



**HAL**  
open science

## Characterizing path graphs by forbidden induced subgraphs

Benjamin Lévêque, Frédéric Maffray, Myriam Preissmann

► **To cite this version:**

Benjamin Lévêque, Frédéric Maffray, Myriam Preissmann. Characterizing path graphs by forbidden induced subgraphs. *Journal of Graph Theory*, 2009, 62 (4), pp.369-384. 10.1002/jgt.20407. hal-00261413v3

**HAL Id: hal-00261413**

**<https://hal.science/hal-00261413v3>**

Submitted on 17 Apr 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Characterizing path graphs by forbidden induced subgraphs

Benjamin Lévêque\*    Frédéric Maffray†    Myriam Preissmann†

April 17, 2008

## 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.

## 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 [6] 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 [11] 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 [7, 13, 15] but no characterization by forbidden subgraphs was known, whereas such results exist for intersection graphs of subpaths of a path (interval graphs [11]), subtrees of a tree (chordal graphs [6]), and also for directed subpaths of a directed tree [14].

---

\*ROSE, EPFL, Lausanne, Switzerland

†C.N.R.S., Laboratoire G-SCOP, Grenoble, France

In 1970, Renz [15] asked for a complete list of graphs that are chordal and not path graphs and are minimal with this property, and he gave two examples of such graphs. Reference [19] extends the list of minimal forbidden subgraphs for path graphs; 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 of  $F_0, \dots, F_{16}$  as an induced subgraph.*

## 2 Special simplicial vertices in chordal graphs

In a graph  $G$ , a *clique* is 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  $v$ . 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, we have  $Q_v \in \mathcal{Q}$ . 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 [10, 1] (see also [8]) 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  $T^v$ . A classical result [6] states that a graph is chordal if and only if it has a clique tree.

For 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 contain a vertex complete to  $S$  [6, 12]. Note that this number is at least one and that it depends only on  $S$  and not on  $T$ , so for a given  $S \in \mathcal{S}$  it is the same in every clique tree.

Given  $X \subseteq \mathcal{Q}$ , 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 whose 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  $G(\mathcal{Q}(T'))$ .

Dirac [5] 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.* We prove the theorem by induction on  $|\mathcal{Q}|$ . By the hypothesis,  $G$  is not a clique, so  $|\mathcal{Q}| \geq 2$  and  $\mathcal{S} \neq \emptyset$ .

*Case 1:  $\mathcal{S}$  has only one maximal element  $S$ .* Let  $Q, Q'$  be two maximal cliques such that  $Q \cap Q' = S$ . Let  $v \in Q \setminus Q'$  and  $v' \in Q' \setminus Q$ . The set  $S$  is the only maximal separator and it does not contain  $v$  or  $v'$ . So  $v$  and  $v'$  do not belong to any element of  $\mathcal{S}$ , and so they are simplicial and  $S_v = S_{v'} = S$ , so they are special.

*Case 2:  $\mathcal{S}$  has two distinct maximal elements  $S, S'$ .* So  $|\mathcal{Q}| \geq 3$ . Let  $T$  be a clique tree of  $G$ . Let  $Q_1, Q_2, Q'_1, Q'_2$  be members of  $\mathcal{Q}$  such that  $S = S_{Q_1 Q_2}$ ,  $S' = S_{Q'_1 Q'_2}$ , and  $Q_2, Q_1, Q'_1, Q'_2$  appear in this order along the path  $T[Q_2, Q_1, Q'_1, Q'_2]$  (possibly  $Q_1 = Q'_1$ ). Let  $Y$  be the subtree of  $T \setminus Q_1$  that contains  $Q_2$ , and let  $Z$  be the tree that consists of  $Y$  plus the vertex  $Q_1$  and the edge  $Q_1 Q_2$ . The subtree  $Z$  does not contain  $Q'_2$ , so  $G(\mathcal{Q}(Z))$  has strictly fewer maximal cliques than  $G$ ; and  $G$  is not a clique. By the induction hypothesis, there exist two non adjacent simplicial vertices  $v, w$  of  $G(\mathcal{Q}(Z))$  such that  $S_v, S_w$  are maximal elements of  $\mathcal{S}(Z)$ . At most one of  $v, w$  is in  $Q_1$  since they are not adjacent, say  $v$  is not in  $Q_1$ . We claim that  $v$  is a simplicial vertex of  $G$  and that  $S_v$  is a maximal element of  $\mathcal{S}$ . Vertex  $v$  does not belong to any element of  $\mathcal{S}(Z)$ . If it belongs to an element of  $\mathcal{S} \setminus \mathcal{S}(Z)$ , then it must also belong to  $Q_1 \cap Q_2 = S \in \mathcal{S}(Z)$ , a contradiction. So  $v$  does not belong to any element of  $\mathcal{S}$  and so it is a simplicial vertex of  $G$ . The set  $S_v$  is a maximal element of  $\mathcal{S}(Z)$ . If it is not a maximal element of  $\mathcal{S}$ , then it is included in  $S \in \mathcal{S}(Z)$ , a contradiction. So  $v$  is a special simplicial vertex of  $G$ . Likewise, let  $Y'$  be the subtree of  $T \setminus Q'_1$  that contains  $Q'_2$ , and let  $Z'$  be the tree that consists of  $Y'$  plus the vertex  $Q'_1$  and the edge  $Q'_1 Q'_2$ . Just like with  $v$ , we can find a simplicial vertex  $v'$  of  $G(\mathcal{Q}(Z'))$  not in  $Q'_1$  that is a simplicial vertex of  $G$  with  $S_{v'}$  being a maximal element of  $\mathcal{S}$ . Vertices  $v$  and  $v'$  are not adjacent since  $S$  is a  $\{v, v'\}$ -separator. So  $v$  and  $v'$  are the desired vertices.  $\square$

