J Comb Optim
DOI 10.1007/s10878-008-9164-y

Linear time construction of 5-phylogenetic roots for tree chordal graphs

William S. Kennedy · Hui Kong · Guohui Lin · Guiying Yan

Abstract Inspired by phylogenetic tree construction in computational biology, Lin et al. (The 11th Annual International Symposium on Algorithms and Computation (ISAAC 2000), pp. 539–551, 2000) introduced the notion of a *k-phylogenetic root*. A k-phylogenetic root of a graph G is a tree T such that the leaves of T are the vertices of G, two vertices are adjacent in G precisely if they are within distance k in T, and all non-leaf vertices of T have degree at least three. The k-phylogenetic root problem is to decide whether such a tree T exists for a given graph G. In addition to introducing this problem, Lin et al. designed linear time constructive algorithms for $k \leq 4$, while left the problem open for $k \geq 5$. In this paper, we partially fill this hole by giving a linear time constructive algorithm to decide whether a given tree chordal graph has a 5-phylogenetic root; this is the largest class of graphs known to have such a construction.

Research of W.S. Kennedy partially supported by NSERC.
Research of H. Kong partially supported by NNSF Grants 10721101 and 10531070.
Research of G. Lin partially supported by CFI and NSERC.
Research of G. Yan partially supported by NSERC and NNSF Grants 10721101 and 10531070.

W.S. Kennedy
Discrete Mathematics Group, Department of Mathematics, McGill University, Montréal, Québec H3A 2A7, Canada
e-mail: kennedy@math.mcgill.ca

H. Kong · G. Yan
Academy of Mathematics and Systems Science, Zhongguancun, Beijing 100080, China

G. Yan
e-mail: yangy@amss.ac.cn

G. Lin (✉)
Department of Computing Science, University of Alberta, Edmonton, Alberta T6G 2E8, Canada
e-mail: ghlin@cs.ualberta.ca

Keywords Graph algorithms · Phylogenetic roots · Leaf roots · Steiner roots · Chordal graphs · Linear time recognition

1 Introduction

A phylogeny describes the development and history through evolution for a set of related species. One natural representation of a phylogeny is a tree where the leaves are labeled by the species and distances in the tree represent their evolutionary closeness. Reconstruction of these phylogenetic trees is a fundamental problem in computational biology.

Motivated by phylogenetic tree construction, Nishimura et al. (2002) introduced *k-leaf roots* and *k-leaf powers*. A k-leaf power of tree T is a graph $G = (V(G), E(G))$ such that the leaves of T are the vertices of G, and two vertices are adjacent in G precisely if they are within distance k in T, measured by the number of edges on the path connecting them. Tree T is a k-leaf root of G. Given graph G, to construct a k-leaf root T we allow for the addition of internal nodes in T to facilitate the distance constraint. These nodes do not correspond to vertices of $V(G)$ and are called *Steiner points*.

Initially, Nishimura et al. (2002) had given an $O(n^3)$ time algorithm for the recognition of 3-leaf and 4-leaf powers. Brandstädt and Le (2006) and Brandstädt et al. (2006) recently, and Rautenbach (2006) independently, gave simpler characterizations for 3-leaf and 4-leaf powers. Brandstädt and Le (2006) and Brandstädt et al. (2006) also showed the use of their characterizations by producing linear time k-leaf root construction algorithms for both $k = 3$ and 4. Particularly, their characterization of 3-leaf powers is the same class of graphs we restrict our attention to in this paper—*tree chordal* graphs. A graph is tree chordal if it is the result of substituting cliques in place of the vertices of a tree.

A Steiner point in a k-leaf root is thought of as an evolutionary event in the phylogeny of a set of species. These events represent genetic splits from common unknown ancestors and motivated Lin et al. (2000) to consider the *k-phylogenetic root* construction problem. A k-phylogenetic root is similar to a k-leaf root, with the extra constraint that all its Steiner points have degree at least 3. Analogously, when a k-phylogenetic root T exists for a given graph G, G is the *k-phylogenetic power* of T. Lin et al. (2000) showed how to recognize k-phylogenetic powers and construct a k-phylogenetic root in linear time for $k \leq 4$. Nevertheless, no polynomial time k-phylogenetic root construction algorithm is known for $k \geq 5$, neither the complexity of the recognition problem.

In this paper, we partially answer this question by constructing a 5-phylogenetic root for a given tree chordal graph, if such a root exists. As our work builds on the work of Lin et al. (2000), we first give some of the key ideas in their construction algorithms while exploring the structure of k-phylogenetic powers. We then proceed in Sect. 3 to present structural restrictions for the input graph G characterizing when 5-phylogenetic roots do not exist. A 5-phylogenetic root construction algorithm for a slightly restricted class of tree chordal graphs is presented in Sect. 4; this algorithm is then used as a sub-routine in the full construction algorithm presented in Sect. 5.

