $\Gamma(U)$ is the set of neighbours of vertices in U, and the closed neighbourhood $\Gamma[U]$ is $\Gamma(U) \cup U$.

Lemma 1.

(a) *For each $1 \leqslant d < n$ and $G \in \mathcal{G}_n(d)$, we have $\alpha_k(G) \leqslant \alpha_k(K^*_{d,n-d})$ for each $3 \leqslant k \leqslant \alpha(G)$.*

(b) *Let $1 \leqslant d \leqslant n^{1/3}/2$. Then, for each $G, K \in \mathcal{G}_n(d)$ with $\alpha(G) < n - d = \alpha(K)$, we have $\alpha_k(G) < \alpha_k(K)$ for each $4 \leqslant k \leqslant \alpha(G)$.*

Proof. **(a)** Let $3 \leqslant k \leqslant \alpha(G)$. Since each vertex degree in G is at least d, each $I \in \mathcal{I}_{k-1}(G)$ can be extended to at most $n-d-k+1$ independent k-sets. Call I *good* if this upper bound is attained, and otherwise call I *bad*. Note that I is good if and only if $|\Gamma(I)| = d$, if and only if each vertex in I has the same set of d neighbours. Also, each I is good if G is $K^*_{d,n-d}$.

Since each independent k-set contains exactly k independent $(k-1)$-sets, we have

$$i_{k-1}(G)(n-d-k+1) \geqslant k i_k(G).$$

Hence, $\alpha_k(G) \leqslant \frac{n-d-k+1}{k}$. But $\alpha_k(K^*_{d,n-d}) = \frac{n-d-k+1}{k}$ for $k = 3, \ldots, n-d$. This establishes part (a).

(b) Let $4 \leqslant k \leqslant \alpha(G)$. Suppose first that $k \geqslant d+2$. Let J be an independent set in G of size $\alpha(G) \leqslant n-d-1$. Let W be a set of $d+1$ vertices outside J, and note that each vertex in W has at least one neighbour in J. Since $k-1 \geqslant d+1$ we may pick a $(k-1)$-subset I of J with $\Gamma(I) \supseteq W$, and so I is bad. Now, since there is a bad independent $(k-1)$-set, $\alpha_k(G) < \frac{n-d-k+1}{k}$. Further, $\alpha_k(K) = \frac{n-d-k+1}{k}$ for each $k = d+2, \ldots, n-d$, so this case is done; and so to prove part (b) we may assume that $4 \leqslant k \leqslant d+1$.

https://doi.org/10.1017/S0963548313000369 Published online by Cambridge University Press

Let us then assume that $4 \leqslant k \leqslant d+1$, and note that $k \leqslant n-d$. We may write $K = H + I_{n-d}$ for some graph H of order d. Then

$$\alpha_k(K) = \frac{\binom{n-d}{k} + i_k(H)}{\binom{n-d}{k-1} + i_{k-1}(H)}.$$

Since $i_{k-1}(H) \leqslant \binom{d}{k-1}$ we have

$$\alpha_k(K) \geqslant \frac{\binom{n-d}{k}}{\binom{n-d}{k-1} + \binom{d}{k-1}} > \frac{n-d-k+1}{k}\left(1 - \frac{\binom{d}{k-1}}{\binom{n-d}{k-1}}\right). \tag{3}$$

Let p and q denote the numbers of good and bad sets in $\mathcal{I}_{k-1}(G)$ respectively, so $p+q = i_{k-1}(G)$. Then

$$ki_k(G) \leqslant p(n-d-k+1) + q(n-d-k) = (p+q)(n-d-k+1) - q,$$

so

$$\alpha_k(G) \leqslant \frac{n-d-k+1}{k} - \frac{q}{k(p+q)}. \tag{4}$$

Assume for a contradiction that $\alpha_k(G) \geqslant \alpha_k(K)$. Then it follows using (3) and (4) that

$$\frac{q}{p+q} \leqslant (n-d-k+1)\binom{d}{k-1}\Big/\binom{n-d}{k-1} < \frac{d^{k-1}}{(n-d-k+1)^{k-2}}.$$

Now $n-d-k+1 \geqslant n-2d \geqslant n/2$. Hence by the last inequality

$$\frac{q}{p+q} < \frac{d^{k-1}}{(n/2)^{k-2}} \leqslant \frac{d^3}{(n/2)^2} \leqslant \frac{1}{n}. \tag{5}$$

Thus certainly $p > 0$.

