Les Cahiers du GERAD ISSN: 0711–2440

On the Maximum Orders of an Induced Forest, an Induced Tree, and a Stable Set

A. Hertz, O. Marcotte,
D. Schindl

G–2011–45

September 2011

On the Maximum Orders of an Induced Forest, an Induced Tree, and a Stable Set

Alain Hertz

GERAD & Département de mathématiques et génie industriel
École Polytechnique de Montréal
C.P. 6079, succ. Centre-ville
Montréal (Québec) Canada, H3C 3A7
alain.hertz@gerad.ca

Odile Marcotte

GERAD & Département d'informatique
UQAM
C.P. 8888, succ. Centre-ville
Montréal (Québec) Canada, H3C 3P8
odile.marcotte@gerad.ca

David Schindl

Haute école de gestion de Genève
Campus Battelle – Bâtiment F
7, route de Drize
1227 Carouge, Suisse
david.schindl@hesge.ch

September 2011

Les Cahiers du GERAD

G–2011–45

Abstract

Let G be a connected graph, n the order of G, and f (resp. t) the maximum order of an induced forest (resp. tree) in G. We show that $f - t$ is at most $n - \lceil 2\sqrt{n-1} \rceil$. In the special case where n is of the form $a^2 + 1$ for some even integer $a \geq 4$, $f - t$ is at most $n - \lceil 2\sqrt{n-1} \rceil - 1$. We also prove that these bounds are tight. In addition, letting α denote the stability number of G, we show that $\alpha - t$ is at most $n + 1 - \lceil 2\sqrt{2n} \rceil$; this bound is also tight.

Key Words: induced forest, induced tree, stability number, extremal graph theory.

Résumé

Soit G un graphe connexe, n l'ordre de G, et f (resp. t) l'ordre maximum d'une forêt induite (resp. d'un arbre induit) dans G. Dans le présent article nous montrons que la différence $f - t$ est au plus égale à $n - \lceil 2\sqrt{n-1} \rceil$. Dans le cas où n est de la forme $a^2 + 1$ pour un entier pair a au moins égal à 4, $f - t$ est au plus égale à $n - \lceil 2\sqrt{n-1} \rceil - 1$. Nous prouvons aussi que ces bornes sont les meilleures possibles pour un graphe G d'ordre n. De plus, si α dénote le nombre de stabilité de G, nous montrons que la différence $\alpha - t$ est au plus $n + 1 - \lceil 2\sqrt{2n} \rceil$; cette borne aussi est la meilleure possible.

Mots clés : forêt induite, arbre induit, nombre de stabilité, théorie des graphes extrémaux.

1 Introduction

In this article we study the relationships between three invariants of undirected graphs, namely, the maximum order of an induced forest, the stability number, and the maximum order of an induced tree. Although bounds on invariants such as these have been studied for a long time by graph theorists, the past few years have seen a surge of interest in the systematic study of linear relations (or other kinds of relations) between graph invariants. We focus our attention on the difference between the maximum order of an induced forest and the maximum order of an induced tree, give an upper bound on this difference, and prove that it is tight. A similar but simpler proof allows us to bound the difference between the stability number and the maximum order of an induced tree; in this case also we show that the bound is tight. In the rest of this section we review the relevant literature and recall some definitions. Section 2 contains our results on forests and Section 3 those on stable sets. We conclude in Section 4.

We now survey published work relevant to the present article. Erdős, Saks, and Sòs [4] addressed the problem of finding maximum induced trees in graphs. In particular, they proved that any graph G with n vertices and m edges contains an induced tree of order at least $2n/(m-n+3)$. Zheng and Lu [6] considered maximum induced forests and proved that in any cubic, connected, and triangle-free graph G, there is an induced forest of order at least $n - \lceil n/3 \rceil$ (provided n, the order of G, is at least 8). Alon, Mubayi, and Thomas [1] investigated the relationship between the order of an induced forest in a connected graph G, the stability number of G (denoted by α), and its maximum degree (denoted by Δ). They proved that a connected graph G of order n contains an induced forest of order at least $\alpha + (n-\alpha)/(\Delta-1)^2$.

DeLaViña and Waller [3] also studied bounds on the orders of an induced tree and an induced forest, respectively. Among other results, they showed that any connected graph G contains an induced tree of order at least $(\alpha+1)/\gamma$ (where γ denotes the domination number of G) and an induced forest of order at least $g + f_1 - 1$ (where g denotes the girth of G and f_1 the number of vertices of degree 1 in G). Recently, Fox, Loh, and Sudakov [5] proved that any connected triangle-free graph G of order n contains an induced tree of order at least $\sqrt{n}$. The authors also discuss the difference between the order of an induced forest and that of an induced tree, showing that the order of a largest guaranteed induced forest in a K_r-free graph grows in a polynomial fashion while the order of a largest guaranteed induced tree grows in a logarithmic fashion.

Let $G = (V, E)$ be a finite undirected graph, where V is the set of vertices of G and E its set of edges. The cardinality of V is also called the *order* of G and will be denoted by n. Two vertices u and v are said to be *adjacent* if $\{u, v\}$ (also denoted by uv or vu) belongs to E; u and v are called the *ends* of uv. A graph G is said to be *complete* if any two of its vertices are adjacent. For any subset U of V, the subgraph of G *induced* by U is the graph $H = (U, E(U))$, where $E(U)$ consists of those edges of G with both ends in U. A *clique* in G is a complete induced subgraph of G.

