Induced Subtrees in Interval Graphs*

Pinar Heggernes[1], Pim van 't Hof[1], and Martin Milanič[2]

[1] Department of Informatics, University of Bergen, Norway
{pinar.heggernes,pim.vanthof}@ii.uib.no
[2] UP IAM and UP FAMNIT, University of Primorska, Slovenia
martin.milanic@upr.si

Abstract. The INDUCED SUBTREE ISOMORPHISM problem takes as input a graph G and a tree T, and the task is to decide whether G has an induced subgraph that is isomorphic to T. This problem is known to be NP-complete on bipartite graphs, but it can be solved in polynomial time when G is a forest. We show that INDUCED SUBTREE ISOMORPHISM can be solved in polynomial time when G is an interval graph. In contrast to this positive result, we show that the closely related SUBTREE ISOMORPHISM problem is NP-complete even when G is restricted to the class of proper interval graphs, a well-known subclass of interval graphs.

1 Introduction and Background

The problems SUBGRAPH ISOMORPHISM and INDUCED SUBGRAPH ISOMORPHISM both take as input two graphs G and H, and the task is to determine whether G has a subgraph or an induced subgraph, respectively, that is isomorphic to H. SUBGRAPH ISOMORPHISM and INDUCED SUBGRAPH ISOMORPHISM are two well-studied and notoriously hard problems in the area of graph algorithms, generalizing classical NP-complete problems such as CLIQUE, INDEPENDENT SET and HAMILTONIAN PATH.

Both SUBGRAPH ISOMORPHISM and INDUCED SUBGRAPH ISOMORPHISM are known to be NP-complete already when each of G and H is a disjoint union of paths [7, 9], and thus both problems are NP-complete on any hereditary graph class that contains arbitrarily long induced paths. In particular, both problems are NP-complete on proper interval graphs and on bipartite permutation graphs. Interestingly, INDUCED SUBGRAPH ISOMORPHISM can be solved in polynomial time on connected proper interval graphs and connected bipartite permutation graphs [13], whereas SUBGRAPH ISOMORPHISM remains NP-complete on these connected graph classes [16]. Both problems can be solved in polynomial time when G is a forest and H is a tree [23], but remain NP-complete when G is a tree and H is a forest [9].

The problems SUBGRAPH ISOMORPHISM and INDUCED SUBGRAPH ISOMORPHISM remain hard also on several classes of graphs that do not contain long induced paths. For example, both problems are NP-complete on connected cographs [1,5,7],

* This work is supported by the Research Council of Norway, by the Slovenian Research Agency, and by the European Science Foundation.

and remain NP-complete even on connected trivially perfect graphs [2, 16], a subclass of cographs. Both Subgraph Isomorphism and Induced Subgraph Isomorphism are also NP-complete when both input graphs are split graphs [7, 11]. Kijima et al. [16] showed that Subgraph Isomorphism is NP-complete when G is restricted to the class of chain graphs, cochain graphs, or threshold graphs, but the problem becomes polynomial-time solvable when, in addition, H is also restricted to the same class as G.

Given the large amount of hardness results on the two problems, even under severe restrictions on G, it makes sense to consider different restrictions on H in an attempt to obtain tractability. A natural candidate for such a restriction is demanding H to be a tree. This brings us to the two problems that we focus on in this paper:

(Induced) Subtree Isomorphism
Input: A graph G and a tree T.
Question: Does G have an (induced) subgraph isomorphic to T?

Since Subtree Isomorphism is a generalization of Hamiltonian Path, this problem is NP-complete on all graph classes on which Hamiltonian Path is NP-complete, like planar graphs, chordal bipartite graphs, and strongly chordal split graphs [9, 26]. Similarly, the fact that finding a longest induced path in a bipartite graph is NP-hard [9] implies that Induced Subtree Isomorphism is NP-complete on bipartite graphs. Both Subtree Isomorphism and Induced Subtree Isomorphism are also NP-complete on graphs of treewidth at most 2 [22], but can be solved in polynomial time when G is a forest [23].

The main result of this paper is a polynomial-time algorithm for Induced Subtree Isomorphism on interval graphs, which gives a nice contrast to the NP-completeness of Induced Subgraph Isomorphism on this graph class. On the negative side, we show that Subtree Isomorphism is NP-complete already on proper interval graphs. Note that the problem of finding a longest path in an interval graph can be solved in polynomial time [14] (and even on cocomparability graphs [24]). Hence, our negative result on proper interval graphs shows that finding a given tree as a subgraph in a proper interval graph is much harder than finding a path of given length in such a graph, despite the linear structure that proper interval graphs possess.

2 Definitions and Notation

All graphs considered in this paper are finite, undirected and simple. We refer to the monograph by Diestel [8] for basic graph terminology not defined below. Detailed information on all the graph classes mentioned in this paper can be found in the books by Golumbic [12] and Brandstädt, Le, and Spinrad [3]. Figure 1 below shows the inclusion relations between most of the graph classes mentioned this paper.

A graph is an *interval graph* if there is a bijection between its vertices and a family of closed intervals of the real line such that two vertices are adjacent if and only if the two corresponding intervals overlap. Such a bijection is called an *interval*

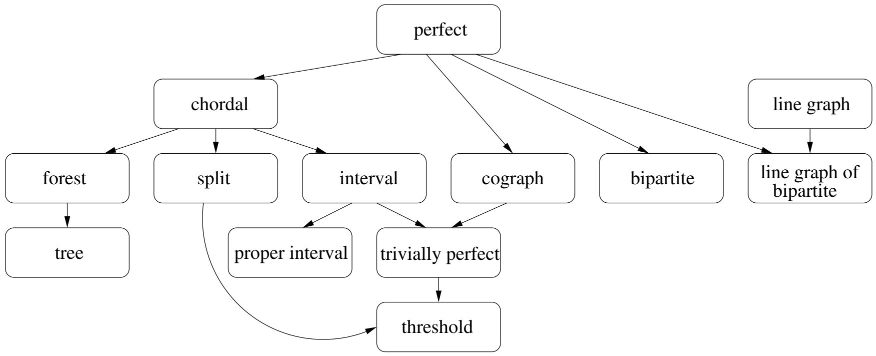

Fig. 1. An overview of most of the graph classes mentioned in this paper. An arrow from a class $\mathcal{G}$ to a class $\mathcal{H}$ indicates that $\mathcal{H}$ is a subset of $\mathcal{G}$.

representation of the graph. A graph is a *proper interval graph* if it has an interval representation where no interval properly contains another interval. Many different characterizations of interval graphs are known in the literature. In order to state the one we will use in our algorithm, we need the following definition.

