Hitting all maximum independent sets

Noga Alon *

Abstract

We describe an infinite family of graphs G_n, where G_n has n vertices, independence number at least $n/4$, and no set of less than $\sqrt{n}/2$ vertices intersects all its maximum independent sets. This is motivated by a question of Bollobás, Erdős and Tuza, and disproves a recent conjecture of Friedgut, Kalai and Kindler. Motivated by a related question of the last authors, we show that for every graph G on n vertices with independence number $(1/4+\varepsilon)n$, the average independence number of an induced subgraph of G on a uniform random subset of the vertices is at most $(1/4+\varepsilon-\Omega(\varepsilon^2))n$.

1 Background and results

The following conjecture appears in a recent paper of Friedgut, Kalai and Kindler.

Conjecture 1.1 ([8], Conjecture 3.1). *For every $\alpha \in (0, 1/2)$ there exists k and $\tau > 0$ such that if G is a graph on n vertices with maximum independent set of size αn, then there exist pairwise disjoint subsets of vertices $A_1, A_2, \ldots, A_r$ in G such that*

1. *$|A_i| = k$ for all i.*
2. *$|\cup_{i=1}^{r} A_i| \geq \tau n$*
3. *Every maximum independent set in G intersects every set A_i.*

Here we show that this is false for every fixed $\alpha \in (0, 1/2)$ even if the requirement (2) is omitted, but that the assertion does hold for any $\alpha > 1/2$. We also discuss several related problems.

For a graph $G = (V, E)$ let $h(G)$ denote the minimum cardinality of a set of vertices that intersects every maximum independent set of G. Bollobás, Erdős and Tuza (see [7], page 224, or [5], page 52) raised the following conjecture.

*Princeton University, Princeton, NJ 08544, USA and Tel Aviv University, Tel Aviv 69978, Israel. Email: nalon@math.princeton.edu. Research supported in part by NSF grant DMS-1855464, BSF grant 2018267 and the Simons Foundation.

Conjecture 1.2 ([7], [5]). *For any positive α, if the size $\alpha(G)$ of a maximum independent set in an n-vertex graph G is at least αn, then $h(G) = o(n)$.*

They formulated the conjectures for cliques and not for independent sets; replacing G by its complement leads to the formulation above.

Theorem 1.3. *For every positive integer k there is a graph $G = G_k$ with $n = 2k(2k-1)$ vertices, independence number $\alpha(G) = k^2 (> n/4)$, and $h(G) = k + 1 (> \sqrt{n}/2)$.*

Theorem 1.4. *For any positive integers m and t, where m is even and $4t^2 \leq m$, there is a graph $G = G_{m,t}$ on $n = 2^m$ vertices with independence number $\alpha(G) = \sum_{i=0}^{m/2-t} \binom{m}{i}$ and $h(G) = \Theta(t^2)$.*

The above two results provide counterexamples to Conjecture 1.1. On the other hand we observe that an old result of Hajnal implies that the assertion of the conjecture (with $k = 1$ and $\tau = 2\alpha - 1$) does hold for any $\alpha > 1/2$. Theorem 1.4 also settles the final open problem raised by Dong and Wu in [6].

The graphs establishing the assertion of Theorem 1.3 are regular. It turns out that for regular graphs (of any degree) the estimates in this theorem are nearly tight, as stated in the next proposition.

Proposition 1.5. *For any fixed $\varepsilon > 0$ and any regular graph G with $n > n_0(\varepsilon)$ vertices satisfying $\alpha(G) \geq (1/4 + \varepsilon)n$, the parameter $h(G)$ satisfies $h(G) < (1/\varepsilon)\sqrt{n \log n} + 1$.*

Conjecture 1.1 is motivated by another conjecture suggested in [8]. For a graph G on n vertices and independence number $\alpha(G) = \alpha n$, let $\alpha' n = \alpha'(G) n$ denote the average value of the independence number of the induced subgraph of G on a uniform random set of vertices.

Conjecture 1.6 ([8], Conjecture 2.9). *For any $\alpha \in (0, 1/2)$ there is an $\varepsilon = \varepsilon(\alpha) > 0$ so that for every graph G with n vertices and independence number αn, $\alpha'(G) \leq \alpha - \varepsilon(\alpha)$.*

The following result shows that the above is true for any $\alpha > 1/4$. The case of positive α close to 0 appears to be significantly more difficult (and interesting).