Given two vertices u and v of G, a simple path (or *path*) of length ℓ between u and v is a sequence $(u_0 = u, u_1, \ldots, u_\ell = v)$ of distinct vertices such that $u_i u_{i+1}$ is an edge of G for all $i \in \{0, 1, \ldots, \ell-1\}$. A *cycle* C is a sequence $(u_0, u_1, \ldots, u_{\ell-1})$ of distinct vertices such that $u_i u_{i+1}$ is an edge of G for all $i \in \{0, 1, \ldots \ell-1\}$ (where the addition is modulo ℓ). We say that G is *connected* if for any pair $\{u, v\}$ of vertices of G, there is a path between u and v.

If G is not connected, its vertex set can be partitioned into *connected components*, i.e., maximal induced subgraphs that are connected. A graph G is a *tree* if it is connected and has exactly $|V|-1$ edges. A graph G is a *forest* if every one of its connected components is a tree. A subset S of V is said to be *stable* if it induces a subgraph with no edges. The *stability number* of G (denoted by $\alpha(G)$ or α) is the maximum cardinality of a stable set in G. We refer the reader to Bondy and Murty [2] for any concept not defined here.

2 Forests and trees

Let G be an undirected graph and f (resp. t) the maximum order of an induced forest (resp. tree) in G. In order to find an upper bound for $f - t$, we must first investigate the relationship between an induced forest F (not necessarily of maximum order) and induced trees in G. In what follows we use F to denote either the induced subgraph of G or the vertex set of that subgraph. The following lemma is useful for bounding the difference between f and t. It is actually very similar to the claim proved in the conclusion of the article of Fox, Loh, and Sudakov [5]. The main differences between our lemma and the claim are that we consider a forest instead of a tree and the complement of the forest includes a single vertex.

Lemma 2.1 *Let $G = (V, E)$ be a connected graph and assume that $F = V\backslash\{u\}$ induces a forest for some vertex u of G. Then there exists an induced tree T in G containing u and whose order is at least $1 + \left\lceil \frac{|F|}{2} \right\rceil = 1 + \left\lceil \frac{|V|-1}{2} \right\rceil$.*

Proof. Let the components of F be denoted by $V_1, \ldots, V_p$. Because G is connected, each of the V_i (for $i \in \{1, \ldots, p\}$) contains at least one neighbour of u. In what follows we call the neighbours of u *black vertices*; the other vertices in $V\backslash\{u\}$ are *white vertices*. Let t_i be the number of black vertices in V_i; we denote these vertices by $u_{i1}, \ldots, u_{it_i}$, for $i \in \{1, \ldots, p\}$.

For each component V_i we construct a subset of vertices V_i' as follows. For $j = 2, \ldots, t_i$, let e_{ij} denote the last edge on the path between u_{i1} and u_{ij}. The removal of $E_i = \{e_{i2}, \ldots, e_{it_i}\}$ from the subgraph induced by V_i produces t_i trees denoted respectively by $T_{i1}, \ldots, T_{it_i}$, each of which containing one black vertex (i.e., u_{ij} belongs to T_{ij}). Let X_{ij} denote the vertex set of T_{ij}. We contract all the edges of all the trees T_{ij}. This amounts to creating a new graph H_i with t_i vertices $v_{i1}, \ldots, v_{it_i}$, where v_{ij} represents the set X_{ij}. In H_i there is an edge between v_{ij} and v_{ik} (for some $k \neq j$) if and only if e_{ij} has one end in X_{ik} or e_{ik} has one end in X_{ij}.

The graph H_i is a tree because the contraction operation cannot create any cycle in an acyclic graph. If we consider $|X_{ij}|$ to be the weight of w_{ij} (for every j), it is obvious that the sum of the weights of the w_{ij} equals $|V_i|$. The graph H_i being bipartite, its vertex set can be partitioned into two stable sets S_{i1} and S_{i2}. Without loss of generality, we assume that $\sum_{v_{ij} \in S_{i1}} |X_{ij}|$ is at least $\frac{|V_i|}{2}$. We then define V_i' as $\bigcup_{v_{ij} \in S_{i1}} X_{ij}$. It follows from our construction that the subgraph of G induced by V_i' is a forest, each connected component of which contains exactly one black vertex.

Let T be the union of $\{u\}$ and all the V_i', for $i = 1, \ldots, p$. We claim that T induces a tree satisfying the conclusion of the lemma. Indeed, adding vertex u and the edges uu_{ij} (for $i = 1, \ldots, p$ and $v_{ij} \in S_{i1}$) to the subgraph induced by the V_i' produces a connected graph without any cycle, i.e., a tree. Moreover, the choice of V_i' implies that

$$|T| = 1 + \sum_{i=1}^{p} |V_i'| \geq 1 + \sum_{i=1}^{p} \frac{|V_i|}{2} = 1 + \frac{|V| - 1}{2}.$$

Since $|T|$ is an integer, this completes the proof of the lemma. □

The construction used in the proof of Lemma 2.1 is illustrated in Figure 1. Graph G appears in 1.a, vertex u being represented by a square while the vertices of F are the black and white circles. The forest F induced by $V\backslash\{u\}$ has four connected components and the edges with bold lines are those in the sets E_i. The graphs $H_1, \ldots, H_4$ are displayed in 1.b along with the respective bipartitions of their vertex sets. The vertices in the sets $S_{11}, \ldots, S_{41}$ are displayed in black while those in the sets $S_{12}, \ldots, S_{42}$ are in grey. The final tree T is displayed in 1.c.

