

# Structure and Linear-Time Recognition of 4-Leaf Powers

ANDREAS BRANDSTÄDT AND VAN BANG LE

*Universität Rostock*

AND

R. SRITHARAN

*The University of Dayton*

**Abstract.** A graph  $G$  is the  $k$ -leaf power of a tree  $T$  if its vertices are leaves of  $T$  such that two vertices are adjacent in  $G$  if and only if their distance in  $T$  is at most  $k$ . Then  $T$  is a  $k$ -leaf root of  $G$ . This notion was introduced and studied by Nishimura, Ragde, and Thilikos [2002], motivated by the search for underlying phylogenetic trees. Their results imply an  $\mathcal{O}(n^3)$ -time recognition algorithm for 4-leaf powers. Recently, Rautenbach [2006] as well as Dom et al. [2005] characterized 4-leaf powers without true twins in terms of forbidden subgraphs. We give new characterizations for 4-leaf powers and squares of trees by a complete structural analysis. As a consequence, we obtain a conceptually simple linear-time recognition of 4-leaf powers.

Categories and Subject Descriptors: F.2.2 [Analysis of Algorithms and Problem Complexity]: Nonnumerical Algorithms and Problems; G.2.2 [Discrete Mathematics]: Graph Theory

General Terms: Algorithms, Theory

Additional Key Words and Phrases: Graph powers, leaf powers, phylogenetic trees, squares of trees, trees

## ACM Reference Format:

Brandstädt, A., Le, V. B., and Sritharan, R. 2008. Structure and linear-time recognition of 4-leaf powers. *ACM Trans. Algor.* 5, 1, Article 11 (November 2008), 22 pages. DOI = 10.1145/1435375.1435386 <http://doi.acm.org/10.1145/1435375.1435386>

## 1. Introduction

One of the fundamental problems in computational biology is to reconstruct the evolutionary history of a set of species, based on quantitative biological data. Typically, the evolutionary history is modeled by an evolutionary tree called the *phylogeny*, which is a tree whose leaves are labeled by species and in which each

---

Authors' addresses: A. Brandstädt, V. B. Le, Institut für Informatik, Universität Rostock, D-18051 Rostock, Germany, e-mail: {ab, le}@informatik.uni-rostock.de; R. Sritharan, Computer Science Department, The University of Dayton, Dayton, OH 45469, e-mail: rst@cps.udayton.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org).  
© 2008 ACM 1549-6325/2008/11-ART11 \$5.00 DOI 10.1145/1435375.1435386 <http://doi.acm.org/10.1145/1435375.1435386>

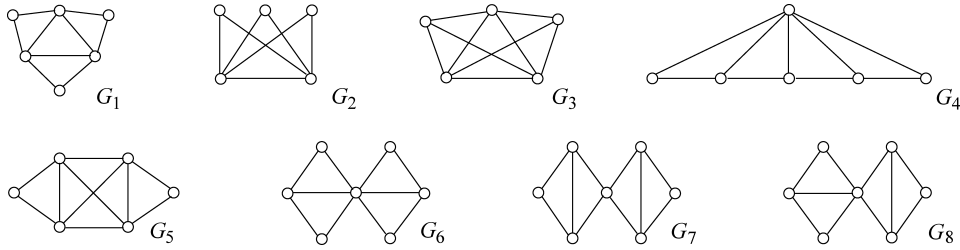


FIG. 1. Forbidden induced subgraphs for chordal basic 4-leaf powers.

internal node represents a speciation event whereby an ancestral species gives rise to two or more child species [Chen et al. 2003].

Motivated by this background, Nishimura et al. [2002] introduce the notions of  $k$ -leaf power and  $k$ -leaf root, defined as follows: Let  $G = (V, E)$  be a finite undirected graph and  $k \geq 2$  an integer.  $G$  is a  $k$ -leaf power if there is a tree  $T$  with  $V$  as its set of leaves such that for all  $x, y \in V, xy \in E$  if and only if their distance in  $T$  is at most  $k$ :  $d_T(x, y) \leq k$ . Such a tree  $T$  is then called a  $k$ -leaf root of  $G$ .

Obviously, a graph is a 2-leaf power if and only if it is the disjoint union of cliques, that is, it contains no induced  $P_3$ .

Nishimura et al. [2002] gave (very complicated)  $\mathcal{O}(n^3)$ -time recognition algorithms for recognizing 3-leaf powers and 4-leaf powers, respectively, and constructing 3-leaf roots (4-leaf roots, respectively), if existent. For  $k \geq 5$ , no characterization and no efficient recognition of  $k$ -leaf powers is known.

In Brandstädt and Le [2006] and Rautenbach [2006], it was recently shown that 3-leaf powers are exactly those graphs which are obtained from a tree by substituting cliques into its vertices. More about 3-leaf powers and a linear-time recognition for them can be found in Brandstädt and Le [2006].

The structure of 4-leaf powers is considerably more complicated than that of 3-leaf powers. Recently, Rautenbach [2006] showed that a graph without true twins (i.e., two adjacent vertices having the same neighborhood) is a 4-leaf power if and only if it is chordal and does not contain any of the graphs  $G_1, \dots, G_8$ , shown in Figure 1, as an induced subgraph. In Dom et al. [2005], essentially the same forbidden subgraph characterization is given in terms of critical cliques of a graph. The concept of critical cliques was introduced in Lin et al. [2000]; it takes  $\mathcal{O}(n^3)$  time to construct the critical clique graph of a given graph (see also the corresponding discussion in Dom et al. [2005]).

We extend the 4-leaf powers without true twins to the following slightly more general notion which turns out to be useful: Graph  $G = (V, E)$  is a *basic 4-leaf power* if  $G$  has a 4-leaf root  $T$  where, to any of the internal nodes of  $T$ , at most one leaf (i.e., vertex of  $V$ ) is attached. Such a 4-leaf root  $T$  is then called a *basic 4-leaf root* for  $G$ . Since the set of leaves attached to the same tree node always forms a clique module, obviously every 4-leaf power  $G$  results from a basic 4-leaf power  $G'$  by substituting cliques into the vertices of  $G'$ . Observe also that 4-leaf powers without true twins are basic 4-leaf powers, but, as we will see, basic 4-leaf powers may have true twins.

An interesting and important feature of basic 4-leaf powers is their relation to the classical concept of the square of a tree, where the square of a tree  $T$  is obtained from  $T$  by adding additional edges between every two vertices at distance two in

$T$ . An internal node in a tree is called *invisible* [Nishimura et al. 2002] if no leaf is attached to it. We will show that a graph is a 2-connected basic 4-leaf power if and only if it has a basic 4-leaf root without invisible vertices if and only if it is the square of a tree.

Then one of our main results, namely Theorem 5.7, states that a graph is a basic 4-leaf power if and only if its 2-connected components are squares of trees, and if two such 2-connected components share a cut vertex, then at least one of them must be a clique. This also generalizes the known characterization of 4-leaf powers without true twins [Dom et al. 2005; Rautenbach 2006] mentioned before.

Finally, based on our characterization of basic 4-leaf powers, we give a conceptually simple linear-time recognition for 4-leaf powers which drastically improves the previously known, very complex,  $\mathcal{O}(n^3)$ -time algorithm in Nishimura et al. [2002].

As a byproduct, we obtain a new structural characterization of squares of trees which extends results by Lin and Skiena [1995] as well as by Kearney and Corneil [1998]. In particular, we show that  $G$  is the square of a tree if and only if it is chordal and 2-connected and does not contain any of the graphs  $G_1, \dots, G_5$  in Figure 1. This also explains the forbidden subgraph characterization by Dom et al. [2005] and Rautenbach [2006] in a new way: The graphs  $G_1, \dots, G_5$  are responsible for the blocks (i.e., 2-connected components), and the graphs  $G_6, G_7, G_8$  represent the gluing conditions for the blocks of basic 4-leaf powers.

Our structural characterization of squares of trees implies that the root tree for the square of a tree can be directly constructed in linear time in a very simple way.

We also believe that our ideas shed new light on the problem of recognizing  $k$ -leaf powers for  $k \geq 5$ .

This article is organized as follows: In Section 2, we present the definitions and notation used in the article. In Section 3 we collect some facts on  $k$ -leaf powers for  $k \geq 2$ . In Section 4 we obtain two new characterizations of squares of trees which are useful tools for the structural analysis of 4-leaf powers. In Sections 5 and 6 we give a full structural analysis of basic 4-leaf powers and, respectively, of general 4-leaf powers. Finally, in Section 7 we describe a linear-time recognition algorithm for 4-leaf powers.

## 2. Definitions and Notation

Throughout this article, let  $G = (V, E)$  be a finite undirected graph without self-loops and multiple edges with vertex set  $V$  and edge set  $E$ , and let  $|V| = n$ ,  $|E| = m$ . For a positive integer  $k$ , a  $k$ -connected component in a graph  $G$  is a maximal (induced)  $k$ -connected subgraph of  $G$ ; the 1-connected components of  $G$  are the usual connected components, and the 2-connected components of  $G$  are also called *blocks* of  $G$ . A  $k$ -cut in a connected graph is a cutset with  $k$  vertices; a 1-cut is also called a *cut vertex*. For a vertex  $v \in V$ , let  $N(v) = \{u \mid uv \in E\}$  denote the (*open*) neighborhood of  $v$  in  $G$ , and let  $N[v] = \{v\} \cup \{u \mid uv \in E\}$  denote the (*closed*) neighborhood of  $v$  in  $G$ . The *degree*  $\deg_G(v)$  of a vertex  $v$  is the number of its neighbors,  $\deg_G(v) = |N(v)|$ . If  $N[v] = V$ ,  $v$  is a *universal vertex* of  $G$ . A *clique* is a set of vertices which are mutually adjacent. A *stable set* is a set of vertices which are mutually nonadjacent.

For  $U \subseteq V$ , let  $G[U]$  denote the subgraph of  $G$  induced by  $U$ . We write  $G - U$  for  $G[V \setminus U]$ . In case  $U = \{u\}$  we also write  $G - u$  instead of  $G - \{u\}$ . Throughout this article, all subgraphs are understood to be induced subgraphs. Let  $\mathcal{F}$  denote a

set of graphs. A graph  $G$  is  $\mathcal{F}$ -free if none of its induced subgraphs is in  $\mathcal{F}$ . For two graphs  $G_i = (V_i, E_i)$ ,  $i = 1, 2$ , the graphs  $G_1 \cap G_2$  and  $G_1 \cup G_2$  have vertex sets  $V_1 \cap V_2$ , respectively,  $V_1 \cup V_2$  and edge sets  $E_1 \cap E_2$ , respectively,  $E_1 \cup E_2$ .

Two vertices  $x, y \in V$  are *true twins* if  $N[x] = N[y]$ , namely if  $x$  and  $y$  have the same neighbors and are adjacent to each other. A vertex subset  $U \subseteq V$  is a *module* in  $G$  if for all  $v \in V \setminus U$ , either  $v$  is adjacent to all vertices of  $U$  or  $v$  is adjacent to none of them; a *nontrivial module* consists of at least two, but not all, vertices. A module of a graph is *maximal* if there is no other nontrivial module of the graph containing it. *Substituting* (or *replacing*) a vertex  $v$  in a graph  $G$  by a graph  $H$  results in the graph obtained from  $(G - v) \cup H$  by adding all edges between vertices in  $N_G(v)$  and vertices in  $H$ . A *clique module* in  $G$  is a module which induces a clique in  $G$ . Obviously, true twins are a clique module. Two disjoint vertex sets  $X, Y \subseteq V$  form a *join* (*cojoin*), denoted by  $X \textcircled{+} Y$  ( $X \textcircled{-} Y$ ), if for all pairs  $x \in X$ ,  $y \in Y$ ,  $xy \in E$  ( $xy \notin E$ ) holds. We also use  $\textcircled{+}$  and  $\textcircled{-}$  as graph operations: For two vertex disjoint graphs  $G_1$  and  $G_2$ ,  $G_1 \textcircled{+} G_2$  is just the union of  $G_1$  and  $G_2$ , and  $G_1 \textcircled{-} G_2$  is obtained from  $G_1$  and  $G_2$  by adding all edges between vertices in  $G_1$  and vertices in  $G_2$ .

