1998

On the Structure of Graphs With Few P4s

Luitpold Babel

Stephan Olariu
Old Dominion University

ELSEVIER

Discrete Applied Mathematics 84 (1998) 1–13

DISCRETE APPLIED MATHEMATICS

On the structure of graphs with few P_4s ☆

Luitpold Babel[a], Stephan Olariu*,[b]

[a] Institut für Mathematik, Technische Universität München, 80290 München, Germany
[b] Department of Computer Science, Old Dominion University, Norfolk, VA 23529, USA

Received 19 December 1996; received in revised form 15 May 1997

Abstract

We present new classes of graphs for which the isomorphism problem can be solved in polynomial time. These graphs are characterized by containing – in some local sense – only a small number of induced paths of length three. As it turns out, every such graph has a unique tree representation: the internal nodes correspond to three types of graph operations, while the leaves are basic graphs with a simple structure. The paper extends and generalizes known results about cographs, P_4-reducible graphs, and P_4-sparse graphs. © 1998 Elsevier Science B.V. All rights reserved.

1. Introduction

In recent years the study of the P_4-structure of graphs turned out to be of considerable importance. The starting point and original motivation for many investigations was the class of graphs where no induced P_4 is allowed to exist (hereinafter P_k denotes a chordless path on k vertices and $k-1$ edges). For these graphs, commonly termed *cographs*, some interesting structural results have been obtained which helped to solve efficiently many graph-theoretic problems which are hard in general (see [7] for a discussion). The study of cographs has been extended by B. Jamison and S. Olariu to graphs which contain a restricted number of paths of length three. Besides *P_4-extendible graphs* [14] and *P_4-lite graphs* [15] they studied *P_4-reducible graphs* [13], defined as those graphs where no vertex belongs to more than one P_4, and *P_4-sparse graphs* [11], which generalize both cographs and P_4-reducible graphs. A graph is P_4-sparse if no set of five vertices induces more than one P_4.

We propose to call a graph a *(q,t) graph* if no set of at most q vertices induces more than t distinct P_4s. In this sense, the cographs are precisely the $(4,0)$ graphs, the P_4-sparse graphs coincide with the $(5,1)$ graphs and P_4-lite graphs turn out to be

☆ This work was supported by NSF Grant CCR-9522093, and by ONR Grant N00014-97-1-0526.
* Corresponding author. E-mail: olariu@cs.odu.edu.

PII S0166-218X(97)00120-0

special $(7,3)$ graphs. The main contribution of this paper is to investigate the structure of $(q,q-4)$ graphs for any fixed $q \geqslant 4$.

Tree representations for special graphs are often the basis for fast solutions of algorithmic problems which are hard in general. One of the best known paradigms is the isomorphism problem whose complexity is still unknown for arbitrary graphs. Using tree representations, polynomial isomorphism tests have been obtained among others for hook-up graphs [16], transitive series parallel digraphs [17], interval graphs [5], rooted directed path graphs [3], cographs [7], P_4-extendible graphs [14] and P_4-sparse graphs [11].

We consider the concept of encoding a graph into a rooted tree whose internal nodes represent certain graph operations and whose leaves correspond to certain basic graphs. If the encoding is unique and can be obtained in polynomial time, and if the basic graphs can efficiently be tested for isomorphism then we are able to solve the isomorphism problem for two such graphs in polynomial time. We will prove that the $(q,q-4)$ graphs admit such a tree representation.

The remainder of the paper is organized as follows. In Section 2 we review the concept of p-connectedness and recall some fundamental facts. Section 3 studies minimally p-connected graphs. The results obtained are used in Section 4 to classify all p-connected $(q,q-4)$ graphs and, furthermore, to prove that $(q,q-4)$ graphs are brittle graphs for $q \leqslant 8$. Thus, as a very interesting by-product, we are provided with new classes of brittle graphs, distinct from all the previously known brittle graphs. Section 5 discusses the tree representation and an efficient isomorphism test for $(q,q-4)$ graphs. Finally, in the last section we summarize the results and pose some open problems.

2. Background and terminology

Let $G=(V,E)$ be a simple graph with vertex-set V and edge-set E. For a vertex v of G define $N(v)$ to be the set of vertices adjacent to v. A vertex of G is said to be an *articulation point* if its removal disconnects G. Given a set A of vertices of G, we let $G(A)$ denote the subgraph of G induced by A. We shall use $G-\{v\}$ as a shorthand for $G(V-\{v\})$.

A chordless path P_4 with vertices u,v,w,x and edges uv,vw,wx is denoted by $uvwx$. The vertices u and x are termed the *endpoints*, while v and w are the *midpoints* of P_4. A graph is a *clique* if its vertices are pairwise adjacent. A *stable set* denotes a set of pairwise non-adjacent vertices. For other graph-theoretic notations we refer to Golumbic [9].