To prepare for the main ingredient of the proof, Lemma 2.2, we introduce some definitions and a system of inequalities. Let G be a connected graph and F any induced forest in G. We let K denote the complement of F (i.e., $V\backslash F$). For any pair $\{u, v\}$ of vertices in K, we choose a shortest path (P_{uv}) between u and v. Note that this path may contain vertices that are in F, since K need not induce a connected graph. For a vertex w in F, we denote by C_w the connected component of F that contains w, and by S_w the *attachment set* of w, i.e.,

$$\{u \in K \mid \exists w' \text{ such that } uw' \in E \text{ and } w' \in C_w\}.$$

Thus S_w is the set of vertices in K that are adjacent to at least one vertex in the component C_w. For any u in S_w, we say that w is *attached* to u. For an illustration, consider the graph in Figure 2, where the vertices

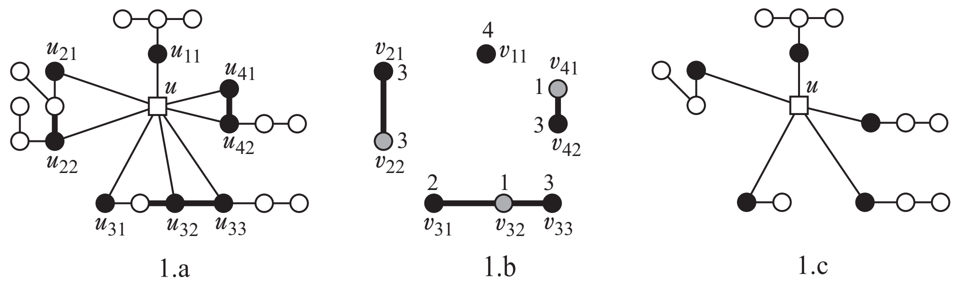

Figure 1: Illustration of Lemma 2.1.

in F are represented by circles and those in $K = \{a, b, c, d\}$ by squares. F has three connected components. The attachment set of every white (resp. grey, black) vertex is $\{b\}$ (resp. $\{a, c\}$, $\{b, c, d\}$).

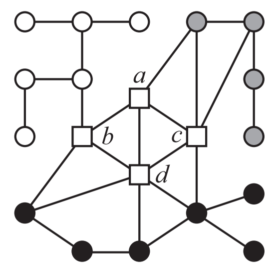

Figure 2: Illustration of attachment sets.

For any non empty subset S of K, we define x_S as the number of vertices w in F verifying $S_w = S$. We consider the x_S as variables appearing in a system of linear inequalities, the system (SLI), and we also introduce the variable Z. We describe two groups of constraints in the system (SLI). The first group contains $|K|$ constraints, each indexed by a vertex in K. The constraint corresponding to vertex $u \in K$ is

$$\sum_{S \text{ contains } u} x_S \leq Z + 2.$$

Note that the left-hand side of this inequality represents the number of vertices w in F that are attached to u. The second group contains $|K|(|K| - 1)/2$ constraints, each indexed by a pair of vertices in K. The constraint corresponding to the pair $\{u, v\}$ is

$$\sum_{S \cap P_{uv} = \{u\}} x_S + \sum_{S \cap P_{uv} = \{v\}} x_S \leq Z.$$

The left-hand side of this inequality represents the number of vertices w in F that are attached to u but no other vertex of P_{uv} **or** are attached to v but no other vertex of P_{uv}.

The system (SLI) consists of the two groups of constraints described above and the following constraint, stating that every vertex in F has a unique attachment set.

$$\sum_{S} x_S = |F|$$

The sum is taken over all non empty subsets S of K.

Lemma 2.2 *For any connected graph G of order n, any forest F in G, and any value Z satisfying the system (SLI), the relation $Z \geq 2(|F| - 2)/(n + 1 - |F|)$ holds. Moreover, if $Z = 2(|F| - 2)/(n + 1 - |F|)$ holds, then every constraint in (SLI) is satisfied at equality.*

Proof. We claim that each variable x_S appears in at least $|K| = n - |F|$ inequality constraints of (SLI). More precisely, it appears exactly $|S|$ times (one time for each vertex in S) in an inequality of the first group and at least $|K| - |S|$ times in an inequality of the second group. Indeed, if u is any vertex in $K \backslash S$, we choose a vertex v in S that minimizes the length of P_{uv}. Then the variable x_S appears in the inequality constraint corresponding to the pair $\{u, v\}$, because the intersection of S and P_{uv} equals $\{v\}$. Hence, for any u in $K \backslash S$, there is at least one constraint in the second group where x_S appears.

Thus if we add all the inequality constraints in the first and second groups, we obtain an inequality whose left-hand side is at least $(n - |F|) \sum_S x_S$ and right-hand side equals

$$(n - |F|)(Z + 2) + \frac{(n - |F|)(n - |F| - 1)}{2} Z.$$

Since the equality $\sum_S x_S = |F|$ holds, we obtain

$$(n - |F|)|F| \leq (n - |F|)(Z + 2) + \frac{(n - |F|)(n - |F| - 1)}{2} Z,$$

which yields

$$Z \geq \frac{2(|F| - 2)}{n + 1 - |F|}.$$

The second part of the lemma follows easily from the above derivation. □

Theorem 2.3 *For any connected graph G of order n and any forest F in G, there exists an induced tree in G whose order is at least equal to*

$$\left\lceil \frac{|F| - 2}{n + 1 - |F|} \right\rceil + 2.$$

Proof. Let us denote by $Z_{\min}$ the smallest value of Z for which all the constraints of (SLI) are satisfied. Then there is at least one "tight" constraint in which $Z_{\min}$ appears.

