

Extremal problems for independent set enumeration

Jonathan Cutler

Montclair State University
New Jersey, U.S.A.

jonathan.cutler@montclair.edu

A.J. Radcliffe

University of Nebraska-Lincoln
Nebraska, U.S.A.

aradcliffe1@math.unl.edu

Submitted: Jun 20, 2011; Accepted: Aug 16, 2011; Published: Aug 26, 2011

Mathematics Subject Classification: 05C35, 05C69

Abstract

The study of the number of independent sets in a graph has a rich history. Recently, Kahn proved that disjoint unions of $K_{r,r}$'s have the maximum number of independent sets amongst r -regular bipartite graphs. Zhao extended this to all r -regular graphs. If we instead restrict the class of graphs to those on a fixed number of vertices and edges, then the Kruskal-Katona theorem implies that the graph with the maximum number of independent sets is the lex graph, where edges form an initial segment of the lexicographic ordering. In this paper, we study three related questions. Firstly, we prove that the lex graph has the maximum number of weighted independent sets for any appropriate weighting. Secondly, we solve the problem of maximizing the number of independent sets in graphs with specified independence number or clique number. Finally, for $m \leq n$, we find the graphs with the minimum number of independent sets for graphs with n vertices and m edges.

1 Introduction

The study of independent sets in graphs has a long and rich history. We write $\mathcal{I}(G)$ for the set of independent sets in a graph G and $i(G) = |\mathcal{I}(G)|$ for the number of independent sets. Much recent research has focused on the problem of maximizing the number of independent sets in graphs with certain restrictions. For example, Kahn [9] gave the following upper bound for regular bipartite graphs.

Theorem 1.1 (Kahn). *Let G be an r -regular bipartite graph on n vertices. Then*

$$i(G) \leq (2^{r+1} - 1)^{n/2r}.$$

The result is sharp for disjoint unions of copies of $K_{r,r}$. Zhao [19] proved a conjecture of Kahn by extending Theorem 1.1 to all r -regular graphs. Kahn was also interested in

weighted independent sets. Assign a weight of λ^k (for some $\lambda > 0$) to an independent set of size k and set $i^\lambda(G)$ to be the total weight of all independent sets in G . Kahn also showed that if G is an r -regular bipartite graph on n vertices, and $\lambda \geq 1$, then

$$i^\lambda(G) \leq i^\lambda(K_{r,r})^{n/2r}.$$

Galvin and Tetali [7] extended this result to all positive weights.

In this paper, we will be concerned with the number of independent sets in graphs from a variety of graph classes. As noted in [4], one can prove, as a consequence of the Kruskal-Katona theorem [11, 10], that the graph on n vertices having m edges that has the largest number of independent sets is the *lex graph*, denoted $L(n, m)$, which we define in the next section.

Theorem 1.2. *If G is a graph on n vertices and m edges, then*

$$i(G) \leq i(L(n, m)).$$

This result was also proved by Wood [17], though he was working in the framework of counting cliques, i.e., complete subgraphs.

This paper will focus on three independent set enumeration problems. In Section 2, we extend Theorem 1.2 to the enumeration of independent sets of fixed size and deduce a theorem about weighted independent set enumeration. We also note some consequences for homomorphism enumeration. In Section 3, we extend our study to classes of graphs determined by bounds on their independence number and clique number. In Section 4, we will study the problem of minimizing the number of independent sets among graphs of fixed order and size.

2 Consequences of the Kruskal-Katona theorem

The Kruskal-Katona theorem [11, 10] has many consequences for the study of independent sets in graphs. We summarize some of them in this section, after introducing the lex orderings.

Definition 1. The *lexicographic* (or *lex*) ordering on subsets of \mathbb{N} is defined as follows. Given $A, B \subseteq \mathbb{N}$ we say A precedes B in lex, written $A \prec_L B$, if $\min(A \triangle B) \in A$.

Further, we define the lex graph with n vertices and m edges, denoted $L(n, m)$, to be the graph with vertex set $[n]$ and edge set the first m elements of $\binom{[n]}{2}$ under the lex ordering. The first few elements of the lex order on $\binom{[n]}{2}$ are

$$\{1, 2\}, \{1, 3\}, \{1, 4\}, \dots, \{1, n\}, \{2, 3\}, \{2, 4\}, \dots, \{2, n\}, \{3, 4\}, \dots,$$

and so the lex graph evolves by successively saturating the vertices $1, 2, 3, \dots, n$ in order.

Our concern will mostly be with the lex ordering restricted to the level sets $\binom{[n]}{k}$. The Kruskal-Katona theorem says that among all subsets of $\binom{[n]}{k}$ of given size the one with the

smallest “upper shadow” on levels above k is the initial segment of lex^1 . The following definition makes the notion of shadow precise.