In the following we shall adopt the terminology introduced by Jamison and Olariu [10]. A graph $G=(V,E)$ is *p-connected* if for every partition of V into nonempty disjoint sets A and B there exists a *crossing* P_4, that is, a P_4 containing vertices from both A and B. The *p-connected components* of a graph are the maximal induced subgraphs which are p-connected. Note that a p-connected component has either one or at least four vertices. Vertices which are not contained in a nontrivial p-connected component are also called *weak*. It is easy to see that each graph has a unique partition into p-connected components. Furthermore, the p-connected components are closed under complementation and are connected subgraphs of G and $\overline{G}$.

A p-connected graph $G=(V,E)$ is called *separable* if there exists a partition of V into nonempty disjoint sets V_1, V_2 such that each P_4 which contains vertices from both sets has its endpoints in V_2 and its midpoints in V_1. We say that (V_1, V_2) is a *separation* of G. Obviously, the complement of a separable p-connected graph is also separable. If (V_1, V_2) is a separation of G then (V_2, V_1) is a separation of $\overline{G}$. We now recall some important facts that form the basis for the results derived in this paper.

Theorem 2.1 (Jamison and Olariu [10]). *Every separable p-connected component H has a unique separation (H_1,H_2). Furthermore, every vertex of H belongs to a crossing P_4 with respect to (H_1,H_2).*

Let $G=(V,E)$ be an arbitrary graph. A set Z of vertices of G is called *homogeneous* if $1<|Z|<|V|$ and each vertex outside Z is either adjacent to all vertices of Z or to none of them. A homogeneous set Z is *maximal* if no other homogeneous set properly contains Z. Let H be a p-connected component. The graph obtained from H by replacing every maximal homogeneous set by one single vertex is called *characteristic* p-connected component of H. Recall that a graph is a *split graph* if its vertex-set can be partitioned into a clique and a stable set.

Theorem 2.2 (Jamison and Olariu [10]). *A p-connected component H is separable if and only if the characteristic p-connected component of H is a split graph.*

The introduction and study of separable p-connected graphs is justified by the following general structure theorem for arbitrary graphs.

Theorem 2.3 (Jamison and Olariu [10]). *Let $G=(V,E)$ be a graph. Exactly one of the following statements holds*:

(i) *G is disconnected.*
(ii) *$\overline{G}$ is disconnected.*
(iii) *There exists a unique proper separable p-connected component H with separation (H_1,H_2) such that every vertex outside H is adjacent to all vertices in H_1 and to no vertex in H_2.*
(iv) *G is p-connected.*

As already pointed out in [10], this structure theorem suggests, in a natural way, a tree representation for every graph G. The leaves of the tree correspond to the p-connected components of G. If these subgraphs have a simple structure then we may hope to solve the isomorphism problem in polynomial time. This observation motivates a further study of p-connected graphs. As a first step in this direction, in the next section of this work, we shall look at graphs that are critical in the sense of p-connectedness.

3. Minimally p-connected graphs

A graph $G=(V,E)$ is *minimally p-connected* if G is p-connected and, for every vertex v of $G, G-\{v\}$ is not p-connected. Following the notation in [11] a p-connected graph $G=(V,E)$ is called a *spider* if V admits a partition into disjoint sets S and K such that:

(i) $|S|=|K|\geqslant 2$, S is stable, K is a clique;
(ii) There exists a bijection $f:S\to K$ such that either

$$N(s)=\{f(s)\} \quad \text{for all vertices } s \text{ in } S,$$

or else

$$N(s)=K-\{f(s)\} \quad \text{for all vertices } s \text{ in } S.$$

If the first of the two alternatives of (ii) holds then G is said to be a spider with *thin* legs, otherwise the spider has *thick* legs (see Fig. 1). As a technicality, a P_4 is considered to be a spider with thin legs. Obviously, the complement of a spider with thin legs is a spider with thick legs and vice versa. The main goal of this section is to prove that each minimally p-connected graph is a spider. Our first result shows that no minimally p-connected graph contains a homogeneous set.

Lemma 3.1. *Let $G=(V,E)$ be a p-connected graph and let Z be a homogeneous set in G. Then, for every vertex v in Z, $G-\{v\}$ is p-connected.*

Proof. Since G is p-connected there is a P_4 containing vertices from both Z and $V-Z$. This P_4 contains exactly one vertex from Z, say u. If u is replaced by any other vertex w from Z then we again get a P_4.

