Characterizing Path Graphs by Forbidden Induced Subgraphs

Benjamin Lévêque,[1] Frédéric Maffray,[2] and Myriam Preissmann[2]

[1] *ROSE, EPFL, LAUSANNE, SWITZERLAND*
E-mail: benjamin.leveque@free.fr

[2] *C.N.R.S., LABORATOIRE G-SCOP, GRENOBLE, FRANCE*
E-mail: frederic.maffray@g-scop.inpg.fr, myriam.preissmann@g-scop.inpg.fr

Received May 6, 2008; Revised January 27, 2009

Published online 6 April 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI 10.1002/jgt.20407

Abstract: A path graph is the intersection graph of subpaths of a tree. In 1970, Renz asked for a characterization of path graphs by forbidden induced subgraphs. We answer this question by determining the complete list of graphs that are not path graphs and are minimal with this property.

Keywords: *intersection graphs; path graphs; forbidden induced subgraphs*

1. INTRODUCTION

All graphs considered here are finite and have no parallel edges and no loop. A *hole* is a chordless cycle of length at least four. A graph is *chordal* (or *triangulated*) if it contains no hole as an induced subgraph. Gavril [7] proved that a graph is chordal if and only if it is the intersection graph of a family of subtrees of a tree. In this paper, whenever we talk about the intersection of subgraphs of a graph we mean that the *vertex sets* of the subgraphs intersect.

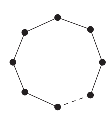
An *interval graph* is the intersection graph of a family of intervals on the real line; equivalently, it is the intersection graph of a family of subpaths of a path. An *asteroidal triple* in a graph G is a set of three non-adjacent vertices such that for any two of them, there exists a path between them in G that does not intersect the neighborhood of the third. Lekkerkerker and Boland [13] proved that a graph is an interval graph if and only if it is chordal and contains no asteroidal triple. They derived from this result the list of minimal forbidden subgraphs for interval graphs.

An intermediate class is the class of path graphs. A graph is a *path graph* if it is the intersection graph of a family of subpaths of a tree. Clearly, the class of path graphs is included in the class of chordal graphs and contains the class of interval graphs. Several characterizations of path graphs have been given [8, 15, 17] but no characterization by forbidden subgraphs was known, whereas such results exist for intersection graphs of subpaths of a path (interval graphs [13]), subtrees of a tree (chordal graphs [7]), and also for directed subpaths of a directed tree (directed path graphs [16]).

In 1970, Renz [17] asked for a complete list of graphs that are chordal, not path graphs, and are minimal with this property, and he gave two examples of such graphs. The list of minimal forbidden subgraphs for path graphs was extended in [21], but that list is incomplete. Here, we answer Renz's question and obtain a characterization of path graphs by forbidden induced subgraphs. We will prove that the graphs presented in Figures 1–5 are all the minimal non-path graphs. In other words:

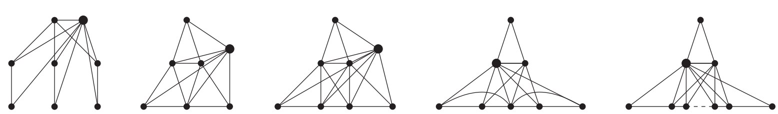
Theorem 1. *A graph is a path graph if and only if it does not contain any members of the families of* $F_0, \ldots, F_{16}$ *as an induced subgraph.*

We could not find a characterization similar to the one found by Lekkerkerker and Boland [13] for interval graphs ("an interval graph is a chordal graph with no asteroidal triple"). We know that in a path graph, the neighborhood of every vertex contains no



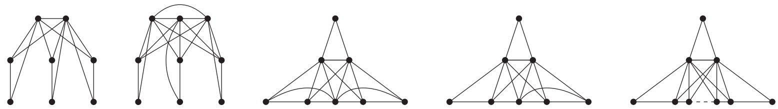
$F_0(n)_{n\geq 4}$

FIGURE 1. Forbidden subgraphs with no simplicial vertices.



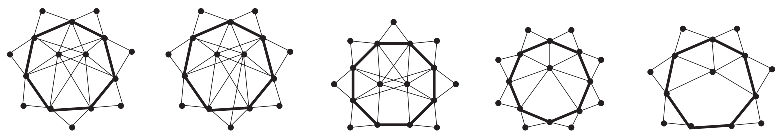
F_1 F_2 F_3 F_4 $F_5(n)_{n\geq 7}$

FIGURE 2. Forbidden subgraphs with a universal vertex.



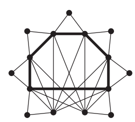
F_6 F_7 F_8 F_9 $F_{10}(n)_{n\geq 8}$

FIGURE 3. Forbidden subgraphs with no universal vertex and exactly three simplicial vertices.



$F_{11}(4k)_{k\geq 2}$ $F_{12}(4k)_{k\geq 2}$ $F_{13}(4k+1)_{k\geq 2}$ $F_{14}(4k+1)_{k\geq 2}$ $F_{15}(4k+2)_{k\geq 2}$

FIGURE 4. Forbidden subgraphs with at least one simplicial vertex that is not co-special (bold edges form a clique).



$F_{16}(4k+3)_{k\geq 2}$

FIGURE 5. Forbidden subgraphs with ≥ 4 simplicial vertices that are all co-special (bold edges form a clique).

asteroidal triple; but this condition is not sufficient. So we prove directly that a graph that does not contain any of the excluded subgraphs is a path graph. The initial proof of the characterization of interval graphs by Lekkerkerker and Boland [13] was fairly complicated. It was simplified by Halin [12] by using the concept of prime graph decomposition. Cameron et al. [3] translated Halin's proof in terms of clique tree. Our proof is, in its principle, a generalization of the proof presented in [3].

2. SPECIAL SIMPLICIAL VERTICES IN CHORDAL GRAPHS

In a graph G, a *clique* is a set of pairwise adjacent vertices. Let $\mathcal{Q}(G)$ be the set of all (inclusionwise) maximal cliques of G. When there is no ambiguity we will write $\mathcal{Q}$ instead of $\mathcal{Q}(G)$.

Given two vertices u, v in a graph G, a $\{u,v\}$-*separator* is a set S of vertices of G such that u and v lie in two different components of $G \setminus S$ and S is minimal with this property. A set is a *separator* if it is a $\{u,v\}$-separator for some u, v in G. Let $\mathcal{S}(G)$ be the set of separators of G. When there is no ambiguity we will write $\mathcal{S}$ instead of $\mathcal{S}(G)$.

The neighborhood of a vertex v is the set $N(v)$ of vertices adjacent to v. For a set X of vertices, let $N(X) = (\bigcup_{v\in X} N(v)) \setminus X$. Let us say that a vertex u is *complete* to a set X

of vertices if $X \subseteq N(u)$. A vertex is *simplicial* if its neighborhood is a clique. It is easy to see that a vertex is simplicial if and only if it does not belong to any separator. Given a simplicial vertex v, let $Q_v = N(v) \cup \{v\}$ and $S_v = Q_v \cap N(V \setminus Q_v)$. Since v is simplicial, Q_v is the unique maximal clique containing v. Remark that S_v is not necessarily in $\mathcal{S}$; for example, in the graph H with vertices a, b, c, d, e and edges ab, bc, cd, de, bd, we have $S_c = \{b, d\}$ and $\mathcal{S}(H) = \{\{b\}, \{d\}\}$.

A classical result [1, 11] (see also [9]) states that, in a chordal graph G, every separator is a clique; moreover, if S is a separator, then there are at least two components of $G \setminus S$ that contain a vertex that is complete to S, and so S is the intersection of two maximal cliques.

A *clique tree* T of a graph G is a tree whose vertices are the members of $\mathcal{Q}$ and such that, for each vertex v of G, those members of $\mathcal{Q}$ that contain v induce a subtree of T, which we will denote by T^v. Note that G is the intersection graph of these subtrees. Gavril [7] proved the classical result that a graph is chordal if and only if it has a clique tree.