Definition 1. *Let $G = (V, E)$ be a graph. An ordering $(u_1, \ldots, u_n)$ of V is called an* interval order *of G if, for every triple (i, j, k) with $1 \leq i < j < k \leq n$, it holds that $u_i u_k \in E$ implies $u_j u_k \in E$.*

Olariu [27] showed that a graph G is an interval graph if and only if G has an interval order.

A *tree* is a connected graph without cycles. Let G be a graph and let T be a tree. If G has a induced subgraph that is isomorphic to T, then we say that T is an *induced subtree* of G; a *subtree* of G is defined analogously. A tree T is a *caterpillar* if it has a path that contains every vertex of degree at least 2 in T; such a path is called a *backbone* of T. A well-known characterization of interval graphs by Lekkerkerker and Boland [20] immediately implies that a tree is an interval graph if and only if it is a caterpillar.

Let G and H be two graphs. A mapping $\varphi : V(H) \to V(G)$ is said to be an *induced subgraph isomorphism*, or ISI *mapping* for short, of H into G, if φ is injective and $uv \in E(H)$ if and only if $\varphi(u)\varphi(v) \in E(G)$ for all $u, v \in V(H)$. Consequently, H is an induced subgraph of G if and only if there exists an ISI mapping of H into G.

3 Induced Subtree Isomorphism on Interval Graphs

In this section, we show that Induced Subtree Isomorphism can be solved in polynomial time on interval graphs. Before presenting our algorithm in the proof of Theorem 1 below, we first prove a sequence of five lemmas, as well as a corollary of these lemmas that forms the main ingredient of our algorithm.

Throughout this section, up to the statement of Theorem 1, let $G = (V, E)$ be a connected interval graph and let T be a caterpillar on at least three vertices. We fix an interval order $\sigma = (u_1, \dots, u_n)$ of G. For any two vertices $x, y \in V$, we write $x \prec_\sigma y$ if x appears before y in the interval order σ, i.e., if $x = u_i$ and $y = u_j$ for some $i < j$.

Suppose there exists an ISI mapping φ of T into G. Then, for any ordered path $P = (t_1, \dots, t_p)$ of T, we say that P is *φ-increasing* if $\varphi(t_1) \prec_\sigma \varphi(t_2) \prec_\sigma \cdots \prec_\sigma \varphi(t_p)$, and P is *φ-decreasing* if $\varphi(t_p) \prec_\sigma \varphi(t_{p-1}) \prec_\sigma \cdots \prec_\sigma \varphi(t_1)$.

Lemma 1. *Let $P = (t_1, \dots, t_p)$ be an ordered path in T whose vertices all have degree at least 2 in T. Then, for any* ISI *mapping φ of T into G, the path P is either φ-increasing or φ-decreasing.*

Proof. The statement is trivially true if $p \leq 2$. Let $p \geq 3$. Suppose, for contradiction, that there exists an ISI mapping φ of T into G such that P is neither φ-increasing nor φ-decreasing. Then, in particular, there exist three consecutive vertices t_i, t_{i+1}, t_{i+2} of P such that the ordered path (t_i, t_{i+1}, t_{i+2}) is neither φ-increasing nor φ-decreasing. Let $u_{j_1} = \varphi(t_i)$, $u_{j_2} = \varphi(t_{i+1})$ and $u_{j_3} = \varphi(t_{i+2})$. Without loss of generality, we may assume that $j_1 < j_3$. Observe that Definition 1 implies that $j_1 < j_2$. Since we assumed the ordered path (t_i, t_{i+1}, t_{i+2}) to be neither φ-increasing nor φ-decreasing, it holds that $j_2 > j_3$.

Let w be a T-neighbor of t_{i+2} other than t_{i+1}; such a vertex w exists since we assume that all vertices of P have degree at least 2 in T. Let $u_{j_w} = \varphi(w)$. We consider three cases according to the value of j_w.

- Suppose that $j_w < j_1$. Since φ preserves adjacencies and $wt_{i+2} \in E(T)$, we have $u_{j_w}u_{j_3} \in E(G)$. On the other hand, since φ preserves non-adjacencies and $t_it_{i+2} \notin E(T)$, we have $u_{j_1}u_{j_3} \notin E(G)$. However, this contradicts Definition 1 applied to the triple $(i, j, k) = (j_w, j_1, j_3)$.
- Suppose that $j_1 < j_w < j_2$. Since φ preserves adjacencies and $t_it_{i+1} \in E(T)$, we have $u_{j_1}u_{j_2} \in E(G)$. On the other hand, since φ preserves non-adjacencies and $wt_{i+1} \notin E(T)$, we have $u_{j_w}u_{j_2} \notin E(G)$. Again, we have a contradiction to Definition 1, this time for the triple $(i, j, k) = (j_1, j_w, j_2)$.
- Suppose that $j_2 < j_w$. Since φ preserves adjacencies and $t_{i+2}w \in E(T)$, we have $u_{j_3}u_{j_w} \in E(G)$. On the other hand, since φ preserves non-adjacencies and $t_{i+1}w \notin E(T)$, we have $u_{j_2}u_{j_w} \notin E(G)$. We have a contradiction to Definition 1 for the triple $(i, j, k) = (j_3, j_2, j_w)$.

This completes the proof of the lemma. $\square$

We will show in Lemma 2 below that the problem of determining whether G has an induced subgraph isomorphic to T can be reduced to computing the values of a certain Boolean-valued function f_B for each $B \in \mathcal{B}$, where $\mathcal{B}$ is a set of three or four so-called *representative backbones* of T. In order to state the lemma, we first need to define the set $\mathcal{B}$ and the function f_B.