Assume that $G^*=G-\{v\}$ is not p-connected. Then there is a partition A,B of the vertex set $V^*=V-\{v\}$ of G^* without a crossing P_4. Let $Z^*=Z-\{v\}$. Z^* is

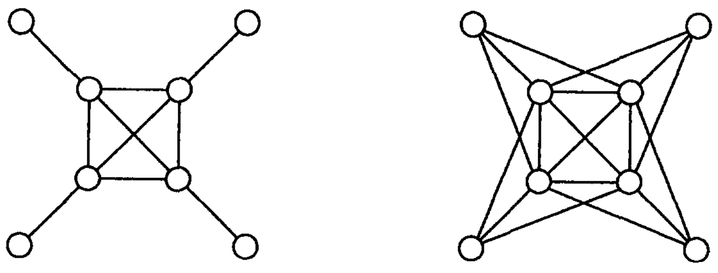

Fig. 1. The spiders with eight vertices.

a subset of one of the sets A, B. This can be seen as follows. Let $Z^* \cap A \neq \emptyset$ and $Z^* \cap B \neq \emptyset$. Take a P_4 with vertices from both Z^* and $V^* - Z^*$ (the existence follows from the above observation). This P_4 is contained in one of the sets A or B, say A. Replace the vertex from $Z^* \cap A$ by a vertex from $Z^* \cap B$. Then we get a crossing P_4, a contradiction. Therefore, let without loss of generality $Z^* \subseteq A$. In G there exists a P_4 containing vertices from both $A \cup \{v\}$ and B. This P_4 contains v but no vertex from Z^*. If v is replaced by any vertex from Z^* then we obtain a new P_4 which is crossing between A and B, contrary to the assumption. □

Let G be p-connected and $G^* = G - \{v\}$ not p-connected. By Theorem 2.3 exactly one of the following statements is true:

(i) G^* is disconnected, i.e. v is an articulation point in G.

(ii) $\overline{G^*}$ is disconnected, i.e. v is an articulation point in $\overline{G}$.

(iii) There is a unique proper separable p-connected component H of G^* with separation (H_1, H_2) such that every vertex outside H is adjacent to all vertices in H_1 and to no vertex in H_2.

According to the different cases we call the vertex v to be of *type* 1, 2 or 3.

Lemma 3.2. *Let $G = (V, E)$ be p-connected. If each vertex of G is of type 1 or 2 then G is a P_4.*

Proof. A connected graph has at most $|V| - 2$ articulation points. Therefore, G contains vertices of both types. In particular, since $|V| \geqslant 4$ there exist at least two vertices which are articulation points in $\overline{G}$. Furthermore, since G is connected there are vertices of different type, say x of type 1 and y of type 2, with $xy \in E$.

Suppose first that $|N(y)| > 1$.

Denote $G(U_1), G(U_2), \ldots, G(U_r)$ the components of $G - \{x\}$ and let $y \in U_1$. Note that under the above assumption we have $U_1 - \{y\} \neq \emptyset$ and $r \geqslant 2$. Since there is no edge in G connecting vertices from different sets $U_1 - \{y\}, U_2, \ldots, U_r$ we conclude that $\overline{G} - \{x, y\}$ is connected. Now let $G(W_1), G(W_2), \ldots,$ be the components of $\overline{G} - \{y\}$. Then we get $W_1 = \{x\}$ and $W_2 = V - \{x, y\}$. This means that x is adjacent to all other vertices in G. However, then there is no P_4 containing x and this contradicts to the fact that G is p-connected. Therefore $|N(y)| = 1$.

Since there exist at least two articulation points in $\overline{G}$ and since G is connected, there is a second vertex y' of type 2 which is adjacent to a vertex x' of type 1. Analogously as above we conclude that $|N(y')| = 1$. Thus we have $N(y) = \{x\}$ and $N(y') = \{x'\}$. Again denote $G(W_1), G(W_2), \ldots$ the components of $\overline{G} - \{y\}$. Since $|N(y')| = 1$ we have $W_1 = \{x'\}$ and $W_2 = V - \{x', y\}$. If $x = x'$ then x would be adjacent to all other vertices in G. This is not possible since G is p-connected. Therefore $x \neq x'$. $x' \in W_1$ and $x \in W_2$ implies $xx' \in E$. Therefore, the vertex set $\{y, x, x', y'\}$ induces a P_4. Each further vertex w is adjacent to x' and also to x (exchange the parts of y and y'), thus exactly to the midpoints of the P_4. As a consequence, there is no crossing P_4 between $\{y, x, x', y'\}$ and the remaining vertices. Therefore no such vertex w exists. This proves the lemma. □

Lemma 3.2 implies that each nontrivial minimally p-connected graph contains a vertex of type 3. If v is of type 3 then we write $H(v)$ for the separable p-connected component and $(H_1(v), H_2(v))$ for the separation. Further we denote $R(v)$ to be the vertices of G^* outside $H(v)$.

Lemma 3.3. *Let* $G=(V,E)$ *be minimally p-connected and let* $x \in V$ *be a vertex of type* 3 *with* $|R(x)|$ *minimal. Then* $|R(x)|=1$.

Proof. Assume that $|R(x)| \geqslant 2$. By virtue of Lemma 3.1, G contains no homogeneous set. Therefore, x is adjacent to some but not to all vertices in $R(x)$. Consequently, we find vertices u and u' in $R(x)$ with $xu \in E$ and $xu' \notin E$.

