5th Phylogenetic Root Construction for Strictly Chordal Graphs

William Kennedy * and Guohui Lin **

Algorithmic Research Group and Bioinformatics Research Group,
Department of Computing Science, University of Alberta.
Edmonton, Alberta T6G 2E8, Canada.

Abstract. Reconstruction of an evolutionary history for a set of organisms is an important research subject in computational biology. One approach motivated by graph theory constructs a relationship graph based on pairwise evolutionary closeness. The approach builds a tree representation equivalent to this graph such that leaves of the tree, corresponding to the organisms, are within a specified distance of k in the tree if connected in the relationship graph. This problem, the kth phylogenetic root construction, has known linear time algorithms for $k \leq 4$. However, the computational complexity is unknown if $k \geq 5$. We present a polynomial time algorithm for strictly chordal relationship graphs if $k = 5$.

Keywords: Computational biology, phylogeny reconstruction, phylogenetic root, Steiner root, chordal, strictly chordal.

1 Introduction

A phylogeny is the development and history through evolution of a set of organisms or evolutionary units. A phylogenetic tree is a visual representation with the leaves labeled by the evolutionary units and distances in the tree representing evolutionary closeness; reconstruction of such trees is a fundamental question in computational biology. One approach, based on graph theory, uses an input graph representing the known relationships of the units; the vertices are labeled by the units and connected if their relative closeness is greater than some prespecified threshold. The approach then constructs a tree, if it exists, where the unit-labeled leaves are within a given path distance in the tree if and only if the evolutionary units' vertices are connected in the input graph. This problem is a variation of the well-studied graph root and graph power problem.

The kth power of a graph, $G = (V, E)$, denoted G^k, is another graph on the same vertex set with an edge between $x, y \in V$ if and only if a path in G of length at most k exists between x, y. Given a graph, it is efficient to compute a kth power, but the reverse direction, finding a kth root, is generally more complex. For example, finding a square root of graph is NP-complete [9]. A few

* Supported partially by NSERC. Email: kennedy@cs.ualberta.ca.
** Supported by NSERC and CFI. Email: ghlin@cs.ualberta.ca.

polynomial time algorithms exist for computing if a graph has a kth root where this root is a tree [3,8]. For a graph G, a *kth leaf root* is a tree T where the vertex set of G corresponds to the leaves of T and the kth power T^k induced on the leaf set of T is isomorphic to G. The internal vertices in such a tree are *Steiner points*. For a Steiner point the number of Steiner points it is adjacent to in T is its *Steiner degree*. All *strictly chordal* graphs are leaf powers, and finding the kth leaf root of a strictly chordal graph for $k \geq 4$ has a linear time solution [4].

The *k-root phylogenetic tree* problem is a restriction of the kth leaf root with, in addition, all Steiner points having degree of at least 3; this follows from the notion of an internal point in the tree representing a genetic split from a common ancestor. The problem of finding a k-root phylogenetic tree has a linear time solution for $k \leq 4$ [6]. [5] has produced an algorithm for the $k = 5$ case that runs in linear time if the critical clique graph $CC(G)$ is a tree.

The *k-root Steiner* problem is a relaxation of the kth leaf root with vertices of G now represented by both the leaves and some internal vertices of the tree. [6] used this problem as an important intermediary step for constructing the 3rd and 4th root phylogenetic tree; we employ an analogous approach in this paper. The main contribution of this paper will be to produce an algorithm that decides if a given strictly chordal graph has a 5-root phylogeny tree, and if such a tree exists, constructs this tree.

2 Preliminaries

A forest is a simple acyclic graph; a tree is a connected forest; we will assume that a tree is undirected. A graph is *chordal* [1] if it contains no induced cycle of length four or more. Chordal graphs can be recognized in linear time [1]. All graphs that have a leaf power representation are chordal graphs [6].

For a graph G, a *clique* is a set of vertices that are pairwise adjacent. A clique is *maximal* if no other vertex of G is adjacent to all vertices in the clique. A *critical clique* [6] is a clique such that all the vertices in the clique share a common neighborhood. The *cardinality* of a maximal clique K, denoted $card(K)$, is the number of critical cliques it contains. For convenience, we define a maximal clique to be *large* if it has cardinality of three or more. The *size* of a critical clique C in a graph G is the number of vertices it contains. A critical clique is *internal* if contained in at least two maximal cliques and *external* otherwise.

The set formed by all the critical cliques of a graph is a partition of the vertex set of the graph. Two critical cliques C_1 and C_2 are adjacent if the vertices of C_1 are adjacent to the vertices of C_2. Define a critical clique graph $CC(G)$ with nodes of $CC(G)$ correpsonding to the critical cliques of G such that two nodes are adjacent if and only if the critical cliques that they represent are adjacent [6]. As with the conventions established in [6], to avoid confusion we will refer the vertices in $CC(G)$ as nodes.