2 Preliminaries

We present some of the key ideas in the k-phylogenetic root construction algorithms by Lin et al. (2000), from there, we explore the structure of the 5-phylogenetic powers. Two facts used in Lin et al. (2000) are that k-phylogenetic powers are necessarily chordal, *i.e.*, they do not contain any induced cycles of length four or more, and that chordal graphs are linear time recognizable.

The first key insight in Lin et al. (2000) is that certain subsets of vertices could be grouped together as in some sense they were *indistinguishable* in any k-phylogenetic root of graph G. In graph $G = (V, E)$, a subset $H \subseteq V$ is *homogeneous* if every vertex in $V \setminus H$ is either adjacent to all vertices of H or none of H. A *clique* $C \subseteq V$ is a subset of pairwise adjacent vertices. A clique C is *critical* if C is a homogeneous subset and C is maximal (in the sense that no vertex v exists such that $C \cup v$ is still a clique and homogeneous). The *size* of a critical clique denotes the number of vertices it contains. Clearly, the set of critical cliques, $\mathcal{V}$, forms a partition of the vertex set V. Moreover, if M is a maximal clique and C is a critical clique, then C is either contained in M ($C \subseteq M$) or disjoint to M ($C \cap M = \emptyset$).

Based on the notion of critical clique, an auxiliary graph associated with G, called the *critical clique graph*, can be defined as $CC(G) = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of critical cliques in G and two critical cliques are adjacent if and only if they are contained in some common maximal clique. That is, a critical clique in graph G one-to-one corresponds to a node in graph $CC(G)$. This construction is invertible, *i.e.*, given $CC(G)$, one can reconstruct graph G by replacing each node of $\mathcal{V}$ by the corresponding critical clique and adding edges to connect vertices of two adjacent critical cliques. In this sense, this associated critical clique graph is equivalent to the original graph. We note that the structure of critical clique graphs is central to our 5-phylogenetic root construction algorithm, to be detailed. Moreover, we assume the input graph G to be chordal, and consequently its critical clique graph $CC(G)$ can be constructed in linear time (Lin et al. 2000). Graph G is tree chordal precisely if $CC(G)$ is a tree. Furthermore, a critical clique in a tree chordal graph G is a *leaf critical clique* if it is a leaf in $CC(G)$; otherwise, an *internal* critical clique.

The second key insight in Lin et al. (2000) is to reduce the k-phylogenetic root problem to an equivalent auxiliary *S-restricted $(k-2)$-Steiner root* problem. An S-restricted k-Steiner root T of graph $G = (V, E)$, where $S \subset V$, is similar to a k-phylogenetic root but now the vertices of graph G may label internal nodes of tree T, more than one vertex of graph G may label the same vertex of T, and the pre-specified subset S of vertices must be internal in tree T. Lin et al. (2000) showed that when graph G has a k-phylogenetic root T, then the leaves adjacent to a common Steiner point in T must belong to a common critical clique in G. It follows that by labeling Steiner points in T using the corresponding critical cliques, and then removing all the leaves from T, we can obtain a new tree T' which is an S-restricted $(k-2)$-Steiner root of $CC(G)$, where S is the set of size-1 critical cliques in G. Conversely, the existence of such an S-restricted $(k-2)$-Steiner root T' of $CC(G)$ with S being the set of size-1 critical cliques in G implies a k-phylogenetic root T of graph G, where T can be constructed, with some care, from T' by attaching the vertices in a critical clique to the corresponding labeled nodes.

3 Structural restrictions

For the remainder of this article, we assume the given graph G is tree chordal, to examine its structure when it is a 5-phylogenetic power. We also assume graph G is non-trivial, *i.e.*, containing at least 3 critical cliques. For ease of presentation, when graph G has a 5-phylogenetic root T, the Steiner points in T that are attached with vertices from a critical clique C are called the *representatives* of C.

Lemma 1 *Assume T is a 5-phylogenetic root of graph G and C is a leaf critical clique in G. Then removing the vertices and the representatives of C from T does not disconnect the other leaves of T.*

Proof Since C is a leaf critical clique, it has only one neighbor in $CC(G)$, denoted as D. Suppose to the contrary that removing the vertices and the representatives of C from T disconnect the other leaves of T. Then in T, at least one representative of C, $r(C)$, must lie on a path, P, connecting one representative of D, $r(D)$, and one representative, $r(X)$, for a third critical clique $X \neq D$. It follows that every leaf in the subtree rooted at $r(C)$ and containing $r(X)$ is strictly closer to $r(C)$ than to any other representatives outside this subtree, including $r(D)$. This implies either C is adjacent to another critical clique other than D or graph G is disconnected, a contradiction. □

The following follows directly from Lemma 1:

Structural Restriction 1 *If G has a size-1 leaf critical clique, then G does not have a 5-phylogenetic root.*

Assume T is a 5-phylogenetic root of graph G and C is a critical clique in G. Let T_C be the subgraph of T induced by the representatives of C and the paths connecting them. Clearly, T_C must be a tree, and must have diameter at most 3.