Clique trees are very useful when studying chordal graphs or subclasses of chordal graphs as they give the structure of graphs for which they are a clique tree. We recall the definitions and properties of clique trees that we need in the article, but the reader who is not familiar with this notion can refer to classical books of graph theory (like [9, 14]). Our proofs are done in the clique tree. Occasionally, we will have to refer to the original graph (for example, to obtain the forbidden subgraphs explicitly) but most of the time everything can be understood just by studying the clique tree.

In a clique tree T, the *label* of an edge QQ' of T is defined as $S_{QQ'} = Q \cap Q'$. Note that every edge QQ' satisfies $S_{QQ'} \in \mathcal{S}$; indeed, there exist vertices $v \in Q \setminus Q'$ and $v' \in Q' \setminus Q$, and the set $S_{QQ'}$ is a $\{v, v'\}$-separator. The number of times an element S of $\mathcal{S}$ appears as a label of an edge is equal to $c-1$, where c is the number of components of $G \setminus S$ that contains a vertex complete to S [7, 14]. As pointed out above, c is at least two; moreover, it depends only on S and not on T; so, for a given $S \in \mathcal{S}$, the number $c-1$ is the same in every clique tree.

Given a set $X \subseteq \mathcal{Q}$ of maximal cliques, let $G(X)$ denote the subgraph of G induced by all the vertices that appear in members of X. If T is a clique tree of G, then $T[X]$ denotes the subtree of T of minimum size such that its set of vertices contains X. Note that if $|X| = 2$, then $T[X]$ is a path.

Given a subtree T' of a clique-tree T of G, let $\mathcal{Q}(T')$ be the set of vertices of T' and $\mathcal{S}(T')$ be the set of separators of $G(\mathcal{Q}(T'))$. It is easy to verify the following important property: T' is a clique tree of $G(\mathcal{Q}(T'))$. Moreover, if $T' \neq T$ there exists a leaf L of T not in T', and a vertex in L that is not in any vertex of T', so $G(\mathcal{Q}(T'))$ is a strict induced subgraph of G.

Dirac [6] proved that a chordal graph that is not a clique contains two non-adjacent simplicial vertices. We need to generalize this theorem to the following. Let us say that a simplicial vertex v is *special* if S_v is a member of $\mathcal{S}$ and is (inclusionwise) maximal in $\mathcal{S}$.

Theorem 2. *In a chordal graph that is not a clique, there exist two non-adjacent special simplicial vertices.*

Proof. By the hypothesis G is not a clique, so $|\mathcal{Q}| \geq 2$ and $\mathcal{S} \neq \emptyset$. Let T be a clique tree of G.

Let us choose, in the set of vertices of T incident to an edge with (inclusionwise) maximal label, two maximal cliques Q_1, Q_2 that are at a maximum distance in T. Since $\mathcal{S} \neq \emptyset$ these maximal cliques are distinct.

For $i = 1, 2$, let Q'_i be the neighbor of Q_i on $T[Q_1, Q_2]$ (possibly $Q'_1 = Q'_2$ or both $Q'_1 = Q_2$ and $Q'_2 = Q_1$). By the choice of Q_1, Q_2, the label $S_{Q_i Q'_i}$ of $Q_i Q'_i$ is maximal and no edge of T_i, the subtree of $T \setminus Q'_i$ that contains Q_i, has a maximal label. So the label of each edge of T_i is included in $S_{Q_i Q'_i}$. Let $v_i \in Q_i \setminus Q'_i$. As v_i is not in $S_{Q_i Q'_i}$, it is not in any label of T_i and so not in any label of T. Thus, v_i is simplicial and $Q_{v_i} = Q_i$. All the labels of the edges incident to Q_i are included in $S_{Q_i Q'_i}$, so $S_{v_i} = S_{Q_i Q'_i}$ and v_i is special. Since Q_{v_1} and Q_{v_2} are distinct cliques, v_1 and v_2 are not adjacent. ■

Algorithms LexBFS [18] and MCS [20] are linear time algorithms that were developed to find a simplicial elimination ordering in a chordal graph. (A *simplicial elimination ordering* is an ordering of the vertices $v_1, \ldots, v_n$ such that, for $1 < i \leq n$, vertex v_i is simplicial in the graph induced by vertices $v_1, \ldots, v_{i-1}$.) The last vertex found by these algorithms is simplicial in the whole graph. This vertex is not necessarily special simplicial. For example, on the graph with vertices a, b, c, d, e, f and edges $ab, bc, cd, eb, ec, fb, fc$, every application of LexBFS or MCS will end on one of the simplicial vertices a, d, which are not special. The proof of Theorem 2 can be turned into a polynomial time algorithm to find a special simplicial vertex in a chordal graph. We leave open the problem of finding a special simplicial elimination ordering in linear time. (A *special simplicial elimination ordering* is a simplicial elimination ordering where vertex v_i is a special simplicial vertex in the graph induced by vertices $v_1, \ldots, v_{i-1}$.)

3. FORBIDDEN INDUCED SUBGRAPHS

A *clique path tree* T of G is a clique tree of G such that, for each vertex v of G, the subtree T^v induced by the cliques that contain v is a path. Note that G is the intersection graph of these subpaths. Gavril [8] proved that a graph is a path graph if and only if it has a clique path tree. A graph G is a *minimal non-path graph* if G is not a path graph but any induced subgraph of G distinct from G is a path graph. Note that any induced subgraph of a path graph is also a path graph, so it is enough to require that $G \setminus v$ is a path graph for every vertex v of G.

Consider graphs $F_0, \ldots, F_{16}$ presented in Figures 1–5. Let us make a few remarks about them. Each graph in Figure 2 is obtained by adding a universal vertex to some minimal forbidden subgraph for interval graphs. Clearly, in a path graph the neighborhood of every vertex is an interval graph; so $F_1, \ldots, F_5$ are not path graphs. Graphs $F_{10}(n)_{n \geq 8}$ are also forbidden in interval graphs. Graphs F_6 and $F_{10}(8)$ are from Renz [17, Figures 1 and 5]. For $i \in \{0, 1, 3, 4, 5, 6, 7, 9, 10, 13, 15, 16\}$, Panda [16] proved that F_i is a minimal non-directed path graph, so $F_i \setminus x$ is a directed path graph for every vertex x (obviously every directed path graph is a path graph). In general, we have the following:

Theorem 3. *$F_0, \ldots, F_{16}$ are families of minimal non-path graphs.*

Proof. Clearly, F_0 is a minimal non-path graph. As pointed out above, $F_1, \ldots, F_5$ are not path graphs and it is easy to verify that by deleting any vertex of these graphs one obtains a path graph.

We do not give a detailed proof for each graph, but we show a general statement that can be used to prove that $F_6, \ldots, F_{16}$ are not path graphs. Let $F = F_i$ for some $i \in \{6, \ldots, 16\}$. A maximal clique of F will be called *peripheral* if it contains only one separator S and $G \setminus S$ has only two connected components. Such a clique must be a leaf in any clique tree of F. A maximal clique that is not peripheral will be called *central*. In any clique tree of F the central cliques will induce a subtree. It is easy to see that there is a set K of vertices of F such that every central clique of F contains K and every peripheral clique intersects K. Therefore, if there is a clique path tree T of F, then the central cliques must induce a path in T, every peripheral clique must be adjacent to some extremity of this path, and if two peripheral cliques share a common vertex of K then they must be adjacent to distinct extremities. Another simple remark is that if a vertex of F is in exactly two maximal cliques, then these two cliques must be adjacent in any clique tree of G.

Let $P_1, \ldots, P_k$ and $C_1, \ldots, C_\ell$ be, respectively, the peripheral and the central cliques of F. The following conditions for a path tree of F are easy to check using our previous remarks:

- When $F = F_6$ or F_7, then $k = \ell = 3$ and P_i must be adjacent to C_i.
- When $F = F_8, F_9, F_{10}, F_{11}, F_{12}, F_{13}$, or F_{16}, then k is odd, for each $i \in \{1, \ldots, k-1\}$, P_i and P_{i+1} must be adjacent to distinct extremities of the path induced by the central cliques, and this holds also for P_1 and P_k.
- When $F = F_{14}$ or F_{15}, then k is even, for each $i \in \{1, \ldots, k-1\}$, P_i and P_{i+1} must be adjacent to distinct extremities of the path induced by the central cliques, and P_1 and P_k must be adjacent to the same extremity.