Recall that a backbone of T is any path B containing all vertices of degree at least 2 in T. The *order* of a backbone B is the number of vertices in B and is denoted by $|B|$. For any ordered backbone $B = (t_1, t_2, \dots, t_p)$, the ordered backbone

$B^{-1} = (t_p, t_{p-1}, \ldots, t_1)$ is called the *reverse* of B. Given two ordered backbones $B = (t_1, t_2, \ldots, t_p)$ and $B' = (t'_1, \ldots, t'_{p'})$ of T, we say that B and B' are *equivalent* if $p = p'$ and $t_i = t'_i$ for all $i \in \{1, 2, \ldots, p-1\}$.

We now define a set $\mathcal{B}$ consisting of three or four non-equivalent ordered backbones of T as follows. Let $B_{\min} = (t_1, t_2, \ldots, t_p)$ be an ordered backbone of T of minimum order, i.e., $B_{\min}$ is an ordered backbone of T whose vertex set consists of exactly those vertices that have degree at least 2 in T. Let $B_{\min}^{-1} = (t_p, t_{p-1}, \ldots, t_1)$ be the reverse of $B_{\min}$, where $B_{\min} = B_{\min}^{-1}$ if $p = 1$. Note that every ordered backbone of T other than $B_{\min}$ or $B_{\min}^{-1}$ contains either $|V(B_{\min})| + 1$ or $|V(B_{\min})| + 2$ vertices. Let us fix a neighbor t'_1 of t_1 and a neighbor t'_p of t_p such that $t'_1 \neq t_p$ and such that both t'_1 and t'_p have degree 1 in T. Such neighbors t'_1 and t'_p exist due to the fact that both t_1 and t_p have degree at least 2 in T and the assumption that T has at least three vertices. We now define two ordered backbones $B^+ = (t_1, t_2, \ldots, t_p, t'_p)$ and $B^- = (t_p, t_{p-1}, \ldots, t_1, t'_1)$. Finally, we define $\mathcal{B} = \{B_{\min}, B_{\min}^{-1}, B^+, B^-\}$. The backbones in $\mathcal{B}$ are called the *representative backbones* of T. Since we assume that T contains at least 3 vertices and $B_{\min} = B_{\min}^{-1}$ if and only if $p = 1$, we have that $|\mathcal{B}| = 3$ if $p = 1$ and $|\mathcal{B}| = 4$ if $p \geq 2$.

Let $B = (t_1, \ldots, t_p) \in \mathcal{B}$ be a representative backbone of T. For every $i \in \{1, \ldots, p\}$, let B_i denote the subgraph of T induced by the first i vertices of B together with their neighbors outside the backbone, that is,

$$B_i = T\left[\left\{t_1, \ldots, t_i\right\} \cup L_1 \cup \ldots \cup L_i\right] ,$$

where L_i denotes the set of neighbors of t_i outside B, i.e., $L_i = N_T(t_i) \setminus V(B)$. The following straightforward property of representative backbones will be useful in later proofs.

Observation 1 *For every representative backbone $B = (t_1, \ldots, t_p) \in \mathcal{B}$ and every $i \in \{1, \ldots, p\}$, vertex t_i has a neighbor in B_i.* □

For $1 \leq j < k \leq n$, let G_j^k denote the subgraph of G induced by $\{u_1, \ldots, u_j, u_k\}$, that is,

$$G_j^k = G\big[\{u_1, \ldots, u_j, u_k\}\big] .$$

With these definitions in mind, we define a Boolean-valued function $f_B : \mathcal{T}_B \to \{0, 1\}$, where

$$\mathcal{T}_B = \{(i, j, k) \mid 1 \leq i \leq p \quad \text{and} \quad 1 \leq j < k \leq n\},$$

as follows: for all $(i, j, k) \in \mathcal{T}_B$, we set $f_B(i, j, k) = 1$ if and only if there exists an ISI mapping φ of B_i into G_j^k such that $\varphi(t_i) = u_k$.

Lemma 2. *Graph G has an induced subgraph isomorphic to T if and only if there exists a backbone $B \in \mathcal{B}$ and an integer $k \in \{2, \ldots, n\}$ such that $f_B(|B|, k-1, k) = 1$.*

Proof. Suppose first that there exists a backbone $B \in \mathcal{B}$ and an integer k with $1 < k \leq n$ such that $f_B(|B|, k-1, k) = 1$. Then, denoting $B = (t_1, \ldots, t_p)$, there exists an ISI mapping φ of $B_p = T$ into $G_{k-1}^k = G[\{u_1, \ldots, u_k\}]$ such that $\varphi(t_p) = u_k$.

Trivially, φ is an ISI mapping of T into G, and hence G has an induced subgraph isomorphic to T.

Conversely, suppose G has an induced subgraph that is isomorphic to T. Then there exists an ISI mapping φ of T into G. Since every vertex of $B_{\min}$ and $B_{\min}^{-1}$ has degree at least 2 in T, either $B_{\min}$ or $B_{\min}^{-1}$ is φ-increasing due to Lemma 1. Among all pairs (φ, B) where φ is an ISI mapping of T into G and B is a φ-increasing backbone in $\mathcal{B}$, choose a pair (φ, B) such that $B = (t_1, \ldots, t_p) \in \mathcal{B}$ is of maximum possible order. Let $\varphi(t_j) = u_{i_j}$ for all $j \in \{1, \ldots, p\}$. Let $k = i_p$.

We claim that, with B and k defined as above, it holds that $f_B(p, k-1, k) = 1$. By definition of f_B, this is equivalent to verifying the existence of an ISI mapping ψ of $B_p = T$ into $G_{k-1}^k = G[\{u_1, \ldots, u_k\}]$ such that $\psi(t_p) = u_k$. We claim that such a mapping is obtained by taking $\psi = \varphi$. Condition $\psi(t_p) = u_k$ follows from the definition of k, and the fact that ψ is an injective mapping that preserves adjacencies and non-adjacencies follows trivially from the corresponding properties of φ. It remains to verify that $\psi(T) = \varphi(T) \subseteq \{u_1, \ldots, u_k\}$.