Let  $d_G(x, y)$  (or  $d(x, y)$  for short if  $G$  is understood) be the length, namely the number of edges, of a shortest path in  $G$  between  $x$  and  $y$ . Let  $G^k = (V, E^k)$  with  $xy \in E^k$  if and only if  $d_G(x, y) \leq k$  denote the *kth power* of  $G$ .

For  $k \geq 1$ , let  $P_k$  denote a chordless path with  $k$  vertices and  $k - 1$  edges, and for  $k \geq 3$ , let  $C_k$  denote a chordless cycle with  $k$  vertices and  $k$  edges. A complete bipartite graph with  $a$  vertices in one color class and  $b$  vertices in the other color class is denoted by  $K_{a,b}$ ; the  $K_{1,3}$  is also called a *claw*. Let  $S_k$  denote the (*complete*) *sun* with  $2k$  vertices  $u_1, \dots, u_k$  and  $w_1, \dots, w_k$  such that  $u_1, \dots, u_k$  is a clique,  $w_1, \dots, w_k$  is a stable set, and for  $i \in \{1, \dots, k\}$ ,  $w_i$  is adjacent to  $u_i$  and  $u_{i+1}$  (index arithmetic modulo  $k$ ).

A graph is *chordal* if it contains no induced  $C_k$ ,  $k \geq 4$ . A graph is *strongly chordal* if it is chordal and sun-free. See Brandstädt et al. [1999] for various characterizations of (strongly) chordal graphs.

### 3. Basic Facts on $k$ -Leaf Powers

In this section we collect some useful facts on  $k$ -leaf powers.

PROPOSITION 3.1.

- (i) Every induced subgraph of a  $k$ -leaf power is a  $k$ -leaf power.
- (ii) A graph is a  $k$ -leaf power if and only if each of its connected components is a  $k$ -leaf power.

PROOF. (i): Let  $T$  be a  $k$ -leaf root of a graph  $G$ , and let  $H$  be an induced subgraph of  $G$ . Then the subtree  $T'$  of  $T$ , induced by all paths in  $T$  connecting the leaves of  $T$  in  $V(H)$ , is a  $k$ -leaf root of  $H$ .

(ii): The only-if part follows from (i). For the if-part, assume that each connected component  $G_i$  of  $G$  has a  $k$ -leaf root  $T_i$ . Take a new vertex  $v$  and connect the trees  $T_i$  and  $v$  by a path of length  $k$ ; the resulting tree is clearly a  $k$ -leaf root of  $G$ .  $\square$

In Dahlhaus and Duchet [1987], Lubiw [1982], and Raychaudhuri [1992], it is shown that the class of strongly chordal graphs is closed under powers, given as follows.

PROPOSITION 3.2. *If  $G$  is strongly chordal then for every  $k \geq 1$ ,  $G^k$  is strongly chordal.*

Let  $T$  be a  $k$ -leaf root of a graph  $G$ . Then, by definition,  $G$  is isomorphic to the subgraph of  $T^k$  induced by the leaves of  $T$ . Since trees are strongly chordal and induced subgraphs of strongly chordal graphs are strongly chordal, this implies by Proposition 3.2 (see also Kearney and Corneil [1998]) the next proposition.

PROPOSITION 3.3. *For every  $k \geq 1$ ,  $k$ -leaf powers are strongly chordal.*

This strengthens the fact that  $k$ -leaf powers are chordal, which is observed in previous papers dealing with  $k$ -leaf powers such as Dom et al. [2005], Nishimura et al. [2002], and Rautenbach [2006].

*Definition 3.4.* A  $k$ -leaf power is *basic* if it admits a  $k$ -leaf root in which to every internal node at most one leaf is attached; we refer to such a root as a *basic  $k$ -leaf root*.

OBSERVATION 3.5.

- (i) *Every induced subgraph of a basic  $k$ -leaf power is a basic  $k$ -leaf power.*
- (ii) *A  $k$ -leaf power without true twins is a basic  $k$ -leaf power.*

PROOF. (i): A basic  $k$ -leaf root for an induced subgraph can be obtained from a basic  $k$ -leaf root for the graph in the same way as in the proof of Proposition 3.1 (i).

(ii): Observe that two leaves with the same parent in a  $k$ -leaf root of  $G$  form true twins in  $G$ .  $\square$

Note that the other direction of Observation 3.5 (ii) does not hold; refer to comments after the proof of Proposition 3.7.

PROPOSITION 3.6. *For every graph  $G$ , and for every  $k \geq 2$ ,  $G$  is a  $k$ -leaf power if and only if  $G$  is obtained from a basic  $k$ -leaf power by substituting the vertices by nonempty cliques.*

PROOF. Let  $T$  be a  $k$ -leaf root of  $G = (V, E)$ . Let  $V = V_1 \cup \dots \cup V_n$  be the partition of  $V$  into the subsets  $V_i$  of leaves in  $T$  having the same parent node  $t_i$  in  $T$ . Obviously,  $V_i$  are nonempty clique modules in  $G$ . For each  $i \in \{1, \dots, n\}$  fix a leaf  $v_i \in V_i$ . Let  $H$  be the subgraph of  $G$  induced by the  $v_i, i \in \{1, \dots, n\}$ . Obviously,  $H$  is a basic  $k$ -leaf power (a basic  $k$ -leaf root for  $H$  is obtained from  $T$  by deleting all leaves in  $V_i \setminus \{v_i\}, 1 \leq i \leq n$ ), and  $G$  is obtained from  $H$  by substituting the cliques  $V_i \neq \emptyset$  into  $v_i$ .

For the other direction, assume  $G$  is obtained from a basic  $k$ -leaf power  $H$  by substituting vertices  $v \in V(H)$  by cliques  $K_v \neq \emptyset$ . Let  $T$  be a  $k$ -leaf root of  $H$ . Then clearly, the tree obtained from  $T$  by replacing each leaf  $v$  by the set  $K_v$  of leaves is a  $k$ -leaf root of  $G$ .  $\square$

PROPOSITION 3.7. *Let  $k \geq 3$ . Then for every graph  $G$ ,*

- (i)  *$G$  is a basic  $k$ -leaf power having a basic  $k$ -leaf root without invisible vertices if and only if  $G$  is the  $(k - 2)$ -th power of some tree.*
- (ii)  *$G$  is a  $k$ -leaf power if and only if  $G$  is obtained from the  $(k - 2)$ -th power of some tree by substituting the vertices by (possibly empty) cliques.*

PROOF. (i): Consider a tree  $T$ . Let  $R$  be the tree obtained from  $T$  by attaching leaf  $x'$  to vertex  $x$ , for every vertex  $x$  in  $T$ . Then for  $u, v \in V(T)$ ,  $uv \in E(T^{k-2})$  if and only if  $d_T(u, v) \leq k - 2$  if and only if  $d_R(u', v') \leq k$ . Therefore,  $R$  is a basic  $k$ -leaf root of  $T^{k-2}$ , and has no invisible vertices.

Conversely, assume that  $G$  has a basic  $k$ -leaf root  $R$  without invisible vertices. Then, clearly,  $G$  is the  $(k - 2)$ -th power of the subtree  $T$  of  $R$  induced by the nonleaf nodes of  $R$ .

(ii): The if-part follows from (i) and Proposition 3.6. For the only-if part, let  $T$  be a  $k$ -leaf root of  $G$ . First we assume that every internal vertex of  $T$  has a leaf attached to it. Let  $B$  be the tree obtained from  $T$  by deleting all the leaves of  $T$ . For a vertex  $b \in B$ , let  $L(b)$  denote the set of all the leaves of  $T$  having  $b$  as a parent, and for a leaf  $x$  in  $T$  let  $v_x$  be the parent of  $x$  in  $T$ . Recall that by our assumption, no  $L(b)$  is empty. Moreover, each  $L(b)$  forms a clique module in  $G$ .

Let  $H$  be the graph obtained from  $G$  by collapsing each  $L(b)$  to a single vertex  $l_b$ . Then, for all  $l_b, l_{b'} \in H$ ,  $l_b l_{b'} \in E(H)$  if and only if  $3 \leq d_T(x, x') \leq k$  for all  $x \in L(b)$ ,  $x' \in L(b')$ , if and only if  $d_B(b, b') \leq k - 2$  if and only if  $bb' \in E(B^{k-2})$ . Therefore  $H$  is isomorphic to  $B^{k-2}$ . Thus,  $G$  is obtained from  $B^{k-2}$  by substituting cliques  $L(b)$  into vertices  $b$ .

In the case that some internal vertex of  $T$  has no leaf attached to it, simply assume that the clique substituted into this vertex is empty.  $\square$

Note that (i) in Proposition 3.7 gives examples for basic  $k$ -leaf powers that may contain true twins. For example, the clique with  $k$  vertices minus an edge  $K_k^-$  (the  $(k - 2)$ -th power of the induced path  $P_k$ ) is a basic  $k$ -leaf power, but contains true twins provided  $k \geq 4$ . Recall that by Observation 3.5 (ii), on the other hand,  $k$ -leaf powers without true twins are basic  $k$ -leaf powers.

COROLLARY 3.8. *Let  $k \geq 3$  and  $G$  be a graph. Then  $G$  is a basic  $k$ -leaf power if and only if  $G$  is an induced subgraph of the  $(k - 2)$ -th power of some tree.*

PROOF. If  $G$  is an induced subgraph of the  $(k - 2)$ -th power of some tree, then  $G$  is a basic  $k$ -leaf power by Proposition 3.7 (i) and Observation 3.5 (i). If  $G$  has a basic  $k$ -leaf root  $T$ , the proof of Proposition 3.7 (ii) shows that  $G$  is (isomorphic to) the subgraph induced in  $T^{k-2}$  by the parents of leaves of  $T$ .  $\square$

#### 4. New Characterizations of Squares of Trees

Squares of trees is a well-studied concept in graph theory [Harary 1972; Ross and Harary 1960]. Efficient algorithms for recognizing squares of trees were studied in Kearney and Corneil [1998], Lau [2006], and Lin and Skiena [1995]. Lin and Skiena [1995], and Lau [2006] give a linear-time algorithm for recognizing whether a given graph is the square of a tree.

In this section, we provide two new structural characterizations for squares of trees which we use later as tools for describing 4-leaf powers. As a direct consequence, we obtain a new linear-time algorithm for recognizing squares of trees and computing the tree root.

Recall that  $K_p^-$  denotes the clique with  $p$  vertices minus an edge;  $K_4^-$  is also called a *diamond*. For tree  $T = (V, E_T)$  let  $T_x$  denote the star with center  $x$  in  $T$ , namely  $T_x = N_T[x]$ .

OBSERVATION 4.1. *Let  $G = T^2$  for a tree  $T$ . Then the following conditions are equivalent:*

- (i)  $\{x, y\}$  is a 2-cut in  $G$ ;
- (ii)  $xy$  is the edge shared by the two triangles in a diamond of  $G$ ;
- (iii)  $x$  and  $y$  are no leaves in  $T$  and  $x, y \in T_x \cap T_y$ ; and
- (iv)  $xy$  is the mid-edge of a  $P_4$  in  $T$ .