In each of these cases, it is easy to see that no clique tree can satisfy all the conditions.

Furthermore, it is not difficult to check that when we delete any vertex of F then one of the above constraints is removed and this is sufficient to make it possible to construct a clique path tree. ■

4. CO-SPECIAL SIMPLICIAL VERTICES

Let us say that a simplicial vertex v is *co-special* if S_v is a separator such that $G \setminus S_v$ has exactly two components. Lekkerkerker and Boland [13] call this type of vertex *strongly simplicial*. Note that in that case S_v is a minimal element of $\mathcal{S}$ and it appears exactly once as a label of any path tree of G. (The fact that S_v is a minimal element of $\mathcal{S}$ for some simplicial vertex v does not imply that v is co-special; for example, consider the graph with vertices a, b, c, d and edges ab, ac, ad; in fact it has no co-special vertex.)

Also note that, in contrast with Theorem 2, a chordal graph does not necessarily have a simplicial vertex v where S_v is a minimal element of $\mathcal{S}$; for example, consider the graph with seven vertices a, b, c, d, e, f, g and edges $bc, cd, ef, fg, ab, ac, ad, ae, af, ag$.

Lemma 1. *Let G be a minimal non-path graph. Then either G is one of* $F_{11}, \dots, F_{15}$ *or every simplicial vertex of G is co-special.*

Proof. Suppose on the contrary that G is a minimal non-path graph, different from $F_{11}, \dots, F_{15}$, and there is a simplicial vertex q of G that is not co-special. All simplicial vertices of $F_0, \dots, F_{10}, F_{16}$ are co-special, so G is not any of these graphs; moreover, it does not contain any of them strictly (for otherwise G would not be minimal). Therefore, G contains none of $F_0, \dots, F_{16}$.

The graph G is not a clique as it is not a path graph. Let us denote by Q the unique maximal clique that contains q. Consider graph $G(\mathcal{Q} \setminus Q)$, which is equal to $G \setminus (Q \setminus S_q)$. Since $Q \setminus S_q \neq \emptyset$, and by the minimality of G, it follows that $G(\mathcal{Q} \setminus Q)$ admits a clique path tree T_0. Let Q' be a vertex of T_0 such that $S_q \subseteq Q'$. If $Q' = S_q$, then by replacing Q' by Q in T_0 we obtain a clique path tree of G, a contradiction. So $Q' \neq S_q$. Let $q' \in Q' \setminus S_q$, then $S_q = Q \cap Q'$ is a $\{q, q'\}$-separator. Let T'_0 be obtained from T_0 by adding vertex Q and edge QQ'. Remark that T'_0 is a clique tree of G but not a clique path tree since G is not a path graph.

Let T' be the maximal subtree of T'_0 that contains Q and Q' and such that the edge QQ' is the only edge whose label is included in S_q. So T' is a clique tree of $G(\mathcal{Q}(T'))$. Since q is not co-special, there is an edge of T_0 whose label is included in S_q, and so T' is a strict subgraph of T'_0. So $G(\mathcal{Q}(T'))$ is a strict subgraph of G and by the minimality of G it is a path graph. Let T be a clique path tree of this graph. From now on, our goal will be to show that either G contains one of the forbidden subgraphs or T can be extended into a clique path tree of G.

We claim that Q is a leaf of T. If not, then there are at least two labels of T that are included in S_q, which contradicts the definition of T' (the number of times a label appears in a clique tree is constant).

Let $T_1, \dots, T_\ell$ be the subtrees of $T'_0 \setminus T'$ $(\ell \geq 1)$. For $1 \leq i \leq \ell$, let $Q_i Q'_i$ be the edge between T_i and T' with $Q_i \in T_i$ and $Q'_i \in T'$. Note that $Q_1, \dots, Q_\ell$ are pairwise disjoint (but $Q', Q'_1, \dots, Q'_\ell$ are not necessarily pairwise disjoint). Let $S_i = Q_i \cap Q'_i$ and $v_i \in Q_i \setminus Q'_i$. Let $\mathcal{H}$ be the intersection graph of $S_1, \dots, S_\ell$, that is, $\mathcal{H}$ has vertex-set $V_{\mathcal{H}} = \{S_1, \dots, S_\ell\}$ and edge-set $E_{\mathcal{H}} = \{S_i S_j | S_i \cap S_j \neq \emptyset\}$.

Claim 1. *$\mathcal{H}$ contains no odd cycle.*

Proof. Suppose on the contrary, without loss of generality, that $S_1 - \cdots - S_p - S_1$ is an odd cycle in $\mathcal{H}$, with length $p = 2r+1$ $(r \geq 1)$. Let $I_j = S_j \cap S_{j+1}$ $(j = 1, \dots, p)$ with $S_{p+1} = S_1$. Suppose that for some $j \neq k$ we have $I_j \cap I_k \neq \emptyset$; then there is a common vertex in the cliques $Q_j, Q_{j+1}, Q_k, Q_{k+1}$, and the number of different cliques among these is at least three, which contradicts the fact that T_0 is a clique path tree as these three cliques do not lie on a common path of T_0. Therefore, we have $I_j \cap I_k = \emptyset$ whenever $j \neq k$. For $1 \leq j \leq p$, let $s_j \in I_j$. By the preceding remark, the s_j's are pairwise distinct. By the definition of T', we have $S_j \subseteq S_q$ for each $1 \leq j \leq p$, so the s_j's are all in Q and Q'. Let us consider the subgraph induced by $q, q', v_1, \ldots, v_p, s_1, \ldots, s_p$. Both q and q' are adjacent to all of the clique formed by the s_j's. Each vertex v_j is adjacent to s_{j-1} and s_j (with $s_0 = s_p$) and not to any other s_i or to q. Vertex q' has no neighbor among the v_j's, for otherwise q' is in some S_j and then also in $S_q \subseteq Q$, a contradiction to its definition. Now $\{q, q', v_1, \ldots, v_p, s_1, \ldots, s_p\}$ induces $F_{11}(4r+4)_{r \geq 1}$, a contradiction. Thus the claim holds. ■

By the preceding claim, $\mathcal{H}$ is a bipartite graph.

For $1 \leq i \leq \ell$, let $\mathcal{R}_i = \{S \in \mathcal{S}(T') | S_i \cap S \neq \emptyset$ and $S_i \setminus S \neq \emptyset\}$. Let $X = \{S_i | \mathcal{R}_i \neq \emptyset\}$. We remark that $S_Q \notin \mathcal{R}_i$.

Claim 2. *There is no odd path between two vertices of X in $\mathcal{H}$.*

Proof. Suppose on the contrary, without loss of generality, that $S_1 - \cdots - S_p$ is an odd path in $\mathcal{H}$ between two vertices S_1, S_p of X (with $p = 2k$, $k \geq 1$), and assume that p is minimum with this property. By the minimality, all interior vertices S_j $(1 < j < p)$ are not in X. For $1 \leq j < p$, let s_j be a vertex in $S_j \cap S_{j+1}$. As in the preceding claim, the s_j's are pairwise distinct and lie in Q and Q'. Let P be the path $T'[Q'_1, Q'_2]$. We note that when $p > 2$, then S_2 is not in X, so $Q'_3 = Q'_1$, for otherwise $T_0^{S_2}$ would not be a path; then S_3 is not in X, so $Q'_4 = Q'_2$, and so on. Thus, the two extremities of P are $Q'_1 = Q'_3 = \cdots = Q'_{p-1}$ and $Q'_2 = Q'_4 = \cdots = Q'_p$. Since S_1 and S_p are in X, the sets $\mathcal{R}_1, \mathcal{R}_p$ are non-empty.