Suppose for a contradiction that there exists a vertex $v \in V(T)$ such that $\varphi(v) = u_r$ for $r > k$. Since B is φ-increasing, we have $i_j \leq i_p = k$ for all $1 \leq j \leq p$, and hence v is not a vertex of B. Therefore, vertex v has a unique neighbor t_j in B. Suppose first that $1 \leq j < p$. Then, $p \geq 2$ and since φ preserves adjacencies and non-adjacencies, we conclude that $u_{i_j}u_r \in E(G)$ and $u_k u_r \notin E(G)$, which contradicts Definition 1 applied to the triple (i_j, k, r) with $i_j < k < r$. Therefore, we may assume that $j = p$, and v is adjacent to t_p. Consider the backbone $B' = (t_1, \ldots, t_p, v)$ of T, and let $B'' \in \mathcal{B}$ be a backbone in $\mathcal{B}$ equivalent to B'. Since B'' is equivalent to B', there exists a neighbor w of t_p such that $B'' = (t_1, \ldots, t_p, w)$. By the maximality of B, it follows that B'' is not φ-increasing, and consequently $\varphi(w) = u_s$ where $s < k$. Consider the mapping $\varphi' : V(T) \to V(G)$ obtained from φ by switching the images of v and w. Formally, for every $t \in V(T)$, set

$$\varphi'(t) = \begin{cases} \varphi(t) & \text{if } t \notin \{v, w\}\,; \\ \varphi(w) & \text{if } t = v\,; \\ \varphi(v) & \text{if } t = w\,. \end{cases}$$

Then φ' is an ISI mapping of T into G. However, since $\varphi'(w) = \varphi(v) = u_r$ and $r > k$, backbone B'' is a φ'-increasing backbone strictly longer than B, contradicting the definition of the pair (φ, B). This shows that if T is isomorphic to an induced subgraph of G, then there exists a backbone $B \in \mathcal{B}$ and an integer k with $1 < k \leq n$ such that $f_B(|B|, k-1, k) = 1$. □

In what follows, we show that for any backbone $B \in \mathcal{B}$, the values of the function f_B can be computed recursively. Definition 1 implies that for every $j \in \{1, \ldots, n\}$, there exists an index $\ell(j) \leq j$ such that $N_{G_j}[u_j] = \{u_{\ell(j)}, u_{\ell(j)+1}, \ldots, u_j\}$, where G_j denotes the subgraph of G induced by $\{u_1, \ldots, u_j\}$. Also recall that for any given backbone $B = (t_1, \ldots, t_p) \in \mathcal{B}$ and any $i \in \{1, \ldots, p\}$, we write L_i to denote the set of neighbors of t_i outside B.

First, we consider the simplest case, namely computing $f_B(i, j, k)$ when $i = 1$.

Lemma 3. *Let $B \in \mathcal{B}$ be a representative backbone of T, and let $(1, j, k) \in \mathcal{T}_B$. Then $f_B(1, j, k) = 1$ if and only if $\alpha(G[\{u_{\ell(k)}, \ldots, u_j\}]) \geq |L_1|$.*

Proof. The definition of f_B implies that $f_B(1,j,k)=1$ if and only if there exists an ISI mapping φ of B_1 into G_j^k such that $\varphi(t_1)=u_k$. Notice that B_1 is isomorphic to a star with $|L_1|$ leaves. Hence, there exists an ISI mapping φ of B_1 into G_j^k such that $\varphi(t_1)=u_k$ if and only if u_k has at least $|L_1|$ pairwise non-adjacent neighbors in the graph G_j^k. But this last condition is equivalent to the condition that the independence number of the subgraph of G induced by $\{u_{\ell(k)},\ldots,u_j\}$ is at least $|L_1|$. □

Now, let us consider the problem of computing the value of $f_B(i,j,k)$ for some $(i,j,k)\in\mathcal{T}_B$ with $i>1$, assuming that we have already computed the values of $f_B(i',j',k')$ for all $(i',j',k')\in\mathcal{T}_B$ with $i'<i$. Lemma 4 states a necessary condition for $f_B(i,j,k)=1$.

Lemma 4. *Let $B\in\mathcal{B}$ be a representative backbone of T, and let (i,j,k) be a triple in $\mathcal{T}_B$ with $i\geq 2$ such that $f_B(i,j,k)=1$. Then there exists an integer $k'\in\{\ell(k),\ldots,j\}$ such that*

$$f_B(i-1,\ell(k)-1,k')=1 \quad \textit{and} \quad \alpha\Big(G\big[\{u_{k'+1},\ldots,u_j\}\setminus N_G(u_{k'})\big]\Big)\geq |L_i|\,.$$

Proof. Suppose that the conditions in the lemma are satisfied. By the definition of f_B, there exists an ISI mapping φ of B_i into G_j^k such that $\varphi(t_i)=u_k$. Let $k'\in\{1,\ldots,n\}$ be the index satisfying $\varphi(t_{i-1})=u_{k'}$. Let us verify that k' has all the desired properties. First of all, it holds that $k'\leq j$, since $k'\neq k$ by the injectivity of φ and since φ maps $V(B_i)$ to $V(G_j^k)=\{u_1,\ldots,u_j,u_k\}$. Second, since φ preserves adjacencies and $t_{i-1}t_i\in E(T)$, we have $u_{k'}u_k=\varphi(t_{i-1})\varphi(t_i)\in E(G)$. Consequently, $u_{k'}\in N_{G_k}[u_k]$ and hence $k'\geq\ell(k)$.