Definition 2. If $\mathcal{A} \subseteq \binom{[n]}{k}$ and $\ell \geq k$ then the *upper ℓ -shadow* of \mathcal{A} is the collection

$$\partial^\ell(\mathcal{A}) = \left\{ B \in \binom{[n]}{\ell} : \text{for some } A \in \mathcal{A} \text{ we have } B \supseteq A \right\}.$$

Theorem 2.1 (Kruskal-Katona). *If $\mathcal{A} \subseteq \binom{[n]}{k}$ has size a and $\mathcal{L}(a) \subseteq \binom{[n]}{k}$ is the initial segment of size a in the lex order on $\binom{[n]}{k}$ then for all $\ell \leq k$ we have*

$$|\partial^\ell(\mathcal{A})| \geq |\partial^\ell(\mathcal{L}(a))|.$$

We derive below some consequences of Theorem 2.1 for extremal problems in independent set enumeration. The study of independent sets of a fixed size has generated interest when the graphs considered are regular, but seem to have not been well-studied if the graphs have a fixed number of edges. Carroll, Galvin, and Tetali [2] gave asymptotics on the maximum number of independent sets of a fixed size in regular graphs. Theorem 2.2 shows that the lex graph maximizes the number of independent sets of a fixed size, and so also maximizes weighted independent set counts for arbitrary weight functions depending only on the size of the set.

Definition 3. If $w : \mathbb{N} \cup \{0\} \rightarrow [0, \infty)$ is any function and G is a graph then we set

$$i_w(G) = \sum_{I \in \mathcal{I}(G)} w(|I|).$$

If $k \in \mathbb{N} \cup \{0\}$ then in particular we define w_k by $w_k(i) = \mathbb{1}(i = k)$, and

$$i_k(G) = i_{w_k}(G) = |\{I \in \mathcal{I}(G) : |I| = k\}|.$$

Theorem 2.2. *Let G be a graph on n vertices with m edges. If $w : \mathbb{N} \cup \{0\} \rightarrow [0, \infty)$ is any function then*

$$i_w(G) \leq i_w(L(n, m)).$$

Proof. It clearly suffices to prove the special case where $w = w_k$ for some $k \in \mathbb{N}$, since any other w can be expressed as a positive linear combination of these. We may also assume that $V(G) = [n]$. Now let $I_k = \left\{ A \in \binom{[n]}{k} : A \text{ is independent in } G \right\}$ and $i_k = |I_k|$. We have $I \in I_k$ iff $I \in \binom{[n]}{k} \setminus \partial^k(E(G))$, so

$$i_k = \binom{n}{k} - |\partial^k(E(G))| \leq \binom{n}{k} - |\partial^k(L(n, m))| = i_k(L(n, m)).$$

□

¹The Kruskal-Katona theorem is more usually stated in terms of lower shadows, and initial segments in the colex order, defined for finite subsets of \mathbb{N} by $A \prec_C B$ if $\max(A \triangle B) \in B$. The two forms are equivalent by a double reversal: complementation of sets and reversing the ordering on $[n]$.

A natural set of weight functions arises by considering independent set enumeration as a special case of homomorphism enumeration. For graphs G and H , let $\text{Hom}(G, H)$ be the set of homomorphisms from G to H and $\text{hom}(G, H) = |\text{Hom}(G, H)|$. Letting H_I be the path on two vertices, with one vertex looped and the other unlooped (and vertices labeled as in Figure 1), we see that $i(G) = \text{hom}(G, H_I)$. This is because there is a one-to-one correspondence between independent sets A and homomorphisms ϕ , given by $\phi^{-1}(\{a\}) = A$. In the context of statistical mechanics, it is natural to weight homomorphisms in the



Figure 1: The graph H_I

following manner: each vertex x in H is assigned a weight $\beta(x)$. Then we “count” the number of homomorphisms from G to a graph H with weight function $\beta : V(H) \rightarrow [0, \infty)$ as follows:

$$\text{hom}_\beta(G, H) = \sum_{\phi \in \text{Hom}(G, H)} \prod_{v \in G} \beta(\phi(v)).$$

This is the partition function of G for the model specified by H . Note that if $\beta \equiv 1$, then $\text{hom}_\beta(G, H) = \text{hom}(G, H)$. Let β^λ be the weight function on H_I defined as follows.

$$\beta^\lambda(x) = \begin{cases} \lambda & \text{if } x = a \\ 1 & \text{if } x = b. \end{cases}$$

Finally, note that $i^\lambda(G) = \text{hom}_{\beta^\lambda}(G, H_I)$. Note that an independent set of size k is assigned weight λ^k . Thus, the following is a corollary of Theorem 2.2.

Corollary 2.3. *If G is a graph on n vertices and m edges and $\lambda > 0$, then*

$$i^\lambda(G) \leq i^\lambda(L(n, m)).$$

For more general image graphs H , the problem of finding the graph G with a given number of vertices and edges that maximizes $\text{hom}(G, H)$ is usually difficult. In [4], the authors solved the case where H is the Widom-Rowlinson graph, a P_3 with every vertex looped. In [5], the similar case where one end vertex of the P_3 is unlooped is solved. For one class of image graphs, it follows from Corollary 2.3 that the lex graph is an extremal graph. Let $S(p, q)$ be the clique-looped split graph $K_p \vee E_q$, in which each vertex of the K_p is looped.

Corollary 2.4. *Let $H = S(p, q)$ for $p, q \geq 1$ and G be a graph with n vertices and m edges. Then*

$$\text{hom}(G, H) \leq \text{hom}(L(n, m), H).$$

Proof. Set $\lambda = p/q$. Then, we have $\text{hom}(G, H) = q^n i^\lambda(G)$ and so the result follows from Corollary 2.3. \square

3 Classes defined by other parameters

While independent sets in regular graphs and graphs of fixed size have been well studied, there has been less work on the problem of maximizing the number of independent sets in classes of graphs defined by other graph parameters. In this section, we investigate classes defined by bounds on the independence number, $\alpha(G)$, and the clique number, $\omega(G)$.