Algorithms LexBFS [16] and MCS [18] are linear time algorithms that were developed to find a simplicial vertex in a chordal graph. But a simplicial vertex found by these algorithms is not necessarily special. 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 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 do not know how to find such a vertex in linear time.

### 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 cliques that contain  $v$  is a path. Gavril [7] proved a graph is a path graph if and only if it has a clique path tree.

Consider graphs  $F_0, \dots, F_{16}$  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  $F_1, \dots, F_5$  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 [15, Figures 1 and 5]. For  $i \in \{0, 1, 3, 4, 5, 6, 7, 9, 10, 13, 15, 16\}$ , Panda [14] 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**  $F_0, \dots, F_{16}$  are minimal non path graphs.

*Proof.* Clearly,  $F_0$  is a minimal non path graph. For the other graphs, we prove the theorem in one case and then show how the same arguments can be applied to all cases.

Consider  $F = F_{11}(4k)$ ,  $k \geq 2$ ; see Figure 4. Name its vertices such that  $u_1, \dots, u_{2k-1}$  are the simplicial vertices of degree 2, clockwise;  $v_{j-1}, v_j$  are the two neighbors of  $u_j$  ( $j = 1, \dots, 2k-1$ ), with subscripts modulo  $2k-1$ ; and  $a, b$  are the remaining vertices. Let  $Q_j$  be the maximal clique that contains  $u_j$  ( $j = 1, \dots, 2k-1$ ), and call these  $2k-1$  cliques “peripheral”. Let  $R_a = \{a, v_1, \dots, v_{2k-1}\}$  and  $R_b = \{b, v_1, \dots, v_{2k-1}\}$  be the maximal cliques that contain respectively  $a$  and  $b$ , and call these two cliques “central”. Thus  $\mathcal{Q}(F) = \{R_a, R_b, Q_1, \dots, Q_{2k-1}\}$ . Since  $F$  is chordal, it admits a clique tree. Let  $T$  be any clique tree of  $F$ . Then  $R_a$  and  $R_b$  are adjacent in  $T$  (for otherwise, there would be at least one interior vertex  $Q$  on the path  $T[R_a, R_b]$ , so we should have  $R_a \cap R_b \subseteq Q$ , but no member  $Q$  of  $\mathcal{Q}(F) \setminus \{R_a, R_b\}$  satisfies this inclusion). By the same argument, each  $Q_j$  ( $j = 1, \dots, 2k-1$ ) must be adjacent to  $R_a$  or  $R_b$  in  $T$ . Suppose that we are trying to build a clique path tree  $T$  for  $F$ . By symmetry, we may assume that  $Q_1$  is adjacent to  $R_a$ . Then, for  $j = 2, \dots, 2k-2$  successively,  $Q_j$  must be adjacent to  $R_b$  (if  $j$  is even) and to  $R_a$  (if  $j$  is odd) in  $T$ , for otherwise, for some  $v \in \{v_{j-1}, v_j\}$  the subtree  $T^v$  induced by the cliques that contain  $v$  would not be a path. Note that in this fashion we obtain a clique path tree  $T'$  of  $F \setminus u_{2k-1}$ . Now if  $Q_{2k-1}$  is adjacent to  $R_a$ , then the subtree  $T'^{v_{2k-1}}$  is not a path, and if  $Q_{2k-1}$  is adjacent to  $R_b$ , then the same holds for  $T'^{v_{2k-2}}$ . This shows that  $F$  is not a path graph.

Now consider any vertex  $x$  of  $F$ . If  $x$  is one of the  $u_j$ 's, then by symmetry we may assume that  $x = u_{2k-1}$ , and we have seen above that  $F \setminus x$  is a path graph with clique

path tree  $T'$ . Suppose that  $x$  is one of the  $v_j$ 's, say  $x = v_{2k-1}$ . Then by adding vertex  $Q_{2k-1}$  and edge  $Q_{2k-1}R_a$  to  $T'$ , it is easy to see that we obtain a clique path tree of  $F \setminus x$ . Finally, suppose that  $x$  is one of  $a, b$ , say  $x = b$ . Then the tree with vertices  $R_a, Q_1, \dots, Q_{2k-1}$  and edges  $R_aQ_1, \dots, R_aQ_{2k-1}$  is a clique path tree of  $F \setminus x$ . So  $F$  is a minimal non path graph.