Claim. For each good independent $(k-1)$-set I in G there is a vertex $w \notin I \cup \Gamma(I)$ such that $\Gamma(w) \neq \Gamma(I)$.

We will prove the claim later: suppose for now that it holds. Then, from each good independent $(k-1)$-set I we may construct a bad independent $(k-1)$-set I' by deleting a vertex u from I and adding a vertex w as in the claim. This gives at least $p(k-1) \geqslant 3p$ constructions. Also, in each bad independent $(k-1)$-set I' which has been constructed, we can identify the vertex w added (since the other $k-2 \geqslant 2$ vertices all have the same neighbourhood). Thus each bad independent $(k-1)$-set I' is constructed at most $n-k+1 \leqslant n-3$ times. Hence

$$q \geqslant 3p/(n-3) > p/(n-1)$$

and so $q/(p+q) > 1/n$, which contradicts (5).

It remains to prove the claim. Recall that $B = \{v \in V(G) : \deg(v) = d\}$. Let I be a good independent $(k-1)$-set. Note that $I \subseteq B$ and $|\Gamma(I)| = d$. If $|A| = a \geqslant d+1$ then for w we may pick any vertex in $A \setminus \Gamma(I)$. So we may assume that $a \leqslant d$.

Let $B_1 = \{v \in B : \Gamma(v) \cap B \neq \emptyset\}$ and $B_2 = B \setminus B_1$. Since $\alpha(G) < n-d \leqslant |B|$ we have $E(B) \neq \emptyset$ and so $B_1 \neq \emptyset$. Either $I \subseteq B_1$ or $I \subseteq B_2$, since each vertex in I has the same set

https://doi.org/10.1017/S0963548313000369 Published online by Cambridge University Press

of d neighbours. If $I \subseteq B_1$ then $I \subseteq \Gamma(v)$ for some $v \in B_1$, and so for w we may pick any vertex not in $\Gamma(I) \cup \Gamma(v)$ (at least $n-2d \geqslant 1$ choices). If $I \subseteq B_2$ then for w we may pick any vertex in B_1. This completes the proof of the claim, and we are done. □

The previous lemma concerns ratios; the next considers the base case. Of graphs in $\mathcal{G}_n(d)$, clearly a d-regular graph has the most independent 2-sets: we look at the number i_3 of independent 3-sets. We first give a formula for $i_3(G)$ for any graph G. Let t_i be the number of induced subgraphs of G on three vertices with i edges. Then

$$\binom{n}{3} = t_0 + t_1 + t_2 + t_3,$$
$$e(G)(n-2) = t_1 + 2t_2 + 3t_3,$$
$$\sum_{v_i \in V(G)} \binom{\deg(v_i)}{2} = t_2 + 3t_3.$$

Hence

$$i_3(G) = \binom{n}{3} - e(G)(n-2) + \sum_{v_i \in V(G)} \binom{\deg(v_i)}{2} - t(G), \tag{6}$$

where $t(G) = t_3$ is the number of triangles. For example, if G is a d-regular graph then

$$\begin{aligned} i_3(G) &= \binom{n}{3} - \frac{1}{2}dn(n-2) + n\binom{d}{2} - t(G) \\ &= \binom{n-d}{3} - \frac{1}{2}dn + \frac{1}{6}d(d+1)(d+2) - t(G). \end{aligned} \tag{7}$$

Lemma 2. *Let $1 \leqslant d \leqslant n^{1/3}/2$. If $G, K \in \mathcal{G}_n(d)$ are such that $\alpha(G) < n-d = \alpha(K)$, then $i_3(G) \leqslant i_3(K) - n/2 + 1$.*

Proof. Our proof relies on (6) and (7). Consider $G \in \mathcal{G}_n(d)$ with $\alpha(G) < n-d$. We first show that we may assume without loss of generality that the set A of vertices of degree $> d$ is a non-empty independent set, and then that it suffices to prove (8) below; then we prove (8) by considering four cases for $a = |A|$.