Lemma 2 *Assume T is a 5-phylogenetic root of graph G and C is an internal critical clique in G. Then T_C must be a star. Moreover, if D is another internal critical clique of G, then T_C and T_D do not overlap.*

Proof A star is a tree whose diameter is less than or equal to 2. Suppose to the contrary that T_C has diameter 3. Let r_1 and r_2 be the two representatives of C such that their distance in T is 3, and denote the path connecting them as r_1xyr_2. Since C is an internal critical clique, it has at least two neighbors, say D_1 and D_2. Let $r(D_1)$ and $r(D_2)$ be representatives of D_1 and D_2, respectively. It follows that $d(r_i, r(D_j)) \leq 3$ for $i = 1, 2$ and $j = 1, 2$. These four inequalities imply that $d(r(D_1), r(D_2)) \leq 3$, contradicting to the fact that D_1 and D_2 are not adjacent in $CC(G)$.

The second half of the lemma follows easily since otherwise both T_C and T_D would be diameter-2 stars and they share a common center, which would imply that C and D were identical. □

Lemma 2 gives a simple 5-phylogenetic root construction algorithm when every leaf critical clique has size at least 4 and every internal critical clique has size at least

2, as follows: Given graph G, in linear time we can construct the associated $CC(G)$; Second, we replace each node C_i in $CC(G)$ with a Steiner point s_i; Third, if s_i is internal, then attach a representative $r(C_i)$ to s_i for critical clique C_i; Otherwise, attach two representatives $r_1(C_i)$ and $r_2(C_i)$ to s_i for leaf critical clique C_i; Lastly, if s_i is internal, then attach all (at least 2) the vertices in C_i to the representative $r(C_i)$; Otherwise, attach all (at least 4) the vertices in C_i to the two representatives $r_1(C_i)$ and $r_2(C_i)$, making sure at least 2 vertices to each representative. The tree produced in this way is easily seen to be a 5-phylogenetic root. It is worth pointing out that this is essentially the construction presented by Brandstädt and Le (2006) for the 3-leaf root problem.

The real difficulty of the 5-phylogenetic root problem lies in satisfying the degree requirements on Steiner points. The above algorithm does not work for all tree chordal graphs; We still have to deal with leaf critical cliques of size 2 and 3 and internal critical cliques of size 1. The complication with these "small" critical cliques is that they need to "borrow" some representatives from adjacent critical cliques to satisfy the degree requirements. This leads to the following further examination on how representatives can be adjacent to one another.

Lemma 3 *Assume T is a 5-phylogenetic root of graph G. Then every Steiner point in T can be adjacent to representatives for at most two different critical cliques.*

Proof If a Steiner point s in T is adjacent to 3 representatives for 3 critical cliques, respectively, then these 3 critical cliques must be adjacent to each other, violating the fact that graph G is tree chordal. □

Similarly, we can prove the next lemma (proof omitted).

Lemma 4 *Assume T is a 5-phylogenetic root of graph G. Then every length-3 path in T can contain representatives for at most two different critical cliques.*

Lemma 5 *Assume T is a 5-phylogenetic root of graph G and C and D are two adjacent critical cliques in G. If both C and D have exactly one representative, $r(C)$ and $r(D)$, respectively, then $d(r(C), r(D)) = 1$ or $d(r(C), r(D)) = 3$. Furthermore, if $d(r(C), r(D)) = 1$, then one of C and D has degree at least 3 in $CC(G)$; if $d(r(C), r(D)) = 3$, then both C and D are internal in $CC(G)$.*

Proof Suppose to the contrary $d(r(C), r(D)) = 2$. Then the Steiner point connecting $r(C)$ and $r(D)$, denoted by s, should be adjacent to a fourth Steiner point, say x. It follows that the vertices in the subtree of T rooted at s and containing x are disconnected from the other vertices in G, a contradiction. Therefore, $d(r(C), r(D)) = 1$ or 3.

If $d(r(C), r(D)) = 1$ and assuming C is internal in $CC(G)$, let X be another critical clique adjacent to C and $r(X)$ a representative of X. Since $r(X)$ has to be at distance at least 4 from $r(D)$ and at most 3 from $r(C)$, it must be at distance exactly 3 from $r(C)$. Let the path in T connecting $r(C)$ and $r(X)$ be $r(C)yzr(X)$. It follows that y is a non-representative Steiner point and it should adjacent to a third Steiner point w to satisfy the degree constraint. To prevent from the same contradiction in the last paragraph, w must not be a representative for any critical clique but it should be adjacent to a representative for some critical clique other than C, D, and X. This critical clique is adjacent to C in $CC(G)$, that is, C has degree at least 3.