PROOF. (i)  $\Rightarrow$  (ii): It is well known that  $G = T^2$  for a tree  $T$  is 2-connected and chordal. Consider two connected components  $A$  and  $B$  of  $G - \{x, y\}$ . As  $G$  is chordal (and every 2-cut in  $G$  is a minimal cutset), there exist vertices  $a \in A$  and  $b \in B$  such that  $a$  and  $b$  are adjacent to  $x$  and  $y$ , hence (ii).

(ii)  $\Rightarrow$  (iii): Let  $a, b, x, y$  induce a diamond in  $G$  with edges  $ax, ay, bx, by$ , and  $xy$ . Since leaves in  $T$  are clearly simplicial vertices in  $G$ ,  $x$  and  $y$  are no leaves in  $T$ . So, (iii) follows if  $xy$  is an edge of  $T$ . Suppose to the contrary that  $xy \notin E(T)$ . Then there exists a vertex  $z$  such that  $zx$  and  $zy$  are edges of  $T$ . Now, since  $a$  and  $b$  are adjacent to  $x$  and  $y$  in  $G$ ,  $a$  and  $b$  must be adjacent to  $z$  in  $T$ , or one of them is adjacent to  $z$  in  $T$  and the other is the vertex  $z$ . In any case,  $a$  and  $b$  would be adjacent in  $G$ , a contradiction. Thus  $xy \in E(T)$ , hence (iii).

(iii)  $\Rightarrow$  (iv): Obvious by definition of  $T_x, T_y$ .

(iv)  $\Rightarrow$  (i): Let  $X$  and  $Y$  be the vertex sets of the subtrees of  $T - xy$  containing  $x$ , respectively,  $y$ . Since  $x$  and  $y$  are no leaves in  $T$ ,  $X - x$  and  $Y - y$  are nonempty. Clearly, any path in  $G$  connecting a vertex in  $X - x$  and a vertex in  $Y - y$  must contain  $x$  or  $y$ , hence  $\{x, y\}$  is a 2-cut of  $G$ .  $\square$

The following fact has also been shown in Lau [2006], but here we give another short proof. A family of subsets of a ground set has the *Helly property* if for every subfamily, pairwise nonempty intersection implies nonempty total intersection. It is well known that subtrees of a tree fulfill the Helly property (see, for example, Brandstädt et al. [1999]).

OBSERVATION 4.2. *Let  $T$  be a tree. Then the maximal cliques in  $T^2$  are exactly the stars  $T_x$ ,  $x \in V(T)$ , for which  $x$  is no leaf in  $T$ .*

PROOF. If  $x$  is a nonleaf vertex of  $T$ , then  $T_x$  clearly induces a maximal clique in  $T^2$ . Conversely, assume that  $Q = \{q_1, \dots, q_k\}$  is a maximal clique in  $T^2$ . We have to show that there is a vertex  $x$  such that  $Q = T_x$ . The property that  $Q$  is a clique means that  $T_{q_i} \cap T_{q_j} \neq \emptyset$  for all  $i, j \in \{1, \dots, k\}$ . Since subtrees of a tree have the Helly property, this means that there is a vertex  $x$  in the intersection of all  $T_{q_i}$ ,  $i \in \{1, \dots, k\}$ , namely  $Q = T_x$ .  $\square$

For the graphs  $G_1, \dots, G_5$  mentioned next, refer to Figure 1.

THEOREM 4.3. *For every graph  $G$ , the following conditions are equivalent.*

- (i)  $G$  is the square of a tree.
- (ii)  $G$  is chordal, 2-connected, and has the following properties:
  - (1) every two distinct maximal cliques have at most two vertices in common;
  - (2) every 2-cut belongs to exactly two maximal cliques of  $G$ ;
  - (3) every pair of nondisjoint 2-cuts belongs to the same maximal clique; and

(4) all 2-cuts contained in the same maximal clique of  $G$  have a common vertex.

(iii)  $G$  is chordal, 2-connected, and  $(G_1, \dots, G_5)$ -free.

PROOF. (i)  $\Rightarrow$  (ii): Let  $G = T^2$  for a tree  $T$ . Then by Observation 4.2, the maximal cliques of  $G$  are exactly the stars in  $T$  whose midpoints are no leaves in  $T$ . Clearly, for  $x \neq y$ ,  $|T_x \cap T_y| \leq 2$ . This shows (1) of (ii).

If  $|T_x \cap T_y| = 2$  then, by Observation 4.1,  $T_x \cap T_y = \{x, y\}$  is a 2-cut and vice versa, which shows (2). Moreover if  $\{u, x\}$  and  $\{x, y\}$  are 2-cuts having the common vertex  $x$ , then both are contained in the maximal clique  $T_x$ , which shows (3). Finally, all 2-cuts in the same maximal clique, say  $T_x$ , have  $x$  in common, which shows (4).

(ii)  $\Rightarrow$  (i): Let  $G$  satisfy (ii). If  $G$  is a clique then  $G$  is the square of a star and (i) holds. Now assume that  $G$  is not a clique but chordal and 2-connected. Then let  $R$  be a clique tree for  $G$  (see Brandstädt et al. [1999] and Spinrad [2003] for more details on clique trees of chordal graphs and Spinrad [2003] for a simple linear-time algorithm for determining such a tree). We construct a spanning tree  $T$  of  $G$  bottom-up along  $R$  as follows.

For each 2-cut  $\{x, y\}$  of  $G$  belonging to the two maximal cliques  $A$  and  $B$ , let  $x$  be the common vertex of all 2-cuts in  $A$  (if  $A$  has more than one; refer condition (4)). Note that for maximal cliques which are adjacent in  $R$ , the intersection contains exactly two vertices, since  $G$  is 2-connected and satisfies condition (1). Then, by condition (3),  $y$  is the common vertex of all 2-cuts in  $B$  (if  $B$  has more than one). Add  $xy$  to  $E(T)$  and add the edges  $xa$ ,  $a \in A - x$ , and the edges  $yb$ ,  $b \in B - y$ , to  $E(T)$ . Clearly,  $T$  is a spanning tree of  $G$ . Moreover, by construction and by condition (2),  $T$  has the following property.

(\*) For each 2-cut  $\{x, y\}$  belonging to maximal cliques  $Q$  and  $Q'$ ,  $T \cap Q$  is a star at a vertex  $z \in \{x, y\}$  and  $T \cap Q'$  is a star at the other vertex in  $\{x, y\} \setminus \{z\}$ .

We now claim that  $G = T^2$ . First, since every edge of  $G$  belongs to a maximal clique, (\*) implies that  $E(G) \subseteq E(T^2)$ . We show now that  $E(T^2) \subseteq E(G)$ . Let  $uv$  be an edge of  $T^2$  not belonging to  $T$ . Then there exists a vertex  $x$  such that  $xu$  and  $xv$  are edges of  $T$ , hence both  $xu$  and  $xv$  are edges of  $G$ . If  $uv$  is also an edge of  $G$ , we are done. If not, as  $G$  is 2-connected, there exists an induced path  $u_1u_2 \dots u_p$ ,  $p \geq 3$ , in  $G - x$  connecting  $u_1 = u$  and  $u_p = v$ . As  $G$  is chordal,  $x$  is adjacent to all  $u_i$ . As  $G$  satisfies condition (1),  $\{x, u_i\}$ ,  $2 \leq i \leq p - 1$ , are 2-cuts of  $G$ . (Indeed, if there exists an induced path  $u_{i-1}v_1 \dots v_qv_{q+1}$  in  $G - \{x, u_i\}$  connecting  $u_{i-1}$  and  $v_{q+1} = u_{i+1}$  then, as  $G$  is chordal,  $x$  and  $u_i$  are adjacent to all  $v_1, \dots, v_q$ . But then the maximal cliques containing  $x, u_{i-1}, u_i, v_1$  and  $x, u_i, v_1, v_2$ , respectively, have more than two vertices in common, contradicting condition (1).) Now, the maximal cliques  $Q$  and  $Q'$  containing  $x, u_1, u_2$  and  $x, u_2, u_3$ , respectively, and the 2-cut  $\{x, u_2\}$  contradict (\*). Thus, we have shown  $E(T^2) \subseteq E(G)$ , hence  $G = T^2$ .

(i)  $\Rightarrow$  (iii): If  $G = T^2$  for some tree  $T$  then  $G$  is chordal and 2-connected. Moreover, every 2-connected induced subgraph of  $T^2$  is also the square of some tree. (Specifically, if the 2-connected induced subgraph  $H$  of  $T^2$  is not a clique, then  $T' = T[V(H)]$  is a subtree of  $T$  with  $H = T'^2$ .) The 2-connected graphs  $G_1, \dots, G_5$  are not the square of any tree, since their maximal cliques do not satisfy one of the conditions (1)–(4) of (ii). Thus, (iii) follows.

(iii)  $\Rightarrow$  (ii): Let  $G$  be a 2-connected  $(G_1, \dots, G_5)$ -free chordal graph. To establish condition (1) of (ii), note that the intersection of two maximal cliques contains at most two vertices, since  $G$  is  $G_3$ -free. If condition (2) does not hold, then  $G$  would contain  $G_2$  as an induced subgraph.

We now prove that  $G$  satisfies condition (3). Assume to the contrary that the 2-cuts  $\{x, y\}$  and  $\{x, z\}$  of  $G$  do not belong to the same maximal clique. In particular,  $y$  and  $z$  are nonadjacent. Let  $A$  be a connected component of  $G - \{x, y\}$  not containing  $z$ , and let  $B$  be a connected component of  $G - \{x, z\}$  not containing  $y$ . Note that  $B \subseteq G - (A \cup \{x, y\})$ . As  $G$  is 2-connected and chordal, there exist vertices  $a \in A$ ,  $b \in B$  such that  $a$  is adjacent to  $x$  and  $y$ , and  $b$  is adjacent to  $x$  and  $z$ . Moreover, there exists an induced path  $v_1 v_2 \dots v_p$ ,  $p \geq 3$ , in  $G - x$  connecting  $v_1 = y$  and  $v_p = z$ , and  $x$  is adjacent to all  $v_i$ ,  $1 \leq i \leq p$ . By the choice of  $A$  and  $B$ , all  $v_i$ ,  $2 \leq i \leq p - 1$ , do not belong to  $A \cup B$ , and  $a$  is nonadjacent to all  $v_i$ ,  $2 \leq i \leq p$ , and  $b$  is nonadjacent to all  $v_i$ ,  $1 \leq i \leq p - 1$ . Thus,  $x, a, v_1, v_2, v_3$ , together with  $v_4$  (if  $p \geq 4$ ) or with  $b$  (if  $p = 3$ ), induce a  $G_4$  in  $G$ , a contradiction. Thus,  $G$  satisfies condition (3).