We consider vertex u and examine the possible types of u:

(i) Assume that u is of type 1, i.e. u is an articulation point in G. Since $G - \{u,x\}$ is connected we conclude that $N(x)=\{u\}$. Obviously, u' is not an articulation point in G and not in $\overline{G}$. Thus, u' is of type 3. x can neither be in $R(u')$ nor in $H_1(u')$ since each vertex from this two sets is adjacent to at least two vertices. Thus $x \in H_2(u')$ and as an immediate consequence $u \in H_1(u')$. Since both $H(x)$ and $H(u')$ are p-connected, we easily see that $H(x) \subset H(u')$. However, now $|R(u')| < |R(x)|$, contradicting the choice of x.

(ii) Assume that u is of type 2, i.e. u is an articulation point in $\overline{G}$. Since $\overline{G} - \{u,x\}$ is connected this would imply $N(x)=V - \{x\}$. However, this is not possible since $xu' \notin E$.

(iii) Assume that u is of type 3. Since $H(x)$ and $H(u)$ are p-connected, either $H(x) \subseteq H(u)$ or $H(x) \subseteq R(u)$ holds. The second case is not possible since some edges between $R(u)$ and $H_1(u)$ would be missing (take vertices $v \in H_2(x) \cap R(u)$ and $w \in R(x) \cap H_1(u)$, then $vw \notin E$).

Therefore $H(x) \subseteq H(u)$. Since, due to the choice of x, $|R(u)| \geqslant |R(x)|$ must hold, we conclude that $H(u)=H(x)$ and, due to the uniqueness of the separation (Theorem 2.1) $(H_1(u), H_2(u)) = (H_1(x), H_2(x))$. However, since we know from above that u is adjacent to all vertices in $H_1(x)$ and to none in $H_2(x)$, this would imply a homogeneous set $R(u) \cup \{u\}$, a contradiction.

This shows that the assumption $|R(x)| \geqslant 2$ is not correct. □

Lemma 3.4. *Let* $G=(V,E)$ *be minimally p-connected and let* $x \in V$ *be a vertex of type* 3 *with* $R(x)=\{v\}$. *Then* $N(x)=R(x)$ *or* $N(x)=H_1(x) \cup H_2(x)$.

Proof. Assume first that $xv \in E$. We distinguish the possible types for v. If v is of type 2, i.e. an articulation point in $\overline{G}$ then $N(x)=V - \{x\}$. This is not possible since no P_4 would exist containing x in contradiction to the p-connectedness of G. If v is of type 3 then obviously $R(v)=\{x\}$ and therefore $N(x)=\{v\} \cup H_1(x)$. Thus $\{v,x\}$ would be a homogeneous set. Therefore v is of type 1, i.e. articulation point in G and $N(x)=\{v\}$. This shows the first part of the statement.

For the second part assume that $xv \notin E$. If v is of type 1 then $N(x)=\emptyset$ which is not possible since G is connected. If v is of type 3 then $R(v)=\{x\}$ and therefore $N(x)=H_1(x)$. Again $\{v,x\}$ would be a homogeneous set. Therefore v is of type 2 and $N(x)=H_1(x)\cup H_2(x)$. □

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

Theorem 3.5. *Every minimally p-connected graph is a spider.*

Proof. If G contains no vertex of type 3 then, by Lemma 3.2, G is a P_4 and therefore a spider. Let x be a vertex of type 3 with $|R(x)|$ as small as possible. By virtue of Lemmas 3.3 and 3.4, we have $R(x)=\{v\}$ and $N(x)=R(x)$ or $N(x)=H_1(x)\cup H_2(x)$. It suffices to consider the case $N(x)=R(x)$, the second case being handled similarly.

Note that, if Z is a homogeneous set in the subgraph $H(x)$ then $Z\subseteq H_1(x)$ or $Z\subseteq H_2(x)$. This can be seen as follows. Assume that $Z\cap H_i(x)\neq\emptyset$ for $i=1,2$. Take a P_4 with vertices from both Z and $H(x)-Z$. Since Z is homogeneous, this P_4 contains exactly one vertex from Z, say z. As we have already seen, z may be replaced by any other vertex from Z to form another P_4. If $z\in H_1(x)$ then replace z by a vertex $z'\in Z\cap H_2(x)$, if $z\in H_2(x)$ then by a vertex $z''\in Z\cap H_1(x)$. It is immediately clear that a P_4 results which is crossing between $H_1(x)$ and $H_2(x)$ and whose midpoints or endpoints are not both in $H_1(x)$ or $H_2(x)$.

We can conclude that Z is also homogeneous in G. However, Lemma 3.1 implies that G contains no homogeneous set. Therefore, no such set Z exists. Using Theorem 2.2 we conclude that $G(H_1(x)\cup H_2(x))$ is a split graph. For convenience denote K the vertex set of the clique induced by $H_1(x)$ and S the stable set $H_2(x)$. Note that each vertex of G is contained in a P_4 $xvks$ with $k\in K$ and $s\in S$.