Let L_1 be the closest vertex to Q'_1 in P such that there exists an edge incident to L_1 with label in $\mathcal{R}_1$, and let L_1K_1 be such an edge and R_1 be its label (such an edge exists because $\mathcal{R}_1 \neq \emptyset$). Similarly, let L_p be the closest vertex to Q'_p in P such that there exists an edge incident to L_p with label in $\mathcal{R}_p$, and let L_pK_p be such an edge and R_p be its label. So $S_1 \subseteq L_1$, $S_1 \nsubseteq K_1$ and $S_p \subseteq L_p$, $S_p \nsubseteq K_p$. Each of K_1, K_p may be in P or not. Since $T' \setminus Q$ is a clique path tree, Q' lies between Q'_1 and L_1 and between L_p and Q'_p along P. So Q'_1, L_p, Q', L_1, Q'_p lie in this order on P, and S_1 is included in all labels between Q'_1 and L_1 in P, and S_p is included in all labels between Q'_p and L_p in P.

Let $v_0 \in K_1 \setminus L_1$ and $v_{p+1} \in K_p \setminus L_p$. Since T'_0 is a clique tree, v_0 and v_{p+1} are distinct from $v_1, \ldots, v_p$ and not adjacent to q.

Let $s_0 \in S_1 \cap R_1$ and $s_p \in S_p \cap R_p$. Then v_0 and s_0 are adjacent, and v_{p+1} and s_p are adjacent. Since T_0 is a clique path tree, if K_1 or K_p is not in P, then s_0 and s_p are different from each other, from $s_1, \ldots, s_{p-1}$ and from $v_0, \ldots, v_{p+1}$. Furthermore, if K_1 is not in P, then v_0 is not adjacent to any of $s_1, \ldots, s_p$; and if K_p is not in P, then v_{p+1} is not adjacent to any of $s_0, \ldots, s_{p-1}$.

Let $s'_0 \in S_1 \setminus R_1$ and $s'_p \in S_p \setminus R_p$. Then v_0 and s'_0 are not adjacent, and v_{p+1} and s'_p are not adjacent. Since T_0 is a clique path tree, if K_1 or K_p is in P, then s'_0 and s'_p are different from each other, from $s_1, \ldots, s_{p-1}$ and from $v_0, \ldots, v_{p+1}$. Furthermore, if K_1 is in P, then v_0 is adjacent to s'_p and to $s_1, \ldots, s_p$; and if K_p is in P, then v_{p+1} is adjacent to s'_0 and to $s_0, \ldots, s_{p-1}$.

Note that $\{q, s'_0, s_0, s_1, s_2, \ldots, s_p, s'_p\}$ induces a clique in G. Moreover, v_1 is adjacent to s'_0, v_p is adjacent to s'_p, for $i = 1, \ldots, p$, v_i is adjacent to s_{i-1} and s_i, and there is no other edge between $v_1, \ldots, v_p$ and that clique.

Suppose that $K_1 = K_p$. Then $L_1 = L_p = Q'$ and K_1 is not in P. By the definition of T', there exists $y \in R_1 \setminus S_q$. Vertex y is distinct from all s_i's as it is not in S_q, and it is adjacent to all of $v_0, s_0, \dots, s_p$ and to none of $q, v_1, \dots, v_p$. Then $\{q, y, v_0, \dots, v_p, s_0, \dots, s_p\}$ induces $F_{12}(4k+4)_{k\geq 1}$, a contradiction. So $K_1 \neq K_p$, and v_0 and v_{p+1} are distinct non-adjacent vertices. We can choose vertices $x_1, \dots, x_r$ $(r \geq 1)$ not in S_q and on the labels of $T'[K_1, K_p]$ such that $v_0 - x_1 - \cdots - x_r - v_{p+1}$ is a chordless path in G. Vertices $x_1, \dots, x_r$ are distinct from and adjacent to $s'_0, s'_p, s_0, \dots, s_p$, and they are distinct from and not adjacent to any of $q, v_1, \dots, v_p$.

Suppose that $L_1 = Q'_p$ and $L_p = Q'_1$. Then K_1 and K_p are not in P. If $r = 1$, then $\{q, v_0, \dots, v_{p+1}, s_0, \dots, s_p, x_1\}$ induces $F_{14}(4k+5)_{k\geq 1}$. If $r = 2$, then $\{q, v_0, \dots, v_{p+1}, s_0, \dots, s_p, x_1, x_2\}$ induces $F_{15}(4k+6)_{k\geq 1}$. If $r \geq 3$, then $\{q, v_0, v_{p+1}, s_0, s_p, x_1, \dots, x_r\}$ induces $F_{10}(r+5)_{r\geq 3}$, a contradiction.

Suppose now that $L_1 \neq Q'_p$ and $L_p = Q'_1$. Then K_p is not in P and we may assume that K_1 is in P. If $r = 1$, then $\{q, v_0, \dots, v_{p+1}, s'_0, s_1 \dots, s_p, x_1\}$ induces $F_{13}(4k+5)_{k\geq 1}$. If $r \geq 2$, then $\{q, v_0, v_{p+1}, x_1, \dots, x_r, s'_0, s_p\}$ induces $F_5(r+5)_{r\geq 2}$, a contradiction.

Suppose finally that $L_1 \neq Q'_p$ and $L_p \neq Q'_1$. Then we may assume that K_1 and K_p are in P. If $r = 1$, then $\{q, v_0, v_{p+1}, s'_0, s_1, s'_p, x_1\}$ induces F_2. If $r = 2$, then $\{q, v_0, v_{p+1}, s'_0, s_1, s'_p, x_1, x_2\}$ induces F_3. If $r \geq 3$, then $\{q, v_0, v_{p+1}, x_1, \dots, x_r, s'_0, s'_p\}$ induces $F_{10}(r+5)_{r\geq 3}$, a contradiction. Thus the claim holds. ■

By the preceding two claims, $\mathcal{H}$ is a bipartite graph, so its vertex-set can be partitioned into two stable sets $A_{\mathcal{H}}, B_{\mathcal{H}}$, and we may assume that $X \subseteq A_{\mathcal{H}}$. Now all the subtrees T_i can be linked to T to get a clique path tree of G as follows. For each $S_i \in A_{\mathcal{H}}$, we add an edge QQ_i between T and T_i. This creates a clique path tree on the corresponding subset of cliques because $A_{\mathcal{H}}$ is a stable set of $\mathcal{H}$ and Q is a leaf of T. For each $S_i \in B_{\mathcal{H}}$, let $Q''_i \in \mathcal{Q}(T)$ be such that $Q''_i \cap S_i \neq \emptyset$ and the length of $T[Q, Q''_i]$ is maximal. Since $S_i \in B_{\mathcal{H}}$, we have $\mathcal{R}_i = \emptyset$, so $S_i \subseteq Q''_i$ and we can add an edge $Q''_i Q_i$ between T and T_i. This creates a clique path tree of G because $B_{\mathcal{H}}$ is a stable set of $\mathcal{H}$ and by the definition of Q''_i, a contradiction. ■

5. CHARACTERIZATION OF PATH GRAPHS

In this section we prove the main theorem, that is, path graphs are exactly the graphs that do not contain any of $F_0, \dots, F_{16}$.

Lemma 2. *In a graph that does not contain any member of the families of $F_0, \dots, F_5, F_{10}$, the neighborhood of every vertex does not contain an asteroidal triple.*

Proof. Suppose that in a graph G the neighborhood of some vertex v contains an asteroidal triple. Then, by [13], the neighborhood contains a minimal forbidden induced subgraph H for interval graphs. Then H and v induce one of $F_0, \dots, F_5, F_{10}$ in G. ■

Given three non-adjacent vertices a, b, c, we say that a is in the *middle* of b, c if every path between b and c contains a vertex from $N(a)$. If a, b, c is not an asteroidal triple, then at least one of them is in the middle of the others.