If $d(r(C), r(D)) = 3$ and assuming the path in T connecting $r(C)$ and $r(D)$ be $r(C)xyr(D)$, note that both x and y are non-representative Steiner points and each of them is adjacent to another non-representative Steiner point, say z and w, respectively. Again, to prevent from the disconnectivity contradiction, each of z and w should be adjacent to a representative for some critical clique other than C and D, that is, both C and D have degree at least 2 in $CC(G)$. □

Structural Restriction 2 *Given a tree chordal graph G, if there are two degree-2 size-1 adjacent internal nodes in $CC(G)$, then G is not a 5-phylogenetic power.*

Proof If graph G has a 5-phylogenetic root T, then these two adjacent critical cliques, denoted as C and D, have exactly one representative each. Moreover, these two representatives must be internal in the associated Steiner root of T. It follows that these two representatives must be adjacent to each other in T, a contradiction to Lemma 5. □

The following lemma is important as it simplifies the 5-phylogenetic root construction considerably.

Lemma 6 *Assume graph G is a 5-phylogenetic power. Then G has a 5-phylogenetic root T in which every leaf critical clique in G of size less than 4 has exactly one representative, and this representative must be adjacent to a representative for the neighboring critical clique.*

Proof Let C be a leaf critical clique in G of size less than 4. Since G is a 5-phylogenetic power, by Structural Restriction 1 we know that C has size either 2 or 3. From Lemma 2, we conclude that at least one representative of C must be a leaf in the 3-Steiner root obtained from a 5-phylogenetic root of G. Consequently, if C has size 2, then it has only one representative. If C has size 3, it might have another representative. Let T be a 5-phylogenetic root of G in which C has two representatives $r_1(C)$ and $r_2(C)$, and assume that $r_1(C)$ is a leaf in the 3-Steiner root obtained from T. Again from Lemma 2, it can be easily proved that $r_1(C)$ and $r_2(C)$ must be adjacent in T; for otherwise either graph G would be disconnected or C would have more than one neighbor. Subsequently, we may merge $r_1(C)$ and $r_2(C)$ into one representative, and the resultant tree is still a 5-phylogenetic root of G.

To prove the second half of the lemma, let T be a 5-phylogenetic root of G in which C has one representative $r(C)$, which appears as a leaf in the 3-Steiner root obtained from T. Let D be the neighboring critical clique of C in G, and $r_1(D)$ be a representative of D in T which is the closest to $r(C)$. A similar proof as in the last part of the proof of Lemma 5 shows that $r(C)$ and $r_1(D)$ must be adjacent in T. □

When graph G is a 5-phylogenetic power, it will have a 5-phylogenetic root satisfying the property stated in Lemma 6. Such a root is called *minimal*. We have the following two more structural restrictions.

Structural Restriction 3 *Given a tree chordal graph G, if there is a degree-2 size-1 internal node in $CC(G)$ which is adjacent to a leaf node of size 2 or 3, then G is not a 5-phylogenetic power.*

Proof This holds directly from Lemmas 5 and 6. □

Structural Restriction 4 *Given a tree chordal graph G, if there is an internal node C in $CC(G)$ for which the number of adjacent leaf nodes of size 2 and 3 is greater than the size of C, then G is not a 5-phylogenetic power.*

Proof From Lemma 6, every leaf node of size 2 or 3 requests for a distinct representative of critical clique C in a minimal 5-phylogenetic root of graph G. This is certainly impossible to satisfy if the size of C is less than the requested number of representatives. □

Lemma 7 *Assume graph G is a 5-phylogenetic power. Then G has a 5-phylogenetic root T in which every leaf critical clique in G of size at least 4 has exactly two representatives, which are adjacent to a common non-representative Steiner point.*

Proof Let C be a leaf critical clique of size at least 4 in graph G, and D be its neighboring critical clique. Let $r(C)$ and $r(D)$ be two representatives of C and D, respectively, such that their distance in a 5-phylogenetic root of G achieves the maximum. Clearly, $d(r(C), r(D)) = 1, 2$ or 3.