Theorem 1.7. *Let $G = (V, E)$ be a graph with n vertices and independence number $\alpha(G) = (1/4 + \varepsilon)n$, where $\varepsilon > 0$ satisfies $\varepsilon < 1/4$. Then $\alpha'(G) \leq 1/4 + \varepsilon - \varepsilon^2/3$.*

The constant $1/3$ above can be easily improved and we make no attempt to optimize it here.

For regular graphs we can show that the assertion of the conjecture holds for all fixed $\alpha > 1/8$.

Proposition 1.8. *For any $\varepsilon > 0$ there is some $g(\varepsilon) > 0$ so that the following holds. Let $G = (V, E)$ be a regular graph with n vertices and independence number $\alpha(G) = (1/8+\varepsilon)n$. Then $\alpha'(G) \leq 1/8 + \varepsilon - g(\varepsilon)$.*

The proofs appear in the next section. All logarithms throughout the paper are in base 2, unless otherwise specified. To simplify the presentation we omit all floor and ceiling signs whenever these are not crucial.

2 Proofs

2.1 Constructions

Proof of Theorem 1.3: The graph $G = G_k$ is the shift graph described as follows. Put $K = \{1, 2, .., 2k\}$. The set of vertices of G_k is the set of all ordered pairs (i, j) with $i \neq j$ and $i, j \in K$. Thus the number of vertices is $n = 2k(2k - 1)$. Two vertices (a, b) and (c, d) are adjacent if $b = c$ or $d = a$. Note that the vertices can be viewed as all directed edges of the complete directed graph on K, where two are adjacent iff they form a directed path of length 2. It is easy to check that the maximum independent sets of this graph are of size k^2. Indeed, for every partition of K into two disjoint parts S and T of equal cardinality, the set of all pairs (s, t) with $s \in S, t \in T$ is a maximum independent set, and these are all the maximum independent sets. Any set H of at most k vertices of G can be viewed as k directed edges of the complete graph on K. Let S be a set of k points in K that does not contain the head of any of these k directed edges, and put $T = K - S$. Then the maximum independent set consisting of all pairs (s, t) with $s \in S, t \in T$ does not intersect H. Therefore $h(G) \geq k + 1$. This is tight as shown by a set of pairs forming a directed cycle of length $k + 1$ in the complete directed graph on K. $\square$

Proof of Theorem 1.4: Let $G = G_{m,t}$ be the graph whose vertices are all binary vectors of length m, where two are adjacent iff the Hamming distance between them exceeds $m-2t$. Note that this is the Cayley graph of Z_2^m with respect to the set of all vectors of Hamming weight at least $m - 2t + 1$. This graph contains as an induced subgraph the Kneser graph $K(m, m/2 + 1 - t)$. By an old result of Kleitman [10], the independence number of this graph is exactly $\sum_{i=0}^{m/2-t} \binom{m}{i}$. The maximum independent sets are the 2^m Hamming balls of radius $m/2 - t$ centered at the vertices of G. Any set of vertices that hits all these independent sets forms a covering code of radius $m/2 - t$ in Z_2^m. By using known results about covering codes in this range of the parameters it is not difficult to prove that the minimum possible size of such a set is $\Omega(t^2)$. Indeed, viewing the vectors of the covering code as vectors with $\{-1, 1\}$ coordinates, if their number is T then by a known result in

Discrepancy Theory (see, e.g., [3], Corollary 13.3.4), there is a $\{-1,1\}$ vector whose inner product with all members of the code is in absolute value at most $12\sqrt{T}$. If $12\sqrt{T} < 2t$ this gives a vector whose Hamming distance from any codeword is larger than $m/2 - t$, contradicting the assumption. This shows that the size of the code is at least $\Omega(t^2)$. This is tight up to the hidden constant in the Ω-notation as can be shown by a random construction of vectors of length $\Theta(t^2)$, extending each such vector in two complementary ways on the remaining coordinates, or by taking the rows of a Hadamard matrix of order $\Theta(t^2)$ and their inverses, extending them in the same way. Note that the fact that the Kneser graph $K(m, m/2+1-t)$ is a subgraph of G also implies a lower bound of $2t$ for the size of the hitting set (as the Hamming balls of radius $m/2 - t$ centered in the points of the hitting set cover all points, providing a proper coloring of the Kneser graph), but the bound obtained this way is weaker than the tight $\Theta(t^2)$ bound. □