1. If this constraint belongs to the first group, there is a vertex u in K such that $Z_{\min} + 2$ vertices in F are attached to u. By Lemma 2.1, there exists a tree in G whose order is at least $1 + \lceil (Z_{\min} + 2)/2 \rceil = 2 + \lceil (Z_{\min}/2) \rceil$.
2. If this constraint belongs to the second group, there is a pair of vertices $\{u, v\}$ such that $Z_{\min}$ vertices in F are attached to u but no other vertex of P_{uv} or to v but no other vertex of P_{uv}. Let C_1 (resp. C_2) denote the set of vertices in F that are attached to u (resp. v) but no other vertex of P_{uv}. By Lemma 2.1 again, the subgraph induced by $C_1 \cup \{u\}$ contains a tree T_1 of order at least $1 + \lceil |C_1|/2 \rceil$ and the subgraph induced by $C_2 \cup \{v\}$ a tree T_2 of order at least $1 + \lceil |C_2|/2 \rceil$. By construction there is no edge joining any vertex in C_1 to any vertex in P_{uv} (except u) and no edge joining any vertex in C_2 to any vertex in P_{uv} (except v). Hence the union of T_1, T_2, and P_{uv} is an induced tree, of order at least $2 + \lceil |C_1|/2 \rceil + \lceil |C_2|/2 \rceil \geq 2 + \lceil (|C_1| + |C_2|)/2 \rceil \geq 2 + \lceil |Z_{\min}|/2 \rceil$.

We conclude that G always contains an induced tree whose order is at least equal to

$$\left\lceil \frac{Z_{\min}}{2} \right\rceil + 2 \geq \left\lceil \frac{|F| - 2}{n + 1 - |F|} \right\rceil + 2,$$

where the inequality follows from Lemma 2.2. □

Corollary 2.4 *The relation $f - t \leq n - \lceil 2\sqrt{n-1} \rceil$ holds for any connected graph G of order n.*

Proof. Assume that F is a forest of maximal order, i.e., of order f. The previous theorem implies that

$$f - t \leq f - \frac{f - 2}{n + 1 - f} - 2,$$

and the maximum value of the right-hand side can be derived by studying an equation in f. Indeed, the derivative of the right-hand side with respect to f equals

$$1 - \frac{(n-1)}{(n + 1 - f)^2}.$$

The only value of f not exceeding n for which the derivative equals 0 is $n+1-\sqrt{n-1}$, which maximizes the value of $f-(f-2)/(n+1-f)$ since this function is concave. Substituting $n+1-\sqrt{n-1}$ for f in $f-(f-2)/(n+1-f)-2$ yields

$$f-t \leq n-2\sqrt{n-1}.$$

The corollary follows by observing that $f-t$ is an integer. □

Theorem 2.5 *Let G be a connected graph of order n, where n is of the form a^2+1 for some even positive integer $a \geq 4$. Then we have $f-t \leq n-\lceil 2\sqrt{n-1}\rceil -1$.*

Proof. Let b denote $n-a-f$ (note that b may be a negative integer). Then we have

$$\begin{aligned} f-t &\leq f-\frac{f-2}{n+1-f}-2 = n-a-b-\left\lceil\frac{a^2-a-b-1}{a+b+1}\right\rceil-2 \\ &= n-a-b-\left\lceil\frac{(a-b-2)(a+b+1)+(b+1)^2}{a+b+1}\right\rceil-2 \\ &= n-2a-\left\lceil\frac{(b+1)^2}{a+b+1}\right\rceil. \end{aligned}$$

Thus if b does not equal -1, the relation $f-t \leq n-2a-1 = n-2\sqrt{n-1}-1$ holds and the theorem is proved.

Assume that b equals -1 (which implies that $(f-2)/(n+1-f)$ equals $a-1$). We know from the proof of Theorem 2.3 that G contains a tree of order $\lceil Z_{\min}/2\rceil+2$. If $Z_{\min} > 2(f-2)/(n-f+1)$ holds, we obtain $\lceil Z_{\min}/2\rceil+2 > (f-2)/(n-f+1)+2 = a+1$, which implies that G contains a tree of order at least $a+2$. Hence $f-t$ is at most $(n-a+1)-(a+2) = n-2a-1$, i.e., at most $n-2\sqrt{n-1}-1$.

Finally assume that $b=-1$ holds and $Z_{\min}$ equals $2(f-2)/(n-f+1) = 2(a-1)$. Since a is at least 4, $n-f$ is at least 3. Therefore the second group of inequalities in the system (SLI) is not empty, and by Lemma 2.2, every inequality in both groups is satisfied at equality. Let $\{u,v\}$ be a pair of vertices in K such that $P_{uv}\cap K$ equals $\{u,v\}$. Let F_u (resp. F_v) denote the set of vertices in F that are attached to u but not v (resp. v but not u), and F_{uv} the set of vertices in F that are attached to u and v. Then we have

$$|F_u|+|F_{uv}| = Z_{\min}+2, \quad |F_v|+|F_{uv}| = Z_{\min}+2, \quad |F_u|+|F_v| = Z_{\min},$$

which implies that $|F_u| = |F_v| = Z_{\min}/2$ and $|F_{uv}| = (Z_{\min}+4)/2$ hold. By Lemma 2.1 again, G contains an induced tree that includes u, v, $\lceil Z_{\min}/4\rceil$ vertices in F_u, and $\lceil Z_{\min}/4\rceil$ vertices in F_v. Therefore we have

$$f-t \leq (n-a+1)-\left(2+2\left\lceil\frac{Z_{\min}}{4}\right\rceil\right) = (n-a+1)-\left(2+2\left\lceil\frac{a-1}{2}\right\rceil\right).$$

Since a is even, the relations

$$f-t \leq (n-a+1)-(2+a) = n-2a-1 = n-2\sqrt{n-1}-1$$

hold. This completes the proof of the theorem. □

We now prove that the above bounds are tight. Note that 0 is a trivial lower bound for $f-t$.