If $d(r(C), r(D)) = 3$, and supposing $r(C)xyr(D)$ is the path connecting them, then all representatives for C must be adjacent to x. It follows that we can revise the root, if necessary, to ensure that x is not a representative for D, and merge/create exactly two representatives for C. If $d(r(C), r(D)) = 2$, and supposing $r(C)xr(D)$ is the path connecting them, then x should not be adjacent with any other representatives for D. That is, x is either a representative for D, or a representative for C, or a non-representative adjacent with another representative for C. In the first two cases, we can always revise the root to ensure that x becomes a non-representative, and merge/create exactly two representatives for C. If $d(r(C), r(D)) = 1$, then both C and D have exactly one representative each. Therefore, we can revise the root by inserting a non-representative Steiner point x into edge $(r(C), r(D))$, and splitting $r(C)$ to create exactly two representatives for C, both adjacent to x. □

Lemma 8 *Assume T is a 5-phylogenetic root of graph G and C and D are two adjacent critical cliques in G. If C has exactly one representative $r(C)$ in T, and D has exactly two representatives $r_1(D)$ and $r_2(D)$ in T and $r_1(D)$ and $r_2(D)$ are adjacent to each other, then $r(C)$ is adjacent to either $r_1(D)$ or $r_2(D)$. Furthermore, if D is a leaf critical clique, then C has degree at least 3 in $CC(G)$.*

Proof First, in T, $r(C)$ is at distance at most two to either $r_1(D)$ or $r_2(D)$. Then, a similar proof as for Lemma 3 shows that there should not be any other Steiner point connecting $r(C)$ and $r_1(D)$ and $r_2(D)$, for otherwise either graph G would not be tree chordal or it would be disconnected. Now assume D is a leaf critical clique, which implies that C is internal. A similar proof as in the middle part of the proof of Lemma 5 shows that C has degree at least 3 in $CC(G)$. □

Structural Restriction 5 *Given a tree chordal graph G containing more than 3 critical cliques, if there is an internal degree-2 and size-2 node C in $CC(G)$ and C is adjacent to a leaf node of size 2 or 3, then G is not a 5-phylogenetic power.*

Proof Assume to the contrary that T is a minimal 5-phylogenetic root of graph G. Let D be the leaf node of size 2 or 3, and $r(D)$ its representative. Lemma 6 implies a representative of C, $r_1(C)$, which is adjacent to $r(D)$ in T. On the other hand, Lemma 5 implies that C has another representative, $r_2(C)$. Because C has size two, both $r_1(C)$ and $r_2(C)$ must be internal in the 3-Steiner root obtained from T. From the fact that C has degree two, $r_1(C)$ and $r_2(C)$ must be adjacent to each other. Let X denote the other neighboring critical clique adjacent to C in $CC(G)$. It follows that all the representatives of X must be at distance exactly 2 from $r_2(C)$, through a common non-representative Steiner point. From Lemma 8, X should have at least two representatives, which however implies that every neighboring critical cliques of X must be adjacent to C as well, a contradiction to the degree assumption on C or the assumption that G contains more than 3 critical cliques. □

Lemma 9 *Assume T is a 5-phylogenetic root of graph G and C is a size-1 critical clique in G. Then C has degree at least 3 in $CC(G)$. Furthermore, if C has degree exactly 3, then one of its neighboring critical cliques must have size strictly greater than the number of adjacent leaf critical cliques of size 2 and 3.*

Proof We prove first that C has degree at least 3 in $CC(G)$. Note that if C is adjacent to a leaf critical clique of size 2 or 3, then by Lemma 5 C has degree at least 3 in $CC(G)$. In the following, assume to the contrary that C is not adjacent to any leaf critical clique of size 2 or 3 and C has degree 2 in $CC(G)$. Let $r(C)$ denote the unique representative for C in root T. Since $r(C)$ must be internal in the associate Steiner root of T, if all the representatives for the two neighboring critical cliques of C are at distance exactly 2 to $r(C)$, then graph G can contain only 3 critical cliques, a trivial case. It follows that there is some representative which is at distance either 1 or 3 to $r(C)$. Note that the existence of a distance-1 representative implies a distance-3 representative for the other neighboring critical clique of C. We conclude that there is one neighboring critical clique of C, say D, such that one of its representatives $r(D)$ is at distance exactly 3 from $r(C)$. Thus, in the same branch rooted at $r(C)$ where $r(D)$ resides, there is another representative which is closer to $r(C)$ than to $r(D)$, a contradiction to either Lemma 2, or Lemma 7, or the fact that C has degree 2 in $CC(G)$. Therefore, C must have degree at least 3 in $CC(G)$.

When C has degree exactly 3, there must be a representative $r(D)$ adjacent to $r(C)$ in root T. It follows that no representative for critical cliques other than C or

D could be adjacent to $r(D)$. Hence, from Lemma 6, we conclude that the size of D is greater than the number of its adjacent leaf critical cliques of size 2 and 3 in $CC(G)$. □

4 Dense tree chordal graphs

A tree chordal graph containing no critical cliques of size 1 is called *dense*. In this short section, we restrict ourselves to consider only dense tree chordal graphs (containing at least 3 critical cliques), and present an algorithm for deciding whether a dense tree chordal graph has a 5-phylogenetic root, and if so to construct such a root. It turns out that Structural Restrictions 4 and 5 proven in the last section are sufficient to guarantee the existence of a 5-phylogenetic root. Therefore, we check where the input graph emits any of Structural Restrictions 4 and 5. For doing so, we need only look at each critical clique and its neighbors once, and thus this step takes line time.