Suppose first that G is d-regular. Then by (7) we have

$$i_3(G) \leqslant \binom{n-d}{3} - \frac{1}{2}dn + \frac{1}{6}d(d+1)(d+2).$$

But $i_3(K) \geqslant \binom{n-d}{3}$. Thus, if $d = 1$ then

$$i_3(G) \leqslant \binom{n-d}{3} - n/2 + 1 \leqslant i_3(K) - n/2 + 1;$$

and if $d \geqslant 2$, then as $(d+1)^3 \leqslant (3/2)^3 d^3 < n/2$,

$$i_3(G) \leqslant \binom{n-d}{3} - \frac{1}{2}dn + \frac{1}{6}(d+1)^3 \leqslant i_3(K) - n/2.$$

Hence we may assume that G is not regular, and so A is non-empty.

https://doi.org/10.1017/S0963548313000369 Published online by Cambridge University Press

Now repeatedly delete edges between vertices of degree $> d$, as long as G keeps satisfying $\alpha(G) < n-d$. We end up with some graph $G' \in \mathcal{G}_n(d)$ with $\alpha(G') < n-d$. Suppose that there is an edge $uv \in E'(A')$ after this step (we use E' and A' to refer to G'). Then there exists an $(n-d)$-set I such that $E'(I) = \{uv\}$. Let $J = V(G') \setminus I$, so $|J| = d$. Since $\deg_{G'}(u), \deg_{G'}(v) > d$ and $\deg_{G'}(w) \geqslant d$ for each other vertex $w \in I$, every possible edge between I and J is present in G'. Therefore, since there are $(n-d-2)$ 3-subsets of I containing u and v,

$$i_3(G) \leqslant i_3(G') \leqslant \binom{n-d}{3} - (n-d-2) + \binom{d}{3} < \binom{n-d}{3} - \frac{n}{2}$$

since $d+2 \leqslant 3d \leqslant \frac{3}{8}n$ and $\binom{d}{3} < \frac{d^3}{6} \leqslant \frac{n}{48}$. Hence, we may assume that A is independent.

For each $v_i \in A$, let $r_i = \deg(v_i)$. Observe that $2e(G) = \sum_i r_i + (n-a)d$. Thus, from (6),

$$\begin{aligned}
&2i_3(G) - 2\binom{n}{3}\\
&\quad = -\left[\sum_{i=1}^{a} r_i + (n-a)d\right](n-2) + \sum_{i=1}^{a} r_i(r_i-1) + (n-a)d(d-1) - 2t(G)\\
&\quad = \sum_{i=1}^{a} r_i(r_i-n+1) - (n-a)d(n-d-1) - 2t(G)\\
&\quad = -dn(n-d-1) + h_d(G),
\end{aligned}$$

where

$$h_d(G) = \sum_{i=1}^{a} r_i(r_i-n+1) + ad(n-1-d) - 2t(G).$$

Thus

$$\begin{aligned}
i_3(G) &= \binom{n}{3} - \frac{1}{2}dn(n-d-1) + \frac{1}{2}h_d(G)\\
&= \binom{n-d}{3} - \frac{1}{2}dn + \frac{1}{6}d(d+1)(d+2) + \frac{1}{2}h_d(G).
\end{aligned}$$

Note that here only $\frac{1}{2}h_d(G)$ varies with $G \in \mathcal{G}_n(d)$. Since $i_3(K) \geqslant \binom{n-d}{3}$, by the last equality

$$h_d(K) \geqslant dn - \frac{1}{3}d(d+1)(d+2) \geqslant \left(d - \frac{1}{4}\right)n.$$

To see the second inequality here, we may check directly for $d=1$, and for $d \geqslant 2$ use

$$\frac{1}{3}d(d+1)(d+2) \leqslant \frac{1}{3}(d+1)^3 \leqslant \frac{3^2}{2^3}\frac{n}{8} < \frac{n}{4}.$$

Thus it suffices to show that

$$h_d(G) \leqslant \left(d - \frac{7}{8}\right)n \tag{8}$$

and the remainder of the proof is devoted to establishing this result.

https://doi.org/10.1017/S0963548313000369 Published online by Cambridge University Press

Recall that we are assuming that in G the set A of vertices of degree $> d$ is independent. Thus $d+1 \leqslant r_i \leqslant n-a$ for each $i = 1, \ldots, a$. Consider the function

$$g(x) = x(x-n+1) = -x(n-1-x)$$

for real x. This is decreasing for $x < (n-1)/2$ and increasing for $x > (n-1)/2$. We now break the proof of (8) into four cases: $a \geqslant d+2$, $a = d+1$, $a = d$, and $1 \leqslant a \leqslant d-1$.