Finally, we show that  $G$  satisfies condition (4). First, consider two 2-cuts  $\{x, y\}$  and  $\{x', y'\}$  in the same maximal clique of  $G$ . Suppose  $\{x, y\} \cap \{x', y'\} = \emptyset$ . Let  $A$  be a connected component of  $G - \{x, y\}$  not containing  $x', y'$ , and let  $B$  be a connected component of  $G - \{x', y'\}$  not containing  $x, y$ . As  $G$  is 2-connected and chordal, there exist vertices  $a \in A$ ,  $b \in B$  such that  $a$  is adjacent to  $x, y$ , and  $b$  is adjacent to  $x', y'$ . By the choice of  $A$  and  $B$ ,  $a$  is nonadjacent to  $x', y'$ , and  $b$  is nonadjacent to  $x, y$ . Hence  $a, b, x, y, x'$ , and  $y'$  induce a  $G_5$  in  $G$ . Thus, we have shown that every two 2-cuts in the same maximal clique must have a common vertex. Next, we show that no three 2-cuts in  $G$  form a triangle. For, if  $\{x, y, z\}$  induces a triangle, where  $\{x, y\}$ ,  $\{x, z\}$ , and  $\{y, z\}$  are 2-cuts of  $G$ , then consider a connected component  $A$  of  $G - \{x, y\}$  not containing  $z$ , a connected component  $B$  of  $G - \{x, z\}$  not containing  $y$ , and a connected component  $C$  of  $G - \{y, z\}$  not containing  $x$ . As before, there exist vertices  $a \in A$ ,  $b \in B$  and  $c \in C$  such that  $a$  is adjacent to  $x, y$  and nonadjacent to  $z$ ,  $b$  is adjacent to  $x, z$  and nonadjacent to  $y$ ,  $c$  is adjacent to  $y, z$  and nonadjacent to  $x$ , and  $a, b, c$  are pairwise nonadjacent. In other words,  $a, b, c, x, y$ , and  $z$  induce a  $G_1$  in  $G$ , a contradiction. Thus, we have shown that no three 2-cuts of  $G$  form a triangle. It follows that  $G$  satisfies condition (4).  $\square$

As testing chordality [Rose et al. 1976; Tarjan and Yannakakis 1984] and determining a clique tree [Spinrad 2003] can be done in linear-time, Theorem 4.3 implies the following linear-time algorithm to test whether a given graph  $G$  is the square of some tree  $T$ : Construct a candidate for tree  $T$  via the proof of (ii)  $\Rightarrow$  (i), test whether  $T$  is a tree, and test whether  $G = T^2$ . The latter can be done in linear time (refer to Lau [2006] and Lin and Skiena [1995]). Note that the linear-time algorithm for recognizing squares of trees given in Lin and Skiena [1995] builds the tree root incrementally, by identifying the leaves and their parents of a possible root tree and repeating the process recursively, while in Lau [2006] this is done by reducing the problem to recognizing the squares of trees with a specified neighborhood. In contrast, our algorithm deduces the tree root directly from the clique structure of the square of a tree.

**COROLLARY 4.4** ([LAU 2006; ROSS AND HARARY 1960]). *The tree roots of squares of trees are unique, up to isomorphism.*

PROOF. Let  $G = (V, E)$  be the square of a tree. If  $G$  is a clique, any tree root of  $G$  must be isomorphic to the star  $K_{1,|V|-1}$ . If  $G$  is not a clique, there is some 2-cut  $\{x, y\}$  in  $G$ . Let  $A$  and  $B$  be the connected components of  $G \setminus \{x, y\}$  (by Theorem 4.3 (ii), condition (2),  $G \setminus \{x, y\}$  has exactly two connected components). Set  $G_A := G[A \cup \{x, y\}]$  and  $G_B := G[B \cup \{x, y\}]$ . By Observation 4.1,  $xy$  is an edge belonging to all tree roots  $T$  of  $G$ , hence  $T_A := T \cap G_A$  and  $T_B := T \cap G_B$  are subtrees of  $T$ , and, as  $G = T^2$ ,  $G_A = (T_A)^2$  and  $G_B = (T_B)^2$ . By induction,  $T_A$ , respectively,  $T_B$ , is the unique tree root of  $G_A$ , respectively,  $G_B$ . Therefore,  $T$  is the unique tree root of  $G$ .  $\square$

### 5. Structure of Basic 4-Leaf Powers

In this section we characterize basic 4-leaf powers. As remarked earlier, every 4-leaf power results from substituting cliques into the vertices of a basic 4-leaf power. Thus, we obtain a new characterization of 4-leaf powers. In Section 7 we will use this for recognizing 4-leaf powers in linear time.

The 2-connected basic 4-leaf powers can be characterized as follows.

PROPOSITION 5.1. *For every graph  $G$ , the following conditions are equivalent:*

- (i)  $G$  has a basic 4-leaf root without invisible vertices;
- (ii)  $G$  is the square of some tree; and
- (iii)  $G$  is a 2-connected basic 4-leaf power.

PROOF. (i)  $\Leftrightarrow$  (ii): follows by Proposition 3.7 (i).

(ii)  $\Rightarrow$  (iii): Let  $G = T^2$  for some tree  $T$ . Then,  $G$  is 2-connected, and by Proposition 3.7 (i),  $G$  is a basic 4-leaf power.

(iii)  $\Rightarrow$  (ii): Let  $G$  be a 2-connected basic 4-leaf power, and  $R$  be a basic 4-leaf root of  $G$ . If  $R$  has no invisible vertices, then (ii) can be deduced using (i). In case  $R$  has an invisible vertex, we claim that  $G$  must be a clique (and we are done). Suppose  $v$  is an invisible vertex of  $R$ . As  $G$  is connected, no neighbor of  $v$  in  $R$  is an invisible vertex. Root  $R$  at  $v$ , and for any neighbor  $w$  of  $v$  in  $R$  let  $R_w$  denote the subtree of  $R$  rooted at  $w$ . Now, if  $G$  is not a clique, then there exists a neighbor  $w$  of  $v$  in  $R$  such that  $R_w$  has a leaf  $x$  different from the leaf  $y$  adjacent to  $w$  (since  $w$  is not an invisible vertex and  $R$  is a basic 4-leaf root,  $w$  has exactly one neighbor that is a leaf). Let  $w' \neq w$  be another neighbor of  $v$  in  $R$ . Clearly, any path in  $G$  connecting  $x \in R_w$  and a vertex that is a leaf in  $R_{w'}$  must pass  $y$ , contradicting the 2-connectedness of  $G$ .  $\square$

COROLLARY 5.2. *2-connected basic 4-leaf powers can be recognized in linear time, and a basic 4-leaf root of a 2-connected basic 4-leaf power can be constructed in linear time.*

PROOF. By Proposition 5.1 (ii), 2-connected basic 4-leaf powers are exactly the squares of trees, hence can be recognized in linear time (refer to Section 4). Moreover, a basic 4-leaf root of the square of a tree can be constructed in linear time according to the proof of Proposition 3.7 (i).  $\square$

COROLLARY 5.3. *The graphs  $G_1, G_2, G_3, G_4$ , and  $G_5$  in Figure 1 are not basic 4-leaf powers. Moreover,  $G_1$  and  $G_4$  are not 4-leaf powers at all.*

PROOF. As  $G_1, G_2, G_3, G_4,$  and  $G_5$  are 2-connected, but not the square of any tree by Theorem 4.3, they are not basic 4-leaf powers by Proposition 5.1. As  $G_1$  and  $G_4$  do not have true twins, by Observation 3.5 (ii), they are not 4-leaf powers.  $\square$

COROLLARY 5.4. *Every 2-connected basic 4-leaf power different from a clique has a unique basic 4-leaf root, up to isomorphism.*

PROOF. Let  $G$  be a 2-connected graph different from a clique, and let  $R$  be a basic 4-leaf root of  $G$ . The proof of Proposition 5.1 (iii)  $\Rightarrow$  (ii) shows that  $R$  cannot have any invisible vertex. The proof of Proposition 5.1 (i)  $\Rightarrow$  (ii) shows that  $G = T^2$ , where  $T$  is the subtree of  $R$  induced by the internal vertices of  $R$ . Since the tree roots of tree squares are unique (refer to Corollary 4.4), the basic 4-leaf root  $R$  of  $G$  is unique.  $\square$

In particular, the diamond has a unique basic 4-leaf root. This gives a simple proof for the following fact.

COROLLARY 5.5. *The graphs  $G_6, G_7,$  and  $G_8$  in Figure 1 are not basic 4-leaf powers. Moreover,  $G_6$  is not a 4-leaf power at all.*

PROOF. Using the unique basic 4-leaf root of a diamond, it can be established via inspection that  $G_6, G_7,$  and  $G_8$  are not basic 4-leaf powers. Moreover, as  $G_6$  does not have true twins, by Observation 3.5 (ii), it is not a 4-leaf power.  $\square$

Having characterized the 2-connected graphs that are basic 4-leaf powers, we characterize the basic 4-leaf powers in general in Theorem 5.7. As a preparing step, we need the next proposition.

OBSERVATION 5.6. *Let  $p \geq 3$  be an integer and  $G$  be a  $(p - 2)$ -connected chordal graph. If  $G$  is not a clique, then every vertex of  $G$  is in some  $K_p^-$ .*

PROOF. (Induction on  $p$ .) The case  $p = 3$  is easily verified. Let  $p \geq 4$ , and let  $G$  be a  $(p - 2)$ -connected chordal graph. Suppose  $G$  is not a clique, and consider an arbitrary vertex  $v$  of  $G$ . If  $v$  was a universal vertex, then  $G - v$  is a  $(p - 3)$ -connected chordal graph that is not a clique. By induction,  $G - v$  contains a  $K_{p-1}^-$ , which together with  $v$  forms a  $K_p^-$ . Therefore, we can assume that  $v$  has a non-neighbor in  $G$ . Let  $A$  be a connected component of  $G - N[v]$ ,  $S \subseteq N(v)$  be the set that minimally separates  $v$  from  $A$ , and  $w \in A$  be a vertex that has the most neighbors in  $S$ . As  $G$  is chordal,  $S$  induces a clique (it is well known that every minimal cutset in a chordal graph induces a clique). We claim that  $w$  is adjacent to every vertex in  $S$ . Suppose to the contrary that  $x \in S$  is a vertex that is not adjacent to  $w$ . Let  $P$  be a shortest path from  $x$  to  $w$  such that every vertex of  $P$ , except  $x$ , is in  $A$ . Clearly,  $P$  has at least 3 vertices. Let  $y$  be the vertex next to  $x$  on  $P$ . By choice of  $w$ , there exists a vertex  $z \in S$  that is adjacent to  $w$  but not to  $y$ . Then, the subgraph of  $G$  induced by the vertices of  $P$  and  $z$  has a chordless cycle on at least 4 vertices, a contradiction. Now, the connectivity of  $G$  implies that  $|S| \geq p - 2$ . Therefore,  $v, w,$  and any  $p - 2$  vertices of  $S$  induce a  $K_p^-$ .  $\square$

In the proof of Theorem 5.7 to follow (and of Theorem 6.9 in Section 6) we need the following tree structure of the blocks in connected graphs. The *block-cut-vertex graph*  $BC(G)$  of a graph  $G$  has all blocks and cut vertices of  $G$  as vertices, and two vertices in  $BC(G)$  are adjacent if one of them is a block in  $G$  and the other is a cut vertex in  $G$  belonging to the block. It is well known (refer to Harary [1972]) that

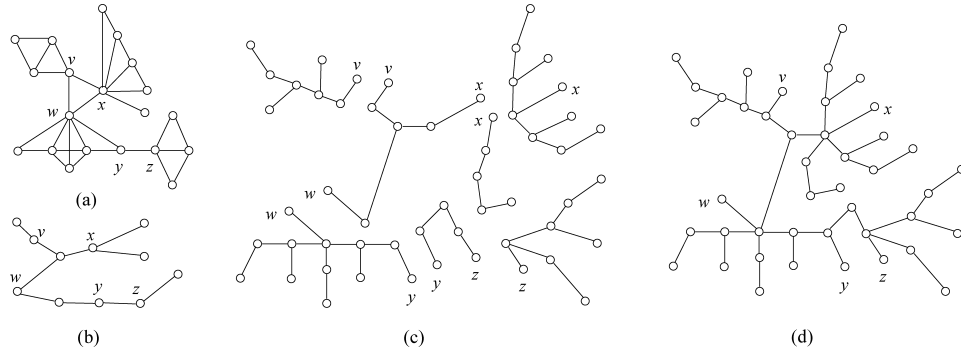


FIG. 2. (a) The graph  $G$ ; (b) the block-cut-vertex tree of  $G$ ; (c) the basic 4-leaf roots of the blocks of  $G$ ; and (d) the basic 4-leaf root for  $G$ .

the block-cut-vertex graph  $BC(G)$  of any connected graph  $G$  is a tree, called the *block-cut-vertex tree* of  $G$ .