In the following, assume that the input graph G does not emit any of the four Structural Restrictions. We start with the critical clique graph $CC(G)$, which is a tree, to construct a 5-phylogenetic root T for graph G. First, we replace each critical clique C_i with a Steiner point s_i. Second, in $CC(G)$, for each degree-2 C_i adjacent to a leaf critical clique of size 2 or 3, denoted as L, we distinguish two cases: C_i has size two or greater. If C_i has size two, then G contains exactly 3 critical cliques with the third critical clique of size at least 4, by Structural Restriction 5; and for this case, T contains one representative for L, two adjacent representatives for C_i, and two non-adjacent representatives, which are adjacent to a common non-representative Steiner point, for the third critical clique of size at least 4. In the following, we assume C_i has size at least three. Let D be the other neighboring critical clique of C_i. If D is also a leaf critical clique of size 2 or 3, then G contains exactly three critical cliques and the skeleton of the 5-phylogenetic root T is a path of length 5, with the middle three representatives for C_i and one leaf representative for each of L and D. If D is not a leaf critical clique, then we create two representatives $r_1(C_i)$ and $r_2(C_i)$ for C_i, such that $r_1(C_i)$ is a leaf adjacent to s_i and $r_2(C_i)$ bisects the edge $(r(L), s_i)$. In the final tree T, one vertex of C_i is attached to $r_2(C_i)$, while the other vertices attached to $r_1(C_i)$.

Third, for each C_i with degree three or more in $CC(G)$, and for each adjacent leaf critical clique of size 2 or 3, denoted as L, place a representative $r(C_i)$ bisecting the edge $(r(L), s_i)$. In the final tree T, at least one vertex of C_i is attached to $r(C_i)$, while making sure all vertices of C_i are properly distributed to its representatives. Fourth, for every internal critical clique C_i in $CC(G)$ not adjacent with any leaf critical clique of size 2 or 3, create exactly one representative $r(C_i)$ and attach it to s_i. In the final tree T, all vertices of C_i are attached to $r(C_i)$. Fifth, for every leaf critical clique C_i of size at least 4, create two representatives $r_1(C_i)$ and $r_2(C_i)$, and attach them to s_i. In the final tree T, at least two vertices of C_i are attached to each of $r_1(C_i)$ and $r_2(C_i)$, while making sure all vertices of C_i are properly distributed to these two representatives. Finally, for every leaf critical clique C_i of size 2 or 3, replace its the Steiner point s_i with its only representative $r(C_i)$; and in the final tree T, all vertices of C_i are attached to $r(C_i)$. One can easily check that all critical cliques have been included in the above six cases, and the resultant tree T is a 5-phylogenetic root for graph G. That is, we have proved the following:

Theorem 1 *A dense tree chordal graph $G = (V, E)$ emitting none of the two Structural Restrictions 4 and 5 is a 5-phylogenetic power, and one of its 5-phylogenetic root T can be constructed in $O(|V| + |E|)$ time.*

5 General tree chordal graphs

Structural Restrictions 4 and 5 summarize all those dense tree chordal graphs that are not 5-phylogenetic powers. For general tree chordal graphs which might contain size-1 critical cliques, Structural Restrictions 1–3 tell that some of them are not 5-phylogenetic powers either. In particular, these size-1 critical cliques have to be internal (by Structural Restriction 1), each of them can be adjacent with at most one leaf critical clique of size 2 or 3 (by Structural Restriction 4), and if one does then it has to have degree at least 3 (by Structural Restriction 3). In this section, we present three reductions each of which eliminates exactly one size-1 internal critical clique from the input graph G to give a new graph G^*, through a (constant time) local transformation, such that 1) G^* contains one less size-1 internal critical clique than G, 2) G^* is a 5-phylogenetic power if and only if G is a 5-phylogenetic power, and 3) Every internal critical clique of size equal to or greater than 2 in G is unchanged in G^*. A reduction satisfying the above three properties is said to be *valid*. At the end, we will obtain a graph G^* that is dense tree chordal, for which a 5-phylogenetic root can be constructed in the way as described in the last section (Theorem 1). A 5-phylogenetic root for graph G can then be constructed from this root of graph G^*. Let C be a size-1 internal critical clique in tree chordal graph G.

Reduction 1 *If C is adjacent to exactly one leaf critical clique D of size 2 or 3, then G^* is the graph created by deleting D and increasing the size of C to 2.*

Reduction 2 *If C has degree 2 and is adjacent to exactly one leaf critical clique D of size at least 8, then G^* is the graph created by deleting D and increasing the size of C to 2.*

Reduction 3 *If C is adjacent to no leaf critical cliques or at least two leaf critical cliques of size at least 4, then G^* is the graph created by increasing the size of C to 2.*

Lemma 10 *Reduction 1 is valid.*

Proof It is easily seen that G^* contains one less size-1 internal critical clique than G, and every internal critical clique of size equal to or greater than 2 in G is unchanged in G^*.