When  $F$  is any other  $F_i$  ( $i = 1, \dots, 16$ ), the same arguments apply as follows. For  $i = 1, \dots, 10$ , call peripheral the three cliques that contain a simplicial vertex. For  $i = 11, \dots, 16$ , call peripheral the cliques that contain a simplicial vertex of degree 2, plus, in the case of  $F_{12}$ , the clique that contain the bottom simplicial vertex (which has degree 3). Call central all other maximal cliques. Then it is easy to prove, as above, that the central cliques must form a subpath in any clique tree of  $F$ , and all the peripheral cliques except one can be appended to either end of that subpath, but whichever way this is done, when the last clique is appended, the subtree  $T^v$  is not a path for some vertex  $v$  of  $F$ . Moreover, when any vertex  $x$  is removed, it is possible to build a clique path tree for  $F \setminus x$ .  $\square$

## 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. 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  $G$ .

**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  $F_0, \dots, F_{16}$ .

Let  $T_0$  be a clique path tree of  $G \setminus q$ . Let  $Q' \in \mathcal{Q}(G \setminus q)$  be such that  $S_q \subseteq Q'$ . If  $Q' = S_q$ , then we can add  $q$  to  $Q'$  to obtain a clique path tree of  $G$ , a contradiction. So  $Q' \neq S_q$ , and  $S_q \in \mathcal{S}$  (as there is a vertex  $q' \in Q' \setminus S_q$  and  $S_q$  is a  $\{q, q'\}$ -separator).

Let  $T'$  be the maximal subtree of  $T_0$  that contains  $Q'$  and such that no label of the edges of  $T_0$  is included in  $S_q$ . Remark that  $T'$  plus vertex  $Q$  and edge  $QQ'$  is a clique tree of  $G(\mathcal{Q}(T') \cup \{Q\})$  (but not necessarily a clique path tree), and in that tree only one label is included in  $S_q$ . Since  $q$  is not co-special, there is an edge of  $T_0$  whose label is included in  $S_q$ , and so  $T'$  has strictly fewer vertices than  $T_0$ . So  $G(\mathcal{Q}(T') \cup \{Q\})$  is a path graph. Let  $T$  be a clique path tree of this graph.

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  $\mathcal{H} = (V_{\mathcal{H}}, E_{\mathcal{H}})$  be the intersection graph of  $S_1, \dots, S_\ell$ , that is,  $V_{\mathcal{H}} = \{S_1, \dots, S_\ell\}$  and  $E_{\mathcal{H}} = \{S_i S_j \mid S_i \cap S_j \neq \emptyset\}$ .

**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  $T_0$ . 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  $q' \in Q' \setminus Q$ . Let us consider the subgraph induced by  $q, q', v_1, \dots, v_p, s_1, \dots, s_p$ . Each of the non-adjacent vertices  $q$  and  $q'$  is 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  $q'$  can have at most two neighbors among the  $v_j$ 's. If  $q'$  has zero or one neighbor among them, then  $q, q', v_1, \dots, v_p, s_1, \dots, s_p$  induce respectively  $F_{11}(4r+4)_{r \geq 1}$  or  $F_{12}(4r+4)_{r \geq 1}$ . If  $q'$  has two consecutive neighbors  $v_j, v_{j+1}$ , then  $q, q', v_j, v_{j+1}, s_{j-1}, s_j, s_{j+1}$  induce  $F_2$ . If  $q'$  has two non-consecutive neighbors  $v_j, v_k$ , then we can assume that  $1 \leq j < j+1 < k \leq p$  and  $k - j$  is odd,  $k - j = 2s + 1$  with  $s \geq 1$ , and then  $q, q', v_j, \dots, v_k, s_j, \dots, s_{k-1}$  induce  $F_{14}(4s+5)_{s \geq 1}$ . In all cases we obtain a contradiction. Thus the claim holds.  $\diamond$

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

For  $1 \leq i \leq \ell$ , let  $\mathcal{R}_i = \{S \in \mathcal{S}(T') \mid S_i \cap S \neq \emptyset \text{ and } S_i \setminus S \neq \emptyset\}$ . Let  $X = \{S_i \mid \mathcal{R}_i \neq \emptyset\}$ .

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

*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  $T'[Q'_1, Q'_2]$ . If  $p \neq 2$ , 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  $Q'_1 = Q'_3 = \dots = Q'_{p-1}$  and  $Q'_2 = Q'_4 = \dots = Q'_p$ . 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  $L_1 K_1$  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  $L_p K_p$  be such an edge and  $R_p$  be its label. So  $S_1 \subseteq L_1$ ,  $S_1 \not\subseteq K_1$  and  $S_p \subseteq L_p$ ,  $S_p \not\subseteq K_p$ . Each of  $K_1, K_p$  may be in  $P$  or not. Since  $T'$  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 path tree,  $v_0$  and  $v_{p+1}$  are distinct from  $v_1, \dots, v_p$  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  $s_1, \dots, s_{p-1}$  and from  $v_0, \dots, v_{p+1}$ . Furthermore, if  $K_1$  is not in  $P$ , then  $v_0$  is not adjacent to any of  $s_1, \dots, s_p$ ; and if  $K_p$  is not in  $P$ , then  $v_{p+1}$  is not adjacent to any of  $s_0, \dots, s_{p-1}$ .