The first problem we attack is that of determining

$$\max \{i(G) : n(G) = n, \alpha(G) \leq \alpha\}.$$

Definition 4. If G is any graph and S is a subset of $V(G)$, we let $i(G; S)$ be the number of independent sets in G containing S , and similarly $i_k(G; S)$ be the number of such independent sets of size k .

We will prove the following theorem.

Theorem 3.1. *If G is a graph with n vertices and $\alpha(G) \leq \alpha$, then*

$$i(G) \leq i(\overline{T_{n,\alpha}}),$$

where $T_{n,\alpha}$ is the Turán graph with α parts. That is,

$$i(G) \leq i(K_{n_1} \cup K_{n_2} \cup \cdots \cup K_{n_\alpha}),$$

where $\sum n_i = n$ and $n_1 \leq n_2 \leq \cdots \leq n_\alpha \leq n_1 + 1$.

This follows from a more detailed result that among all graphs with independence number at most α , the complement of the Turán graph $\overline{T_{n,\alpha}}$ has the maximum number of independent sets of each fixed size. This result is equivalent, by complementation, to the result of Zykov [20] which states that among all graphs with $\omega(G) \leq \alpha$, the Turán graph $T_{n,\alpha}$ has the maximum number of cliques of each fixed size. This result has been independently proved by many people, including Erdős [6], Sauer [15], Hadžiivanov [8], and Roman [14]. We give here a direct proof that is substantially shorter than earlier ones. Our proof is based on the fifth proof of Turán's theorem (the origin of which seems to be unknown) in *Proofs from the Book* [1].

Theorem 3.2. *If G is a graph with n vertices and $\alpha(G) \leq \alpha$ and $0 \leq k \leq n$, then*

$$i_k(G) \leq i_k(K_{n_1} \cup K_{n_2} \cup \cdots \cup K_{n_\alpha}),$$

where $\sum n_i = n$ and $n_1 \leq n_2 \leq \cdots \leq n_\alpha \leq n_1 + 1$.

Proof. Let G have n vertices, $\alpha(G) \leq \alpha$, $i_k(G)$ maximal, and, subject to these conditions, $e(G)$ maximal. It suffices to show that G is of the form $K_{m_1} \cup K_{m_2} \cup \cdots \cup K_{m_\alpha}$ for some $m_i \geq 0$ since for such G ,

$$i_k(G) = \sum_{A \subseteq \binom{[\alpha]}{k}} \prod_{j \in A} m_j \leq i_k(K_{n_1} \cup K_{n_2} \cup \cdots \cup K_{n_\alpha}).$$

We begin by showing that adjacency is an equivalence relation on $V(G)$. The symmetry and reflexivity of \sim are obvious, so transitivity is all that needs to be proved. Suppose that $u \sim v \sim w$, but $u \not\sim w$. We consider two cases depending on the relative sizes of $i_k(G; u)$, $i_k(G; v)$, and $i_k(G; w)$.

If $i_k(G; u) > i_k(G; v)$ (or $i_k(G; w) > i_k(G; v)$), then we change G by deleting v and then cloning u . To be precise, we form G' by deleting v from G and adding a new vertex u' where $N_{G'}(u') = N_G(u) \cup \{u\} \setminus \{w\}$. No independent set in G' contains both u and u' , since $u \sim u'$, and so $\alpha(G) \leq \alpha$. Further,

$$i_k(G) = i_k(G) - i_k(v) + i_k(u) > i_k(G),$$

contradicting the maximality of G .

On the other hand, if $i_k(G; v) \geq i_k(G; u), i_k(G; w)$, then we consider the graph in which we delete u and w and clone v twice. That is, we form G'' from $G - u - w$ by adding vertices v' and v'' joined to each other, v , and $N_G(v) \setminus \{u, w\}$. Again, no independent set in G'' contains any two of v, v' , and v'' since they are mutually adjacent, so $\alpha(G'') \leq \alpha$. We have

$$\begin{aligned} i_k(G'') &= i_k(G - u - w) + 2i_k(G; v) \\ &= i_k(G) - i_k(G; u) - i_k(G; w) + i_k(G; u, w) + 2i_k(G; v). \end{aligned}$$

If $i_k(G; u, w) > 0$, then $i_k(G'') > i_k(G)$, a contradiction. On the other hand, if $i_k(G; u, w) = 0$, then $G + uw$ has the same number of k -independent sets as G , but more edges, contradicting the edge-maximality of G . \square

The next natural problem is that of computing

$$\max \{i(G) : n(G) = n, \omega(G) \geq \omega\}.$$

Unfortunately, this question is trivial since once your graph has an ω -clique, it need not have any other edges, and so the extremal graph is $K_\omega \cup E_{n-\omega}$. However, the question becomes interesting if one insists that G has copies of K_ω “everywhere.” For a graph G and a vertex $v \in V(G)$, we define

$$\omega(G; v) = \max \{|K| : K \text{ is a clique in } G \text{ containing } v\}.$$

Theorem 3.3. *If G is a graph on n vertices such that $\omega(G; v) \geq \omega$ for every $v \in V(G)$, then*

$$i(G) \leq i(K_{\underbrace{1, 1, \dots, 1}_{\omega-1}, n-\omega+1}) = 2^{n-\omega+1} + \omega - 1.$$

Before proving Theorem 3.3, we need to introduce quasi-threshold graphs and compressions that make our graphs “more quasi-threshold.”