Let $s'\in S$ with $N(s')=\{k'\}$. If $|N(k')\cap S|\geqslant 2$ then each vertex of $G-\{s'\}$ is contained in a path $xvks$ with $s\neq s'$, thus $G-\{s'\}$ would be p-connected, contradicting the minimality of G. Therefore $|N(k')\cap S|=1$. Analogously, let $k''\in K$ with $N(k'')\cap S=\{s''\}$. Then $|N(s'')|=1$, otherwise $G-\{k''\}$ would be p-connected. Clearly, the vertices $k'\in K$ and $s'\in S$ with $|N(k')\cap S|=1$ and $|N(s')|=1$ together with x and v induce a spider with thin legs.

For all further vertices $k'''\in K$ and $s'''\in S$ which are not in the spider $|N(k''')\cap S|\geqslant 2$ resp. $|N(s''')|\geqslant 2$ holds. Assume that any of this vertices, say s''', is deleted. For each $k'''\in K$ with $s'''\in N(k''')$ there is at least one additional vertex in S which is adjacent to k'''. Therefore each vertex of $G-\{s'''\}$ is contained in a P_4 $xvks$ with $s\neq s'''$ and $G-\{s'''\}$ remains p-connected. Consequently, no further vertices exist and the proof is complete. □

Theorem 3.5 implies the following very useful property of p-connected graphs that may be the starting point for more and deeper results concerning the structure of arbitrary graphs.

Theorem 3.6. *Let G be p-connected. Then there is an ordering $(v_n, v_{n-1}, \dots, v_1)$ of the vertices of G and an integer $k \in \{4, 5, \dots, n\}$ such that the following holds:*
$G(\{v_i, v_{i-1}, \dots, v_1\})$ is p-connected for $i = k, \dots, n$ and a spider for $i = k$.

4. On p-connected $(q, q-4)$ graphs

We start with some properties concerning minimally p-connected graphs.

Observation 4.1. In a spider each P_4 has its midpoints in the clique K and its endpoints in the stable set S, i.e. a spider is separable. For each pair $s, s' \in S$ $(k, k' \in K)$ there is exactly one P_4 containing both vertices.

Observation 4.2. A spider with $|K| = |S| = r$ contains exactly $\frac{1}{2}r(r-1)P_4s$.

Observation 4.3. If H and G are spiders with thin (thick) legs and H has fewer vertices than G, then H is isomorphic to an induced subgraph of G.

Fact 4.4. *If q is even and G is a spider with q vertices then G is not a $(q, q-4)$ graph. If q is odd, $q \geqslant 9$, and G is a spider with $q-1$ vertices then G is not a $(q, q-4)$ graph.*

Proof. Let q be even. By virtue of Observation 4.2, the spider G contains $\frac{1}{2}r(r-1)$ P_4s with $r = \frac{q}{2}$. Since $\frac{1}{8}q(q-2) > q-4$ holds, G does not satisfy the definition of a $(q, q-4)$ graph.

Let q be odd. Then $r = \frac{1}{2}(q-1)$ and G contains $\frac{1}{8}(q-1)(q-3)P_4s$. For $q \geqslant 9$ we get $\frac{1}{8}(q-1)(q-3) > q-4$. Therefore G is not a $(q, q-4)$ graph. □

The following theorem characterizes p-connected $(q, q-4)$ graphs. Part (a) already implicitly appeared in [11]. For the sake of completeness we restate it, giving, however, a completely different proof.

Theorem 4.5. *Let $G = (V, E)$ be p-connected.*
(a) *If G is a $(5, 1)$ graph then G is a spider.*
(b) *If G is a $(7, 3)$ graph then $|V| < 7$ or G is a spider.*
(c) *If G is a $(q, q-4)$ graph, $q = 6$ or $q \geqslant 8$, then $|V| < q$.*

Proof. By Theorem 3.6 there is an ordering $(v_n, \dots, v_1)$ of the vertices of G and an integer $k \in \{4, 5, \dots, n\}$ such that $G_i := G(\{v_i, v_{i-1}, \dots, v_1\})$ is p-connected for $i = k, \dots, n$ and G_k is a spider.

(a) Let G be a $(5, 1)$ graph. It can easily be verified that each spider is a $(5, 1)$ graph. Assume that $k < n$, i.e. there is a vertex v_{k+1} which is not in the spider G_k.

Let X be the vertex set of an arbitrary P_4 in G_k. There are no three vertices in X such that v_{k+1} together with these vertices induces a P_4. Otherwise $G(X \cup \{v_{k+1}\})$

would be a graph with five vertices and at least two P_4s, thus not a $(5,1)$ graph. Therefore, v_{k+1} is either adjacent to all vertices in X, to no vertex in X, or exactly to the two midpoints.