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  $s_1, \dots, s_{p-1}$  and from  $v_0, \dots, v_{p+1}$ . Furthermore, if  $K_1$  is in  $P$ , then  $v_0$  is adjacent to  $s'_p$  and to  $s_1, \dots, s_p$ ; and if  $K_p$  is in  $P$ , then  $v_{p+1}$  is adjacent to  $s'_0$  and to  $s_0, \dots, s_{p-1}$ .

Note that the set  $\{q, s'_0, s_0, s_1, s_2, \dots, s_p, s'_p\}$  induces a clique in  $G$ . Moreover,  $v_1$  is adjacent to  $s'_0$ ,  $v_p$  is adjacent to  $s'_p$ , for  $i = 1, \dots, p$ ,  $v_i$  is adjacent to  $s_{i-1}$  and  $s_i$ , and there is no other edge between  $v_1, \dots, v_p$  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  $q, y, v_0, \dots, v_p, s_0, \dots, s_p$  induce  $F_{12}(4k+4)_{k \geq 1}$ , 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  $v_0-x_1-\dots-x_r-v_{p+1}$  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  $v_1, \dots, v_p$ .

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  $q, v_0, \dots, v_{p+1}, s_0, \dots, s_p, x_1$  induce  $F_{14}(4k+5)_{k \geq 1}$ . If  $r = 2$ , then  $q, v_0, \dots, v_{p+1}, s_0, \dots, s_p, x_1, x_2$  induce  $F_{15}(4k+6)_{k \geq 1}$ . If  $r \geq 3$ , then  $q, v_0, v_{p+1}, s_0, s_p, x_1, \dots, x_r$  induce  $F_{10}(r+5)_{r \geq 3}$ , 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  $q, v_0, \dots, v_{p+1}, s'_0, s_1, \dots, s_p, x_1$  induce  $F_{13}(4k+5)_{k \geq 1}$ . If  $r \geq 2$ , then  $q, v_0, v_{p+1}, x_1, \dots, x_r, s'_0, s_p$  induce  $F_5(r+5)_{r \geq 2}$ , 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  $q, v_0, v_{p+1}, s'_0, s_1, s'_p, x_1$  induce  $F_2$ . If  $r = 2$ , then  $q, v_0, v_{p+1}, s'_0, s_1, s'_p, x_1, x_2$  induce  $F_3$ . If  $r \geq 3$ , then  $q, v_0, v_{p+1}, x_1, \dots, x_r, s'_0, s'_p$  induce  $F_{10}(r+5)_{r \geq 3}$ , a contradiction. Thus the claim holds.  $\diamond$



By the preceding two claims,  $\mathcal{H}$  is a bipartite graph  $(A, B, E_{\mathcal{H}})$  such that  $X \subseteq A$ . Now all the subtrees  $T_i$  can be linked to  $T$  to get a clique path tree of  $G$  as follows. For each  $S_i \in A$ , we add an edge  $QQ_i$  between  $T$  and  $T_i$ . This creates a clique path tree on the corresponding subset of cliques because  $A$  is a stable set of  $\mathcal{H}$  and  $Q$  is a leaf of  $T$ . For each  $S_i \in B$ , 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  $S_i \in B$ , 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  $B$  is a stable set of  $\mathcal{H}$  and by the definition of  $Q_i''$ , a contradiction.  $\square$

## 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  $F_0, \dots, F_{16}$ . We could not find a characterization similar to the one found by Lekkerkerker and Boland [11] 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 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.

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

*Proof.* It suffices to check that when a universal vertex is added to a minimal forbidden induced subgraph for interval graphs ([11]), then one obtains a graph that contains one of  $F_0, \dots, F_5, F_{10}$ . The easy details are left to the reader.  $\square$

Given three non adjacent vertices  $a, b, c$ , we say that  $a$  is 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 the middle of the others.

**Lemma 3** *In a chordal graph  $G$  with clique tree  $T$ , a vertex  $a$  is the middle of two vertices  $b, c$  if and only if for all cliques  $Q_b$  and  $Q_c$  such that  $b \in Q_b$  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 the middle of  $b, c$ . Let  $Q_b$  and  $Q_c$  be cliques such that  $b \in Q_b$  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 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  $1 \leq i \leq r$ , let  $Q_i$  be a maximal clique containing  $x_{i-1}, x_i$ . Then  $Q_1, \dots, Q_r$  appear in this order along a subpath of  $T$ . On each  $T[Q_i, Q_{i+1}]$  ( $1 \leq i \leq r-1$ ), vertex  $a$  is not adjacent to  $x_i$ , so  $a$  is not complete to any label of  $T[Q_1, \dots, Q_r]$ , but  $Q_1$  contains  $b$  and  $Q_r$  contains  $c$ , a contradiction.  $\square$

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

**Proof of Theorem 1** By Theorem 3, a path graph does not contain any of  $F_0, \dots, F_{16}$ . Suppose now that there exists a graph  $G$  that does not contain any of  $F_0, \dots, F_{16}$  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  $Q = Q_q$  and  $S_Q = S_q \in \mathcal{S}$ . Let  $T_0$  be a clique path tree of  $G(\mathcal{Q} \setminus Q)$ . Let  $Q' \in \mathcal{Q} \setminus Q$  be such that  $S_Q \subseteq Q'$ . We add the edge  $QQ'$  to  $T_0$  to obtain a clique tree  $T'_0$  of  $G$ .