Assume graph G is a 5-phylogenetic power. From Lemma 6, we know that there is a 5-phylogenetic root T of G in which $r(D)$, the unique representative for D, appears as a leaf in the corresponding Steiner root obtained from T, and it is adjacent to $r(C)$, the unique representative for C. Since no Steiner points other than $r(C)$ can be within distance 3 to $r(D)$ in T, removing $r(D)$ and the vertex in D, and attaching the newly added vertex to C to its representative $r(C)$ gives 5-phylogenetic root of graph G^*.

Conversely, assume graph G^* is a 5-phylogenetic power. Note that critical clique C has size 2, and is still internal. From the 5-phylogenetic root construction algorithm presented in the last section for dense tree chordal graphs, we see that neither the second nor the third step apply, by Structural Restriction 4 on graph G. Therefore, the fourth step must apply, and thus there is a 5-phylogenetic root T for graph G^* in which critical clique C has exactly one representative $r(C)$, $r(C)$ is adjacent to exactly one other, non-representative, Steiner point, and every representative Steiner point is at distance at least 3 away from $r(C)$. It follows that, if we create a representative Steiner point $r(D)$ and attach it to $r(C)$, then reducing the size of C to 1 (removing one vertex in C from tree T) and attaching the vertices in D to $r(D)$ give a feasible 5-phylogenetic root for graph G. Therefore, Reduction 1 is valid. □

Lemma 11 *Reduction 2 is valid.*

Proof It is easily seen that G^* contains one less size-1 internal critical clique than G, and every internal critical clique of size equal to or greater than 2 in G is unchanged in G^*.

Assume graph G is a 5-phylogenetic power and T is a 5-phylogenetic root. First of all, we claim that the other neighboring critical clique of C, denoted as X, must have a representative adjacent to $r(C)$, the unique representative for C, in T. For otherwise, a similar proof as in the proof of Lemma 5 leads to a contradiction that either graph G is disconnected or C has degree at least 3. In more details, assume to the contrary that no representative of X is adjacent to $r(C)$, but one representative, denoted as $r(X)$, is at distance 2 to $r(C)$, and let $r(X)sr(C)$ denote the path connecting them, where s is a non-representative Steiner point. Then, to satisfy the degree constraint, there must be another Steiner point y adjacent to s. Since C is internal and has degree 2, we conclude that X is internal too (since graph G is non-trivial) and y must not be a representative for X. It follows that y must be a non-representative Steiner point and consequently the vertices in the subtree of T rooted at s and containing y are disconnected from the other vertices in G, a contradiction. Similarly, it is impossible that no representative of X is within distance 2 to $r(C)$ but one is at distance exactly 3 to $r(C)$. That is, our claim holds, which says there must be a representative of X adjacent to $r(C)$. It follows from this claim that every representative for the leaf critical clique D must be at distance exactly 3 to $r(C)$ (otherwise D would be adjacent to X; and therefore, there must be at least 4 leaf Steiner points in T which are representatives for D, implying that the size of D is at least 8). Moreover, removing the whole branch of T containing the representatives for D and all vertices in D gives a 5-phylogenetic root for G^*, in which $r(C)$ is now attached with the two vertices in C.

Conversely, assume graph G^* is a 5-phylogenetic power. Note that critical clique C has size 2, and is a leaf critical clique in graph G^*. From the 5-phylogenetic root construction algorithm presented in the last section for dense tree chordal graphs, we see that the last (the sixth) step applies, and thus there is a 5-phylogenetic root T for graph G^* in which critical clique C has exactly one representative $r(C)$, $r(C)$ is attached with the two vertices in C, and $r(C)$ is adjacent to a representative of its neighboring critical clique in G^*. We create a branch of 4 representatives (r_1, r_2, r_3, r_4) and 3 non-representative Steiner points (s_1, s_2, s_3), in which s_1 is the root, s_2 and s_3 are both adjacent to s_1, r_1 and r_2 are both adjacent to s_2, r_3 and r_4 are both adjacent to s_3, and each representative is attached with at least two vertices in D. It follows that if we connect this branch to $r(C)$ by adding an edge $(s_1, r(C))$, then subsequently reducing the size of C to 1 (removing one vertex in C from tree T) gives a feasible 5-phylogenetic root for graph G. Therefore, Reduction 2 is valid. □

Lemma 12 *Reduction 3 is valid.*

Proof It is easily seen that G^* contains one less size-1 internal critical clique than G, and every internal critical clique of size equal to or greater than 2 in G is unchanged in G^*.