Suppose that $a \geqslant d+2$. Then each $d+1 \leqslant r_i \leqslant n-d-2$, so $g(r_i) \leqslant (d+1)(d+2-n)$. Hence,

$$\begin{aligned} h_d(G) &\leqslant a(d+1)(d+2-n) + ad(n-1-d) \\ &= a(-n+2d+2), \end{aligned} \tag{9}$$

and so (8) holds.

Suppose that $a = d+1$. Then $d+1 \leqslant r_i \leqslant n-d-1$ for each i, and

$$\sum_{i=1}^{a} r_i \leqslant d(n-a) = d(n-d-1).$$

Thus at most $d-1$ of the r_i are equal to $n-d-1$, and so

$$\begin{aligned} \sum_{i=1}^{a} g(r_i) &\leqslant -(d-1)d(n-d-1) - 2(d+1)(n-d-2) \\ &= -n(d^2+d+2) + d^3 + 2d^2 + 5d + 4. \end{aligned}$$

Hence

$$\begin{aligned} h_d(G) &\leqslant \sum_{i=1}^{a} g(r_i) + (d+1)d(n-1-d) \\ &\leqslant -2n + d^3 + d^2 + 3d + 3 \leqslant -n, \end{aligned}$$

and so (8) holds.

Suppose that $a = d$. Since $\alpha(G) < n-d$, we have $e(B) > 0$. It follows that

$$\sum_{i=1}^{a} r_i \leqslant d(n-d) - 2 \leqslant d(n-d) - 1.$$

Hence not all d of the r_i are equal to $n-d$, so

$$\begin{aligned} h_d(G) &\leqslant (d-1)(n-d)(1-d) + (n-d-1)(-d) + d^2(n-1-d) \\ &= (d-1)n - 2d^2 + 2d \leqslant (d-1)n, \end{aligned}$$

and so (8) holds.

Finally, suppose that $1 \leqslant a \leqslant d-1$. Consider $v_i \in A$. Suppose that $r_i \geqslant n-d-1$. Then the edge-boundary of $\Gamma(v_i)$ has size at most $r_i a + (n-a-r_i)d \leqslant r_i a + d(d+1-a)$, and so $2e(\Gamma(v_i)) \geqslant r_i(d-a) - d(d+1-a)$. Hence, twice the number of triangles containing v_i is at least $r_i(d-a) - d(d+1-a)$. Also, using first that $r_i \leqslant n-a$ and then that $r_i \geqslant n-d-1$, we have

$$g(r_i) - r_i(d-a) = r_i(r_i - n + 1 - d + a) \leqslant r_i(1-d) \leqslant (n-d-1)(1-d).$$

https://doi.org/10.1017/S0963548313000369 Published online by Cambridge University Press

On the other hand, if $r_i \leqslant n-d-2$, then $g(r_i) \leqslant (d+1)(d-n+2)$. Let

$$l = |\{i : r_i \geqslant n-d-1\}| \leqslant a.$$

Then $h_d(G)$ is at most

$$\begin{aligned}
&\sum_{i:r_i \geqslant n-d-1} [g(r_i) - r_i(d-a) + d(d+1-a)] + \sum_{i:r_i \leqslant n-d-2} g(r_i) + ad(n-1-d) \\
&\quad \leqslant l[(n-d-1)(1-d) + d(d+1-a)] + (a-l)(d+1)(d-n+2) + ad(n-1-d) \\
&\quad = (2l-a)n + l(d^2-2d-3) + a(2d+2) \\
&\quad \leqslant (d-1)(n+d^2-1) \quad \text{since } l \leqslant a \leqslant d-1 \\
&\quad < (d-1)n + d^3 \leqslant \left(d - \frac{7}{8}\right)n
\end{aligned}$$

as required. □

With the last two lemmas, we may now prove Theorem 1, establishing a stronger version of the conjecture of Galvin [5] mentioned earlier.

Proof of Theorem 1. If $\alpha(G) = n-d$ then G is (isomorphic to) a supergraph of $K_{d,n-d}$ and the result is trivial: so we may assume that $\alpha(G) < n-d$. Let $K \in \mathcal{G}_n(d)$ with $\alpha(K) = n-d$. We want to show that

$$i_2(G) + i_4(G) < i_2(K) + i_4(K). \tag{10}$$

Note that $e(G) \geqslant dn/2$, and so

$$i_2(G) - i_2(K) = e(K) - e(G) \leqslant d(n-d) + \binom{d}{2} - \frac{dn}{2} = \frac{dn}{2} - \binom{d+1}{2}. \tag{11}$$

Also, $i_3(K) - i_3(G) \geqslant n/2 - 1$ by Lemma 2, and $\alpha_4(K) > \alpha_4(G)$ by Lemma 1.