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

For each clique  $L \in \mathcal{Q} \setminus \{Q, Q'\}$ , let  $L'$  be the neighbor of  $L$  along  $T_0[L, Q']$ . Let  $S_L = L \cap L'$ . Let  $\mathcal{S}_L$  be the set of labels of edges incident to  $L$  in  $T_0$ . Let  $\bar{L}$  be the clique in  $T_0[L, Q'] \setminus \{Q'\}$  such that  $S_{\bar{L}} \subseteq S_L$  and no other edge of  $T_0[\bar{L}, Q']$  has a label included in  $S_L$ . (Possibly  $\bar{L} = L$ .)

Let  $\mathcal{L}$  be the set of cliques  $L$  of  $\mathcal{Q} \setminus \{Q, Q'\}$  such that no element of  $\mathcal{S}_L \setminus S_L$  contains  $S_{\bar{L}}$ . For each clique  $L \in \mathcal{L}$ , we define a subtree  $T_L$  of  $T'_0$ , where  $T_L$  is the biggest subtree of  $T'_0$  that contains  $Q'$  and for which no label is included in  $S_L$ . Note that  $\bar{L}'$  is in  $T_L$  and  $\bar{L}$  is not in  $T_L$ . Since  $q$  is special and co-special we have  $S_Q \not\subseteq S_L$ , so  $T_L$  contains  $Q$ .

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

*Proof.* Suppose on the contrary that  $L' \notin T_L$ . Then  $\bar{L} \neq L$ . When we remove the edges  $LL'$  and  $\bar{L}\bar{L}'$  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  $\bar{L}$ , and  $T_3$  is the subtree that contains  $\bar{L}', Q', Q$ . Let  $T_4$  be the tree formed by  $T_1, T_3$  plus the edge  $L\bar{L}'$ . Then, since  $S_{\bar{L}} \subseteq S_L$ ,  $T_4$  is a clique tree of  $G(\mathcal{Q}(T_4))$ . The set  $\mathcal{Q}(T_4)$  contains strictly fewer maximal cliques than  $\mathcal{Q}$ , so there exists a clique path tree  $T_5$  of  $G(\mathcal{Q}(T_4))$ . Label  $S_{\bar{L}}$  is on the edge  $L\bar{L}'$  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  $S_{\bar{L}} \subseteq R \subseteq L$ . (Possibly  $L'' = \bar{L}'$ ). Suppose that  $R \neq S_{\bar{L}}$ . 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  $S_{\bar{L}}$ , so no label of  $T_3$  can be  $R$ , a contradiction. So  $R = S_{\bar{L}}$ . Now if we remove the edge  $LL''$  from  $T_5$  and replace it by the subtree  $T_2$  and edges  $LL'$  and  $\bar{L}L''$ , we obtain a clique path tree of  $G$ , a contradiction. Thus the claim holds.  $\diamond$

Let  $\mathcal{L}^*$  be the set of all  $L \in \mathcal{L}$  such that  $T_L$  is a strict subtree of  $T'_0 \setminus L$ . For every vertex  $x$  of  $G(\mathcal{Q} \setminus Q)$  let  $T_0^x$  be the subtree of  $T_0$  induced by the cliques that contain  $x$ . 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  $G$ . Moreover:

**Claim 3** *For any  $a \in A$ , the two leaves of  $T_0^a$  are in  $\mathcal{L}$  and at least one of them is in  $\mathcal{L}^*$ .*

*Proof.* Let  $L_1, L_2$  be the leaves of  $T_0^a$ , and, for  $i = 1, 2$ , let  $\ell_i \in L_i \setminus S_{L_i}$ . We have  $a \in S_{L_1}$ , and  $a$  is not in any member of  $\mathcal{S}(L_1) \setminus S_{L_1}$ . Thus  $L_1 \in \mathcal{L}$ . Similarly  $L_2 \in \mathcal{L}$ . The three vertices  $q, \ell_1, \ell_2$  are adjacent to  $a$ , so they do not form an asteroidal triple by Lemma 2, and so one of them is the middle of the other two. Vertex  $q$  cannot be the middle of  $\ell_1, \ell_2$  by Claim 1. So we may assume up to symmetry that  $\ell_1$  is the middle of  $q, \ell_2$ . 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 so  $L_1 \in \mathcal{L}^*$ . Thus the claim holds.  $\diamond$

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  $T_0[L, Q']$  that is 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  $T'_L$  is a strict subtree of  $T'_0$ , we can consider 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  $T_L$ .