**THEOREM 5.7.** *For every connected graph  $G$ , the following conditions are equivalent:*

- (i)  $G$  is a basic 4-leaf power;
- (ii) every block of  $G$  is the square of some tree, and for every nondisjoint pair of blocks, at least one of them is a clique;
- (iii)  $G$  is an induced subgraph of the square of some tree; and
- (iv)  $G$  is chordal and  $(G_1, \dots, G_8)$ -free.

**PROOF.** (i)  $\Rightarrow$  (ii): Let  $G$  be a basic 4-leaf power. Then, by Proposition 5.1, every block of  $G$  is the square of some tree. Suppose  $G$  has nondisjoint blocks  $A$  and  $B$  such that neither  $A$  nor  $B$  is a clique. Let  $u$  be the cut vertex shared by  $A$  and  $B$ . By Observation 5.6,  $u$  belongs to a diamond in each of  $A$  and  $B$  but then  $G$  contains  $G_6, G_7$ , or  $G_8$  as an induced subgraph, a contradiction to Corollary 5.5.

(ii)  $\Rightarrow$  (i): We will construct a basic 4-leaf root  $R$  for  $G$  from appropriately constructed basic 4-leaf roots  $R_B$  for blocks  $B$  of  $G$ . See also Figure 2 for an illustration. Let  $B$  be a block of  $G$  with vertices  $x_1, \dots, x_k, k \geq 2$ .

If  $B$  is a clique, take new vertices  $S_B, X_1, \dots, X_k$  and form a star with center vertex  $S_B$  and edges  $S_B X_i, 1 \leq i \leq k$ . Then  $R_B$  is obtained from this star by attaching leaf  $x_i$  to node  $X_i, 1 \leq i \leq k$ .

If  $B$  is not a clique, then consider the tree  $T_B$  (with new vertices) such that  $B$  is isomorphic to  $T_B^2$ . Let  $X_i$  be the node of  $T_B$  corresponding to the vertex  $x_i$  of  $B$ . Then  $R_B$  is obtained from  $T_B$  by attaching leaf  $x_i$  to node  $X_i$  (refer to the proof of Proposition 3.7).

The basic 4-leaf root  $R$  for  $G$  is now obtained by putting all  $R_B$  together as follows. For each cut vertex  $v$  of  $G$ , let  $B_1, \dots, B_k$  be the neighbors of  $v$  in the block-cut-vertex tree  $BC(G)$  of  $G$ . Identify the parents of  $v$  from the roots  $R_{B_i}$  and delete all but one copy of leaf  $v$ . Let the resulting graph be  $R$ . As  $BC(G)$  is a tree,  $R$  is a tree.

We claim that  $R$  is a basic 4-leaf root for  $G$ . It is clear from the construction of  $R$  that the adjacencies within a block of  $G$  are preserved by  $R$ . Consider vertices

$x \in B \setminus B'$  and  $y \in B' \setminus B$ , where  $B$  and  $B'$  are two distinct nondisjoint blocks of  $G$ . As at least one of  $B$ ,  $B'$  is a clique, from the construction of  $R$ ,  $d_R(x, y) \geq 5$ .

(i)  $\Leftrightarrow$  (iii): Follows from Corollary 3.8.

(i)  $\Rightarrow$  (iv): By Corollaries 5.3 and 5.5, and Proposition 3.3.

(iv)  $\Rightarrow$  (ii): By Theorem 4.3, every block of  $G$  is the square of a tree. The rest of the implication follows from Observation 5.6 and the fact that  $G$  is  $(G_6, G_7, G_8)$ -free.  $\square$

**COROLLARY 5.8.** *Basic 4-leaf powers can be recognized in linear time, and a basic 4-leaf root of a basic 4-leaf power can be constructed in linear time.*

**PROOF.** It is well known that the blocks of a graph can be computed in linear time (refer to Tarjan [1972]). So, because of Theorem 5.7 (ii), recognizing basic 4-leaf powers reduces to the 2-connected case. The latter can be done in linear time by Corollary 5.2. Finally, once the basic 4-leaf roots of the blocks of a basic 4-leaf power are available, the basic 4-leaf root for the whole graph can be easily constructed according to the proof of part (ii)  $\Rightarrow$  (i) in Theorem 5.7.  $\square$

Rautenbach [2006] and Dom et al. [2005] have characterized the 4-leaf powers without true twins via the forbidden induced subgraphs listed in Figure 1. Note that each 4-leaf power without true twins is a basic 4-leaf power, but not vice versa (refer to the comments after Proposition 3.7). The equivalence (i)  $\Leftrightarrow$  (iv) in Theorem 5.7 extends the characterization in Rautenbach [2006] and Dom et al. [2005] to the larger class of basic 4-leaf powers. Our result can be interpreted as follows: The graphs  $G_1, \dots, G_5$  are responsible for the structure of the 2-connected components, whereas the graphs  $G_6, G_7, G_8$  represent the gluing conditions of the 2-connected components.

## 6. Structure of 4-Leaf Powers

In this section we characterize 4-leaf powers. Recall that, by Proposition 3.6, a graph  $G$  is a 4-leaf power if and only if it is obtained from a basic 4-leaf power  $H$  by substituting cliques  $K_v$  into vertices  $v \in V(H)$ . If  $v$  is a cut vertex of  $H$ , the clique  $K_v$  must be a maximal module in  $G$ . This observation leads to the following notion.

**Definition 6.1.** Let  $G$  be a graph different from a clique. The graph  $G^*$  is obtained from  $G$  by collapsing each maximal module  $M$  in  $G$  that is a clique to a single vertex  $m$ .

Note that if  $G$  is not a clique, the maximal modules in  $G$  that are cliques are pairwise disjoint. Thus, for such graphs  $G$ ,  $G^*$  is well defined. Moreover, as the maximal modules in a graph can be computed in linear time (see, for example, McConnell and Spinrad [1999]), we can compute  $G^*$  in linear time. Furthermore,  $G^*$  is (isomorphic to) an induced subgraph of  $G$  and has no nontrivial maximal modules that are cliques. Note also that our notion  $G^*$  differs from the critical clique graph of  $G$  in Lin et al. [2000]. Indeed, a clique module is not necessarily a maximal module.

**OBSERVATION 6.2.** *Let  $G$  be a connected graph different from a clique. If  $G$  is a 4-leaf power, then:*

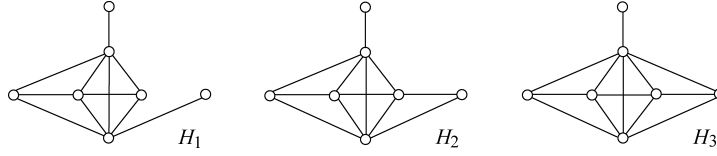


FIG. 3. Some forbidden induced subgraphs for 4-leaf powers.

- (i) every block of  $G^*$  is obtained from the square of a tree by substituting vertices by nonempty cliques; and
- (ii) for every nondisjoint pair of distinct blocks of  $G^*$ , at least one of them is a clique.

PROOF. Assume that  $G$  is a 4-leaf power. Then, as an induced subgraph of  $G$ ,  $G^*$  is also a 4-leaf power, hence  $G^*$  is obtained from a basic 4-leaf power  $H$  by substituting vertices by nonempty cliques. Then, because no nontrivial maximal module in  $G^*$  is a clique, the cut vertices in  $G^*$  and the cut vertices in  $H$  correspond to each other. Therefore, every block of  $G^*$  arises from a block of  $H$  by substituting vertices by nonempty cliques. Since the blocks of  $H$  satisfy condition (ii) of Theorem 5.7, and since graphs obtained from a clique by substituting vertices by (nonempty) cliques are again cliques, the blocks of  $G^*$  must satisfy (i) and (ii).  $\square$

The following three graphs show that the conditions (i) and (ii) in Observation 6.2 are not sufficient for being a 4-leaf power.

OBSERVATION 6.3. *The graphs  $H_1$ ,  $H_2$ , and  $H_3$  depicted in Figure 3 are not 4-leaf powers.*

PROOF. Each of the graphs  $H_1$ ,  $H_2$ , and  $H_3$  contains an induced  $G_3$ , hence none of them is a basic 4-leaf power (refer to Theorem 5.7). None of them contains true twins, hence none of them is a 4-leaf power (refer to Observation 3.5).  $\square$

Observation 6.2 (i) suggests the following notion.

*Definition 6.4.* A *plump square of a tree*, or *pst* for short, is a graph obtained from the square of some tree by substituting vertices by nonempty cliques.

Note that, unlike the case of squares of trees, not every 2-connected induced subgraph of a pst is again a pst. For example, the graph  $G = K_1 \circlearrowleft H$  where  $H$  is the graph  $G_5$  in Figure 1, or more generally a plump spider as defined to follow, is a pst, but  $H$  is not.

Now we are going to characterize pst's. For this purpose, we need the following notion (taken from Jamison and Olariu [1992]).

*Definition 6.5.* A *thin spider*  $(U, Q, R)$  is a graph whose vertex set can be partitioned into three disjoint subsets  $U, Q, R$  such that:

- $U$  is a stable set,  $Q$  is a clique,  $|U| = |Q| \geq 1$ ,  $|R| \leq 1$ ; and
- $R \circlearrowleft U$ ,  $R \circlearrowleft Q$ , and every vertex in  $U$  is adjacent to exactly one vertex in  $Q$ , and vice versa.

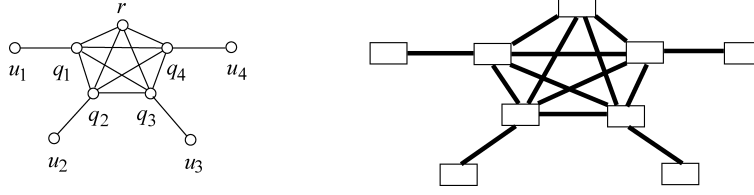


FIG. 4. A thin spider  $H_0 = (\{u_1, u_2, u_3, u_4\}, \{q_1, q_2, q_3, q_4\}, \{r\})$  (left) and a plump spider  $H$  obtained from  $H_0$  (right) illustrated. Each rectangular box in  $H$  is a nonempty clique and each bold edge presents a join between the cliques.

A *plump spider*  $H$  is a (connected) graph obtained from a thin spider  $H_0$  by substituting the vertices by nonempty cliques; we also say that the plump spider  $H$  is obtained from the thin spider  $H_0$ . See Figure 4 for an example.

By Theorem 5.7, thin spiders are basic 4-leaf powers. Hence plump spiders are 4-leaf powers.

Consider now a pst  $G$  different from a clique (we call such pst's *nontrivial*), and let  $T$  be a tree such that  $G$  is obtained from  $G' = T^2$  by substituting vertices  $v \in V(G') = V(T)$  by nonempty cliques  $K_v$ . Clearly, if  $x$  and  $y$  are two leaves in  $T$  having the same parent, then  $G$  is also obtained from  $(T - y)^2$  by substituting vertices by nonempty cliques. Thus, we may choose such a tree  $T$  with the additional property that no two leaves in  $T$  have the same parent. Such trees are called *basic trees*; note that the smallest basic tree of nontrivial pst's is the  $P_4$ .