Suppose $d = 1$. Then $\alpha_4(K) = \frac{n}{4} - 1 \geqslant 1$. Thus if $i_3(G) > 0$ then

$$i_4(K) - i_4(G) > \alpha_4(K)(i_3(K) - i_3(G)) \geqslant \frac{n}{2} - 1;$$

and if $i_3(G) = 0$ then

$$i_4(K) - i_4(G) \geqslant i_3(K) - i_3(G) = \binom{n-1}{3} > \frac{n}{2} - 1.$$

Hence $i_4(K) - i_4(G) > \frac{n}{2} - 1$ in each case; and so using also (11) we obtain (10).

Suppose $d \geqslant 2$. Then $n \geqslant 32d$. Thus by (3), noting for example that $3 + d \leqslant \frac{5}{2}d < \frac{n}{12}$, we find

$$\alpha_4(K) > \frac{11n}{48}\left(1 - \frac{(d-1)^3}{(n-d-2)^3}\right) > \frac{11n}{48}\left(1 - \frac{(\frac{n}{2})^3}{(\frac{15n}{16})^3}\right) > \frac{n}{6},$$

and so

$$i_4(K) - i_4(G) > \frac{n}{6}\left(\frac{n}{2} - 1\right) > \frac{5n^2}{64}.$$

https://doi.org/10.1017/S0963548313000369 Published online by Cambridge University Press

But by (11),

$$i_2(G) - i_2(K) \leqslant \frac{dn}{2} \leqslant \frac{n^2}{64},$$

and (10) follows. □

To prove Theorem 2, as well as the two corollaries, we need one further lemma, which is a general result on growth rates and stochastic domination, adapted from Lemma 2.4 of [9]. Given a finite sequence of positive real numbers $x = (x_0, x_1, \ldots, x_s)$, let $S(x) = \sum_{k \geqslant 0} x_k$. Define a random variable $X = X(x)$ by $\mathbb{P}(X = k) = x_k / S(x)$.

Lemma 3. *Let $x_0, y_0 > 0$, let $1 \leqslant a \leqslant b$ be integers, and let $\alpha_1, \ldots, \alpha_a > 0$ and $\beta_1, \ldots, \beta_b > 0$. For $i = 1, \ldots, a$, let $x_i = x_0 \prod_{0<j\leqslant i} \alpha_j$; and for $i = 1, \ldots, b$, let $y_i = y_0 \prod_{0<j\leqslant i} \beta_j$. Let $x = (x_0, x_1, \ldots, x_a)$ and $y = (y_0, y_1, \ldots, y_b)$, and denote $X(x)$ by X and $X(y)$ by Y. If $\alpha_i \leqslant \beta_i$ for each $i = 1, \ldots, a$, then $X \leqslant_s Y$. Further, if these conditions hold, and $(\alpha_1, \ldots, \alpha_a) \neq (\beta_1, \ldots, \beta_b)$, then*

$$\mathbb{P}(X \geqslant t) < \mathbb{P}(Y \geqslant t) \quad \textit{for each } t = 1, \ldots, b.$$

Proof. By replacing y_a by $\sum_{j>a} y_j$, we may assume that $b = a$. It suffices to consider the case when $\alpha_i = \beta_i$ for all i except j_0, where $\alpha_{j_0} < \beta_{j_0}$. Since $\mathbb{P}(X \leqslant a) = \mathbb{P}(Y \leqslant a) = 1$, it suffices to prove $\mathbb{P}(X \leqslant t) > \mathbb{P}(Y \leqslant t)$ for $t = 0, \ldots, a-1$. Note that we may rescale x_i, y_i without changing the distribution.

Suppose t satisfies $0 \leqslant t \leqslant j_0 - 1$. Rescale to $x_0 = y_0 = 1$. Then $x_i = y_i$ for all $i \leqslant t$ and $S(x) < S(y)$. So

$$\mathbb{P}(X \leqslant t) = \frac{\sum_{i\leqslant t} x_i}{S(x)} > \frac{\sum_{i\leqslant t} y_i}{S(y)} = \mathbb{P}(Y \leqslant t).$$

For t such that $j_0 \leqslant t \leqslant a-1$, we rescale to $x_{j_0} = y_{j_0}$. Then $x_i = y_i$ for all $i = j_0, j_0 + 1, \ldots, a$ and $S(x) > S(y)$. Hence, $\mathbb{P}(X > t) < \mathbb{P}(Y > t)$ and so $\mathbb{P}(X \leqslant t) > \mathbb{P}(Y \leqslant t)$. □

Proof of Theorem 2. There are two cases, depending on whether $\alpha(G) < n - d$ or $\alpha(G) = n - d$.