For a vertex set V and *hyperedge* set $\mathcal{E} = \{E_i | i = 0, 1, 2, ..., n\}$ such that $E_i \subseteq V$, the pair $\mathcal{H} = (V, \mathcal{E})$ is a *hypergraph* [1]. A *twig* in a hypergraph $\mathcal{H}(V, \mathcal{E})$, denoted E_t, is a hyperedge such that there exists another hyperedge, E_b called a branch, with the property that $E_t \cap (\cup_{E \in \mathcal{E} - E_t} E) = E_t \cap E_b$. A hypergraph is a *hypertree* if there exists an ordering $(E_1, E_2, \ldots, E_m)$ of the hyperedges such that E_i is a twig in $\mathcal{H}(V, \mathcal{E}_i)$ for $1 \leq i \leq m$ given that $\mathcal{E}_i = (E_1, E_2, \ldots, E_i)$. This definition of hypertree corresponds to [4,2] and differs from the definition in [1]. Given a set of hyperedges $\mathcal{E}' \subseteq \mathcal{E}$, We say the intersection I of $\mathcal{E}'$ is *maximal* [4] if no other hyperedges exist which contain all the vertices of I. We say that an intersection is *strict* [4] if for every pair of edges, $E', E'' \in \mathcal{E}$, $E' \cap E'' = I$ and $E^* \cap I = \emptyset$ for all $E^* \in \mathcal{E} - \mathcal{E}'$. A hypertree is *strict* [4] if all its intersections are strict. Define a *clique hypergraph* of a graph $G = (V, E)$ with vertex set V and hyperedges set as the maximal cliques of G. A graph is *strictly chordal* if it is chordal and its clique hypergraph is a strict hypertree. There exists a linear time algorithm to recognize strictly chordal graphs [4].

A k-root Steiner tree T of a graph G is S*-restricted* for a set of critical cliques S if T has no degree 2 Steiner points and critical cliques in S are internal in T.

Lemma 1. *[7] For a graph G, let S be the critical cliques of size 1 in G. If no (k-2)-root Steiner tree T is associated with G such that T is a S-restricted Steiner tree, then no k-root phylogenetic tree is equivalent to G.*

We now give a brief overview of the strategy we employ to produce a 5-root phylogeny tree from a graph G. Starting with G, check if G is strictly chordal, and if yes, we build the critical clique graph $CC(G)$. Using $CC(G)$, we produce the set of tree chordal graphs $\mathcal{T}$ by removing edges in large maximal cliques (see Section 2.1). For each tree chordal graph $T_i \in \mathcal{T}$, we apply $Alg(T_i)$, a modification of the algorithm for producing a 5-root phylogeny tree for tree chordal graphs given by Lin et al. [7]. If $Alg(T_i)$ fails to produce a 3-root Steiner tree we will find no phylogeny tree by Lemma 2. If for all $T_i \in \mathcal{T}$, $Alg(T_i)$ produces a Steiner tree S_i we then continue to consider the edges removed from the large maximal cliques. We combine each 3-root Steiner tree S_i until either we come to a contradiction or we have a valid S-restricted Steiner tree where S is the critical cliques of size 1 in G. If we construct such a Steiner tree, by Lemma 1 we are always able to produce a corresponding 5-root phylogeny tree. The Figure 1 shows a simplified flow chart of the algorithms steps.

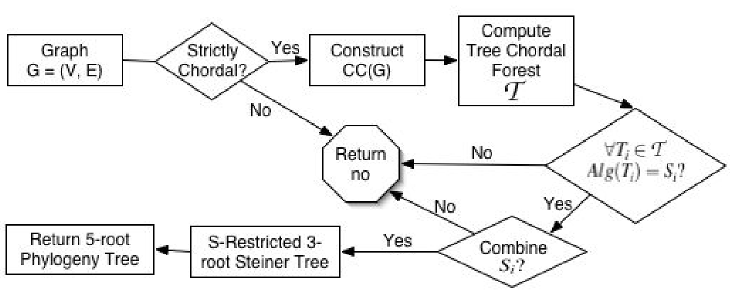


Fig. 1. A flow chart of the 5-root phylogeny tree construction algorithm.