Now, let us show that $f_B(i-1,\ell(k)-1,k')=1$. This is equivalent to showing the existence of an ISI mapping φ' of B_{i-1} into $G_{\ell(k)-1}^{k'}$ such that $\varphi'(t_{i-1})=u_{k'}$. Let φ' be the restriction of φ to $V(B_{i-1})$. We will verify that φ' is an ISI mapping of B_{i-1} into $G_{\ell(k)-1}^{k'}$ such that $\varphi'(t_{i-1})=u_{k'}$. Since φ is an injective mapping that preserves adjacencies and non-adjacencies, so is φ'. The condition $\varphi'(t_{i-1})=u_{k'}$ is also satisfied, by the definition of k'. It remains to verify that for every $w\in V(B_{i-1})$, we have $\varphi'(w)\in V(G_{\ell(k)-1}^{k'})$, or, equivalently, that $\varphi(w)\in\{u_1,\ldots,u_{\ell(k)-1},u_{k'}\}$. For $w=t_{i-1}$, this is clear, and we only need to check that $\varphi(w)\in\{u_1,\ldots,u_{\ell(k)-1}\}$ for all $w\in V(B_{i-1})\setminus\{t_{i-1}\}$. Suppose for a contradiction that there exists a vertex $w\in V(B_{i-1})\setminus\{t_{i-1}\}$ with $\varphi(w)=u_r$ for some $r\in\{\ell(k),\ldots,j\}$. Then $u_ru_k\in E(G)$, and since φ preserves non-adjacencies, this implies $wt_i\in E(T)$. This contradiction to the fact that the only neighbor of t_i in B_{i-1} is t_{i-1} implies that $f_B(i-1,\ell(k)-1,k')=1$, as claimed.

It remains to show that the independence number of the graph $G'=G\big[\{u_{k'+1},\ldots,u_j\}\setminus N_G(u_{k'})\big]$ satisfies $\alpha(G')\geq|L_i|$. The vertices of L_i form an independent set of size $|L_i|$ in B_i, and since φ is an injective mapping preserving non-adjacencies, its image $\varphi(L_i)$ is an independent set of size $|L_i|$ in G_j^k. Since φ preserves non-adjacencies and $L_i\cap N_{B_i}(t_{i-1})=\emptyset$, we have $\varphi(L_i)\cap N_G(u_{k'})=\emptyset$. Hence, it is enough to show that $\varphi(L_i)\subseteq\{u_{k'+1},\ldots,u_j\}$. Clearly, $\varphi(L_i)\subseteq\{u_1,\ldots,u_j\}$, so the only way the condition $\varphi(L_i)\subseteq\{u_{k'+1},\ldots,u_j\}$ could fail is if there exists a vertex $w\in L_i$ such that $\varphi(w)=u_{i_w}$ for some integer $i_w\leq k'$. Since φ maps t_{i-1}

to $u_{k'}$ and $w \neq t_{i-1}$, we have $i_w < k'$. Moreover, since φ preserves adjacencies and $wt_i \in E(B_i)$, we have $u_{i_w}u_k \in E(G)$. By Observation 1, vertex t_{i-1} has a neighbor, say z, in B_{i-1}. Clearly $z \neq t_i$; moreover, $u_{i_z}u_{k'} \in E(G)$, where $u_{i_z} = \varphi(z)$. Furthermore, Definition 1 implies that $i_z < k'$. Since φ preserves non-adjacencies and $wt_{i-1} \notin E(B_i)$, we have $u_{i_w}u_{k'} \notin E(G)$. Similarly, since $zt_i \notin E(B_i)$, we have $u_{i_z}u_k \notin E(G)$. If $i_z < i_w$ then $i_z < i_w < k'$, and we have a contradiction to Definition 1 for the triple $(i, j, k) = (i_z, i_w, k')$. Hence, $i_w < i_z$, and consequently $i_w < i_z < k$, and we have a contradiction to Definition 1 for the triple $(i, j, k) = (i_w, i_z, k)$. This shows that $\varphi(L_i) \subseteq \{u_{k'+1}, \ldots, u_j\}$. We conclude that $\varphi(L_i)$ is an independent set in the graph G', implying $\alpha(G') \geq |L_i|$, as claimed. $\square$

We now show that the necessary condition in Lemma 4 is also a sufficient condition for $f_B(i, j, k) = 1$.

Lemma 5. *Let $B \in \mathcal{B}$ be a representative backbone of T, and let (i, j, k) be a triple in $\mathcal{T}_B$ with $i \geq 2$. If there exists an integer $k' \in \{\ell(k), \ldots, j\}$ such that*

$$f_B(i-1, \ell(k)-1, k') = 1 \quad \textit{and} \quad \alpha\Big(G\big[\{u_{k'+1}, \ldots, u_j\} \setminus N_G(u_{k'})\big]\Big) \geq |L_i|\,,$$

then $f_B(i, j, k) = 1$.

Proof. Suppose that the conditions in the lemma are satisfied. By the definition of f_B, there exists an ISI mapping φ of B_{i-1} into $G^{k'}_{\ell(k)-1}$ such that $\varphi(t_{i-1}) = u_{k'}$. We need to show that there exists an ISI mapping φ' of B_i into G^k_j such that $\varphi'(t_i) = u_k$. Let I be an independent set of size $|L_i|$ in the graph $G' = G\big[\{u_{k'+1}, \ldots, u_j\} \setminus N_G(u_{k'})\big]$. We fix a bijection $\psi : L_i \to I$, and we define a mapping $\varphi' : V(B_i) \to V(G)$ as follows: for every $v \in V(B_i)$, we have

$$\varphi'(v) = \begin{cases} \varphi(v) & \text{if } v \in V(B_{i-1})\,; \\ u_k & \text{if } v = t_i\,; \\ \psi(v) & \text{if } v \in L_i\,. \end{cases}$$

Notice that since the vertex set of B_i is the disjoint union of sets $V(B_{i-1})$, $\{t_i\}$ and L_i, the mapping φ' is well-defined. In order to complete the proof, we will verify that φ' is an ISI mapping of B_i into G^k_j such that $\varphi'(t_i) = u_k$. In what follows, we will use the fact that $G^{k'}_{\ell(k)-1}$ is an induced subgraph of G^k_j.

(i) Since

$$\begin{aligned} \varphi'(V(B_i)) &= \varphi(V(B_{i-1})) \cup \{u_k\} \cup \psi(L_i) \\ &\subseteq V(G^{k'}_{\ell(k)-1}) \cup \{u_k\} \cup I \\ &\subseteq \{u_1, \ldots, u_{k'}\} \cup \{u_k\} \cup \{u_{k'+1}, \ldots, u_j\} \\ &= V(G^k_j)\,, \end{aligned}$$

mapping φ' is indeed a mapping from $V(B_i)$ to $V(G^k_j)$.