Definition 5. A graph is *quasi-threshold* if, inductively, it is a single vertex, a disjoint union of two non-empty quasi-threshold graphs, or the join of a vertex and a quasi-threshold graph.

Theorem 3.4 (Chvátal, Hammer [3]). *A graph is quasi-threshold if for every pair of adjacent vertices x and y , the closed neighborhoods of x and y are nested, i.e., either $N(x) \cup \{x\} \subseteq N(y) \cup \{y\}$ or $N(y) \cup \{y\} \subseteq N(x) \cup \{x\}$.*

Our proof of Theorem 3.3 proceeds by repeatedly applying a compression operator that produces nested closed neighborhoods.

Definition 6. Let G be a graph with vertex set $V(G)$, and suppose $x \sim y$. The choice of x and y defines a natural partition of $V(G) \setminus \{x, y\}$ into four parts: vertices which are adjacent only to x , vertices adjacent only to y , vertices adjacent to both and vertices adjacent to neither. We write

$$\begin{aligned} A_{x\bar{y}} &= \{v \in V(G) \setminus \{x, y\} : v \sim x, v \not\sim y\}, \\ A_{xy} &= \{v \in V(G) \setminus \{x, y\} : v \sim x, v \sim y\}, \text{ and} \\ A_{\bar{x}y} &= \{v \in V(G) \setminus \{x, y\} : v \not\sim x, v \sim y\}. \end{aligned}$$

The *compression of G from x to y* , denoted $G_{x \rightarrow y}$, is the graph obtained from G by deleting all edges between x and $A_{x\bar{y}}$ and adding all edges from y to $A_{x\bar{y}}$.

Lemma 3.5. *If G is a graph and $x \sim y$, then $i(G) \leq i(G_{x \rightarrow y})$. Also, if x and y do not have nested closed neighborhoods, we have $d_2(G) < d_2(G_{x \rightarrow y})$, where $d_2(G) = \sum_{v \in V(G)} (d(v))^2$.*

Proof. See [4]. □

Proof of Theorem 3.3. If $n = \omega$, the result is trivial, so we assume $n > \omega$. We will select G from the class of graphs with $\omega(G; v) \geq \omega$ for all $v \in V(G)$ and $i(G)$ maximal. Among these, we select a graph with the fewest number of edges. Among all such graphs, we pick one with $d_2(G)$ maximal. We first note that every edge of G is in a K_ω , for if e were not, then $G - e$ would be a candidate with fewer edges.

We now show that G is quasi-threshold. Consider a pair x, y with $x \sim y$ (unless $\omega = 0$ in which case the result is trivial). If the closed neighborhoods of x and y are not nested, then applying Lemma 3.5 we have $i(G_{x \rightarrow y}) \geq i(G)$, $e(G_{x \rightarrow y}) = e(G)$, and $d_2(G_{x \rightarrow y}) > d_2(G)$. Thus, if we can show that $\omega(G_{x \rightarrow y}; v) \geq \omega$ for all $v \in V(G_{x \rightarrow y}) = V(G)$, we have contradicted the choice of G . Let $v \in V(G)$ with $v \neq x$. Certainly there exists $K \subset V(G)$ inducing a clique with $v \in K$. If $x \notin K$, then K is also a clique in $G_{x \rightarrow y}$. If both x and y are in K , then $K \setminus \{x, y\} \subseteq A_{xy}$, so K is also a clique in $G_{x \rightarrow y}$. Also, if $x \in K$ but $y \notin K$, then $K' = K - x + y$ is a clique in $G_{x \rightarrow y}$. Finally, if $v = x$, by our earlier observation, there is a clique K'' containing the edge xy . Since all of the elements of K'' are in A_{xy} , we have that K'' is a clique in $G_{x \rightarrow y}$. Thus, G is quasi-threshold.

By the definition of quasi-threshold, G is either a union of non-empty quasi-threshold graphs or the join of a vertex and a quasi-threshold graph. In the first case, where

$G = G_1 \cup G_2$ with $n(G_1) = n_1$ and $n(G_2) = n_2$, we have

$$\begin{aligned} i(G) &= i(G_1)i(G_2) \\ &\leq (2^{n_1-\omega+1} + \omega - 1)(2^{n_2-\omega+1} + \omega - 1) \\ &= 2^{n-2\omega+2} + (\omega - 1)(2^{n_1-\omega+1} + 2^{n_2-\omega+1}) + (\omega - 1)^2 \\ &\leq 2^{n-2\omega+2} + (\omega - 1)(2^{n-2\omega+1} + 2) + (\omega - 1)^2 \\ &= (\omega + 1)(2^{n-2\omega+1} + \omega - 1). \end{aligned}$$

For $\omega > 2$, it is straightforward to show that this final expression is strictly less than $2^{n-\omega+1} + \omega - 1$. For $\omega = 2$, we still have strict inequality unless $n = 4$. In this case, $i(2K_2) = i(K_{1,3}) = 9$.

In the case where $G = x \vee G'$, every vertex in G' is contained in an $(\omega - 1)$ -clique. Thus,

$$i(G) = i(G') + 1 \leq 2^{n-1-(\omega-1)+1} + (\omega - 1) - 1 + 1 = 2^{n-\omega+1} + \omega - 1,$$

completing the proof. □

4 Minimizing the number of independent sets

Another natural question to ask relates to minimizing the number of independent sets in a graph of fixed order and size.