**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  $K \notin T_L, K' \in T_L$  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  $K \notin T_L, K' \in T_L$  satisfies  $S_K \subseteq S_{L_a}$ . Thus the claim holds.  $\diamond$

**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:

$$\begin{aligned}\mathcal{U} &= \{U \in \mathcal{Q} \setminus \mathcal{Q}(T'_L) \mid UL \text{ is an edge of } T_0\} \\ \mathcal{V} &= \{V \in \mathcal{Q} \setminus \mathcal{Q}(T'_L) \mid V' \in \mathcal{Q}(T_L)\}.\end{aligned}$$

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:

$$\begin{aligned}\mathcal{V}_0 &= \{W \in \mathcal{V} \mid W \cap Q \neq \emptyset\} \\ \mathcal{U}_p &= \{U \in \mathcal{U} \setminus (\mathcal{U}_1 \cup \dots \cup \mathcal{U}_{p-1}) \mid \exists V \in \mathcal{V}_{p-1} \text{ such that } U \cap V \neq \emptyset\} \quad (p \geq 1) \\ \mathcal{V}_p &= \{V \in \mathcal{V} \setminus (\mathcal{V}_1 \cup \dots \cup \mathcal{V}_{p-1}) \mid \exists U \in \mathcal{U}_p \text{ such that } V \cap U \neq \emptyset\} \quad (p \geq 1).\end{aligned}$$

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  $k = \infty$ . The claim 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  $S_U \subseteq Q'$ ; let  $U'' \in \mathcal{Q}(T)$  be such that  $U'' \cap S_U \neq \emptyset$  and the length of  $T[L, U'']$  is maximal. Remark that  $S_U$  is included in  $U''$  if and only if all members of  $\mathcal{Q}(T)$  that intersect  $S_U$  contain  $S_U$ . Let us prove that:

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

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

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  $S_{U_p} \not\subseteq U''$ . 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  $T[L, Z']$  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  $z-y_1-\dots-y_t-q$  is a chordless path in  $G$ . Let  $\ell \in L \setminus S_L$ . By Claim 1, there exists a path  $P$  between  $z$  and  $\ell$  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  $S_Q \not\subseteq S_Z$ , so let  $c \in S_Q \setminus S_Z$ . Then  $z, \ell, q$  form an asteroidal triple (because of paths  $P$ ,  $z-y_1-\dots-y_t-q$  and  $\ell-b-c-q$ ), and they lie in the neighborhood of  $x_0$ , a contradiction. So  $Z \notin T_0^{x_0}$ . Let  $x_{r+1} \in Z \cap U_p$ . If  $x_{r+1} \in Q$ , then  $z, \ell, q$  form an asteroidal triple (because of paths  $P$ ,  $z-y_1-\dots-y_t-q$  and  $\ell-x_0-q$ ), and they lie in the neighborhood of  $x_{r+1}$ , a contradiction. 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  $\ell$  is not in any of the  $S_{U_i}$ 's, and  $\ell$  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 Claim 4, the leaf  $L_{x_0}$  of  $T_0^{x_0}$  that is in  $\mathcal{L}^*$  is such that  $L'_{x_0}$  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  $S_L \not\subseteq S_{V_0}$ , 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  $u_1, \dots, u_p, v_0, \dots, v_{p-1}, x_0, \dots, x_{r+1}, y_1, q, z, \ell$  induce  $F_{14}(4p+5)_{p \geq 1}$ . If  $t = 2$ , then  $u_1, \dots, u_p, v_0, \dots, v_{p-1}, x_0, \dots, x_{r+1}, y_1, y_2, q, z, \ell$  induce  $F_{15}(4p+6)_{p \geq 1}$ . If  $t \geq 3$ , then  $\ell, x_0, x_{r+1}, z, y_1, \dots, y_t, q$  induce  $F_{10}(s+5)_{t \geq 3}$ , 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  $\mathcal{V}$ , 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 Claim 4, the leaf  $L_{x_0}$  of  $T_0^{x_0}$  that is in  $\mathcal{L}^*$  is such that  $L'_{x_0}$  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  $S_{V_1} \not\subseteq S_{V_0}$ , 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  $u_1, \dots, u_k, v_0, \dots, v_{k-1}, x_0, \dots, x_r, s_{U_k}, s_Q, q$  induce  $F_{16}(4k+3)_{k \geq 2}$ , a contradiction.

Now  $k$  is infinite. Then the members of  $\bigcup_{p \geq 0} \mathcal{U}_p$  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 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  $V \in \mathcal{V} \setminus (\bigcup_{p \geq 1} \mathcal{V}_p)$ , we define  $V'' \in \mathcal{Q}(T)$  such that  $V'' \cap S_V \neq \emptyset$  and the length of  $T[L, V'']$  is maximal. 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.  $\diamond$

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  $w \in W \setminus S_W$ .