Lemma 3. *In a chordal graph G with clique tree T, a vertex a is in the middle of two vertices b, c if and only if for all maximal cliques Q_b and Q_c with $b \in Q_b$ and $c \in Q_c$, there is an edge of the path $T[Q_b, Q_c]$ such that a is complete to its label.*

Proof. Suppose that a is in the middle of b, c. Let Q_b and Q_c be maximal cliques with $b \in Q_b$ and $c \in Q_c$, and suppose there is no edge of $T[Q_b, Q_c]$ such that a is complete to its label. For each edge on $T[Q_b, Q_c]$, one can select a vertex that is not adjacent to a. Then the set of selected vertices forms a path from b to c that uses no vertex from $N(a)$, a contradiction.

Suppose now that for all maximal cliques Q_b and Q_c with $b \in Q_b$ and $c \in Q_c$, there is an edge of the path $T[Q_b, Q_c]$ such that a is complete to its label. Suppose that there exists a path $x_0 - \cdots - x_r$, with $b = x_0$ and $c = x_r$ and none of the x_i's is in $N(a)$. We may assume that this path is chordless. For $1 \le i \le r$, let Q_i be a maximal clique containing x_{i-1}, x_i. Then $Q_1, \ldots, Q_r$ appear in this order along a subpath of T. On each $T[Q_i, Q_{i+1}]$ $(1 \le i \le r-1)$, vertex a is not adjacent to x_i, so a is not complete to any label of $T[Q_1, \ldots, Q_r]$, but Q_1 contains b and Q_r contains c, a contradiction. ■

Now we are ready to prove the main theorem. Part of the proof was done in the previous section. Lemma 1 deals with the case where there exists a simplicial vertex that is in the middle of two other vertices; now we have to look at the case where all simplicial vertices are not in the middle of any pair of vertices.

Proof of Theorem 1. By Theorem 3, a path graph does not contain any of $F_0, \ldots, F_{16}$. Suppose now that there exists a graph G that does not contain any of $F_0, \ldots, F_{16}$ and is a minimal non-path graph. Since G contains no F_0, it is chordal. By Theorem 2, there is a special simplicial vertex q of G. By Lemma 1, q is co-special. Let us denote by Q the unique maximal clique containing q. It will be convenient to denote by S_Q the separator S_q.

The graph G is not a clique as it is not a path graph. Consider graph $G(\mathcal{Q} \setminus Q)$, which is equal to $G \setminus (Q \setminus S_q)$. Since $Q \setminus S_q \neq \emptyset$, and by the minimality of G, it follows that $G(\mathcal{Q} \setminus Q)$ admits a clique path tree T_0. Let Q' be a vertex of T_0 such that $S_q \subset Q'$ (by the fact that S_Q is a separator Q' does exist). Let T_0' be obtained from T_0 by adding vertex Q and edge QQ'. Remark that T_0' is a clique tree of G but not a clique path tree since G is not a path graph.

Claim 1. *For all non-adjacent vertices $u, w \notin Q$, there exists a path between u and w that avoids the neighborhood of q.*

Proof. Suppose the contrary. Let $U, W \in \mathcal{Q}$ be such that $u \in U$ and $w \in W$. We have $U \neq W$ since u, w are not adjacent. By Lemma 3, there is an edge of $T_0[U, W]$ whose label is included in S_Q, contradicting that q is co-special. Thus the claim holds. ■

For each clique $L \in \mathcal{Q} \setminus \{Q, Q'\}$ we will use the following notation. Let L' be the neighbor of L along $T_0[L, Q']$ and S_L be the label $L \cap L'$ of the edge LL'. Let T_L be the largest subtree of T_0' that contains Q' and in which no label is included in S_L. Let S_L' be the label of the edge of $T_0[L, Q']$ that has exactly one extremity in T_L.

Since q is special and co-special we have $S_Q \not\subseteq S_L$, so T_L contains Q. Note that $S'_L \subseteq S_L$ by the definition of T_L.

Let $\mathcal{L}$ be the set of cliques L of $\mathcal{Q}\setminus\{Q, Q'\}$ such that LL' is the only edge incident to L whose label contains S'_L. In particular, for a vertex $x \in Q'$, any leaf of T_0^x which is not equal to Q' is in $\mathcal{L}$. Recall that T_0^x is a path because T_0 is a clique path tree. Let A be the set of vertices a of Q such that Q' is a vertex of T_0^a that is not a leaf. Then A is not empty, for otherwise T'_0 would be a clique path tree of G.

Claim 2. *For each clique $L \in \mathcal{L}$ we have $L' \in T_L$.*

Proof. Suppose on the contrary that $L' \notin T_L$. Let $\overline{L}$ be the clique in $T_0[L, Q']$ such that $\overline{L} \notin T_L$ and $\overline{L}' \in T_L$. Then $\overline{L} \neq L$ and the edge $\overline{L}\overline{L}'$ has label S'_L (possibly $\overline{L} = L'$). When we remove the edges LL' and $\overline{L}\overline{L}'$ from T'_0, there remain three subtrees T_1, T_2, T_3, where T_1 is the subtree that contains L, T_2 is the subtree that contains L' and $\overline{L}$, and T_3 is the subtree that contains $\overline{L}', Q', Q$. Let T_4 be the tree formed by T_1 and T_3 plus the edge $L\overline{L}'$. Then, since $S'_L \subseteq S_L$, T_4 is a clique tree of $G(\mathcal{Q}(T_4))$. Let x be any vertex in $L' \setminus L$. Vertex x does not belong to any vertex of T_1 as it is not in L. Since $S'_L \subseteq L$, vertex x does not belong to any vertex of T_3. So $G(\mathcal{Q}(T_4))$ is a strict subgraph of G and there exists a clique path tree T_5 of $G(\mathcal{Q}(T_4))$. Label S'_L is on the edge $L\overline{L}'$ of T_4, so it is also a label of T_5. Consequently, there is an edge LL'' of T_5 with a label R such that $S'_L \subseteq R \subseteq L$. (Possibly $L'' = \overline{L}'$). Suppose that $R \neq S'_L$. Then there is an edge of T_1 or T_3 with label R. But no label of T_1 can be R by the definition of $\mathcal{L}$; and all the labels of T_3 that are included in L are also included in S'_L, so no label of T_3 can be R, a contradiction. So $R = S'_L$. Now if we remove the edge LL'' from T_5 and replace it by the subtree T_2 and edges LL' and $\overline{L}L''$, we obtain a clique path tree of G, a contradiction. Thus the claim holds. ■

By the preceding claim, every $L \in \mathcal{L}$ satisfies $S'_L = S_L$.

Let $\mathcal{L}^*$ be the set of all $L \in \mathcal{L}$ such that T_L is a strict subtree of $T'_0 \setminus L$.

Claim 3. *For any $a \in A$, at least one leaf of T_0^a is in $\mathcal{L}^*$.*

Proof. Let L_1, L_2 be the leaves of T_0^a; as already noted, both are in $\mathcal{L}$. For $i = 1, 2$, let $\ell_i \in L_i \setminus S_{L_i}$. The three vertices q, ℓ_1, ℓ_2 are adjacent to a, so they do not form an asteroidal triple by Lemma 2, and so one of them is in the middle of the other two. Vertex q cannot be in the middle of ℓ_1, ℓ_2 by Claim 1. So we may assume up to symmetry that ℓ_1 is in the middle of q, ℓ_2. So, by Lemma 3, there is an edge of $T'_0[Q, L_2]$ with a label included in S_{L_1}. So T_{L_1} is a strict subtree of $T'_0 \setminus L_1$ and $L_1 \in \mathcal{L}^*$. Thus the claim holds. ■

The preceding claim implies that $\mathcal{L}^*$ is not empty. We choose $L \in \mathcal{L}^*$ such that the subtree T_L is maximal. Let $S_{Q'}$ be the label of the edge of $T_0[L, Q']$ incident to Q'. Vertex q is special and co-special, so there exists s_Q in $S_Q \setminus S_{Q'}$, and we have $s_Q \notin S_L$. Therefore, no clique of $\mathcal{Q} \setminus \mathcal{Q}(T_L)$ contains s_Q. We add the edge LL' to T_L to obtain a clique tree T'_L of $G(\mathcal{Q}(T_L) \cup \{L\})$. Since $L \in \mathcal{L}^*$, we have $T'_L \neq T'_0$, and by the minimality of G, there exists a clique path tree T of $G(\mathcal{Q}(T'_L))$. Note that L is a leaf of T, for otherwise there are at least two labels of T that are included in S_L, which contradicts the definition of T_L. From now on, our goal will be to show that either G contains one of the forbidden subgraphs, or T can be extended into a clique path tree of G.