Plump squares of trees can be described as follows.

LEMMA 6.6. *Let  $G$  be a nontrivial pst, and let  $T$  be a basic tree such that  $G$  is obtained from  $T^2$  by substituting vertices  $v \in V(T)$  by nonempty cliques  $K_v$ . Then exactly one of the following conditions holds.*

- (1)  $T$  is the  $P_4$ ,  $G = K \circledast H$ , where  $K$  is a clique with at least two vertices and  $H$  is the disjoint union of two nonempty cliques.
- (2)  $T$  contains a  $P_5$  but no  $P_6$ ,  $G = K \circledast H$ , where  $K$  is a nonempty clique and  $H$  is a plump spider obtained from a thin spider  $(U, Q, R)$  with  $|U| = |Q| \geq 2$ .
- (3)  $T$  contains a  $P_6$ ,  $\overline{G}$  is connected, and every maximal nontrivial module of  $G$  is a clique  $K_v$  for some  $v \in V(T)$  and  $G^* = T^2$ .

PROOF. Let  $G' = T^2$ . Then  $G$  is obtained from  $G'$  by substituting vertices  $v \in V(G') = V(T)$  by nonempty cliques  $K_v$ .

If  $T$  is  $P_5$ -free,  $T$  is the  $P_4$   $abcd$ , and hence  $G'$  is a diamond and  $G = K \circledast H$ , where  $K = K_b \circledast K_c$  is a clique with at least two vertices, and  $H$  is the disjoint union of the two nonempty cliques  $K_a$  and  $K_d$ . Assume that  $T$  has a  $P_5$ , but no  $P_6$ . Let  $x$  be the midpoint of a  $P_5$  in  $T$ . Then  $G' - x$  is the thin spider  $(U, Q, R)$ , where  $|U| \geq 2$  is the set of all leaves in  $T$  nonadjacent to  $x$ ,  $Q$  is the set of parents of leaves in  $T$  in  $U$  (as  $T$  is basic,  $|Q| = |U|$ ), and  $R$ , possibly empty, consists of the leaf in  $T$  adjacent to  $x$  (if any). Thus  $G = K \circledast H$ , where  $K = K_x$  is a nonempty clique, and  $H$  is a plump spider obtained from the thin spider  $G' - x$ .

Suppose  $T$  contains a  $P_6$ . In this case we will show that every maximal nontrivial module in  $G$  is a clique  $K_v$  for some  $v \in V(G')$ , hence  $G^* = G' = T^2$ . Let  $P = abcdef$  be a  $P_6$  in  $T$ , and let  $L$  and  $R$  be the two trees of  $T - cd$  containing  $c$ , respectively,  $d$ . Consider a maximal module  $M$  of  $G$ . Note that for all  $u, v \in V(T)$

and all  $x \in K_u, y \in K_v, x$  and  $y$  are adjacent in  $G$  if and only if  $d_T(u, v) \leq 2$ . Hence

$$\text{if } M \cap K_u \neq \emptyset \text{ and } M \cap K_v \neq \emptyset, \text{ then for all vertices } w \text{ of } T \text{ with } d_T(w, u) > 2: d_T(w, v) > 2 \text{ or } K_w \subset M. \quad (1)$$

For otherwise, some vertex in  $K_w \setminus M$  would distinguish a vertex in  $M \cap K_u$  and a vertex in  $M \cap K_v$ .

Suppose now  $M \cap K_c \neq \emptyset$ . If  $M \cap K_d \neq \emptyset$  then by (1)  $K_a \subset M, K_f \subset M$ , and then  $K_b \subset M, K_e \subset M$ . Therefore  $K_c \subset M$  (otherwise some vertex in  $K_c \setminus M$  would distinguish a vertex in  $M \cap K_c$  and a vertex in  $K_f$ ), and similarly,  $K_d \subset M$ . Thus  $M = V(G)$ ; otherwise there would be a vertex in  $L \setminus M$  (or in  $R \setminus M$ ) adjacent to a vertex in  $M \cap L$  (in  $M \cap R$ , respectively) but nonadjacent to a vertex in  $M \cap R$  (in  $M \cap L$ , respectively).

Thus, we may assume that  $M \cap K_d = \emptyset$ . By (1) with  $w = d$  and  $v = c$ ,  $M \cap K_x = \emptyset$  for all  $x \in L \setminus N_T[c]$ , and by (1) with  $w = a, M \cap K_y = \emptyset$  for all  $y \in R$ . Hence also  $M \cap K_z = \emptyset$  for all  $z \in N_T(c)$  (by (1) with  $w = e$ ). Thus  $M \subseteq K_c$ , hence  $M = K_c$  because  $K_c$  is a module and  $M$  is a maximal module in  $G$ .

By symmetry, if  $M \cap K_d \neq \emptyset$  then  $M = V(G)$  or  $M = K_d$ . So, we may assume that  $M \cap K_c = M \cap K_d = \emptyset$ . Let, without loss of generality,  $M \cap K_x \neq \emptyset$  for some  $x \in L - c$ . Among such vertices  $x$ , choose one such that the path  $xx_1x_2 \cdots x_k$  in  $L$  from  $x$  to  $x_k := c$  is as short as possible. In particular,  $M \cap K_{x_i} = \emptyset, i = 1, 2, \dots, k$ . Set  $x_{k+1} := d$ . We claim that  $M \cap K_z = \emptyset$  for all  $z \in R$ : To the contrary, let  $M \cap K_z \neq \emptyset$  for some  $z \in R - d$ . Then  $k = 1$  (otherwise  $u = z, v = x$ , and  $w = x_1$  contradict (1)). Hence  $M \cap K_a = \emptyset$  (by (1) with  $u = a, v = x$ , and  $w = d$ ) and  $M \cap K_b = \emptyset$  (by (1) with  $u = z, v = b$ , and  $w = a$ ). In particular,  $x \neq b$ . But then the vertices in  $K_b$  distinguish a vertex in  $M \cap K_x$  and a vertex in  $M \cap K_z$ . This contradiction shows  $M \cap K_z = \emptyset$  for all  $z \in R$ .

Now, if  $M \cap K_y \neq \emptyset$  for some other vertex  $y \in L - x$ , then, by (1) with  $w = x_1$  (if  $d_T(x_1, y) > 2$ ) or with  $w = x_2$  (if  $d_T(x_1, y) = 2$ ),  $x_1$  and  $y$  must be adjacent in  $T$ . By the choice of  $T$ , at least one of  $x, y$  is a nonleaf of  $T$ . Let  $z \neq x_1$  be a neighbor of  $x$  or  $y$  in  $T$ . By (1) with  $w = x_2, M \cap K_z = \emptyset$ . But then the vertices in  $K_z$  distinguish a vertex in  $M \cap K_x$  and a vertex in  $M \cap K_y$ . Thus  $M \cap K_y = \emptyset$  for all  $y \in T - x$ , hence  $M = K_x$ .

It remains to show that if  $T$  contains a  $P_6$  then  $\overline{T^2}$  is connected. Consider a  $P_6 abcdef$  in  $T$ . In  $\overline{T^2}$ , the vertices  $a, c, d, f$  induce a  $P_4$  and all other vertices are adjacent to a vertex of this  $P_4$ . Thus  $\overline{T^2}$  is connected.  $\square$

**COROLLARY 6.7.** *The basic trees for nontrivial pst's are unique, up to isomorphism.*

**PROOF.** This follows from Lemma 6.6 and Corollary 4.4.  $\square$

For a subset  $S$  of vertices in a graph  $G$ , set  $N_G(S) := \bigcup_{v \in S} N_G(v) \setminus S$ . A module  $M$  of  $G$  is called *internal* if  $N_G(M)$  is not a clique in  $G$ ; otherwise,  $M$  is a *leaf module*. The meaning of this notion is given in the next observation.

**OBSERVATION 6.8.** *Let  $G$  be a nontrivial pst, and let  $T$  be a tree such that  $G$  is obtained from  $T^2$  by substituting vertices  $v \in V(T)$  by nonempty cliques  $K_v$ . Then  $K_v$  is an internal module in  $G$  if and only if  $v$  is an internal vertex in  $T$ .*

PROOF. Since  $G$  is not a clique,  $T^2$  is not a clique. Hence (see also Lin and Skiena [1995])  $v$  is a leaf in  $T$  if and only if  $v$  is a simplicial vertex in  $T^2$ . The latter is clearly equivalent to the fact that  $N_G(K_v)$  is a clique.  $\square$

Now 4-leaf powers can be characterized as follows. This characterization allows us to recognize 4-leaf powers in linear time, which is described in the next section.

**THEOREM 6.9.** *A connected graph  $G$  different from a clique is a 4-leaf power if and only if the following holds.*

- (i) *For every nondisjoint pair of distinct blocks of  $G^*$ , at least one of them is a clique; and*
- (ii) *for every nonclique block  $B$  of  $G^*$ , exactly one of the following three conditions holds:*
  - (1)  *$B = K \oplus H$ , where  $K$  is a clique with at least two vertices and  $H$  is the disjoint union of two nonempty cliques. Moreover, if  $|K| \geq 3$ ,  $K$  contains at most one cut vertex of  $G^*$ .*
  - (2)  *$B = K \oplus H$ , where  $K$  is a nonempty clique and  $H$  is a plump spider obtained from a thin spider  $(U, Q, R)$  with  $|U| = |Q| \geq 2$ . Moreover, no maximal nontrivial internal module  $M$  of  $H$  contains a cut vertex of  $G^*$ , and if  $|K| \geq 2$ ,  $K$  contains no cut vertices of  $G^*$ .*
  - (3)  *$\overline{B}$  is connected and  $B^*$  is the square of a tree. Moreover, no maximal nontrivial internal module  $M$  of  $B$  contains a cut vertex of  $G^*$ .*

PROOF. *Necessity.* Let  $G$  be a connected 4-leaf power different from a clique. As we have seen in Observation 6.2, (i) must hold, and every nonclique block  $B$  of  $G^*$  must be a nontrivial pst. Note that  $G^*$  is also a 4-leaf power, hence by Observation 6.3,  $G^*$  is  $(H_1, H_2, H_3)$ -free (see Figure 3). By Lemma 6.6, we have exactly one of the following cases.

(1)  $B = K \oplus H$ , where  $K$  is a clique with at least two vertices and  $H$  is the disjoint union of two nonempty cliques. Now, if  $x, y \in K$  are two distinct cut vertices of  $G^*$  and  $K \setminus \{x, y\} \neq \emptyset$ , then  $x, y$ , a vertex  $z \in K \setminus \{x, y\}$ , two nonadjacent vertices in  $H$ , together with a neighbor of  $x$  not in  $B$  and a neighbor of  $y$  not in  $B$  would induce an  $H_1$  in  $G^*$ , a contradiction. Thus, if  $|K| \geq 3$ ,  $K$  contains at most one cut vertex of  $G^*$ .

(2)  $B = K \oplus H$ , where  $K$  is a nonempty clique and  $H$  is a plump spider obtained from a thin spider  $(U, Q, R)$  with  $|U| = |Q| \geq 2$ . Observe that such a thin spider  $(U, Q, R)$  always contains an induced  $P_4$  and thus  $H$  also contains an induced  $P_4$ . Now, if  $x \in K$  is a cut vertex of  $G^*$  and  $K \setminus \{x\} \neq \emptyset$ , then  $x$ , a vertex  $y \in K \setminus \{x\}$ , an induced  $P_4$  in  $H$ , together with a neighbor of  $x$  not in  $B$ , would induce an  $H_3$  in  $G^*$ , a contradiction. Thus, if  $|K| \geq 2$ ,  $K$  contains no cut vertex of  $G^*$ .