**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 Claim 4, the leaf  $L_a$  of  $T_0^a$  that is in  $\mathcal{L}^*$  is such that  $L'_a$  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  $x_1, \dots, x_r$  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  $u-x_1-\dots-x_r-q$  is a path from  $u$  to  $q$  that avoids  $N(w)$ . If  $r = 1$ , then  $x_1$  is different from  $s_U$  and  $s_Q$ , and  $w, b, c, u, s_U, x_1, s_Q, q$  induce  $F_8$ . If  $r = 2$ , then, if  $x_1$  is adjacent to  $s_Q$ , vertices  $w, b, c, u, s_U, x_1, s_Q, q$  induce  $F_9$ , and if  $x_1$  is not adjacent to  $s_Q$ , vertices  $w, b, c, u, x_1, x_2, s_Q, q$  induce  $F_9$ . Finally, if  $r \geq 3$ , then  $w, b, c, u, x_1, \dots, x_r, q$  induce  $F_{10}(r+5)_{r \geq 3}$ . In all cases we obtain a contradiction. Thus the claim holds.  $\diamond$

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

*Proof.* If  $W \in \mathcal{L}$ , then, by Claim 6, we have  $T_W = T_L$  and  $W \in \mathcal{L}^*$ , as desired. So suppose  $W \notin \mathcal{L}$ . By the definition of  $W$ , there is a vertex  $a \in W \cap Q$ , and so  $a \in L$ . Let  $L_1, L_2 \in \mathcal{L}$  be the leaves of  $T_0^a$  such that  $L_1, L, Q', W, L_2$  lie in this order on that path. Let  $K$  be the member of  $\mathcal{L}$  that is closest to  $W$  on  $T_0[L_2, W]$ . Clearly  $W \neq K$ . The edges of  $T_L$  are not included in  $S_L$ , so they are also not in  $S_W$  and not in  $S_K$ . So  $T_K$  contains  $T_L$ . If  $K \in \mathcal{L}^*$ , then  $T_K = T_L$  by the maximality of  $T_L$ , so  $K' \notin T_K$ , which contradicts Claim 2. Thus  $K \notin \mathcal{L}^*$ . This means that  $T_K = T'_0 \setminus K$ , and so the labels of  $T'_0 \setminus K$  are not included in  $S_K$ , in particular  $S_W \not\subseteq S_K$ . Let  $XX'$  be the edge of  $T_0[K, W]$  such that  $X'$  contains  $S_W$  and  $X$  does not (maybe  $X' = W$ ,  $X = K$ ). The set  $S_X$  contains  $a$  but not all of  $S_{X'}$ , and the members of  $S_{X'} \setminus \{S_{X'}, S_X\}$  do not contain  $a$ . So no element of  $S_{X'} \setminus S_X$  contains  $S_{X'}$ , which means that  $X' \in \mathcal{L}$ , a contradiction to the definition of  $K$ . Thus the claim holds.  $\diamond$

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  $q, u, x, s_Q, s_U, s_X, a, b$  induce  $F_6$ , 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  $x, a, b, u, s_U, c, s_Q, q$  induce  $F_8, F_9$  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  $q, u, x, s_Q, s_U, s_X, a, b$  induce  $F_6$ , 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  $q, u, x, s_Q, s_U, s_X, c, d$  induce  $F_6$ , else  $q, u, x, s_Q, s_U, s_X, a, c, d$  induce  $F_7$ , a contradiction. So  $a \in Q$ . Let  $e \in X \cap S_W$ ; so  $e \in L$ . If  $e \notin Q$ , then  $q, u, x, s_Q, s_U, s_X, a, e$  induce  $F_6$ , 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  $q, u, x, s_U, s_X, a, e, f$  induce  $F_9$  or  $F_{10}(8)$ , a contradiction. This completes the proof of Theorem 1.  $\square$

## 6 Recognition algorithm

The proof that we give 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  $F_0, \dots, F_{16}$ . 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 [7], Schäffer [17] and Chaplick [3], whose complexity is respectively  $O(n^4)$ ,  $O(nm)$  and  $O(nm)$  for graphs with  $n$  vertices and  $m$  edges. Another algorithm was proposed in [4] and claimed to run in  $O(n + m)$  time, but it has only appeared as an extended abstract (see comments in [3, Section 2.1.4]).

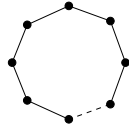
There are classical linear time recognition algorithms for triangulated graphs [16], and, following [2], there have been several linear time recognition algorithms for interval graphs, of which the most recent is [9]. We hope that the work presented here will be helpful in the search for a linear time recognition algorithm for path graphs.

## 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, G.S. Lueker. Testing for the consecutive ones property, interval graphs and graph planarity using PQ-tree algorithm. *J. Comput. Syst. Sci.* 13 (1976) 335–379.
- [3] S. Chaplick. *PQR-trees and undirected path graphs*. M.Sc. Thesis, Dept. of Computer Science, University of Toronto, 2008.
- [4] E. Dahlhaus, G. Bailey. Recognition of path graphs in linear time. 5th Italian Conference on Theoretical Computer Science (Revello, 1995) World Sci. Publishing, River Edge, NJ, 1996, 201–210.
- [5] G.A. Dirac. On rigid circuit graphs. *Abh. Math. Sem. Univ. Hamburg* 38 (1961) 71–76.