Problem 1. *Which graphs have the minimum number of independent sets amongst graphs on n vertices and m edges?*

Somewhat unexpectedly, the problem is related to the problem of maximizing the number of *maximal* independent sets. Historically, this problem was studied in the complement, following a question first posed by Erdős and Moser, see [13]. Let $f(n)$ be the maximum number of maximal independent sets in an n -vertex graph. Note that there is no restriction on the number of edges in the graph. The question was resolved independently by Miller and Muller [12] and by Moon and Moser [13]. Short proofs of this result were recently given by Vatter [16] and Wood [18].

Theorem 4.1. *If $n \geq 2$, then*

$$f(n) = \begin{cases} 3^{n/3} & \text{if } n \equiv 0 \pmod{3}, \\ 4 \cdot 3^{\lfloor n/3 \rfloor - 1} & \text{if } n \equiv 1 \pmod{3}, \\ 2 \cdot 3^{\lfloor n/3 \rfloor} & \text{if } n \equiv 2 \pmod{3}. \end{cases}$$

Further, they showed that the extremal graphs are unions of triangles, with possibly at most one K_2 or K_4 . It turns out that in the case where 3 divides n , the graph that maximizes the number of maximal independent sets actually minimizes the number of independent sets over graphs with n edges.

The aim of this section is to solve Problem 1 when $m \leq n$. We begin by proving a series of lemmas that will essentially define the extremal graphs. Our first lemma will be used to show that all components of an extremal graph must be unicyclic. We do this by eliminating dense components, which imply the existence of isolated vertices in this regime. We extend our previous notation and write $i(G; \bar{x})$ for the number of independent sets in G not containing x . Similarly, if G is a graph, and x and y are vertices of G , we write $i(G; x, \bar{y})$ for the number of independent sets containing x , but not y , and so on. We begin by proving a simple proposition.

Proposition 4.2. *If G is a graph, $S \subset V(G)$, and $x \notin S$ is a vertex of G , then*

$$i(G; S, x) \leq i(G; S, \bar{x}).$$

Proof. Let \mathcal{I}_0 be the collection of independent sets containing S but not x , and \mathcal{I}_1 be the collection of independent sets containing S and x . The map $\phi : \mathcal{I}_1 \rightarrow \mathcal{I}_0$ defined by $\phi(I) = I \setminus \{x\}$ is an injection. \square

Armed with this proposition, we can prove a lemma that implies that graphs with at most n edges cannot have components with more than one cycle.

Lemma 4.3. *Suppose that G contains an edge $e = xy$ and an isolated vertex z . Then $i(G) \geq i(G - xy + xz)$.*

Proof. Let $H = G - z - e$ and note that $i(G) = 2(i(H; \bar{x}, \bar{y}) + i(H; x, \bar{y}) + i(H; \bar{x}, y))$. Also in terms of G , we note that $i(G - xy + xz) = 2i(H; \bar{x}, \bar{y}) + 2i(H; \bar{x}, y) + i(H; x, \bar{y}) + i(H; x, y)$. Thus,

$$\begin{aligned} i(G) - i(G - xy + xz) &= 2(i(H; \bar{x}, \bar{y}) + i(H; x, \bar{y}) + i(H; \bar{x}, y)) \\ &\quad - 2i(H; \bar{x}, \bar{y}) - 2i(H; \bar{x}, y) - i(H; x, \bar{y}) - i(H; x, y) \\ &= i(H; x, \bar{y}) - i(H; x, y) \\ &\geq 0, \end{aligned}$$

since $i(H; x, y) \leq i(H; x, \bar{y})$ by Proposition 4.2. \square

Thus, by repeatedly applying Lemma 4.3, we may assume that our extremal graphs do not have components with more than one cycle. In fact, this lemma is enough to show the result for $m \leq n/2$. The extremal example consists of independent edges and isolated vertices. In the regime $n/2 < m \leq n$, things get a bit more interesting. In the proof, we will show that our components are unicyclic. The components, then, each consist of a cycle with trees off of the cycle. We now prove a series of lemmas to narrow the possibilities for the structure of these trees. The vertex labels in each lemma refer to the associated figure. Recall that a *cutvertex* for a graph G is a vertex x such that $G - x$ has more components than G .

Lemma 4.4. *Suppose G is graph containing a cutvertex x such that $G - x$ contains a new P_3 -component of the type in Figure 2. Then $i(G) \geq i(G - xy + wy)$.*

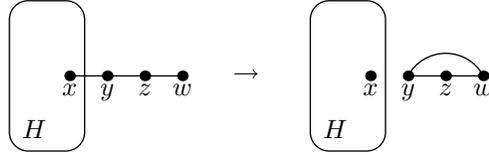


Figure 2: Lemma 4.4

Proof. We compare G to $G - xy + wy$, effectively disconnecting the pendant path and replacing it with a disjoint triangle. Let $H = G - w - y - z$. We have

$$\begin{aligned}
 i(G) - i(G - xy + wy) &= 3i(H; x) + 5i(H; \bar{x}) - 4i(H) \\
 &= i(H; \bar{x}) - i(H; x) \\
 &\geq 0,
 \end{aligned}$$

by Proposition 4.2. □

Lemma 4.5. *Suppose G is graph containing a cutvertex x such that $G - x$ contains a new P_3 -component of the type in Figure 3. Then $i(G) \geq i(G - wx + yz)$.*