Using Observation 4.1 we conclude that v_{k+1} is either adjacent to all vertices of G_k, to none of them, or exactly to the vertices of the clique of G_k. However, in all three cases G_{k+1} is not p-connected since there is no P_4 in G_{k+1} containing v_{k+1}. This is a contradiction. Therefore, $k=n$ and G is a spider.

(b) Let G be a $(7,3)$ graph. Again, it can easily be verified that each spider is a $(7,3)$ graph. If $k=4$ then the spider G_k is a P_4. Since G_i is p-connected for $i=k,\dots,n$, adding v_{i+1} to G_i increases the number of P_4s by at least one. Since G is a $(7,3)$ graph no more than two vertices can be added. Therefore we get $|V|<7$.

Let $k>4$ and assume that $k<n$, i.e. there is a vertex v_{k+1} which is not in the spider G_k. Since G_{k+1} is p-connected there exists a P_4 in G_{k+1} containing v_{k+1}. Let $X=\{x,y,z,v_{k+1}\}$ be the vertex set of this P_4. Further let H be the spider with smallest number of vertices which is a subgraph of G_k and which contains x, y and z. Obviously, H has four or six vertices. In the first case extend H to a spider with six vertices. Now adding v_{k+1} to H results in a graph with seven vertices and at least four P_4s. This is a contradiction. Therefore we have $k=n$ and G is a spider.

(c) Let G be a $(q,q-4)$ graph with $q=6$ or $q\geqslant 8$. We know from Observation 4.3 and Fact 4.4 that $k<q$, i.e. the spider G_k has less than q vertices. By Observation 4.2 G_k contains exactly $\frac{1}{8}k(k-2)$ P_4s. Since G_i is p-connected for $i=k,\dots,n$, adding v_{i+1} to G_i strictly increases the number of P_4s. Therefore, G_i contains at least $\frac{1}{8}k(k-2)+(i-k)$ P_4s.

Assume that G has at least q vertices, i.e. $n\geqslant q$. This would imply that the number of P_4s which are contained in the graph G_q is at least

$$\tfrac{1}{8}k(k-2)+(q-k)=q+\tfrac{1}{8}k(k-10)\geqslant q-3>q-4.$$

As a consequence, G_q would not be a $(q,q-4)$ graph, a contradiction. Therefore we have $|V|<q$.

This completes the proof. □

This characterization can be used to derive interesting properties of $(q,q-4)$ graphs. A graph G is called *brittle* if each induced subgraph H of G contains a vertex which is either not the endpoint or not the midpoint of any P_4 in H. It is well known that brittle graphs are perfectly orderable. A graph G is *perfectly orderable* in the sense of Chvatal [6] if there exists a linear order on the set of vertices of G such that no induced path with vertices u,v,w,x and edges uv, vw, wx has $u<v$ and $x<w$. The importance of perfectly orderable graphs stems from the fact that these are precisely the graphs for which the coloring heuristic “always use the first available color” based on the linear order yields a coloring using the minimum number of colors. Chvatal has shown that perfectly orderable graphs are perfect.

It is easy to see that $(q,q-4)$ graphs, $q\geqslant 9$, are not brittle and not even perfect since the induced cycle of length five belongs to these classes. On the other side the following holds.

Theorem 4.6. *Every $(q,q-4)$ graph, $4 \leqslant q \leqslant 8$, is brittle.*

Proof. If a vertex v is not endpoint (midpoint) of any P_4 in a p-connected component of G then v is not endpoint (midpoint) of any P_4 in G. Therefore, it suffices to prove that p-connected $(q,q-4)$ graphs, $4 \leqslant q \leqslant 8$, are brittle.

Let $q=8$ and $G=(V,E)$ be a p-connected $(8,4)$ graph with maximal number of vertices, i.e. $|V|=7$. Further let $(v_7, v_6, \ldots, v_1)$ be an ordering of the vertices of V defined by Theorem 3.6. It is easy to see that v_7 is contained in exactly one P_4. For that reason v_7 is either not the endpoint or not the midpoint of any P_4 in G.

If we have at most six vertices, the conclusion follows by an exhaustive search. For $q \leqslant 7$ use Obsevation 4.1 to see that spiders are brittle. Then, as above, an exhaustive search should convince the reader that $(q,q-4)$ graphs, $q \leqslant 7$, with no more than six vertices are brittle. □

5. The tree structure of $(q, q-4)$ graphs

Theorem 2.3 enables us to give for any graph a tree representation. The tree associated with a graph G carries labels on the interior nodes and is constructed by the obvious recursive procedure. The labels correspond to the cases in the theorem. Thus, label (1) indicates that the graph associated with this node as a root is the disjoint union of the graphs defined by its children. Label (2) defines the operation which we will call disjoint sum. All pairs of vertices belonging to different children are linked by an edge. Operation (3) adjoins the midpoints of the leftmost son – which has to represent a separable p-connected component – to all vertices of its other children. The leaves of the tree represent the p-connected components of the graph G along with its weak vertices.