Assume graph G is a 5-phylogenetic power and T is a 5-phylogenetic root. Trivially, attaching the newly added vertex to critical clique C to the representative of C in T gives a 5-phylogenetic root for graph G^*.

Conversely, assume graph G^* is a 5-phylogenetic power. Note that critical clique C has size 2, is internal, and is not adjacent with any leaf critical clique of size 2 or 3. From the 5-phylogenetic root construction algorithm presented in the last section for dense tree chordal graphs, we see that neither the second nor the third step apply, by Structural Restriction 4 on graph G. Therefore, the fourth step must apply, and thus there is a 5-phylogenetic root T for graph G^* in which critical clique C has exactly one representative $r(C)$, $r(C)$ is adjacent to exactly one other, non-representative, Steiner point, and every representative Steiner point is at distance at least 3 away from $r(C)$.

Let T denote this 5-phylogenetic root. If the non-representative Steiner point $s(C)$ adjacent to representative $r(C)$ for C has degree at least 5 (that is, at least 4 edges connecting $r(C)$ to non-representative Steiner points), then by splitting $s(C)$ into two non-representative Steiner points $s_1(C)$ and $s_2(C)$ such that each of them inherits at least two edges out of $s(C)$, and removing one vertex from C, we obtain from T a 5-phylogenetic root for graph G. In the other case, Steiner point $s(C)$ has degree only 4, which implies that C has degree 3 in $CC(G^*)$. Therefore, from Lemma 9, we have at least one neighboring critical clique D of C whose size is strictly larger than the number of adjacent leaf critical cliques of size 2 and 3 in $CC(G^*)$. Consequently, either we have one representative for D that is a leaf in the associated Steiner root, or we can create a new representative $r(D)$ for D and attach it to $s(D)$ (that is, again a leaf in the Steiner root). The obtained tree is not a 5-phylogenetic root for graph G yet, but by removing the edge connecting $s(C)$ and $s(D)$, adding the edge connecting $r(C)$ and $r(D)$, and removing a vertex from C, it becomes one. Therefore, Reduction 3 is valid. □

The following main theorem summarizes the above efforts:

Theorem 2 *Given a tree chordal graph $G = (V, E)$, whether or not G is a 5-phylogenetic power can be determined in $O(|V| + |E|)$ time. When G is a 5-phylogenetic power, one 5-phylogenetic root can be constructed in $O(|V| + |E|)$ time too.*

6 Concluding remarks

We have shown how to recognize 5-phylogenetic powers for tree chordal graphs, and if the input graph is a 5-phylogenetic power, then we construct one root in linear time. This algorithmic result extends the results of Lin *et al.* (2000) and tree chordal graph is the largest known class of graphs to have such a root construction. The complexity of recognizing k-phylogenetic powers for $k \geq 5$ is still unknown. We believe that for the $k = 5$ case there exists a polynomial time algorithm, though it is likely to be quite complex.

Our root construction algorithm is much more complicated than the characterization of 3-leaf roots, which is exactly the class of tree chordal graphs (Brandstädt and Le 2006). We see that the difficulty of the k-phylogenetic root construction lies in the satisfaction of the Steiner point degree requirements. Note that both the leaf root and the phylogenetic root problems were formulated out of the phylogeny reconstruction in computational biology. Because in a phylogeny, internal nodes correspond to the speciation events and thus are at least bifurcations, the phylogenetic root problem seems more natural. Nevertheless, the leaf root problem has attracted much attention in the last few years. To a larger extent, it would be interesting to study the more general (k, ℓ)-leaf root (Brandstädt and Wagner 2007) or (k, ℓ)-phylogenetic root problem, where a tree T is a (k, ℓ)-leaf/phylogenetic root of a graph $G = (V, E)$ if for every edge $(x, y) \in E$, the distance between x and y in T is at most k, and for every $(x, y) \notin E$, the distance between x and y in T is at least ℓ. The k-leaf/phylogenetic root problem is the special case where $\ell = k + 1$.

References

Brandstädt A, Le VB (2006) Structure and linear time recognition of 3-leaf powers. Inf Process Lett 98:133–138

Brandstädt A, Le VB, Sritharan R (2006) Structure and linear time recognition of 4-leaf powers (submitted)

Brandstädt A, Wagner P (2007). On (k, l)-leaf powers. In: Proceedings the 32nd international symposium of mathematical foundations of computer science (MFCS 2007). LNCS, vol 4708, pp 525–535

Lin G-H, Jiang T, Kearney PE (2000) Phylogenetic k-root and Steiner k-root. In: The 11th annual international symposium on algorithms and computation (ISAAC 2000), LNCS, vol 1969, pp 539–551

Nishimura N, Ragde P, Thilikos DM (2002) On graph powers for leaf-labeled trees. J Algorithms 42:69–108

Rautenbach D (2006) Some remarks about leaf roots. Discrete Math 306:1456–1461