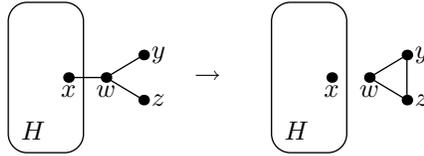


Figure 3: Lemma 4.5

Proof. Again, we compare the graphs in the lemma, and note that we are replacing a pendant tree with a disjoint triangle. Let $H = G - w - y - z$. We see that

$$\begin{aligned}
 i(G) - i(G - wx + yz) &= 4i(H; x) + 5i(H; \bar{x}) - 4i(H) \\
 &= i(H; \bar{x}) \\
 &\geq 0.
 \end{aligned}$$

□

Lemma 4.6. *Suppose G is a graph containing a cutvertex x such that $G - x$ contains a new isolated vertex and a new P_2 -component as in Figure 4. Then $i(G) \geq i(G - xy - xw + wy + wz)$.*

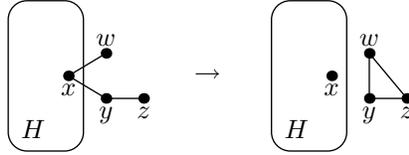


Figure 4: Lemma 4.6

Proof. We are again essentially replacing the two pendant paths by a disjoint triangle. Let $H = G - y - z - w$. We see that

$$\begin{aligned} i(G) - i(G - xy - xw + wy + wz) &= 2i(H; x) + 6i(H; \bar{x}) - 4i(H) \\ &= 2i(H; \bar{x}) - 2i(H; x) \\ &\geq 0, \end{aligned}$$

by Proposition 4.2. □

Lemma 4.7. *Suppose G is a graph with leaves u and v having a common neighbor. (See Figure 5.) Then $i(G) \geq i(G - xv + uv)$.*

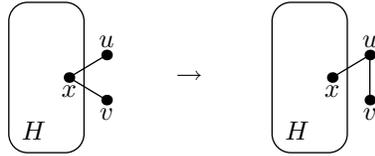


Figure 5: Lemma 4.7

Proof. Here, we replace two short leaf paths with one longer leaf path. Let $H = G - u - v$. We have

$$\begin{aligned} i(G) - i(G - xv + uv) &= 4i(H; \bar{x}) + i(H; x) - 3i(H; \bar{x}) - 2i(H; x) \\ &= i(H; \bar{x}) - i(H; x) \\ &\geq 0, \end{aligned}$$

by Proposition 4.2. □

Lemma 4.8. *Let G be a graph with leaf v adjacent to a vertex u of degree three as in Figure 6. Then $i(G) \geq i(G - uy + vy)$.*

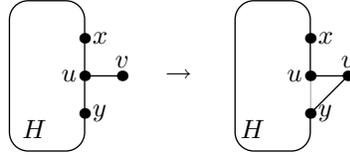


Figure 6: Lemma 4.8

Proof. Let $H = G - u - v$. We see that

$$\begin{aligned}
 i(G) - i(G - uy + vy) &= 3i(H; \bar{x}, \bar{y}) + 2(i(H; x, \bar{y}) + i(H; \bar{x}, y) + i(H; x, y)) \\
 &\quad - 3i(H; \bar{x}, \bar{y}) - 2(i(H; \bar{x}, y) + i(H; x, \bar{y})) - i(H; x, y) \\
 &= i(H; x, y) \\
 &\geq 0.
 \end{aligned}$$

□

Lemma 4.9. *Suppose G contains the structure from Figure 7 where x and y are allowed to be adjacent or non-adjacent. Then $i(G) \geq i(G - xu - yw + uw + vw)$.*

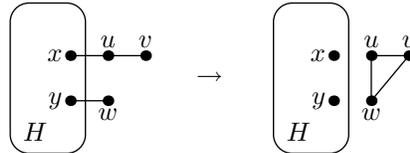


Figure 7: Lemma 4.9

Proof. Again, we are replacing leaf paths with a disjoint triangle. Let $H = G - u - v - w$. We see that

$$\begin{aligned}
 i(G) - i(G - xu - yw + uw + vw) &= 6i(H; \bar{x}, \bar{y}) + 3i(H; x, \bar{y}) + 2i(H; \bar{x}, y) \\
 &\quad + 2i(H; x, y) - 3i(H) \\
 &= 3i(H; \bar{x}, \bar{y}) - i(H; x, \bar{y}) - i(H; x, y) \\
 &\geq 0,
 \end{aligned}$$

by Proposition 4.2. □

Our next lemma states that there is a relationship between the number of independent sets in paths and cycles and the Fibonacci numbers. We let F_n be the n^{th} Fibonacci number so that $F_1 = 1$, $F_2 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for $n \geq 3$. Similarly, L_n is the n^{th} Lucas number, i.e., $L_1 = 1$, $L_2 = 3$, and $L_n = L_{n-1} + L_{n-2}$ for $n \geq 3$. It is easy to see that $L_n = F_{n-1} + F_{n+1}$ for $n \geq 2$. The following lemma is well-known. We include a proof for completeness.

Lemma 4.10. For any integer $n \geq 1$, we have $i(P_n) = F_{n+2}$. For any integer $n \geq 3$, it is the case that $i(C_n) = L_n$.