It is well known that each cograph arises from single vertices by a sequence of operations disjoint union and disjoint sum. Thus, in this special case the leaves of the tree represent vertices and the labels of the interior nodes are (1) and (2).

Let $\mathscr{G}(q,t)$ denote the set of all (q,t) graphs. In particular, $\mathscr{G}(4,0)$ corresponds to the set of cographs, $\mathscr{G}(5,1)$ to the set of P_4-sparse graphs. The following theorem reflects the containment relations between the different classes.

Theorem 5.1. (a) $\mathscr{G}(4,0) \subset \mathscr{G}(5,1)$, $\mathscr{G}(6,2) \subset \mathscr{G}(7,3)$.
(b) $\mathscr{G}(6,2) \subset \mathscr{G}(q,q-4) \subset \mathscr{G}(q+1,q-3)$ *for* $q \geqslant 8$. *All inclusions are strict.*

Proof. It is clear from the tree representation that it suffices to consider the p-connected components of the graphs. With this in mind all inclusions can immediately be deduced from Theorem 4.5.

Examples to confirm the strict inclusions are in case (a) the P_4 respectively the graph consisting of a P_4 $uvwx$ extended by two vertices y,z which are adjacent to w. In case (b) take the path P_6 with 6 vertices for the first and the path P_q with q vertices for the second inclusion. The classes $\mathscr{G}(5,1)$ and $\mathscr{G}(6,2)$ are not comparable (take the path P_5 respectively a spider with 6 vertices). □

As already indicated in Section 1 it is known from [13] that *P_4-reducible graphs* belong to the class $\mathscr{G}(5,1)$. We would like to mention another interesting set of graphs. A graph G is called *P_4-lite* [15] if every induced subgraph of G with at most six vertices either contains at most two P_4s or is isomorphic to a spider with six vertices. It is an easy observation that P_4-lite graphs are a proper superclass of $\mathscr{G}(5,1)$ and $\mathscr{G}(6,2)$ and a proper subclass of $\mathscr{G}(7,3)$. Up to now no polynomial isomorphism test for P_4-lite graphs was known.

It follows immediately from Theorem 2.3 that for any graph G the tree representation given above is unique up to isomorphism. It is known from [10] that it can be obtained in time polynomial in the number of vertices in G. Note that in our special case of $(q,q-4)$ graphs the nontrivial leaves of the tree represent
- spiders if $q=5$;
- graphs with less than seven vertices or spiders if $q=7$;
- graphs with less than q vertices if $q=6$ or $q\geqslant 8$.

With this information we are able to give an efficient isomorphism test. Here is an informal description. The algorithm tests whether two $(q,q-4)$ graphs are isomorphic or not. In the positive case, it stops in state "true", otherwise in state "false".

Algorithm ISOMORPH($G_1, G_2, Boole$)

Input: Two $(q,q-4)$ graphs G_1, G_2.

Output: A boolean variable *Boole*, which is true or false depending on whether G_1 and G_2 are isomorphic.

Step 1: Construct the representing trees T_1, T_2 for G_1 and G_2.

Step 2: Test all pairs of graphs corresponding to leaves in T_1 and T_2 for isomorphism and assign two leaves the same label if and only if the corresponding graphs are isomorphic. As a result we obtain two labeled trees T_1^*, T_2^* (with integer labels on the internal nodes and on the leaves).

Step 3: Perform a labeled tree isomorphism test for T_1^* and T_2^*. If T_1^* is isomorphic to T_2^* then set $Boole := true$ else set $Boole := false$.

The correctness of the algorithm is obvious. It is well known that labeled tree isomorphism can be tested in time linear in the number of vertices of the tree (see e.g. [1]). Therefore, it remains to ensure that the task of transforming the trees of G_1, G_2 into labeled trees can be done in polynomial time.

The crucial point is that the subgraphs associated with the leaves are very simple. If the number of vertices is restricted by the constant q then isomorphism testing for each pair of subgraphs requires only constant time. If the subgraphs are spiders then isomorphism testing can be done in time linear in the size of the spiders (note that the stable set of the spider consists of all vertices with minimal number of neighbors). These considerations imply the following statement.

Theorem 5.2. *For every fixed q the isomorphism of $(q,q-4)$ graphs can be tested in polynomial time.*

6. Conclusions and open problems

In this work we proved that, for any fixed $q \geqslant 4$, $(q,q-4)$ graphs admit a tree representation which enables a polynomial isomorphism test. This generalizes known results about cographs, P_4-reducible graphs and P_4-sparse graphs.