Consider a maximal nontrivial module  $M$  in  $H$  with nonclique  $N_H(M)$ . Recall that  $H$  was obtained from  $(U, Q, R)$  by substituting cliques  $K_v \neq \emptyset$  into vertices  $v \in U \cup Q \cup R$ . Then, by definition of thin spiders,  $M$  is a clique  $K_q$  for some  $q \in Q$ , and  $N_H(M)$  consists of two disjoint nonempty cliques, namely  $K_u$  with  $u \in U$  adjacent to  $q$  and  $\bigcup_{v \in (Q-q) \cup R} K_v$ . Now, if  $x \in M$  is a cut vertex of  $G^*$ , then a vertex in  $K_u$ ,  $x$ , a vertex  $y \in M \setminus \{x\}$ , a vertex  $z \in K_{q'}$  for some  $q' \in Q - q$ , a vertex  $t \in K_{u'}$  with  $u' \in U - u$  adjacent to  $q'$ , a vertex in  $K$ , and a neighbor of  $x$

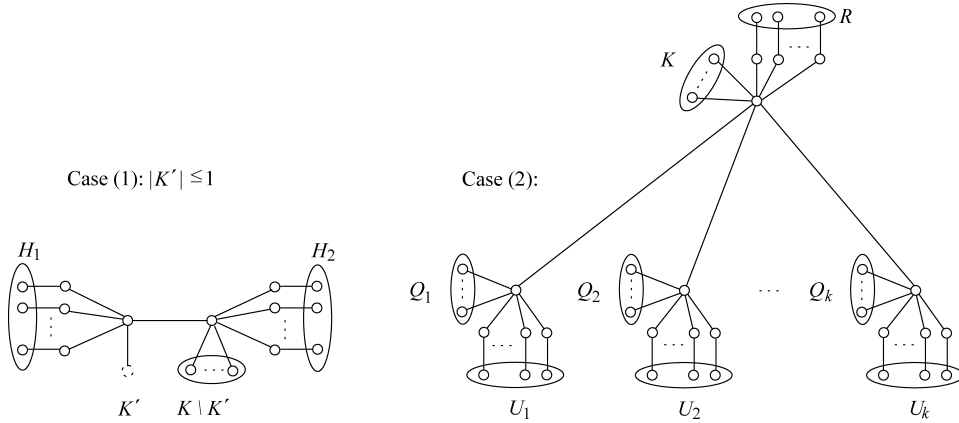


FIG. 5. The 4-leaf roots for cases (1) and (2).

not in  $B$  would induce an  $H_2$  in  $G^*$ , a contradiction. Thus, no maximal nontrivial internal module of  $H$  contains a cut vertex of  $G^*$ .

(3)  $\overline{B}$  is connected and  $B^*$  is the square of a tree. In this case, we know from Lemma 6.6 that there is a (basic) tree  $T$  such that  $B$  is obtained from  $T^2$  by substituting vertices  $v \in V(T)$  by nonempty cliques  $K_v$  and  $B^* = T^2$ , and every maximal nontrivial module  $M$  of  $B$  is a  $K_v$ .

Consider a maximal nontrivial internal module  $M$  of  $B$ . Then the vertex  $v \in V(T)$  with  $K_v = M$  must be an internal vertex of  $T$  (refer to Observation 6.8.) Hence  $v$  is an internal vertex of some  $P_5 P$  in  $T$  (as, by Lemma 6.6,  $T$  has an induced  $P_6$ ). Now, if  $x \in M = K_v$  is a cut vertex of  $G^*$ , then  $x$ , a vertex  $y \in M \setminus \{x\}$ , a neighbor of  $x$  not in  $B$ , and four vertices  $z_i \in K_{w_i}$ ,  $1 \leq i \leq 4$ , where  $P - v = \{w_1, w_2, w_3, w_4\}$  would induce an  $H_2$  (if  $v$  is not the midpoint of  $P$ ) or an  $H_3$  (otherwise) in  $G^*$ , a contradiction. Thus, no maximal nontrivial internal module in  $B$  contains a cut vertex of  $G^*$ . We have proved (ii).

*Sufficiency.* Suppose that the connected graph  $G$  satisfies (i) and (ii). We will show that  $G$  is a 4-leaf power by constructing a 4-leaf root for it. As  $G$  is connected,  $G^*$  is also connected, hence the block-cut-vertex graph of  $G^*$  is a tree. We first construct a 4-leaf root  $R$  for  $G^*$ , similar to the proof of Theorem 5.7. Let  $B$  be a block of  $G^*$ .

If  $B$  is a clique with vertices  $x_1, \dots, x_k$ ,  $k \geq 2$ , let  $R_B$  be the 4-leaf root for  $B$  obtained from a star on  $k + 1$  vertices with leaves  $x_1, \dots, x_k$  by subdividing each edge with one new vertex.

If  $B$  is not a clique, we construct a 4-leaf root  $R_B$  for  $B$  in each of the three cases in (ii) as follows.

(1)  $B = K \circledast H$ , where  $K$  is a clique with at least two vertices and  $H = H_1 \cup H_2$  is the disjoint union of two nonempty cliques  $H_1, H_2$ . Let  $K' \subseteq K$  be such that  $K' = \emptyset$  if  $K$  contains no cut vertex of  $G^*$ ; otherwise  $K' = \{x\}$ , where  $x$  is an arbitrary cut vertex of  $G^*$  belonging to  $K$ . Then  $R_B$  is as depicted in Figure 5 (left).

It is important to note that, by the construction and by the cut vertex condition, the 4-leaf root  $R_B$  for  $B$  has the property that for all cut vertices  $x$  of  $G^*$  belonging to  $B$ ,  $d_{R_B}(x, v) \geq 3$  for all  $v \in B \setminus \{x\}$ .

(2)  $B = K \circledast H$ , where  $K$  is a nonempty clique and  $H$  is a plump spider obtained from a thin spider  $(U, Q, R)$  with  $|U| = |Q| \geq 2$ . Write  $U = \{u_1, \dots, u_k\}$ ,

$Q = \{q_1, \dots, q_k\}$  such that  $u_i$  and  $q_j$  are adjacent if and only if  $i = j$ . We use  $U_i$ , respectively,  $Q_i$  to denote the nonempty cliques in  $H$  that stem from the vertex  $u_i \in U$ , respectively,  $q_i \in Q$ . We use  $R$  to denote the clique in  $H$  substituted for the vertex of  $R$  (if  $R \neq \emptyset$ ); thus, the clique  $R$  in  $H$  may be empty. Then  $R_B$  is as depicted in Figure 5 (right).

Note that the maximal nontrivial internal modules in  $H$  are exactly the  $Q_i$ , the construction of  $R_B$ , and the cut vertex condition in this case ensuring that for all cut vertices  $x$  of  $G^*$  belonging to  $B$ ,  $d_{R_B}(x, w) \geq 3$  for all  $w \in B \setminus \{x\}$ .

(3)  $\overline{B}$  is connected and  $B^*$  is the square of some tree. In particular,  $B$  is a nontrivial pst, and by Lemma 6.6,  $B^* = T^2$ , where  $T$  is a (basic) tree such that  $B$  is obtained from  $B^*$  by substituting cliques  $K_v$  into vertices  $v \in V(T)$ . Moreover, every maximal nontrivial module in  $B$  stems from a  $K_v$ . Now, the 4-leaf  $R_B$  root for  $B$  can be obtained from  $T$  as follows: For each maximal internal module  $M = K_v$  in  $B$  attach the set of leaves  $M$  to the vertex  $v$ . For each maximal leaf module  $M = K_v$  in  $B$ , replace first the leaf  $v$  by the set of leaves  $\{v_m : m \in M\}$ , then attach leaf  $m \in M$  to each vertex  $v_m$ . Note again, that by the cut vertex condition in this case,  $d_{R_B}(x, v) \geq 3$  for all cut vertices  $x$  of  $G^*$  in  $B$  and all  $v \in B \setminus \{x\}$ .

Now, a 4-leaf root  $R$  for  $G^*$  can be obtained from all  $R_B$  by identifying, for each cut vertex  $x$  of  $G^*$ , the parent of  $x$  in  $R_B$  and deleting all but one copy of leaf  $x$ . Recall that the cut vertex conditions in (1), (2), and (3) of (ii), and the construction of 4-leaf roots  $R_B$ 's guarantee that  $d_{R_B}(x, v) \geq 3$  for all cut vertices  $x$  of  $G^*$  belonging to  $B$  and for all  $v \in B \setminus \{x\}$ . Moreover, if  $B$  is a clique,  $d_{R_B}(x, y) = 4$  for all distinct vertices  $x, y$  in  $B$ . Using these facts, it can be seen (as in the proof of Theorem 5.7) that  $R$  is a 4-leaf root of  $G^*$ .

Finally, a 4-leaf root of  $G$  can be obtained from  $R$  by replacing each leaf  $m$  of  $R$  by the set of leaves  $M$ , where  $m$  is the vertex of  $G^*$  corresponding to the maximal module  $M$  of  $G$  that is a clique.  $\square$

## 7. Recognizing 4-Leaf Powers in Linear Time

Let  $G$  be the input graph. Given Proposition 3.1, it is sufficient to test if each connected component of  $G$  is a 4-leaf power. Hence, we assume that  $G$  is connected. Moreover, if  $G$  is a clique, the problem becomes trivial.

---

### Algorithm. 4-Leaf-Power

---

**Input:** Connected graph  $G$  different from a clique.  
**Output:** When  $G$  is a 4-leaf power, a 4-leaf root for  $G$ , and “no” otherwise.

1. Let  $M_1, \dots, M_p$  be the maximal modules of  $G$  that are cliques
2. Let  $G^*$  be the graph obtained from  $G$  by shrinking each  $M_i$  to vertex  $m_i$
3. Let  $B_1, \dots, B_q$  be the blocks of  $G^*$
4. **for** nondisjoint and distinct blocks  $B_i, B_j$  **do**
  - if** none of  $B_i, B_j$  is a clique **then**
    - Output “no”, and Stop
  - endif**
- endfor**
5. **for** each  $B_i$  **do**
  - if**  $B_i$  is a clique **then**
    - 5.1. Construct 4-leaf root  $T_i$  for  $B_i$  as in the proof of Theorem 6.9
  - else**

- 5.2.            Test whether  $B_i$  is as in (1), or (2), or (3) in Theorem 6.9  
                   If so, construct a 4-leaf root  $T_i$  for  $B_i$  as in the proof of  
                   Theorem 6.9  
                   If not, output “no”, and Stop  
                   **endif**  
**endfor**  
 6. Construct a 4-leaf root  $R$  for  $G^*$  from  $T_1, \dots, T_q$  according to the proof of  
 Theorem 6.9  
 7. Replace each leaf  $m_i$  of  $R$  by the set of leaves  $M_i$  to obtain a 4-leaf root for  $G$   
 8. Output the resulting tree  
**end of 4-leaf-power.**
- 

By Theorem 6.9, algorithm 4-leaf-power correctly recognizes if the given nonclique graph  $G$  is a 4-leaf power, and if so, constructs a 4-leaf root for  $G$ .

We claim that the running time of our algorithm is linear. Recall that all maximal modules of  $G$ , all blocks and cut vertices of  $G^*$ , as well as recognizing tree squares and determining tree roots of tree squares, can be computed in linear time; it remains to consider the time bound in step 5.2. In other words, it remains to consider how to test condition (2) in Theorem 6.9 quickly.

To do this, we first give several characterizations for plump spiders. A *gem* consists of a  $P_4$  and a vertex adjacent to all vertices in the  $P_4$ . Recall that  $K_{1,3}$  is also called a claw.