Theorem 2.6 *Let n be an integer at least equal to 2. If n is of the form a^2+1 for some a with a even and $a \geq 4$, there exists a connected graph of order n for which $f-t = n-\lceil 2\sqrt{n-1}\rceil-1$ holds. Otherwise, there exists a connected graph of order n for which $f-t = n-\lceil 2\sqrt{n-1}\rceil$ holds.*

Proof. Let a denote $\lfloor\sqrt{n-1}\rfloor$. We describe a construction assuming that the value of f is known and smaller than n; we will give below the precise relation between f and a. We define a graph G whose vertex set includes $n-f$ vertices $u_1, u_2, \ldots, u_{n-f}$ and the vertices of a forest whose connected components are

$P_1, P_2, \ldots, P_{n+1-f}$. The set $\{u_1, u_2, \ldots, u_{n-f}\}$ induces a clique that is disjoint from the forest, and each component of the forest (i.e., each P_i) is a path. Each vertex of P_i (for $1 \le i \le n-f$) is joined by an edge to the vertex u_i. Each vertex of P_{n+1-f} is joined by an edge to every vertex in $\{u_1, u_2, \ldots, u_{n-f}\}$.

We now address the question of the cardinality of the P_i. Let q denote the largest even integer such that $q(n+1-f) \le f-2$ holds.

$$q = 2\left\lfloor \frac{f-2}{2(n+1-f)} \right\rfloor$$

Let r denote $f-2-q(n+1-f)$. By definition of q, r is strictly smaller than $2(n+1-f)$. In the first round we allocate q vertices to each of $P_1, P_2, \ldots, P_{n-f}$ and $q+2$ vertices to P_{n+1-f}. In the second round we add 2 vertices to $P_1, P_2, \ldots, P_{\lfloor r/2 \rfloor}$ and the last vertex (if r is odd) to $P_{\lceil r/2 \rceil}$. Let $|P_i|$ denote the order of P_i and s_i the number of vertices included into P_i during the second round. Clearly $|P_{n+1-f}| - |P_i|$ is at most 2 for any $i = 1, 2, \ldots, n-f$, and if $|P_1| - |P_2|$ is greater than 0, every $|P_i|$ for $i = 3, 4, \ldots, n-f$ is equal to $|P_1| - 1$ or $|P_1| - 2$.

We now observe that f is the maximum cardinality of a forest in G. Indeed, P_{n+1-f} contains at least two vertices and the union of these vertices and $\{u_1, u_2, \ldots, u_{n-f}\}$ induces a clique in G. Since a forest cannot contain more than 2 vertices of a clique, we conclude that a forest in G is of order at most f. We now consider the value of t. A maximum induced tree in G must be contained in

- the subgraph induced by P_{n+1-f}, P_i (for some $i \le n-f$), and u_i, or
- the subgraph induced by P_i and P_j (for $i < j \le n-f$) and u_i and u_j,
- the subgraph induced by P_{n+1-f}.

We have $|P_1| \ge |P_2| \ge |P_i| \ge |P_{n+1-f}| - 2$ for any i in $\{3, \ldots, n-f\}$. Thus when $f < n-1$ holds, a maximum induced tree in the subgraph induced by P_1, P_2, and $\{u_1, u_2\}$ is of maximal order among all the induced trees of G, and its order equals

$$\lceil (|P_1| + |P_2|)/2 \rceil + 2 = q + \lceil (s_1 + s_2)/2 \rceil + 2.$$

When $f = n-1$ holds, the vertex set of G is the union of P_{n+1-f}, P_1, and u_1, and the order of its maximum induced tree is given by the same formula as above, since

$$\lceil (|P_1| + |P_{n+1-f}|)/2 \rceil + 1 = q + 1 + \lceil (s_1 + s_2)/2 \rceil + 1.$$

Let us define b by the relation $n - 1 = a^2 + b$. Recall that a denotes $\lfloor \sqrt{n-1} \rfloor$, so that b is comprised between 0 and $2a$. One easily verifies that $\lceil 2\sqrt{n-1} \rceil$ equals $2a$ when $b = 0$ holds, $2a+1$ when b belongs to $\{1, 2, \ldots, a\}$, and $2a+2$ when b belongs to $\{a+1, a+2, \ldots, 2a\}$. We first consider the case where a is even.

1. If $0 \le b \le a+2$ holds, we choose f to be $n+1-a$. Then $f-2$ equals a^2+b-a and $n+1-f$ equals a.
   (a) If $0 \le b \le a-1$ holds, then we have $q = a-2$ and $r = a+b$. If $a \ge 4$ or $b \ge 1$ holds, then we have $s_1 + s_2 \ge 3$ and thus $f - t = (n+1-a) - (a-2+2+2) = n-2a-1$. If $a = 2$ and $b = 0$ hold, then we have $s_1 = 2$ and $s_2 = 0$ and we obtain $f - t = (n+1-a) - (a-2+1+2) = n-2a$.
   (b) If $b = a$ holds, then we have $q = a$ and $r = s_1 = s_2 = 0$. Therefore $f-t$ equals $(n+1-a)-(a+2) = n-2a-1$.
   (c) If $b = a+1$ or $b = a+2$ holds, then we have $q = a$ and $r = b-a$, which means that r equals 1 or 2. Hence s_1 equals 1 or 2 and s_2 equals 0. We obtain $f-t = (n+1-a)-(a+1+2) = n-2a-2$.
2. If $a \ge 4$ and $a+3 \le b \le 2a$ hold, we choose f to be $n-a$. Then we have $f-2 = a^2+b+1-a-2$, $n+1-f = a+1$, $q = a-2$, and $r = b+1 \ge a+4$. Hence both s_1 and s_2 are equal to 2 and we obtain $f-t = (n-a)-(a-2+2+2) = n-2a-2$.