Claim 4. *Let $a \in A$ be such that both leaves of T_0^a are not in T_L. Let L_a be a leaf of T_0^a that belongs to $\mathcal{L}^*$. Then L'_a is in T_L, and every edge KK' of T_0 with $K \notin T_L$ and $K' \in T_L$ satisfies $S_K \subseteq S_{L_a}$.*

Proof. By Claim 3, L_a exists. Since the labels of the edges of T_L are not included in S_L, they are also not included in S_{L_a}. So T_L is a subtree of T_{L_a}. By the maximality of T_L, we have $T_L = T_{L_a}$. By Claim 2, L'_a is in T_L. By the definition of T_{L_a}, every edge KK' of T_0 with $K \notin T_L$ and $K' \in T_L$ satisfies $S_K \subseteq S_{L_a}$. Thus the claim holds. ■

Claim 5. *There exist $U, W \in \mathcal{Q} \setminus \mathcal{Q}(T'_L)$ such that UL is an edge of T_0, $S_U \setminus Q' \neq \emptyset$, $U \cap W \neq \emptyset$, $W' \in \mathcal{Q}(T_L)$ and $W \cap Q \neq \emptyset$.*

Proof. We define sets $\mathcal{U}, \mathcal{V}$ as follows:

$$\mathcal{U} = \{U \in \mathcal{Q} \setminus \mathcal{Q}(T'_L) \mid UL \text{ is an edge of } T_0\}$$

$$\mathcal{V} = \{V \in \mathcal{Q} \setminus \mathcal{Q}(T'_L) \mid V' \in \mathcal{Q}(T_L)\}.$$

We observe that the members of $\mathcal{V}$ are pairwise disjoint. For if there is a vertex v in $V_1 \cap V_2$ for some $V_1, V_2 \in \mathcal{V}$, then v is on three labels (namely S_{V_1}, S_{V_2}, and S_L) of T_0 that do not lie on a common path, contradicting that T_0 is a clique path tree.

We define sets $\mathcal{U}_p$ $(p \geq 1)$ and $\mathcal{V}_p$ $(p \geq 0)$ as follows:

$$\mathcal{V}_0 = \{W \in \mathcal{V} \mid W \cap Q \neq \emptyset\}$$

$$\mathcal{U}_p = \{U \in \mathcal{U} \setminus (\mathcal{U}_1 \cup \cdots \cup \mathcal{U}_{p-1}) \mid \exists\ V \in \mathcal{V}_{p-1} \text{ such that } U \cap V \neq \emptyset\}\ (p \geq 1)$$

$$\mathcal{V}_p = \{V \in \mathcal{V} \setminus (\mathcal{V}_0 \cup \cdots \cup \mathcal{V}_{p-1}) \mid \exists\ U \in \mathcal{U}_p \text{ such that } V \cap U \neq \emptyset\}\ (p \geq 1).$$

Consider the smallest $k \geq 1$ such that there exists $U \in \mathcal{U}_k$ with $S_U \setminus Q' \neq \emptyset$. If no such U exists, then let $k = \infty$. Claim 5 to be proved states that $k = 1$, so let us suppose on the contrary that $k \geq 2$. For all $1 \leq p \leq k-1$ and all $U \in \mathcal{U}_p$, we have $S_U \subseteq Q'$; for each such U we denote by U'' the vertex of $\mathcal{Q}(T)$ such that $U'' \cap S_U \neq \emptyset$ and the length of $T[L, U'']$ is maximum. Remark that S_U is included in U'' if and only if all vertices of T that intersect S_U contain S_U. Let us prove that:

$$S_U \subseteq U'' \text{ for every } U \in \mathcal{U}_p, \quad 1 \leq p \leq k-1. \tag{1}$$

Suppose that there exists $U_p \in \mathcal{U}_p$, $1 \leq p \leq k-1$, such that $S_{U_p} \nsubseteq U''_p$, and let p be minimum with this property. Let $V_0, \ldots, V_{p-1}, U_1, \ldots, U_p$ be such that $V_i \in \mathcal{V}_i$, $U_i \in \mathcal{U}_i$, $V_{i-1} \cap U_i \neq \emptyset$, and $U_i \cap V_i \neq \emptyset$. We claim that $V'_0 = V'_1 = \cdots = V'_{p-1}$. For otherwise there exists $i \in \{1, \ldots, p-1\}$ such that $V'_{i-1} \neq V'_i$. Then V'_i contains elements of S_{U_i} but not all, and so $S_{U_i} \nsubseteq U''_i$, which contradicts the minimality of p. Pick $u_i \in U_i \setminus S_{U_i}$ and $v_i \in V_i \setminus S_{V_i}$. Let $x_1, \ldots, x_r$ be such that $x_1 \in V_0 \cap U_1$, $x_2 \in U_1 \cap V_1, \ldots, x_r \in V_{p-1} \cap U_p$ with $r = 2p-1$.

By the definition of the $\mathcal{V}_i$'s, none of $x_2, \dots, x_r$ is in Q. Let $x_0 \in V_0 \cap Q$ (maybe $x_0 = x_1$). So $x_0 \in S_{V_0} \subseteq S_L \subset L$. None of $U_2, \dots, U_p$ can contain x_0 by the definition of $\mathcal{U}_1$. Note that x_r is in U_p and $V'_{p-1} = V'_0$; on the other hand we have $S_{U_p} \nsubseteq U''_p$. So there exists a clique Z of T_L such that $Z' \in T_0^{x_0}$, $S_{U_p} \subseteq Z'$, $S_{U_p} \cap Z \neq \emptyset$ and $S_{U_p} \setminus Z \neq \emptyset$. Vertex Q' is on $T_0[L, Z']$ as $S_{U_p} \subseteq Q'$. Let $z \in Z \setminus Z'$. We can find vertices $y_1, \dots, y_t$ on the labels of $T'_0[Z, Q]$ such that none of them is in S_L and $z - y_1 - \cdots - y_t - q$ is a chordless path in G. Let $\ell \in L \setminus S_L$. By Claim 1, there exists a path P between z and ℓ whose vertices are not neighbors of q.

If $Z \in T_0^{x_0}$, then let $b \in S_{U_p} \setminus Z$. As q is special and co-special, we have $S_Q \nsubseteq S_Z$, so let $c \in S_Q \setminus S_Z$. Then z, ℓ, q form an asteroidal triple (because of the three paths P, $z - y_1 - \cdots - y_t - q$, and $\ell - b - c - q$), and they lie in the neighborhood of x_0, a contradiction to Lemma 2. So $Z \notin T_0^{x_0}$. Let $x_{r+1} \in Z \cap U_p$. If $x_{r+1} \in Q$, then z, ℓ, q form an asteroidal triple (because of paths P, $z - y_1 - \cdots - y_t - q$, and $\ell - x_0 - q$), and they lie in the neighborhood of x_{r+1}, a contradiction again. So $x_{r+1} \notin Q$. The S_{U_i}'s are all included in Q' and so in S_L too. They are pairwise disjoint, for otherwise T_0 is not a clique path tree. Vertex ℓ is not in any of the S_{U_i}'s, and ℓ is adjacent to all of $x_0, \dots, x_{r+1}$ and to none of $u_1, \dots, u_p, v_0, \dots, v_{p-1}, y_1, \dots, y_t, z, q$.