(ii) Condition $\varphi'(t_i) = u_k$ is satisfied by definition.

(iii) The injectivity of φ' follows immediately from the injectivity of φ and the bijectivity of ψ.

(iv) *φ' preserves adjacencies:*
Let $uv \in E(B_i)$. If $u, v \in V(B_{i-1})$, then

$$\varphi'(u)\varphi'(v) = \varphi(u)\varphi(v) \in E(G^{k'}_{\ell(k)-1}) \subseteq E(G^k_j)\,,$$

where the fact that $\varphi(u)\varphi(v)$ is an edge of $G^{k'}_{\ell(k)-1}$ holds since φ preserves adjacencies. If $u = t_{i-1}$ and $v = t_i$ then $\varphi'(u)\varphi'(v) = u_{k'}u_k$, which is an edge of G^k_j since $\ell(k) \le k' \le j$. Finally, if $u = t_i$ and $v \in L_i$, then $\varphi'(u)\varphi'(v) = u_k\psi(v)$, which is an edge of G^k_j, since $\psi(v) = u_r$ for some $r \in \{k'+1, \ldots, j\} \subseteq \{\ell(k)+1, \ldots, j\}$, implying $u_r \in N_{G^k_j}(u_k)$.

(v) *φ' preserves non-adjacencies:*
Let u, v be a pair of distinct non-adjacent vertices of B_i.
If $u, v \in V(B_{i-1})$, then, since φ preserves non-adjacencies, φ' maps $\{u, v\}$ to a pair of non-adjacent vertices in $G^{k'}_{\ell(k)-1}$, and hence in G^k_j.

Suppose that $u \in V(B_{i-1})$ and $v = t_i$. Then $u \neq t_{i-1}$ and hence $\varphi'(u) \in \{u_1, \ldots, u_{\ell(k)-1}\}$. Consequently, by the definition of $\ell(k)$, vertex $\varphi(u)$ is not adjacent to $u_k = \varphi(v)$ in G^k_j.

Suppose that $u = t_{i-1}$ and $v \in L_i$. Then $\varphi'(u) = u_{k'}$, and $\varphi'(v)$ is not adjacent to $\varphi'(u) = u_{k'}$ since $\varphi'(v) \in I \subseteq \{u_{k'+1}, \ldots, u_j\} \setminus N_G(u_{k'})$.

Finally, suppose that $u \in V(B_{i-1}) \setminus \{t_{i-1}\}$ and $v \in L_i$. Then $\varphi'(u) = u_r$ for some $r \in \{1, \ldots, \ell(k)-1\}$, and $\varphi'(v) = u_s$ for some $s \in \{k'+1, \ldots, j\}$. Suppose for a contradiction that u_r and u_s are adjacent in G^k_j. Since G^k_j is an induced subgraph of G, u_r and u_s are adjacent in G. On the other hand, the definition of I implies that $u_{k'}$ and u_s are non-adjacent in G. However, since $r < k' < s$, this contradicts Definition 1 applied to the triple $(i, j, k) = (r, k', s)$.

The above properties imply that φ' is an ISI mapping of B_i into G^k_j such that $\varphi'(t_i) = u_k$. Consequently $f_B(i, j, k) = 1$, completing the proof of the lemma. □

The results of Lemmas 3–5 can be summarized as follows.

Corollary 1. *For any representative backbone $B = (t_1, \ldots, t_p) \in \mathcal{B}$ of T, the values of the function $f_B : \mathcal{T}_B \to \{0, 1\}$ can be computed recursively as follows:*

- *for $i = 1$ and all $1 \le j < k \le n$, we have $f_B(1, j, k) = 1$ if and only if*

$$\alpha(G[\{u_{\ell(k)}, \ldots, u_j\}]) \ge |L_1|\,;$$

- *for all $i \in \{2, \ldots, p\}$ and all $1 \le j < k \le n$, we have $f_B(i, j, k) = 1$ if and only if there exists an integer $k' \in \{\ell(k), \ldots, j\}$ such that*

$$f_B(i-1, \ell(k)-1, k') = 1 \quad \text{and} \quad \alpha\Big(G\big[\{u_{k'+1}, \ldots, u_j\} \setminus N_G(u_{k'})\big]\Big) \ge |L_i|\,.$$

We are now ready to prove the main result of this paper.

Theorem 1. Induced Subtree Isomorphism *can be solved in polynomial time on interval graphs.*

Proof. Let (G,T) be an instance of Induced Subtree Isomorphism, where $G = (V,E)$ is an interval graph and T is a tree. We assume that T has at least three vertices, as the problem can trivially be solved otherwise. We also assume that $|V(T)| \leq |V(G)|$, as otherwise we have a trivial no-instance. Finally, we assume that G is connected; if G is disconnected, then the polynomial-time algorithm described below can be applied to each of the connected components of G within the same overall time bound.

Let $t = |V(T)|$, $n = |V(G)|$ and $m = |E(G)|$. We start by checking whether T is a caterpillar, which can easily be done in time linear in the size of T. As mentioned in Section 2, every induced subtree of an interval graph is a caterpillar due to a characterization of interval graphs by Lekkerkerker and Boland [20]. Hence, if T is not a caterpillar, then we output "no". Suppose T is a caterpillar. Then we compute a set $\mathcal{B} = \{B_{\min}, B_{\min}^{-1}, B^+, B^-\}$ of at most four representative backbones of T in the way described just below Lemma 1. It is clear that such a set $\mathcal{B}$ can be computed in time $O(t)$. We also compute an interval order $\sigma = (u_1, u_2, \ldots, u_n)$ of G, which can be done in $O(n+m)$ time [27]. Using this interval order σ, we then compute, for all $i \in \{1, \ldots, n\}$, the indices $\ell(i)$ that were defined just above Lemma 3; this takes $O(n+m)$ time in total.

Lemma 2 and Corollary 1 imply that we can determine whether or not T is isomorphic to an induced subgraph of G by computing, for each backbone $B \in \mathcal{B}$, the value of $f_B(|B|, k-1, k)$ for every $k \in \{2, \ldots, n\}$. We will now describe how this can be done in polynomial time for a fixed backbone $B \in \mathcal{B}$. Since $\mathcal{B}$ contains at most four backbones, this suffices to complete the proof of Theorem 1.