(a) Let $G \in \mathcal{G}_n(d)$ with $\alpha(G) < n - d$. For $k \geqslant 1$, let α_k^* denote $\alpha_k(K_{d,n-d}^*)$. Then $\alpha_1(G) = \alpha_1^* = n$. By Lemma 1(a), $\alpha_k(G) \leqslant \alpha_k^*$ for $3 \leqslant k \leqslant \alpha(G)$.

If $\alpha_2(G) \leqslant \alpha_2^*$ then directly from Lemma 3 we have $\mathbb{P}(X(G) \geqslant t) < \mathbb{P}(X(K_{d,n-d}^*) \geqslant t)$ for each $t = 1, \ldots, n-d$, and we are done. So we may suppose that $\alpha_2(G) > \alpha_2^*$; that is $i_2(G) > i_2^*$, where i_k^* denotes $i_k(K_{d,n-d}^*)$.

Let x be the i_k-vector for G (up to x_{n-d}), let z be the i_k-vector for $K_{d,n-d}^*$, and let y agree with x in the first three places, and agree with z in the remaining places; that is,

$$x = (x_0, x_1, \ldots, x_{n-d}) = (1, n, i_2(G), i_3(G), i_4(G), \ldots, i_{n-d}(G)),$$
$$y = (y_0, y_1, \ldots, y_{n-d}) = (1, n, i_2(G), i_3^*, i_4^*, \ldots, i_{n-d}^*)$$

https://doi.org/10.1017/S0963548313000369 Published online by Cambridge University Press

and

$$z = (z_0, z_1, \ldots, z_{n-d}) = (1, n, i_2^*, i_3^*, i_4^*, \ldots, i_{n-d}^*).$$

Let $3 \leqslant t \leqslant n-d$. By Lemma 1(b) with $K = K_{d,n-d}^*$, for each $4 \leqslant k \leqslant \alpha(G)$ we have

$$\frac{x_k}{x_{k-1}} \leqslant \frac{y_k}{y_{k-1}}.$$

Moreover, by Lemma 2, $i_3(G) < i_3^*$ so that

$$\frac{x_3}{x_2} < \frac{y_3}{y_2}.$$

Then by Lemma 3,

$$\mathbb{P}(X(G) \geqslant t) = \mathbb{P}(X(x) \geqslant t) < \mathbb{P}(X(y) \geqslant t).$$

Also

$$\mathbb{P}(X(y) \geqslant t) < \mathbb{P}(X(z) \geqslant t) = \mathbb{P}(X(K_{d,n-d}^*) \geqslant t)$$

since $S(y) > S(z)$. Hence $\mathbb{P}(X(G) \geqslant t) < \mathbb{P}(X(K_{d,n-d}^*) \geqslant t)$ as required.

To complete the proof for this case, note that by Theorem 1, $i(G) < i(K_{d,n-d}^*)$, so that

$$\mathbb{P}(X(G) \leqslant 0) = 1/i(G) > 1/i(K_{d,n-d}^*) = \mathbb{P}(X(K_{d,n-d}^*) \leqslant 0),$$

and similarly

$$\mathbb{P}(X(G) \leqslant 1) = (1+n)/i(G) > (1+n)/i(K_{d,n-d}^*) = \mathbb{P}(X(K_{d,n-d}^*) \leqslant 1).$$