Suppose that $V_0 \cap U_1 \cap Q \neq \emptyset$. Then we may assume that $x_0 = x_1$, so x_0 is in A and the two leaves of $T_0^{x_0}$ are not in T_L. By Claims 3 and 4, there exists a leaf L_{x_0} of $T_0^{x_0}$ that belongs to $\mathcal{L}^*$ and L'_{x_0} is in T_L, so $L_{x_0} = V_0$. But x_{r+1} is in $Z \cap U_p$, so it is not in S_{V_0}; thus $S_L \nsubseteq S_{V_0}$, which contradicts the end of Claim 4. Therefore $V_0 \cap U_1 \cap Q = \emptyset$, so $x_0 \neq x_1$, $x_0 \notin U_1$, $x_1 \notin Q$. Now, if $t = 1$, then $\{u_1, \dots, u_p, v_0, \dots, v_{p-1}, x_0, \dots, x_{r+1}, y_1, q, z, \ell\}$ induces $F_{14}(4p+5)_{p \geq 1}$. If $t = 2$, then $\{u_1, \dots, u_p, v_0, \dots, v_{p-1}, x_0, \dots, x_{r+1}, y_1, y_2, q, z, \ell\}$ induces $F_{15}(4p+6)_{p \geq 1}$. If $t \geq 3$, then $\{\ell, x_0, x_{r+1}, z, y_1, \dots, y_t, q\}$ induces $F_{10}(s+5)_{t \geq 3}$, a contradiction. Therefore (1) holds.

Suppose that k is finite. Let $V_0, \dots, V_{k-1}, U_1, \dots, U_k$ be such that $V_i \in \mathcal{V}_i$, $U_i \in \mathcal{U}_i$, $V_{i-1} \cap U_i \neq \emptyset$, and $U_i \cap V_i \neq \emptyset$. Let $u_i \in U_i \setminus S_{U_i}$ and $v_i \in V_i \setminus S_{V_i}$. Pick vertices $x_1 \in V_0 \cap U_1$, $x_2 \in U_1 \cap V_1, \dots, x_r \in V_{k-1} \cap U_k$ with $r = 2k-1$. By the definition of the $\mathcal{V}_i$'s, none of $x_2, \dots, x_r$ is in Q. Let $x_0 \in V_0 \cap Q$. Suppose that $V_0 \cap U_1 \cap Q \neq \emptyset$. Then we can assume that $x_0 = x_1$, so x_0 is in A and the two leaves of $T_0^{x_0}$ are not in T_L. By Claims 3 and 4, a leaf L_{x_0} of $T_0^{x_0}$ is in $\mathcal{L}^*$ and L'_{x_0} is in T_L, so $L_{x_0} = V_0$. But x_2 is in S_{V_1} and not in S_{V_0}, so $S_{V_1} \nsubseteq S_{V_0}$, which contradicts the end of Claim 4. Therefore $V_0 \cap U_1 \cap Q = \emptyset$, and $x_0 \neq x_1$, $x_0 \notin U_1$, $x_1 \notin Q$. Let $s_{U_k} \in S_{U_k} \setminus Q'$. Vertex s_{U_k} is not adjacent to any of $q, s_Q, v_0, \dots, v_{k-1}$ because $s_{U_k} \notin Q'$, and by the minimality of k, vertex s_{U_k} is not adjacent to $u_1, \dots, u_{k-1}$. Then $\{u_1, \dots, u_k, v_0, \dots, v_{k-1}, x_0, \dots, x_r, s_{U_k}, s_Q, q\}$ induces $F_{16}(4k+3)_{k \geq 2}$, a contradiction.

Now k is infinite. Then the members of $\bigcup_{p \geq 1} \mathcal{U}_p$ are included in Q' and pairwise disjoint, for otherwise T_0 is not a clique path tree. For each member M of $\mathcal{U} \cup \mathcal{V}$, let $T'_0(M)$ be the component of $T'_0 \setminus T'_L$ that contains M. Starting from the clique path tree T and the trees $T'_0(M)$ ($M \in \mathcal{U} \cup \mathcal{V}$), we build a new tree as follows. For each $V \in \bigcup_{p \geq 0} \mathcal{V}_p$, we add the edge VL between $T'_0(V)$ and T. For each $U \in \bigcup_{p \geq 1} \mathcal{U}_p$, we add the edge UU'' between $T'_0(U)$ and T. For each $U \in \mathcal{U} \setminus (\bigcup_{p \geq 1} \mathcal{U}_p)$, we add the edge UL between $T'_0(U)$ and T. For each $V \in \mathcal{V} \setminus (\bigcup_{p \geq 0} \mathcal{V}_p)$, we define $V'' \in \mathcal{Q}(T)$ such that $V'' \cap S_V \neq \emptyset$ and the length of $T[L, V'']$ is maximum. By the definition of $\mathcal{V}_0$, we have $S_V \cap Q = \emptyset$, so $V'' \neq Q$, so V'' is a vertex of T_L on $T_0[L, V]$ and it contains S_V as $S_V \subseteq S_L$. Then we can add the edge VV'' between $T'_0(V)$ and T. Thus we obtain a clique path tree of G, a contradiction. So $k=1$, and there exist $U \in \mathcal{U}_1$ and $W \in \mathcal{V}_0$ such that $S_U \setminus Q' \neq \emptyset$, $U \cap W \neq \emptyset$, and $W \cap Q \neq \emptyset$. Thus the claim holds. ■

Let U, W be as in the preceding claim. Let $s_U \in S_U \setminus Q'$. Vertex s_U is not adjacent to s_Q. Let $u \in U \setminus S_U$ and $w \in W \setminus S_W$. We have $W' \in \mathcal{Q}(T_L)$, so $S_W \subseteq S_L$. Moreover $W \cap Q \neq \emptyset$, so $W \cap Q' \cap L \neq \emptyset$, so Q' is on $T_0[W, L]$ as T_0 is a clique path tree.

Claim 6. $S_W = S_L$.

Proof. Assume on the contrary that $S_W \neq S_L$. Then S_W is a proper subset of S_L. Suppose that there exists $a \in U \cap W \cap Q \neq \emptyset$. Then a is in A and the two leaves of T_0^a are not in T_L. By Claims 3 and 4, a leaf L_a of T_0^a is in $\mathcal{L}^*$ and L'_a is in T_L, so $L_a = W$. But $S_L \not\subseteq S_W$, so Claim 4 is contradicted. Therefore $U \cap W \cap Q = \emptyset$. By the definition of U and W, there exists $b \in W \cap Q$ and $c \in U \cap W$. So $b \notin U$, $c \notin Q$, $b \neq c$. Since s_U is in $S_U \setminus Q'$, we have $S_U \not\subseteq S_W$. The labels of the edges of T_L are not included in S_L, so they are also not in S_W. Thus, we can choose vertices $x_1, \ldots, x_r$ on the labels of $T'_0[U, Q]$ such that none of the x_i's is in S_W, $x_1 \in U$, $x_r \in Q$, and $u - x_1 - \cdots - x_r - q$ is a path from u to q that avoids $N(w)$. Suppose $r=1$. Then x_1 is different from s_U and s_Q, and $\{w, b, c, u, s_U, x_1, s_Q, q\}$ induces F_8. Suppose $r=2$. If x_1 is adjacent to s_Q, then $\{w, b, c, u, s_U, x_1, s_Q, q\}$ induces F_9, and if x_1 is not adjacent to s_Q, then $\{w, b, c, u, x_1, x_2, s_Q, q\}$ induces F_9. Finally, suppose $r \geq 3$. Then $\{w, b, c, u, x_1, \ldots, x_r, q\}$ induces $F_{10}(r+5)_{r\geq 3}$. In all cases we obtain a contradiction. Thus the claim holds. ■

Claim 7. $W \in \mathcal{L}^*$.