2.2 Induced subgraphs on random subsets

An old result of Hajnal [9] (see also [11]) asserts that for every graph G the cardinality of the intersection of all maximum independent sets plus the cardinality of the union of all these sets is at least $2\alpha(G)$. If $\alpha(G) = \alpha n$ where $\alpha > 1/2$ and n is the number of vertices of G, this implies that there is a set of at least $(2\alpha - 1)n$ vertices contained in all maximum independent sets. (The result in [9] is formulated in terms of cliques, and not in terms of independent sets, but this is clearly equivalent as shown by replacing the graph with its complement). This shows that the assertion of Conjecture 1.1 holds for $\alpha > 1/2$ (with $k = 1$ and $\tau = 2\alpha - 1$).

Using Hajnal's result we next describe the proof of Theorem 1.7 obtaining a (modest) progress in the study of Conjecture 1.6.

Proof of Theorem 1.7: Without loss of generality we may assume that n is arbitrarily large, as we can replace G by a union of many vertex disjoint copies of itself and use linearity of expectation. Assuming n is large, almost every random subset of vertices is of cardinality $(1/2 + o(1))n$, hence it suffices to show that for almost every set W of $m = (1/2 + o(1))n$ vertices, the independence number of the induced subgraph of G on W is smaller than $(1/4 + \varepsilon - \varepsilon^2/2)n$. Construct the random set W of size m by removing from G vertices, one by one. Starting with $V = V_0$, let V_{i+1} be the set obtained from V_i by removing a uniform random vertex of V_i. The set W is thus V_{n-m}. Let G_i be the induced subgraph of G on V_i. Call a step i, $1 \leq i \leq n - m$ of the random process above successful if either the independence number of G_{i-1} is already smaller than $(1/4 + \varepsilon - \varepsilon^2/2)n$ (note that in this case this will surely be the case in the final graph G_{n-m}), or the independence number of G_i is strictly smaller than that of G_{i-1}. Put $i_0 = (1/2 - \varepsilon)n$ and consider the graph G_{i_0}. For any $i > i_0$, the number of vertices of G_{i-1} is at most $(1/2 + \varepsilon)n$. If its independence number is smaller than $(1/4 + \varepsilon - \varepsilon^2/2)n$ then, by definition, step number i is successful. Otherwise, by the result of Hajnal mentioned above, the number of vertices of G_{i-1} that lie in all the maximum independent sets in it is at least $(\varepsilon - \varepsilon^2)n$. Since $\varepsilon < 1/4$ this is a fraction of at least ε of the vertices of G_{i-1}. Therefore, in this case, the probability that the next chosen vertex lies in all maximum independent sets of G_{i-1} is at least ε. We have thus shown that for every i satisfying $i_0 < i \leq n-m$ the probability that step number i is successful is at least ϵ. Therefore, the probability that there are at least $\varepsilon^2 n/2$ successful steps during the $n - m - i_0 = (\varepsilon - o(1))n$ steps starting with G_{i_0} until we reach G_{n-m} is at least the probability that a binomial random variable with parameters $(\varepsilon - o(1))n$ and ε is at least $\varepsilon^2 n/2$. This probability is $1 - o(1)$ for any fixed positive ε as n tends to infinity. Since having that many successful steps ensures that the independence number of the induced subgraph of G on W is at most $(1/4 + \varepsilon - \varepsilon^2/2)n$, this completes the proof. □

2.3 Regular graphs

In the proofs of Propositions 1.5 and 1.8 we apply the following early version of the container theorem of [4] and [12].

Theorem 2.1 (c.f. [3], Theorem 1.6.1). *Let $G = (V, E)$ be a d-regular graph on n vertices and let $\delta > 0$ be a positive real. Then there is a collection $\mathcal{C}$ of subsets of V of cardinality*

$$|\mathcal{C}| \leq \sum_{i \leq n/\delta d} \binom{n}{i}$$

so that each $C \in \mathcal{C}$ is of size at most $\frac{n}{\delta d} + \frac{n}{2-\delta}$ and every independent set in G is fully contained in a member $C \in \mathcal{C}$.