LEMMA 7.1. *Let  $G$  be a connected graph different from a clique. Then the following conditions are equivalent:*

- (i)  $G$  is a plump spider;
- (ii)  $G$  is a  $(P_5, \text{gem}, \text{claw})$ -free chordal graph;
- (iii)  $G^*$  is a thin spider; and
- (iv)  $K_1 \textcircled{1} G^*$  is the square of a tree.

PROOF. By definition, a thin spider different from a clique has no nontrivial maximal modules that are cliques, and a plump spider is connected, chordal, and cannot have an induced  $P_5$ , gem, or claw. Thus, the implication (i)  $\Rightarrow$  (ii) and the equivalence (i)  $\Leftrightarrow$  (iii) are obvious.

(ii)  $\Rightarrow$  (iii): Let  $G$  be a connected  $(P_5, \text{gem}, \text{claw})$ -free chordal graph. If  $G$  is  $P_4$ -free,  $G$  contains an induced  $P_3$   $P$ , and it is easy to see that  $G$  is obtained from the thin spider  $P$  by substituting vertices by cliques. Assume  $G$  contains an induced  $P_4$   $P = v_1v_2v_3v_4$ . For every subset  $S \subseteq \{1, 2, 3, 4\}$  let  $M_S$  be the set of vertices outside  $P$  adjacent exactly to  $v_i$ ,  $i \in S$ . For  $X \subset V(G)$  and  $v \in V(G)$ , we write shortly  $X + v$  for  $X \cup \{v\}$ . Then  $M_{12} + v_1$ ,  $M_{34} + v_4$ ,  $M_{123} + v_2$ ,  $M_{234} + v_3$  are cliques (otherwise  $G$  would have a  $C_4$  or a claw) and maximal modules in  $G$  (otherwise  $G$  would have a  $P_5$ , a claw, a gem, or a  $C_5$ ). In particular,  $(M_{12} + v_1) \textcircled{1} (M_{123} + v_2)$ ,  $(M_{34} + v_4) \textcircled{1} (M_{234} + v_3)$ ,  $(M_{123} + v_2) \textcircled{1} (M_{234} + v_3)$ ,  $M_{23} \textcircled{1} (M_{123} + v_2)$ ,  $M_{23} \textcircled{1} (M_{234} + v_3)$ ,  $M_{23} \textcircled{0} (M_{12} + v_1)$ , and  $M_{23} \textcircled{0} (M_{34} + v_4)$ . Moreover,  $M_{23}$  is a clique (otherwise  $G$  would have a claw), and each vertex in  $M_\emptyset$  has only neighbors in  $M_{23}$  (otherwise  $G$  would have a  $P_5$  or a claw). Furthermore, each connected component of  $G[M_\emptyset]$  is a maximal module (otherwise  $G$  would have a  $P_5$ ) and hence a clique (otherwise  $G$  would have a claw), and every two distinct connected components of  $G[M_\emptyset]$  have disjoint neighborhoods in  $M_{23}$  (otherwise

$G$  would have a claw). Other sets  $M_S$  not mentioned here are empty (as  $G$  is  $P_5$ -, gem-, and claw-free chordal).

It follows from these facts that  $G^*$  is a thin spider  $(U, Q, R)$ , where  $U$  consists of  $v_1, v_4$  and a vertex  $a \in A$  of each connected component  $A$  of  $G[M_\emptyset]$ ,  $Q$  consists of  $v_2, v_3$  and a vertex  $b \in N(a) \cap M_{23}$  of each vertex  $a \in U \setminus \{v_1, v_4\}$ , and  $R$  consists of a vertex in  $M_{23} \setminus (\bigcup_{a \in U \setminus \{v_1, v_4\}} N(a))$  (if  $M_{23} \setminus (\bigcup_{a \in U \setminus \{v_1, v_4\}} N(a)) \neq \emptyset$ ) or  $R = \emptyset$  (otherwise).

(iii)  $\Rightarrow$  (iv): Let  $G^* = (U, Q, R)$  be a thin spider with  $U = \{u_1, \dots, u_k\}$ ,  $Q = \{q_1, \dots, q_k\}$ , such that  $q_i$  is the neighbor of  $u_i$ . Let  $T$  be the tree on vertex set  $\{x\} \cup U \cup Q \cup R$  and edges  $xq_i, q_iu_i, i = 1, \dots, k$ , and  $xr$  (if  $R = \{r\}$ ). Then obviously,  $K_1 \textcircled{1} G^* = T^2$  where  $K_1$  is the clique consisting of the vertex  $x$ .

(iv)  $\Rightarrow$  (iii): If  $K_1 \textcircled{1} G^* = T^2$  for some tree  $T$ , then  $T$  is  $P_6$ -free (otherwise  $T^2$  cannot have a universal vertex; note that  $K_1$  consists of a universal vertex). If  $T$  is  $P_5$ -free,  $T$  is the  $P_4$  and  $G^*$  must be a  $P_3$ . If  $T$  has a  $P_5$ ,  $T^2$  has exactly one universal vertex  $x$  that is the midpoint of a  $P_5$  in  $T$ , and  $G^*$  is the thin spider  $(U, Q, R)$  as pointed out in the proof of Lemma 6.6.  $\square$

**COROLLARY 7.2.** *Plump spiders can be recognized in linear time. Further, for given plump spider  $G$ , a corresponding thin spider  $(U, Q, R)$  of  $G$  can be constructed in linear time.*

**PROOF.** If the input graph  $G$  is a clique, it is a plump spider of the thin spider  $(K_1, K_1, \emptyset)$ . Otherwise, we have to check whether  $K_1 \textcircled{1} G^*$  is the square of a tree, and this is an easy task. Further, if  $G$  is indeed a plump spider, a corresponding thin spider  $(U, Q, R)$  can be found quickly as in the proof (iv)  $\Rightarrow$  (iii) of the Lemma 7.1 before.  $\square$

Now, given a graph  $G$  different from a clique, we can check in linear time whether  $G$  is as in condition (2) of Theorem 6.9 as follows. First, check in linear time whether  $G$  is chordal and whether  $\overline{G}$  is disconnected. If not,  $G$  cannot be of the type as in condition (2). Next, compute the set  $K$  of all isolated vertices in  $\overline{G}$ , and set  $H := G - K$ . Note that  $K$  is a nonempty clique in  $G$  because, at this step,  $G$  is chordal and  $\overline{G}$  is disconnected. Now,  $G$  is of the type as in condition (2) if and only if  $H$  is a plump spider containing an induced  $P_4$ . The latter is equivalent to the fact that  $K_1 \textcircled{1} H^* = T^2$  for some tree  $T$  containing a  $P_5$  but no  $P_6$  (refer to the proofs of Lemmas 7.1 and 6.6), and can be done in linear time by Corollary 7.2.

In summary, we have the following.

**THEOREM 7.3.** *Given graph  $G$ , whether  $G$  is a 4-leaf power can be tested in  $\mathcal{O}(m + n)$  time. Further, when  $G$  is a 4-leaf power, a 4-leaf root for  $G$  can be constructed in  $\mathcal{O}(m + n)$  time.*

## 8. Concluding Remarks

In this article we introduce the concept of basic  $k$ -leaf powers and show their relation to the classical concept of the  $k$ th power of a tree. In case of basic 4-leaf powers, we are able to describe them (Theorem 5.7) in terms of squares of trees and certain gluing conditions between the blocks. Based on this, we give a structural characterization of all 4-leaf powers (Theorem 6.9), which implies a simple linear-time recognition algorithm, improving the previous  $\mathcal{O}(n^3)$  algorithm.

We also give two new characterizations of tree squares (Theorem 4.3); one of them is in terms of forbidden induced subgraphs, and the other via the structure of the maximal cliques. The latter implies a new and simple linear-time recognition (and computation of the tree root) for the square of a tree.

We believe that the notion of basic  $k$ -leaf powers will play an important role in characterizing and recognizing  $k$ -leaf powers for  $k \geq 5$ .

ACKNOWLEDGMENT. We are grateful to the anonymous referees for their careful reading and helpful suggestions.

## REFERENCES

- BRANDSTÄDT, A., AND LE, V. B. 2006. Structure and linear time recognition of 3-leaf powers. *Inf. Proc. Lett.* 98, 133–138.
- BRANDSTÄDT, A., LE, V. B., AND SPINRAD, J. P. 1999. *Graph Classes: A Survey*. SIAM Monographs on Discrete Mathematics Applications, vol. 3, SIAM, Philadelphia, PA.
- CHEN, Z.-Z., JIANG, T., AND LIN, G. 2003. Computing phylogenetic roots with bounded degrees and errors. *SIAM J. Comput.* 32, 864–879.
- DAHLHAUS, E., AND DUCHET, P. 1987. On strongly chordal graphs. *Ars Combin.* 24 B, 23–30.
- DOM, M., GUO, J., HÜFFNER, F., AND NIEDERMEIER, R. 2005. Extending the tractability border for closest leaf powers. In *Proceedings of the 31st Workshop on Graph-Theoretic Concepts in Computer Science (WG)*. Lecture Notes in Computer Science, vol. 3787. Springer, 397–408.
- HARARY, F. 1972. *Graph Theory*. Addison-Wesley, MA.
- JAMISON, B., AND OLARIU, S. 1992. A unique tree representation for  $p_4$ -sparse graphs. *Discrete Appl. Math.* 35, 115–129.
- KEARNEY, P. E., AND CORNEIL, D. G. 1998. Tree powers. *J. Algor.* 29, 111–131.
- LAU, L. C. 2006. Bipartite roots of graphs. *ACM Trans. Algor.* 2, 178–208.
- LIN, G.-H., KEARNEY, P. E., AND JIANG, T. 2000. Phylogenetic  $k$ -root and Steiner  $k$ -root. In *Proceedings of the International Symposium on Algorithms and Computation (ISAAC)*. Lecture Notes in Computer Science, vol. 1969. Springer, 539–551.
- LIN, Y.-L., AND SKIENA, S. S. 1995. Algorithms for square roots of graphs. *SIAM J. Discrete Math.* 8, 99–118.
- LUBIW, A. 1982.  $\Gamma$ -free matrices. M.S. thesis, University of Waterloo, Canada.
- MCCONNELL, R. M., AND SPINRAD, J. 1999. Modular decomposition and transitive orientation. *Discrete Math.* 201, 189–241.
- NISHIMURA, N., RAGDE, P., AND THILIKOS, D. M. 2002. On graph powers for leaf-labeled trees. *J. Algor.* 42, 69–108.
- RAUTENBACH, D. 2006. Some remarks about leaf roots. *Discrete Math.* 306, 1456–1461.
- RAYCHAUDHURI, A. 1992. On powers of strongly chordal and circular arc graphs. *Ars Combin.* 34, 147–160.
- ROSE, D. J., TARJAN, R. E., AND LUEKER, G. S. 1976. Algorithmic aspects of vertex elimination on graph. *SIAM J. Comput.* 5, 266–283.
- ROSS, I. C., AND HARARY, F. 1960. The square of a tree. *Bell Syst. Tech. J.* 39, 641–647.
- SPINRAD, J. P. 2003. *Efficient Graph Representations*. Fields Institute Monographs, Toronto.
- TARJAN, R. E. 1972. Depth first search and linear time algorithms. *SIAM J. Comput.* 1, 146–160.
- TARJAN, R. E., AND YANNAKAKIS, M. 1984. Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM J. Comput.* 13, 566–579 (Addendum: *SIAM J. Comput.* 14 (1985) 254–255).

RECEIVED OCTOBER 2005; REVISED SEPTEMBER 2006; ACCEPTED MAY 2008