Proof. The first half follows from the fact that $i(P_1) = 2 = F_3$, $i(P_2) = 3 = F_4$, and the fact that

$$i(P_n) = i(P_{n-1}) + i(P_{n-2}),$$

which follows by considering independent sets containing an endpoint, and those that do not. For the second half, note that

$$i(C_n) = i(P_{n-1}) + i(P_{n-3}) = F_{n+1} + F_{n-1} = L_n.$$

□

Lemma 4.11. If n is an integer with $n \geq 6$, then $i(C_n) \geq i(C_3 \cup C_{n-3})$.

Proof. We have

$$i(C_n) = L_n = 3L_{n-3} + 2L_{n-4} > 4L_{n-3} = i(C_3)i(C_{n-3}) = i(C_3 \cup C_{n-3}).$$

□

Lemma 4.12. Let G consist of a C_n , with $n \geq 4$, together with a pendant path of length two. (See Figure 8). Then $i(G) \geq i(C_3 \cup C_{n-1})$.

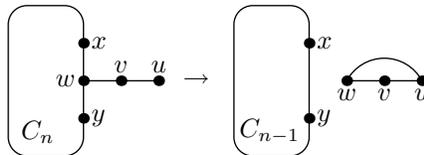


Figure 8: Lemma 4.12

Proof. We have

$$i(G) = 2L_n + F_{n+1} = 8F_{n-1} + 3F_{n-2},$$

and

$$i(C_3 \cup C_{n-1}) = 4L_{n-1} = 4F_{n-1} + 8F_{n-2}.$$

Thus,

$$i(G) - i(C_3 \cup C_{n-1}) = 4F_{n-1} - 5F_{n-2} > 0.$$

□

G	$i(G)$	G'	$i(G')$
$C_4 \cup C_5$	77	$3C_3$	64
$C_4 \cup C_4$	49	$C_3 \cup C_5$	44
$C_5 \cup C_5$	121	$2C_3 \cup C_4$	112
$C_5 \cup P_2$	33	$C_3 \cup P_4$	28
$C_4 \cup P_2$	21	$C_3 \cup P_3$	20
$C_4 \cup P_3$	35	$C_3 \cup P_4$	28
$C_4 \cup P_4$	56	$2C_3 \cup P_2$	48
$C_5 \cup P_3$	55	$2C_3 \cup P_2$	48
$C_5 \cup P_4$	88	$2C_3 \cup P_3$	80
$2P_3$	25	$P_4 \cup P_2$	24
$P_3 \cup P_4$	40	$C_3 \cup 2P_2$	36
$P_4 \cup P_4$	64	$C_3 \cup P_3 \cup P_2$	60

Table 1: A comparison of small graphs. In each case, $i(G) > i(G')$.

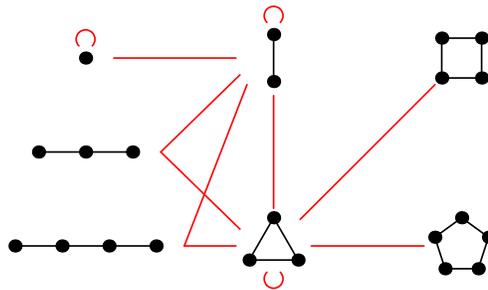


Figure 9: The compatibility graph for components of an extremal graph

Before stating the main theorem, we must define the extremal graphs. Our lemmas will allow us to show that each component of the extremal graph is either a vertex or one of P_2 , P_3 , P_4 , C_3 , C_4 , or C_5 . Some combinations of these graphs are ruled out. Table 1 lists the incompatible combinations of these graphs, giving in the second column a replacement having fewer independent sets. Figure 9 shows the compatibility graph for these, where edges (including loops) are between small graphs that can exist together as components of an extremal graph. Let \mathcal{Z} be the set of all graphs, each component of which is either a vertex or one of P_2 , P_3 , P_4 , C_3 , C_4 , or C_5 , such that each pair of components is compatible according to the graph in Figure 9. We prove that for each n and $0 \leq m \leq n$, there is a unique element of \mathcal{Z} with n vertices and m edges, which we denote $Z(n, m)$. Essentially, if $0 \leq m < n/2$, then $Z(n, m)$ consists of independent edges and isolated vertices. As m increases from $n/2$ to n , there are no longer any isolated vertices, and $Z(n, m)$ consists mainly of independent edges and triangles, along with at most one P_3 or P_4 . When $m = n$, the graph $Z(n, m)$ is disjoint triangles, along with at most one C_4 or C_5 .

Proposition 4.13. *For each n and $0 \leq m \leq n$, there is a unique element of \mathcal{Z} with n vertices and m edges.*

Proof. The claim is trivial when $m \leq n/2$ or $m = n$. For $n/2 < m < n$ the number of path components is equal to $n - m$, so, counting vertices, we have

$$3c + 2(n - m) + \epsilon = n, \quad \text{i.e.,} \quad 3c + \epsilon = 2m - n,$$

where c is the number of triangles and $\epsilon = 0, 1, 2$ corresponds to the presence of only P_2 's, one P_3 , or one P_4 . This determines c and ϵ . \square

We are finally able to state and prove the main theorem of this section.