Proof. In the connected component of $T'_0 \setminus W'$ that contains W, let $X \in \mathcal{Q}$ be such that $S_W \subseteq X$ and the length of $T'_0[X, W]$ is maximum (possibly $X = W$). Then $S_W \subseteq S_X$ and XX' is the only edge of T'_0 incident to X that contains S_W, so $X \in \mathcal{L}$. Since $S_W \subseteq S_X$ we have that $W \notin T_X$. Then, by Claim 2 we have $X = W$ and by Claim 6 we have $T_W = T_L$; so $W \in \mathcal{L}^*$. Thus the claim holds. ■

By Claim 7, we have $W \in \mathcal{L}^*$. By Claim 6, we have $T_W = T_L$, so T_W is also maximal and what we have proved for L can be done for W. Thus, by Claim 5, there exists $X \notin T_W$ such that XW is an edge of T_0 with $S_X \setminus Q' \neq \emptyset$ and $X \cap S_W \neq \emptyset$. Let $x \in X \setminus W$ and $s_X \in S_X \setminus Q'$. Vertex s_X is not in S_W, for otherwise, it would also be in S_L and in Q'. Vertex s_U is not in S_L, for otherwise, it would also be in S_W and in Q'. Vertex s_Q is not in S_W $(= S_L)$. So s_Q, s_X, s_U are pairwise non-adjacent.

Suppose that there exists a vertex $a \in U \cap X \cap Q \neq \emptyset$. So $a \in A$, but none of the two leaves of T_0^a can satisfy Claim 4, a contradiction. Therefore $U \cap X \cap Q = \emptyset$.

Suppose that $U \cap X \neq \emptyset$, and let $a \in U \cap X$. So a is not in Q. Let $b \in S_W \cap Q$ $(= S_L \cap Q)$. So b is not in $U \cap X$. If $b \notin X \cup U$, then $\{q, u, x, s_Q, s_U, s_X, a, b\}$ induces F_6, a contradiction. So b is in one of U, X, say $b \in X \setminus U$ (if b is in $U \setminus X$ the argument is similar). Since W is in $\mathcal{L}$, there is a vertex $c \in S_W \setminus S_X$. Vertex c is adjacent to a, b, s_U, s_Q and not to x. Then $\{x, a, b, u, s_U, c, s_Q, q\}$ induces F_8, F_9, or $F_{10}(8)$, a contradiction. Therefore $U \cap X = \emptyset$.

Let $a \in U \cap W$, so $a \notin X$. Suppose $a \notin Q$. If there exists $b \in X \cap Q$, then b is also in L and $\{q, u, x, s_Q, s_U, s_X, a, b\}$ induces F_6, a contradiction. So $X \cap Q = \emptyset$. Let $c \in W \cap Q$. Then $c \in L$ and $c \notin X$. Let $d \in X \cap S_W$; so $d \in L$, $d \notin Q$, $d \notin U$. If c is adjacent to u, then $\{q, u, x, s_Q, s_U, s_X, c, d\}$ induces F_6, else $\{q, u, x, s_Q, s_U, s_X, a, c, d\}$ induces F_7, a contradiction. So $a \in Q$. Let $e \in X \cap S_W$; so $e \in L$. If $e \notin Q$, then $\{q, u, x, s_Q, s_U, s_X, a, e\}$ induces F_6, a contradiction. So $e \in Q$. Let $f \in S_W \setminus S_Q$ (f exists because q is special and co-special). Since $U \cap X = \emptyset$, f is adjacent to at most one of u, x, and then $\{q, u, x, s_U, s_X, a, e, f\}$ induces F_9 or $F_{10}(8)$, a contradiction. This completes the proof of Theorem 1. ■

6. RECOGNITION ALGORITHM

Our proof above yields a new recognition algorithm for path graphs, which takes any graph G as input and either builds a clique path tree for G or finds one of $F_0, \ldots, F_{16}$ as an induced subgraph of G. We have not analyzed the exact complexity of such a method but it is easy to see that it is polynomial in the size of the input graph. More efficient algorithms were already given by Gavril [8], Schäffer [19], and Chaplick [4], with complexity, respectively, $O(n^4)$, $O(nm)$, and $O(nm)$ for graphs with n vertices and m edges. Another algorithm was proposed in [5] and claimed to run in $O(n+m)$ time, but it has only appeared as an extended abstract (see comments in [4, Section 2.1.4]).

There are classical linear time recognition algorithms for triangulated graphs [18], and, following [2], there have been several linear time recognition algorithms for interval graphs, of which the most recent is [10]. We hope that the work presented here will be helpful in the search for a linear time recognition algorithm for path graphs.

ACKNOWLEDGMENTS

The authors thank the anonymous referees for their useful remarks. We are especially grateful to a referee whose suggestions lead to a significant simplification of the proof of Theorem 2.

REFERENCES

[1] C. Berge, Les problèmes de coloration en théorie des graphes, Publ Inst Stat Univ Paris 9 (1960), 123–160.

[2] K. S. Booth and G. S. Lueker, Testing for the consecutive ones property, interval graphs and graph planarity using PQ-tree algorithm, J Comput System Sci 13 (1976), 335–379.

[3] K. Cameron, C. T. Hoàng, and B. Lévêque, Asteroidals in rooted and directed path graphs, Manuscript, http://hal.archives-ouvertes.fr/hal-00347163, 2008.

[4] S. Chaplick, PQR-trees and undirected path graphs, M.Sc. thesis, Department of Computer Science, University of Toronto, 2008.

[5] E. Dahlhaus and G. Bailey, Recognition of path graphs in linear time, 5th Italian Conference on Theoretical Computer Science (Revello, 1995) World Scientific Publishing, River Edge, NJ, 1996, pp. 201–210.
[6] G. A. Dirac, On rigid circuit graphs, Abh Math Sem Univ Hamburg 38 (1961), 71–76.
[7] F. Gavril, The intersection graphs of subtrees in trees are exactly the chordal graphs, J Combin Theory B 16 (1974), 47–56.
[8] F. Gavril, A recognition algorithm for the intersection graphs of paths in trees, Discrete Math 23 (1978), 211–227.
[9] M. C. Golumbic, Algorithmic graph theory and perfect graphs (2nd edition). Annals of Discrete Mathematics 57. Elsevier B.V., Amsterdam, 2004.
[10] M. Habib, R. McConnell, C. Paul, and L. Viennot, Lex-BFS and partition refinement, with applications to transitive orientation, interval graph recognition and consecutive ones testing, Theoret Comput Sci 234 (2000), 59–84.
[11] A. Hajnal and J. Surányi, Über die Auflösung von Graphen in vollständige Teilgraphen, Ann Univ Sci Budapest Eötvös Sect Math 1 (1958), 113–121.
[12] R. Halin, Some remarks on interval graphs, Combinatorica 2 (1982), 297–304.
[13] C. Lekkerkerker and D. Boland, Representation of finite graphs by a set of intervals on the real line, Fund Math 51 (1962), 45–64.
[14] T. A. McKee and F. R. McMorris, Topics in intersection graph theory, SIAM Monographs on Discrete Mathematics and Applications, SIAM, Philadelphia, 1999.
[15] C. L. Monma and V. K. Wei, Intersection graphs of paths in a tree, J Combin Theory B 41 (1986), 141–181.
[16] B. S. Panda, The forbidden subgraph characterization of directed vertex graphs, Discrete Math 196 (1999), 239–256.
[17] P. L. Renz, Intersection representations of graphs by arcs, Pacific J Math 34 (1970), 501–510.
[18] D. J. Rose, R. E. Tarjan, and G. S. Lueker, Algorithmic aspects of vertex elimination of graphs, SIAM J Comput 5 (1976), 266–283.
[19] A. A. Schäffer, A faster algorithm to recognize undirected path graphs, Discrete Appl Math 43 (1993), 261–295.
[20] R. E. Tarjan and M. Yannakakis, Simple linear time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs, SIAM J Comput 13 (1984), 566–579.
[21] S. Tondato, M. Gutierrez, and J. Szwarcfiter, A forbidden subgraph characterization of path graphs Electron Notes Discrete Math 19 (2005), 281–287.