We conclude that in all subcases, $f-t$ equals $n - \lceil 2\sqrt{n-1} \rceil$, except when $n-1 = a^2$ and $a \ge 4$ hold and a is even. In this special case, we have $f - t = n - \lceil 2\sqrt{n-1} \rceil - 1$.

We now consider the case where a is odd and at least 3.

1. If $0 \le b \le 2$ or $a+1 \le b \le 2a$ holds, we choose f to be $n+1-a$. Then $f-2$ equals a^2+b-a and $n+1-f$ equals a.
   (a) If b equals 0, then $q=a-1$ and $r=s_1=s_2=0$ hold. Therefore $f-t$ equals $(n+1-a)-(a-1+2)=n-2a$.
   (b) If b equals 1 or 2, then $q=a-1$ and $r=b$ hold and thus s_1 equals 1 or 2 and s_2 equals 0. Therefore $f-t$ equals $(n+1-a)-(a-1+1+2)=n-2a-1$.
   (c) If $a+1 \le b \le 2a-1$ holds, then we have $q=a-1$ and $r=b$ and thus $s_1=s_2=2$. Therefore $f-t=(n+1-a)-(a-1+2+2)=n-2a-2$ holds.
   (d) If $b=2a$ holds, then we have $q=a+1$ and $r=s_1=s_2=0$. Therefore $f-t$ equals $(n+1-a)-(a+1+2)=n-2a-2$.
2. If $3 \le b \le a$ holds, we choose f to be $n-a$. Then $f-2$ equals $a^2+b-a-1$ and $n+1-f$ equals $a+1$.
   (a) If $3 \le b \le a-1$ holds, then we have $q=a-3$ and $r=a+b+2$ and thus $s_1=s_2=2$. Therefore $f-t$ equals $(n-a)-(a-3+2+2)=n-2a-1$.
   (b) If $b=a$, then $q=a-1$ and $r=s_1=s_2=0$. We obtain $f-t=(n-a)-(a-1+2)=n-2a-1$.

We conclude that in all subcases, $f-t$ equals $n-\lceil 2\sqrt{n-1}\rceil$.

To complete the proof, we observe that if $a=1$ holds, n must be comprised between 2 and 4. It is easy to verify that $f-t=n-\lceil 2\sqrt{n-1}\rceil=0$ holds for all graphs of order n in $\{2,3,4\}$, and the theorem holds in that case also. □

An extremal graph with $n=22$ is displayed in Figure 3. In that case, we have $a=\lfloor\sqrt{22-1}\rfloor=4$, and thus $b=n-1-a^2=5$ holds. Since a is even and b equals $a+1$, we have $f=n+1-a=19$ and $f-t=n-2a-2=12$. The subgraph induced by the black vertices is a tree of maximum order (i.e., of order 7).

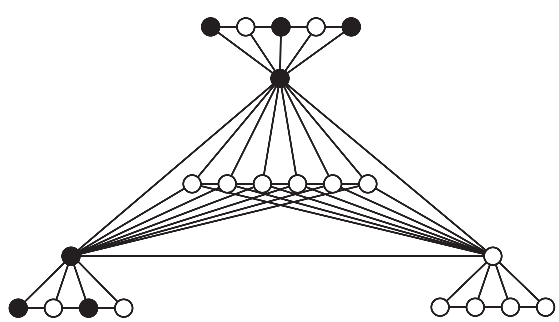

Figure 3: Extremal graph for $n=22$.

Let lf denote the maximum order of an induced linear forest, i.e., a forest in which every connected component is a path (possibly of length 0). We note that "forest" can be replaced by "linear forest" in the previous theorems; indeed the forest introduced at the beginning of the proof of Theorem 2.6 is linear.

Corollary 2.7 *Let n be an integer at least equal to 2. If n is of the form a^2+1 for some a with a even and $a \ge 4$, the relation $lf-t \le n-\lceil 2\sqrt{n-1}\rceil-1$ holds for any connected graph of order n and this bound is tight. Otherwise, the relation $lf-t \le n-\lceil 2\sqrt{n-1}\rceil$ holds for any connected graph of order n and the bound is again tight.*

3 Stable sets and trees

In this section we study the difference between α, the stability number of the graph G, and t, the maximum order of an induced subtree of G. We prove theorems similar to those of the previous section; indeed, the proofs of these theorems follow the same lines as in Section 2. Let G be a connected graph of order n, A any stable set in G, and $K = V\backslash A$ the complement of A. As in Section 2, we choose a shortest path P_{uv} between u and v for any pair $\{u, v\}$ of vertices in K. For any non empty subset S of K, we define x_S as the number of vertices w in A whose set of neighbours is S (note that a vertex w in A does not have any neighbour in A).

As before, we also introduce the variable Z and a system of constraints denoted by (SLI). For each $u \in K$ the system (SLI) includes the constraint

$$\sum_{S \text{ contains } u} x_S \leq Z + 1.$$

The left-hand side of this constraint is actually the number of vertices in A that are adjacent to u. For each pair $\{u, v\}$ of vertices in K, (SLI) includes the constraint

$$\sum_{S \cap P_{uv} = \{u\}} x_S + \sum_{S \cap P_{uv} = \{v\}} x_S \leq Z.$$

The left-hand side of this inequality represents the number of vertices in A that are adjacent to u but no other vertex of P_{uv} **or** to v but no other vertex of P_{uv}. Finally, the system (SLI) includes the constraint

$$\sum_S x_S = |A|,$$

where the sum is taken over all non empty subsets S of K.