- [6] F. Gavril. The intersection graphs of subtrees in trees are exactly the chordal graphs. *J. Combin. Theory B* 16 (1974) 47–56.
- [7] F. Gavril. A recognition algorithm for the intersection graphs of paths in trees. *Discrete Math.* 23 (1978) 211–227.
- [8] M. C. Golumbic. *Algorithmic graph theory and perfect graphs*. Annals Disc. Math. 57, Elsevier, 2004.
- [9] M. Habib, R. McConnell, C. Paul, L. Viennot. Lex-BFS and partition refinement, with applications to transitive orientation, interval graph recognition and consecutive ones testing. *Theoretical Computer Science* 234 (2000) 59–84.
- [10] A. Hajnal and J. Surányi. Über die Auflösung von Graphen in vollständige Teilgraphen. *Ann. Univ. Sci. Budapest Eötvös, Sect. Math.* 1 (1958) 113–121.
- [11] C. Lekkerkerker, D. Boland. Representation of finite graphs by a set of intervals on the real line. *Fund. Math.* 51 (1962) 45–64.
- [12] T.A. McKee and F.R. McMorris. *Topics in intersection graph theory*. SIAM Monographs on Discrete Mathematics and Applications, Philadelphia, 1999.
- [13] C.L. Monma, V.K. Wei. Intersection graphs of paths in a tree. *J. Combin. Theory B* 41 (1986) 141–181.
- [14] B. S. Panda. The forbidden subgraph characterization of directed vertex graphs. *Discrete Mathematics* 196 (1999) 239–256.
- [15] P.L. Renz. Intersection representations of graphs by arcs. *Pacific J. Math.* 34 (1970) 501–510.
- [16] D.J. Rose, R.E. Tarjan, G.S. Lueker. Algorithmic aspects of vertex elimination of graphs. *SIAM J. Comput.* 5 (1976) 266–283.
- [17] A.A. Schäffer. A faster algorithm to recognize undirected path graphs. *Discrete Appl. Math.* 43 (1993) 261–295.
- [18] R.E. Tarjan, 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.
- [19] S. Tondato, M. Gutierrez, J. Szwarcfiter. A forbidden subgraph characterization of path graphs. *Electronic Notes in Discrete Mathematics* 19 (2005) 281–287.





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

Figure 1: Forbidden subgraphs with no simplicial vertices

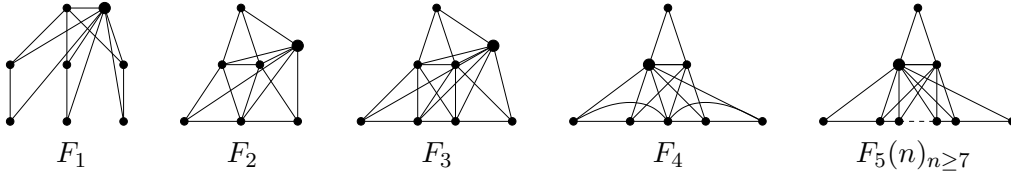


Figure 2: Forbidden subgraphs with a universal vertex

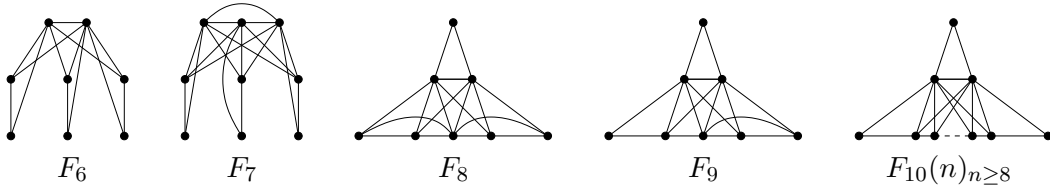


Figure 3: Forbidden subgraphs with no universal vertex and exactly three simplicial vertices

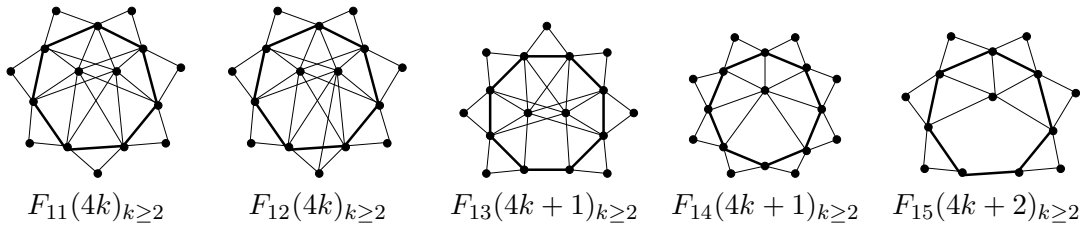
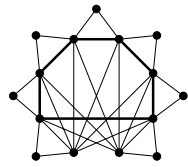


Figure 4: Forbidden subgraphs with at least one simplicial vertex that is not co-special. (bold edges form a clique)



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

Figure 5: Forbidden subgraphs with  $\geq 4$  simplicial vertices that are all co-special. (bold edges form a clique)