(b) It remains to consider the case when $\alpha(G) = n-d$ and G is not $K_{d,n-d}^*$. Then G may be obtained from $K_{d,n-d}^*$ by deleting at least one edge from the K_d part. Thus $i(G) > i(K_{d,n-d}^*)$; and the i_k-vector x of G may be obtained from the i_k-vector z for $K_{d,n-d}^*$ by adding positive integers to some entries amongst the first $d+1$ including adding at least 1 to z_2. It is immediate that $\mathbb{P}(X(x) \geqslant t) < \mathbb{P}(X(z) \geqslant t)$ for each $t = d+1, \ldots, n-d$. Let $2 \leqslant t \leqslant d-1$. Then

$$\mathbb{P}(X(z) \leqslant t) = \frac{\sum_{i=0}^{t} z_i}{S(z)}.$$

To obtain $\mathbb{P}(X(x) \leqslant t)$ from the last ratio we add at least 1 to the numerator and at most 2^d to the denominator. Observe that $S(z) = 2^{n-d} + d \geqslant 2^{n-d}$. Since $z_i \leqslant \binom{n}{i}$ and $t < d$,

$$\mathbb{P}(X(z) \leqslant t) < \frac{dn^d}{2^{n-d}} < \frac{(2n)^{2d}}{2^n 2^d} \leqslant \frac{(2n)^{n^{1/3}}}{2^n 2^d} \leqslant 2^{-d}$$

(where the last inequality holds as $\log_2(2n) \leqslant n^{2/3}$), so overall the ratio increases, that is, $\mathbb{P}(X(z) \leqslant t) < \mathbb{P}(X(x) \leqslant t)$, as required. □

We noted earlier that Corollary 1 follows directly from Theorem 2, so it remains only to prove Corollary 2.

Proof of Corollary 2. If $\alpha(G) < n-d$, the result follows from Theorem 2, since

$$\mathbb{E}[X(G)] = \sum_{t\geqslant 1} \mathbb{P}(X(G) \geqslant t) < \sum_{t\geqslant 1} \mathbb{P}(X(K) \geqslant t) = \mathbb{E}[X(K^*_{d,n-d})].$$

Suppose then that $\alpha(G) = n-d$, so that G may be obtained from $K^*_{d,n-d}$ by deleting at least one edge from the K_d part. Then the average size of the sets which are independent in G but not in $K^*_{d,n-d}$ is at most d, which is less than $\mathbb{E}[X(K^*_{d,n-d})]$, and so $\mathbb{E}[X(G)] < \mathbb{E}[X(K^*_{d,n-d})]$. □

We remark that with an analogous method, a weighted version of the statements can be proved. Let $I(G,\lambda) = \sum_{k\geqslant 0} i_k(G)\lambda^k$ be the independent set polynomial of G ([6], [11]). Instead of a uniform sampling of independent sets of $\mathcal{I}(G)$, we fix $\lambda > 0$ and pick a given independent k-set with probability $\lambda^k/I(G,\lambda)$. Then under this sampling, the analogous versions of Theorem 2 and its corollaries hold.

Acknowledgement

Thanks to a referee for helpful comments, which included pointing out some recent references.

References

[1] Alexander, J., Cutler, J. and Mink, T. (2012) Independent sets in graphs with given minimum degree. *Electron. J. Combin.* **19** #37.
[2] Carroll, T., Galvin, D. and Tetali, P. (2009) Matchings and independent sets of a fixed size in regular graphs. *J. Combin. Theory Ser. A* **116** 1219–1227.
[3] Cutler, J. and Radcliffe, A. J. (2011) Extremal problems for independent set enumeration. *Electron. J. Combin.* **18** #169.
[4] Engbers, J. and Galvin, D. (2013) Counting independent sets of a fixed size in graphs with a minimum degree. *J. Graph Theory* DOI 10.1002/jgt.21756
[5] Galvin, D. (2011) Two problems on independent sets in graphs. *Discrete Math.* **311** 2105–2112.
[6] Gutman, I. and Harary, F. (1983) Generalizations of the matching polynomial. *Utilitas Math.* **24** 97–106.
[7] Kahn, J. (2001) An entropy approach to the hard-core model on bipartite graphs. *Combin. Probab. Comput.* **10** 219–237.
[8] Lin, S. B. and Lin, C. (1995) Trees and forests with large and small independent indices. *Chinese J. Math.* **23** 199–210.
[9] McDiarmid, C., Steger, A. and Welsh, D. (2005) Random planar graphs. *J. Combin. Theory Ser. B* **93** 187–205.
[10] Prodinger, H. and Tichy, R. F. (1982) Fibonacci numbers of graphs. *Fibonacci Quart.* **20** 16–21.
[11] Scott, A. D. and Sokal, A. D. (2005) The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma. *J. Statist. Phys.* **118** 1151–1261.
[12] Zhao, Y. (2010) The number of independent sets in a regular graph. *Combin. Probab. Comput.* **19** 315–320.