Lemma 3.1 *For any connected graph G of order n, any stable set A in G, and any value Z satisfying the system (SLI), the relation $Z \geq 2(|A| - 1)/(n + 1 - |A|)$ holds.*

Proof. As in the proof of Lemma 2.2, we observe that each variable x_S appears exactly $|S|$ times (one time for each vertex in S) in an inequality of the first group and at least$|K| - |S|$ times in an inequality of the second group. Adding all the inequalities in the first and second groups, we obtain an inequality whose left-hand side is at least $(n - |A|) \sum_S x_S$ and right-hand side equals

$$(n - |A|)(Z + 1) + \frac{(n - |A|)(n - |A| - 1)}{2} Z.$$

Since we also have $\sum_S x_S = |A|$, the relation

$$(n - |A|)|A| \leq (n - |A|)(Z + 1) + \frac{(n - |A|)(n - |A| - 1)}{2} Z,$$

holds, yielding

$$Z \geq \frac{2(|A| - 1)}{n + 1 - |A|}.$$

□

Theorem 3.2 *For any connected graph G of order n and any stable set A of G, there exists an induced tree in G whose order is at least equal to*

$$\left\lceil \frac{2(|A| - 1)}{n + 1 - |A|} \right\rceil + 2.$$

Proof. Let us denote by $Z_{\min}$ the smallest value of Z for which all the constraints of (SLI) are satisfied. Then there is at least one “tight” constraint in which $Z_{\min}$ appears.

1. If this constraint belongs to the first group, there is a vertex u in K such that $Z_{\min} + 1$ vertices in A are adjacent to u. Hence there exists a tree in G whose order is at least $Z_{\min} + 2$.

2. If this constraint belongs to the second group, there is a pair of vertices $\{u, v\}$ such that $Z_{\min}$ vertices in F are adjacent to u but no other vertex of P_{uv} or to v but no other vertex of P_{uv}. Let C_1 (resp. C_2) denote the set of vertices in A that are adjacent to u (resp. v) but no other vertex of P_{uv}. Hence there exists a tree in G consisting of the vertices of P_{uv} and $Z_{\min}$ other vertices. The order of this tree is at least $Z_{\min} + 2$.

The statement of the theorem follows from this case analysis and Lemma 3.1. □

Corollary 3.3 *The relation $\alpha - t \leq n - \lceil 2\sqrt{2n} \rceil + 1$ holds for any connected graph G of order n.*

Proof. Let A be a stable set of maximal cardinality, i.e., of cardinality α. Theorem 3.2 implies that

$$\alpha - t \leq \alpha - \frac{2(\alpha - 1)}{n + 1 - \alpha} - 2.$$

The derivative of the right-hand side of this inequality with respect to α is

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

The only value of α not exceeding n for which the derivative equals 0 is $n + 1 - \sqrt{2n}$, which maximizes the value of $\alpha - 2(\alpha - 1)/(n + 1 - \alpha) - 2$ since this function is concave. Replacing α by $n + 1 - \sqrt{2n}$ in $\alpha - 2(\alpha - 1)/(n + 1 - \alpha) - 2$ yields

$$\alpha - t \leq n - 2\sqrt{2n} + 1.$$

The corollary follows by observing that $\alpha - t$ is an integer. □

We now prove that the above bound is tight.

Theorem 3.4 *Let n be an integer at least equal to 2. There exists a connected graph of order n for which $\alpha - t = n - \lceil 2\sqrt{2n} \rceil + 1$ holds.*

Proof. We first observe that if n is comprised between 2 and 5, the star of order n verifies $\alpha - t = -1 = n - \lceil 2\sqrt{2n} \rceil + 1$. In what follows we thus assume that n is at least 6 and α at most $n - 2$ (the precise value of α will be given below). We now describe the construction of a graph $G = (V, E)$ of order n including a stable set A of cardinality α. The complement of A, $V \backslash A = \{v_1, v_2, \ldots, v_{n-\alpha}\}$, is a clique of cardinality $n - \alpha$. The set A is partitioned into $n + 1 - \alpha$ subsets $C_1, C_2, \ldots, C_{n-\alpha}, C_{n+1-\alpha}$ such that $|C_{n+1-\alpha}| \geq |C_1| \geq |C_2| \geq \cdots \geq |C_{n-\alpha}|$ and $|C_{n+1-\alpha}| - |C_{n-\alpha}| \leq 1$. This implies that

$$|C_{n+1-\alpha}| = \left\lceil \frac{\alpha}{n + 1 - \alpha} \right\rceil \text{ and } |C_{n-\alpha}| = \left\lfloor \frac{\alpha}{n + 1 - \alpha} \right\rfloor.$$

For $1 \leq i \leq n - \alpha$, there is an edge between v_i and each vertex in C_i. Also there is an edge between any vertex in $C_{n+1-\alpha}$ and any vertex in the clique $V \backslash A$. We observe that the union of $V \backslash A$ and any singleton $\{u\}$ with $u \in C_{n+1-\alpha}$ induces a clique. Since a stable set of G has at most one member in any clique, its cardinality is at most $|V| - (|V| - |A| + 1) + 1 = \alpha$. We conclude that α is indeed the stability number of G. Also t clearly equals $n_1 + n_2 + 2$, where $n_i = |C_i|$ for $i = 1, 2$.

We define b as $\lfloor \sqrt{2n} \rfloor$ and c as $2n - b^2$, which implies that $0 <= c <= 2b$ holds and b is at least 2. Let q and r be such that $\alpha = q(n + 1 - \alpha) + r$ and $0 <= r < n + 1 - \alpha$ hold. We observe that $\lceil 2\sqrt{2n} \rceil$ equals $2b$ if c equals 0, $2b + 1$ if $1 \leq c \leq b$, and $2b + 2$ if $b + 1 \leq c \leq 2b$. Note also that b and c always have the same parity (i.e., either b and c are both even or they are both odd).