Proof of Proposition 1.5: Let $G = (V, E)$ be a d-regular graph on n vertices with independence number at least $(\frac{1}{4} + \varepsilon)n$, and assume that n is sufficiently large as a function of ε. The closed neighborhood of any vertex of G intersects every maximum independent set of G, implying that $h(G) \leq d + 1$. If $d \leq (1/\varepsilon)\sqrt{n \log n}$ this implies the desired result, hence we may and will assume that d is larger. By Theorem 2.1 with $\delta = \epsilon$ there is a collection $\mathcal{C}$ of at most

$$\sum_{i \leq \sqrt{n}/\sqrt{\log n}} \binom{n}{i} \leq 2^{\sqrt{n \log n}}$$

subsets of V, each of size at most

$$\frac{n}{\sqrt{n\log n}}+\frac{n}{2-\varepsilon}<(\frac{1}{2}+\varepsilon)n$$

so that every independent set of G is fully contained in one of them.

Let X be a random set of $\frac{1}{\varepsilon}\sqrt{n\log n}$ vertices chosen uniformly (with repetitions) among all vertices of G. Fix a container $C\in\mathcal{C}$. By Hajnal's result there are at least εn vertices contained in all maximum independent sets of G that are contained in C. The probability that X does not contain any of these vertices is at most

$$(1-\varepsilon)^{\frac{1}{\varepsilon}\sqrt{n\log n}}\leq e^{-\sqrt{n\log n}}.$$

The desired result follows by applying the union bound over all $C\in\mathcal{C}$. $\square$

Proof of Proposition 1.8: Let $G=(V,E)$ be as in the proposition, where $|V|=n$. As before we may assume without loss of generality that n is sufficiently large as a function of ε. Without trying to optimize the function $g(\varepsilon)$, let d denote the degree of regularity of G. Note that G contains a set S of at least $n/(d^2+1)$ vertices no two of which are adjacent or have a common neighbor. Let W be a uniform random set of vertices of G. If the complement of W fully contains the closed neighborhoods of s vertices of S, then the independence number of the induced subgraph of G on W is at most $\alpha(G)-s$. The random variable counting the above number s is a Binomial random variable with expectation $|S|/2^{d+1}\geq\frac{n}{(d^2+1)2^{d+1}}$. Thus if, say, d is at most $50/\varepsilon^4$ we get that the expected independence number of an induced subgraph of G on a uniform random set of vertices is at most

$$\alpha(G)-\frac{n}{(d^2+1)2^{d+1}}$$

supplying a lower bound (of the form $2^{-\Theta(\varepsilon^{-4})}$) for $g(\varepsilon)$. We thus may and will assume that $d\geq\frac{50}{\varepsilon^4}$. By Theorem 2.1 with $\delta=\varepsilon$ there is a collection $\mathcal{C}$ of subsets of V, satisfying

$$|\mathcal{C}|\leq\sum_{i\leq\varepsilon^3 n/50}\binom{n}{i}\leq 2^{H(\varepsilon^3/50)n},$$

where $H(x)=-x\log x-(1-x)\log(1-x)$ is the binary entropy function. Each member C of $\mathcal{C}$ is of size at most

$$\frac{n}{\delta d}+\frac{n}{2-\delta}\leq\frac{\varepsilon^3 n}{50}+\frac{n}{2-\varepsilon}<(\frac{1}{2}+\varepsilon)n$$

and every independent set of G is contained in a member $C\in\mathcal{C}$.

As in the proof of Theorem 1.7 we can generate a random subset W of V by omitting vertices one by one, starting with V. Since n is large almost all sets W are of size $n/2+o(n)$. Moreover, for almost all of them the size of $W \cap C$ deviates from $|C|/2$ by at most, say, $\frac{\varepsilon}{100}n$ for all $C \in \mathcal{C}$, provided ε is sufficiently small. It suffices to show that with high probability the independence number of the induced subgraph on W is at most, say, $\alpha(G) - 2g(\varepsilon)n$. Since every independent set is contained in at least one of the members C of $\mathcal{C}$ it suffices to show that with high probability the independence number of the induced subgraph of G on $W \cap C$ is at most $\alpha(G) - 2g(\varepsilon)n$ for every $C \in \mathcal{C}$. Fix $C \in \mathcal{C}$. Without loss of generality its size is at least $n/8$ (since otherwise it cannot contain a large independent set at all). Recall that $|C| \leq (1/2+\varepsilon)n$. If the independence number of the induced subgraph of G on C is smaller than $(1/4+\varepsilon)|C|$ then so is the independence number of the induced subgraph on $W \cap C$, and this is smaller than $\alpha(G) - 0.1\varepsilon n$, implying the desired result. Otherwise, as in the proof of Theorem 1.7, in the random process that omits vertices of C one by one to get $W \cap C$, the number of times the independence number drops dominates stochastically a binomial random variable with parameters $\frac{\varepsilon}{2}|C|$ and ε. By the standard estimates for Binomial distributions (c.f., e.g., [3], Theorem A.1.13), the probability this variable is less than half its expectation is at most