It is an open question whether a tree representation for arbitrary graphs can be found in time linear in the size of the graph. If this is true then it would immediately imply a linear isomorphism test and also a linear recognition algorithm for $(q,q-4)$ graphs (essentially, we have to check the leaves of the representing tree for membership in the class $\mathscr{G}(q,q-4)$). Note that the naive method "examine all subsets $U \subseteq V$ of cardinality q and count the P_4s in $G(U)$" shows that the recognition problem is polynomial. Both the isomorphism and the recognition problem are known to be solvable in linear time for cographs (see [8]) and for P_4-sparse graphs (see [12]). We conjecture that this is also possible for $(q,q-4)$ graphs with $q \geqslant 6$, using similar techniques.

Each $(q,q-4)$ graph is also a $(q,q-3)$ graph, therefore $\mathscr{G}(q,q-4) \subseteq \mathscr{G}(q,q-3)$ holds. Obviously $\mathscr{G}(4,1)$ is the set of all graphs. It is easy to see that $\mathscr{G}(5,2)$ coincides with the class of graphs which contain no induced cycle of length five. We conclude with an isomorphism completeness result (a problem is *isomorphism complete* if it is polynomial time equivalent to graph isomorphism).

Lemma 6.1. *The task of testing the isomorphism of $(q,q-3)$ graphs, $q \in \{4,5,6\}$, is isomorphism complete.*

Proof. The statement is trivial for $q=4$. For $q=5$ it follows from the fact that $\mathscr{G}(5,2)$ contains all bipartite graphs, where the isomorphism problem is known to be isomorphism complete (see [4]).

Let $q=6$. We give a polynomial reduction from the set of all graphs to the class $\mathscr{G}(6,3)$ such that two graphs are isomorphic if and only if the corresponding $(6,3)$ graphs are isomorphic. Let $G=(V,E)$ be an arbitrary graph and $v \in V$. Assume that $N(v)=\{u_1,u_2,\ldots,u_r\}$. Replace each nonisolated vertex $v \in V$ by a clique with $|N(v)|=r$ vertices, say $w_1,\ldots,w_r$, and join all r pairs u_i, w_i by an edge. Furthermore, replace each edge which connects vertices from two different such cliques by a path of length two. It is an easy task to verify that the resulting graph is a $(6,3)$ graph. $\square$

The complexity of the isomorphism problem remains unknown for the classes $\mathscr{G}(q,q-3)$, $q \geqslant 7$.

Acknowledgements

The authors would like to thank Professor Peter Hammer for his timely handling of our submission. Our thanks go also to Professor Gottfried Tinhofer for many stimulating discussions and insightful remarks.

References

[1] A.V. Aho, J.E. Hopcroft, J.D. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, MA, 1974.
[2] L. Babel, I.N. Ponomarenko, G. Tinhofer, The isomorphism problem for directed path graphs and for rooted directed path graphs, J. Algorithms 21 (1996) 542–564.
[3] L. Babel, Tree-like P_4-connected graphs, Discrete Math., to appear.
[4] K.S. Booth, C.J. Colbourn, Problems polynomially equivalent to graph isomorphism, Report No. CS-77-04, Computer Science Department, University of Waterloo, 1979.
[5] K.S. Booth, G.S. Lueker, A linear time algorithm for deciding interval graph isomorphism, J. ACM 26 (1979) 183–195.
[6] V. Chvatal, Perfectly ordered graphs, in: C. Berge, V. Chvatal (Eds.), Topics on Perfect Graphs, North-Holland, Amsterdam, 1984, pp. 63–65.
[7] D.G. Corneil, H. Lerchs, L.S. Burlingham, Complement reducible graphs, Discrete Appl. Math. 3 (1981) 163–174.
[8] D.G. Corneil, Y. Perl, L.K. Stewart, A linear recognition algorithm for cographs, SIAM J. Comput. 14 (1985) 926–934.
[9] M.C. Golumbic, Algorithmic Graph Theory and Perfect Graphs, Academic Press, New York, 1980.
[10] B. Jamison, S. Olariu, p-Components and the homogeneous decomposition of graphs, SIAM J. Discrete Math. 8 (1995) 448–463.
[11] B. Jamison, S. Olariu, A unique tree representation for P_4-sparse graphs, Discrete Appl. Math. 35 (1992) 115–129.
[12] B. Jamison, S. Olariu, Recognizing P_4-sparse graphs in linear time, SIAM J. Comput. 21 (1992) 381–406.
[13] B. Jamison, S. Olariu, P_4-reducible graphs, a class of uniquely tree representable graphs, Stud. Appl. Math. 81 (1989) 79–87.
[14] B. Jamison, S. Olariu, On a unique tree representation for P_4-extendible graphs, Discrete Appl. Math. 34 (1991) 151–164.
[15] B. Jamison, S. Olariu, A new class of brittle graphs, Stud. in Appl. Math. 81 (1989) 89–92.
[16] M.M. Klawe, M.M. Corneil, A. Proskurowski, Isomorphism testing in hook-up graphs, SIAM J. Algebraic Discrete Methods 3 (1982) 260–274.
[17] E.L. Lawler, Graphical algorithms and their complexity, Math. Center Tracts 81 (1976) 3–32.