Let $B = (t_1, \ldots, t_p) \in \mathcal{B}$ be a representative backbone of T. For every $i \in \{1, \ldots, p\}$, we compute the number $|L_i| = |N_T(t_i) \setminus V(B)|$. Then, for every pair (j,k) with $1 \leq j < k \leq n$, we compute the independence number of the graph $G\big[\{u_{\ell(k)}, \ldots, u_j\}\big]$. Since $G\big[\{u_{\ell(k)}, \ldots, u_j\}\big]$ is an induced subgraph of G, and thus an interval graph, its independence number can be computed in time $O(n+m)$ [10]. Hence, computing $\alpha\big(G\big[\{u_{\ell(k)}, \ldots, u_j\}\big]\big)$ for all pairs (j,k) takes $O(n^2(n+m))$ time in total. We also compute the independence number of the graph $G\big[\{u_{k'+1}, \ldots, u_j\} \setminus N_G(u_{k'})\big]$ for every pair (k', j) with $1 \leq k' \leq j \leq n$, which can be done in $O(n^2(n+m))$ time in total for similar reasons.

Having precomputed these independence numbers and values of $|L_i|$, we can now use the recursions from Corollary 1 to compute the value of $f_B(i,j,k)$ for every $(i,j,k) \in \mathcal{T}_B$, in increasing order of $i \in \{1, \ldots, |B|\}$, in time $O(tn^3)$ in total: each of the values $f_B(i,j,k)$ can be computed in constant time for $i = 1$ and in time $O(n)$ for $i \geq 2$ from the already computed values. The overall time complexity of the algorithm is $O(n^2(n+m)) + O(tn^3) = O(n^2(tn+m))$.

The algorithm can be easily extended so that it also produces an isi mapping of T into G, in case such a mapping exists. We just need to store, for each $B \in \mathcal{B}$ and each $(i,j,k) \in \mathcal{T}_B$ such that $f_B(i,j,k) = 1$, an isi mapping φ of B_i into G_j^k such that $\varphi(t_i) = u_k$. The proofs of Lemmas 3 and 5 show that such mappings can efficiently be computed in a recursive way. $\square$

4 Subtree Isomorphism on Interval Graphs

To complement our positive result on interval graphs in the previous section, we show in this section that Subtree Isomorphism is NP-complete on interval graphs. In fact, we prove that Subtree Isomorphism is NP-complete already on proper interval graphs, a well-known subclass of interval graphs.

We first need to introduce some additional terminology. Let $G = (V, E)$ be a graph. The *width* of an ordering $(u_1, \ldots, u_n)$ of V is $\max\{|i - j| : u_i u_j \in E\}$. The *bandwidth* of G is the minimum width of any ordering of the vertices of G. The Bandwidth problem takes as input a graph G and an integer k, and the task is to decide whether the bandwidth of G is at most k. An ordering $(u_1, \ldots, u_n)$ of V is a *proper interval order* of G if, for every triple (i, j, k) with $1 \leq i < j < k \leq n$, it holds that $u_i u_k \in E$ implies $u_i u_j \in E$ and $u_j u_k \in E$. A graph is a proper interval graph if and only if it has a proper interval order [21].

It follows from the definition of a proper interval order that for any proper interval graph G, the width of a proper interval order of G is exactly the bandwidth of G. Since a proper interval order of a proper interval graph can be computed in linear time [21], Bandwidth is solvable in linear time on proper interval graphs. However, Bandwidth is NP-complete on trees [25], and it is from this problem that we reduce in the proof of the following result.

Theorem 2. Subtree Isomorhism *is NP-complete on proper interval graphs.*

Proof. We give a reduction from Bandwidth on trees. Let T be a tree on n vertices which, together with an integer k, constitutes an instance of Bandwidth. We construct a proper interval graph G as follows: start from a simple path $(v_1, v_2, \ldots, v_n)$, and add edges so that v_i is adjacent to v_j if and only if $j - i \leq k$, for all $1 \leq i < j \leq n$. Such a graph is called a *k-path power* on n vertices, and is well-known to be a proper interval graph. We show that T is a subgraph of G if and only if T has bandwidth at most k.

If T is a subgraph of G, then clearly the bandwidth of T is at most k, since $|j - i| \leq k$ for every edge $v_i v_j$ in G. If the bandwidth of T is at most k, then we can take an ordering of the vertices of T of width at most k, and add edges to make it a k-path power on n vertices. Since no original edge of T has endpoints that are more than k apart in the ordering, it is indeed possible to obtain a k-path power G in this way, which means that T is a subgraph of G. The proof is completed by observing that the problem is trivially in NP. $\square$

Observe that we in fact proved a stronger result than the statement of Theorem 2. A tree is a *spanning subtree* of a graph G if it is a subtree of G and it has the same number of vertices as G. The above proof shows that Spanning Subtree Isomorphism is NP-complete on path powers, which form a subclass of proper interval graphs.

5 Concluding Remarks

As a consequence of our results in this paper and previously known results, the following boundaries are now established on subgraph problems on interval graphs.

The INDUCED SUBGRAPH ISOMORPHISM problem is NP-complete even if both input graphs are connected interval graphs [2, 7], whereas it becomes polynomial-time solvable if G is an interval graph and H is a tree. The SUBGRAPH ISOMORPHISM problem is NP-complete even if G is a proper interval graph and H is a tree, but it becomes polynomial-time solvable if G is an interval graph and H is a path.

The problem of deciding, given a graph G and an integer k, whether there exists a (not necessarily induced) path of length k in G, is NP-complete on bipartite graphs [19]. An easy reduction from this problem, using the observation that a graph contains a path of length k if and only if its line graph contains an induced path of length $k-1$, shows that INDUCED SUBTREE ISOMORPHISM is NP-complete on line graphs of bipartite graphs, a well-known subclass of perfect graphs. This contrasts our positive result on interval graphs in the following sense. Line graphs are claw-free, implying that every induced tree of a line graph is a path. The restricted nature of induced subtrees of interval graphs allowed us to obtain a polynomial-time algorithm for INDUCED SUBTREE ISOMORPHISM on interval graphs. However, although line graphs have even more restricted induced subtrees than interval graphs, this does not imply tractability for INDUCED SUBTREE ISOMORPHISM on line graphs.