$$e^{-\varepsilon^2|C|/16} \leq e^{-\varepsilon^2 n/128}.$$

By the union bound over all $C \in \mathcal{C}$ the probability this happens even for a single $C \in \mathcal{C}$ is at most

$$2^{H(\varepsilon^3/50)n} \cdot e^{-\varepsilon^2 n/128}$$

which, for small ε, tends to 0 as n tends to infinity. This shows that in this case ($d \geq \frac{50}{\varepsilon^4}$), with high probability the independence number of the induced subgraph of G on $W \cap C$ is smaller than $\alpha(G)$ by at least, say, $\varepsilon^2 n/40$, completing the proof of the proposition. $\square$

3 Remarks

- Conjecture 1.2 remains open for n-vertex graphs with independence number at most $n/2$ and for such regular graphs of independence number at most $n/4$. Similarly, Conjecture 1.6 remains open for n-vertex graphs with independence number at most $n/4$ and for such regular graphs with independence number at most $n/8$. Both conjectures appear to be significantly more difficult for graphs with independence number βn when $\beta > 0$ is a fixed small positive real.

- A conjecture I raised more than ten years ago motivated by some of the results in [2] is that the chromatic number of the graph $G_{m,t}$ described in the proof of Theorem

1.4, where $4t^2 \leq m$, is $\Theta(t^2)$. This has been mentioned in several lectures, see, for example, [1]. By the arguments described in the proof of Theorem 1.4 this chromatic number is at least $2t$ and at most $O(t^2)$.

Acknowledgment I thank Ehud Friedgut for helpful discussions and Zichao Dong and Zhuo Wu for telling me about [6].

References

[1] N. Alon, Graph Coloring: Local and Global, Public Lecture, Harvard, 2017, https://www.youtube.com/watch?v=lFD_DeWodn8

[2] N. Alon, A. Hassidim, E. Lubetzky, U. Stav and A. Weinstein, Broadcasting with side information, Proc. of the 49^{th} IEEE FOCS (2008), 823-832.

[3] N. Alon and J. H. Spencer, The Probabilistic Method, Fourth Edition, Wiley, 2016, xiv+375 pp.

[4] J. Balogh, R. Morris, and W. Samotij, Independent sets in hypergraphs, J. Amer. Math. Soc. 28(2015), 669–709.

[5] F. Chung and R. L. Graham, Erdős on Graphs, His Legacy of Unsolved Problems, A K Peters, Ltd., Wellesley, MA, 1998. xiv+142 pp.

[6] Z. Dong and Z. Wu, On the stability of graph independence number, arXiv:2102.13306v2, 2021.

[7] P. Erdős, Problems and results on set systems and hypergraphs, Extremal problems for finite sets (Visegrád, 1991), 217–227, Bolyai Soc. Math. Stud., 3, János Bolyai Math. Soc., Budapest, 1994

[8] E. Friedgut, G. Kalai and G. Kindler, The success probability in Lionel Levine's hat problem is strictly decreasing with the number of players, and this is related to interesting questions regarding Hamming powers of Kneser graphs and independent sets in random subgraphs, arXiv:2103.01541v1, March 2, 2021.

[9] A Hajnal, A theorem on k-saturated graphs, Canadian J. Math., 17 (1965), 720–724.

[10] D. J. Kleitman, On a combinatorial conjecture of Erdős, J. Combinatorial Theory 1 (1966), 209–214.

[11] L. Rabern, On hitting all maximum cliques with an independent set, arXiv:0907.3705, 2009.

[12] D. Saxton and A. Thomason, Hypergraph containers, Invent. Math. 201 (2015), 925–992.