We now consider five cases.

1. If b is even and c equals 0, we choose the value $n + 1 - b$ for α. Then $n + 1 - \alpha$ equals b, q equals $b/2 - 1$ and r equals 1. Then $n_1 + n_2 + 2$ equals $2q + 2 = b$ and $\alpha - t$ equals $n - 2b + 1$, which is equal to $n - \lceil 2\sqrt{2n} \rceil + 1$ in this case.

2. If b is even and $2 <= c <= b$ holds, we choose the value $n - b$ for α. Then $n + 1 - \alpha$ equals $b + 1$ and there are two subcases:
   - $q = b/2 - 2$ and $r = (c + b)/2 + 2$ if c is smaller than $b - 2$, and
   - $q = b/2 - 1$ and $r = (c - b)/2 + 1$ if c equals $b - 2$ or b.

   In both cases, $n_1 + n_2 + 2$ is equal to b and thus $\alpha - t = n - 2b = n - (2b+1) + 1$ is equal to $n - \lceil 2\sqrt{2n} \rceil + 1$.
3. If b is even and $b + 2 <= c <= 2b$ holds, we choose the value $n + 1 - b$ for α. Then $n + 1 - \alpha$ equals b and there are two subcases again:
   - $q = b/2 - 1$ and $r = c/2 + 1$ if c is smaller than $2b - 2$, and
   - $q = b/2$ and $r = c/2 - b + 1$ if c equals $2b - 2$ or $2b$.

   In both cases, $n_1 + n_2 + 2$ is equal to $b + 2$ and thus $\alpha - t = (n + 1 - b) - (b + 2) = n - (2b + 2) + 1$ is equal to $n - \lceil 2\sqrt{2n} \rceil + 1$.
4. If b is odd and $1 <= c <= b$ holds, we choose the value $n + 1 - b$ for α. Then $n + 1 - \alpha$ equals b and there are two subcases:
   - $q = (b - 3)/2$ and $r = (c + b)/2 + 1$ if c is smaller than $b - 2$, and
   - $q = (b - 1)/2$ and $r = (c - b)/2 + 1$ if c equals $b - 2$ or b.

   In both cases, $n_1 + n_2 + 2$ is equal to $b + 1$ and thus $\alpha - t = (n + 1 - b) - (b + 1) = n - (2b + 1) + 1$ is equal to $n - \lceil 2\sqrt{2n} \rceil + 1$.
5. If b is odd and $b + 2 <= c <= 2b - 1$ holds, we choose the value $n - b$ for α. Then $n + 1 - \alpha$ equals $b + 1$ and there are two subcases:
   - $q = (b - 3)/2$ and $r = (c + 3)/2$ if c is smaller than $2b - 1$, and
   - $q = (b - 1)/2$ and $r = 0$ if c equals $2b - 1$.

   In both cases, $n_1 + n_2 + 2$ is equal to $b + 1$ and thus $\alpha - t = (n - b) - (b + 1) = n - (2b + 2) + 1$ is equal to $n - \lceil 2\sqrt{2n} \rceil + 1$.

This completes the proof of the theorem. □

An extremal graph with $n = 12$ is displayed in Figure 4. In that case we have $b = 4$, $c = 8$, $\alpha = 9$, and $t = 6$. The subgraph induced by the black vertices is a tree of maximum order.

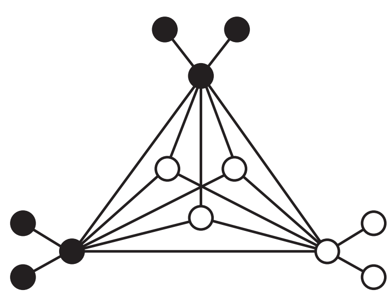

Figure 4: Extremal graph for $n = 12$.

4 Conclusion

In this article we have investigated the difference between the maximum order of an induced forest and that of an induced tree, on one hand, and the difference between the stability number and the maximum order of an induced tree, on the other. In light of the work by Fox, Loh, and Sudakov [5], it would be interesting to extend our results by finding bounds for $f - t$ and $\alpha - t$ in certain families of graphs, for instance triangle-free graphs or, more generally, K_r-free graphs. Note that the extremal graphs presented in this article contain triangles, and that forbidding triangles will likely make the construction of extremal graphs challenging.

References

[1] Alon, N., Mubayi, D., Thomas, R. 2001. Large induced forests in sparse graphs. *Journal of Graph Theory* **38** 113–123.

[2] Bondy, A., Murty, U.S.R., 2008. Graph Theory. *Graduate Texts in Mathematics* **244** Springer.

[3] DeLaViña, E., Waller, B. 2004. On some conjectures of Graffiti.pc on the maximum order of induced subgraphs. *Proceedings of the Thirty-Fifth Southeastern International Conference on Combinatorics, Graph Theory and Computing, Congressus Numerantium* **166** 11-32.

[4] Erdős, P., Saks, M., Sòs, V. 1986. Maximum induced trees in graphs. *Journal of Graph Theory* **41** 61–79.

[5] Fox, J., Loh, P.-S., Sudakov, B. 2009. Large induced trees in K_r-free graphs. *Journal of Combinatorial Theory B* **99** 494–501.

[6] Zheng, M., Lu, X. 1990. On the Maximum Induced Forests of a Connected Cubic Graph without Triangles. *Discrete Mathematics* **85** 89–96.