We would also like to mention that INDUCED SUBTREE ISOMORPHISM is trivially solvable in polynomial time on cographs and on split graphs, since every induced subtree of a cograph or a split graph is a caterpillar that has a backbone on at most two vertices. By similar arguments, the problem can also be solved in polynomial time on $3K_2$-free graphs, a superclass of split graphs.

We conclude with the following two questions regarding the complexity of INDUCED SUBTREE ISOMORPHISM problem on two graph classes generalizing interval graphs:

- What is the computational complexity of INDUCED SUBTREE ISOMORPHISM on chordal graphs, a superclass of both interval graphs and split graphs? Note that this problem is NP-complete on perfect graphs, a superclass of chordal graphs, due to the aforementioned NP-completeness results on bipartite graphs [9] and on line graphs of bipartite graphs. Also note that the easier problem of finding a longest induced path can be solved in polynomial time on chordal graphs, or more generally, in $O(n^k)$ time on k-chordal graphs, i.e., on graphs having no induced cycles on more than k vertices [15].
- What is the computational complexity of INDUCED SUBTREE ISOMORPHISM on AT-free graphs? This is a superclass of interval graphs, which also generalizes the cocomparability graphs. AT-free graphs share many features with interval graphs that were used by our algorithm in Section 3: they have some kind of linear structure [6, 17], the only possible induced subtrees in an AT-free graph are caterpillars, and computing the independence number of an AT-free graph is a polynomially solvable task [4]. Also note that the problem of finding a longest induced path can be solved in polynomial time on AT-free graphs [15, 18].

References

1. Agarwal, R.K.: An investigation of the subgraph isomorphism problem. M.Sc. Thesis, Dept. of Computer Science, University of Toronto, TR 138180 (1980)

2. Belmonte, R., Heggernes, P., van 't Hof, P.: Edge contractions in subclasses of chordal graphs. Discrete Appl. Math. 160:999–1010 (2012)
3. Brandstädt, A., Le, V.B., Spinrad, J.: Graph Classes: A Survey. SIAM, Philadelphia (1999)
4. Broersma, H., Kloks, T., Kratsch, D., Müller, H.: Independent sets in asteroidal triple-free graphs. SIAM J. Discrete Math. 12: 276–287 (1999)
5. Corneil, D.G., Lerchs, H., Stewart Burlingham, L.: Complement reducible graphs. Discrete App. Math. 3:163–174 (1981)
6. Corneil, D.G., Olariu, S., Stewart L.: Asteroidal triple-free graphs, SIAM J. Discrete Math. 10:399–430 (1997)
7. Damaschke, P.: Induced subgraph isomorphism for cographs is NP-complete. In: Proceedings WG 1991, LNCS 484, pp. 72–78. Springer (1991)
8. Diestel, R.: Graph Theory. Electronic Edition (2005)
9. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-completeness. W.H. Freeman & Co. (1979)
10. Gavril, F.: Algorithms for minimum coloring, maximum clique, minimum covering by cliques, and maximum independent set of a chordal graph. SIAM Journal on Computing, 1:180–187 (1972)
11. Golovach, P.A., Kamiński, M., Paulusma, D., Thilikos, D.M.: Containment relations in split graphs. Discrete Appl. Math. 160:155–163 (2012)
12. Golumbic, M.C.: Algorithmic Graph Theory and Perfect Graphs, Second Edition. Annals of Discrete Mathematics, Volume 57. North Holland (2004)
13. Heggernes, P., van 't Hof, P., Meister, D., Villanger, Y.: Induced subgraph isomorphism on proper interval and bipartite permutation graphs. Submitted.
14. Ioannidou, K., Mertzios, G.B., Nikolopoulos, S.D.: The longest path problem has a polynomial solution on interval graphs. Algorithmica 61:320–341 (2011)
15. Ishizeki, T., Otachi, Y., Yamazaki, K.: An improved algorithm for the longest induced path problem on k-chordal graphs. Discrete Appl. Math. 156:3057–3059 (2008)
16. Kijima, S., Otachi, Y., Saitoh, T., Uno, T.: Subgraph isomorphism in graph classes. Discrete Math. 312:3164–3173 (2012)
17. Kratsch, D.: Domination and total domination on asteroidal triple-free graphs. Discrete Appl. Math. 99: 111–123 (2000)
18. Kratsch, D., Müller, H., Todinca, I.: Feedback vertex set and longest induced path on AT-free graphs. In: Proceedings WG 2003, LNCS 2880, pp. 309–321. Springer (2003)
19. Krishnamoorthy, M.S.: An NP-hard problem in bipartite graphs. ACM SIGACT News 7(1) 26 (1975)
20. Lekkerkerker, C.G., Boland J.C., Representation of a finite graph by a set of intervals on the real line. Fund. Math. 51:45–64 (1962)
21. Looges, P.J., Olariu, S.: Optimal greedy algorithms for indifference graphs. Computers Math. Applic. 25:15–25 (1993)
22. Matousek, J., Thomas, R.: On the complexity of finding iso- and other morphisms for partial k-trees. Discrete Math. 108:343–364 (1992)
23. Matula, D.W.: An algorithm for subtree identification. SIAM Rev. 10:273–274 (1968)
24. Mertzios, G.B., Corneil, D.G.: A simple polynomial algorithm for the longest path problem on cocomparability graphs. SIAM J. Discrete Math. 26: 940–963 (2012)
25. Monien, B.: The bandwidth minimization problem for caterpillars with hair length 3 is NP-complete. SIAM J. Alg. Discr. Meth. 7:505–512 (1986)
26. Müller, H.: Hamiltonian circuits in chordal bipartite graphs. Discrete Math. 156:291–298 (1996)
27. Olariu, S.: An optimal greedy heuristic to color interval graphs. Inform. Proc. Lett. 37:21–25 (1991)