Theorem 4.14. *Let n be a positive integer and m be an integer with $0 \leq m \leq n$. If G is a graph on n vertices and m edges, then*

$$i(G) \geq i(Z(n, m)).$$

Proof. Suppose that G is a graph on n vertices and m edges. If $m \leq n/2$ and G contains any component that is not an edge or an isolated vertex, then we can apply Lemma 4.3 repeatedly to any edge of a component with more than one edge until we are left with $Z(n, m)$.

If $n/2 < m \leq n$, then we can apply Lemma 4.3 repeatedly so that every component of G is unicyclic or a tree, since the existence of a component with more than one cycle implies the existence of an isolated vertex in this regime. Each unicyclic component of G consists of a cycle with pendant trees off of the cycle. We define a *leaf path* as a path starting at a leaf with all internal vertices having degree two. If any leaf path is of length three or more, we can apply Lemma 4.4. Thus, every leaf path has length one or two. If there are two of length two with the same endpoint, we can apply Lemma 4.9. If there is one of length two and one of length one, we can apply Lemma 4.6. If there are two of length one, we can apply either Lemma 4.5 or Lemma 4.7. When applying Lemma 4.7, we may create one of the earlier situations again, and so can go through the same process again. At every stage, the number of vertices in the 2-core minus the number of leaves goes strictly up, so the process must eventually terminate; we show that we end up with $Z(n, m)$.

At this point, we are left with a graph in which every unicyclic component consists of a cycle with pendant paths of length one or two. If there are any cycle vertices of degree three with one pendant path of length one, then we can apply Lemma 4.8. If there is more than one pendant path of length two, then we can apply Lemma 4.9. Thus, the unicyclic components that are left are either cycles or have one pendant path of length two. In the latter case, we can apply Lemma 4.12. So, every unicyclic component left is a cycle and we can apply Lemma 4.11 until all cycles are of length at most five.

To finish, we simply use the results compiled in Table 1 to get that the graph contains no pair of incompatible components. This implies that G is $Z(n, m)$, which completes the proof of the theorem. \square

The problem of minimizing $i(G)$ over graphs with n vertices and m edges remains open for $m > n$. We conjecture that if $m = \frac{n}{r} \binom{r}{2}$ and r divides n , then the extremal graph is $\frac{n}{r} K_r$.

References

- [1] Martin Aigner and Günter M. Ziegler, *Proofs from The Book*, fourth ed., Springer-Verlag, Berlin, 2010.
- [2] Teena Carroll, David Galvin, and Prasad Tetali, *Matchings and independent sets of a fixed size in regular graphs*, J. Combin. Theory Ser. A **116** (2009), 1219–1227.
- [3] Václav Chvátal and Peter L. Hammer, *Aggregation of inequalities in integer programming*, Studies in integer programming (Proc. Workshop, Bonn, 1975), North-Holland, Amsterdam, 1977, pp. 145–162. Ann. of Discrete Math., Vol. 1.
- [4] Jonathan Cutler and A. J. Radcliffe, *Extremal graphs for homomorphisms*, J. Graph Theory **67** (2011), no. 4, 261–284.
- [5] ———, *Extremal graphs for homomorphisms II*, preprint (2011).
- [6] P. Erdős, *On the number of complete subgraphs contained in certain graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **7** (1962), 459–464.
- [7] David Galvin and Prasad Tetali, *On weighted graph homomorphisms*, Graphs, morphisms and statistical physics, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 63, Amer. Math. Soc., Providence, RI, 2004, pp. 97–104.
- [8] N. G. Hadžiivanov, *A generalization of Turán’s theorem on graphs*, C. R. Acad. Bulgare Sci. **29** (1976), no. 11, 1567–1570.
- [9] Jeff Kahn, *An entropy approach to the hard-core model on bipartite graphs*, Combin. Probab. Comput. **10** (2001), no. 3, 219–237.
- [10] G. Katona, *A theorem of finite sets*, Theory of graphs (Proc. Colloq., Tihany, 1966), Academic Press, New York, 1968, pp. 187–207.
- [11] Joseph B. Kruskal, *The number of simplices in a complex*, Mathematical optimization techniques, Univ. of California Press, Berkeley, Calif., 1963, pp. 251–278.
- [12] R. E. Miller and D. E. Muller, *A problem of maximum consistent subsets*, IBM Research Report RC-240, J. T. Watson Research Center, 1960.
- [13] J. W. Moon and L. Moser, *On cliques in graphs*, Israel J. Math. **3** (1965), 23–28.
- [14] Steven Roman, *The maximum number of q -cliques in a graph with no p -clique*, Discrete Math. **14** (1976), no. 4, 365–371.
- [15] N. Sauer, *A generalization of a theorem of Turán*, J. Combinatorial Theory Ser. B **10** (1971), 109–112.
- [16] Vincent Vatter, *Maximal independent sets and separating covers*, Amer. Math. Monthly **118** (2011), 418–423.

- [17] David R. Wood, *On the maximum number of cliques in a graph*, Graphs Combin. **23** (2007), no. 3, 337–352.
- [18] ———, *On the number of maximal independent sets in a graph*, arXiv:1104.1243v2 [math.CO], 2011.
- [19] Yufei Zhao, *The number of independent sets in a regular graph*, Combin. Probab. Comput. **19** (2010), no. 2, 315–320.
- [20] A. A. Zykov, *On some properties of linear complexes*, Mat. Sbornik N.S. **24(66)** (1949), 163–188.