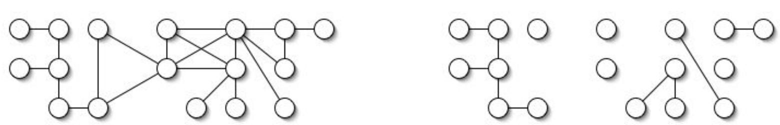
Before we present the algorithms, we first discuss the tree chordal algorithm presented by Lin et al. in [7] (Section 2.1) and discuss the structure of large maximal cliques (Section 2.2). We will present the algorithm in three progressively less restrictive parts. Section 3.1, will assume G contains no small leaves and at most one critical cliques is constrained (see Section 3.1). Section 3.2 will restrict G to not contain small leaves and Section 3.3 will show the entire construction for strictly chordal graphs.

2.1 Tree chordal graphs

A graph G is *tree chordal* [7] if $CC(G)$ is a tree. [7] showed a polynomial time algorithm to construct a 5-root phylogenetic tree from a tree chordal graph. Starting with $CC(G)$, this algorithm produces a S-restricted 3-root Steiner tree where the set S contains external nodes of size 1 in $CC(G)$; from this Steiner tree a 5-root phylogenetic tree is easily produced. The non-trivial cases for this algorithm are internal nodes of size 1 and external nodes of size 2 and 3.

By Lemma 1, for any 5-root phylogeny tree a corresponding S-restricted 3-root Steiner tree exists. To construct such a tree, for each Steiner point we record the number of leaves it is adjacent to and then remove all leaves in the tree. Steiner points that were not adjacent to leaves will remain as Steiner points; notice that they will still have degree of three or more. The remaining Steiner points now correspond to vertices in the input graph; let the set S be the internal representatives in the graph. This simple retraction takes linear time to produce a S-restricted 3-root Steiner tree.

To produce the forest of tree chordal graphs $\mathcal{T}$ we use the critical clique graph $CC(G)$; remove the edges from large maximal cliques and re-substitute each critical clique in for the node that it represents in the decomposed critical clique graph. An example is given in Figure 2.



(a) Critical clique graph $CC(G)$. (b) Forest $\mathcal{T}$.

Fig. 2. An example decomposition from a critical clique graph to a forest of tree chordal critical clique graphs.

If follows from the input graph being strictly chordal that each node is part of exactly one tree chordal graph. In the algorithm's final step of recombining the corresponding Steiner trees for each tree chordal graph, the final state of the critical cliques part of large maximal clique is necessary information. In addition, these critical cliques must be treated as different as they are internal in $CC(G)$. A critical clique C is *constrained* in K if C has two or more representatives in the Steiner tree T, C has a single representative adjacent to a Steiner point with Steiner degree 1, or C has a single representative adjacent to the representative of another critical clique. As will be shown in Section 2.2, at most one critical clique can be constrained. Therefore, the algorithm aims to produced constrained critical cliques only when necessary. We now describe and justify a modification of the [7] algorithm to minimize constrained critical cliques.

A tree chordal graph T is *trivial* when CC(T) is a single node. The algorithm in [7] assumes that $CC(G)$ contains at least three nodes for a tree chordal graphs G. Therefore, we describe an algorithm to handle trivial tree chordal graphs. Trivial tree chordal graphs T can arise from a decomposition in two ways: T was part of only large maximal cliques in G, or T was part of a large maximal clique and decomposed into a tree chordal graph of exactly two critical cliques. The corresponding S-restricted 3-root Steiner tree with $S = \emptyset$ to the first situation will be a single representative vertex. For the later case, we present to following algorithm $TrivAlg(T)$, such that S is the critical cliques of size 1 in G.

1. if both c_1 and c_2 are internal critical cliques in G then:
   a. if either $|c_1| = 1$ or $|c_2| = 1$ in G then represent c_1 and c_2 by two single adjacent representatives,
   b. otherwise, represent c_1 and c_2 by two single representatives connected by a path of two Steiner points.
2. if only one is an internal critical clique, assume c_1, then:
   a. if $|c_1| = 1$ and $1 < |c_2| < 4$ then represent c_1 and c_2 by two single adjacent representatives,
   b. if $|c_1| > 1$ and $1 < |c_2| < 4$ then represent c_1 by r'_{c_1} and $r_{c''_1}$, c_2 by r'_{c_2}, and create path $r'_{c_1} - r''_{c_1} - r'_{c_2}$,
   c. if $|c_2| > 3$ then represent c_1 by a single representative, represent c_2 by two representatives of sizes $\lceil |c_2|/2 \rceil$ and $\lfloor |c_2|/2 \rfloor$, and make all adjacent to a common Steiner point,
   d. otherwise ($|c_2| = 1$) no S-restricted 3-root Steiner tree exists.

For Case 2d no S-restricted 3-root Steiner tree exist for G, as c_2's representative will always be external and have size 1. The remaining cases represent each trivial tree chordal graph as a Steiner tree that can be part of a valid S-restricted 3-root Steiner tree when recombined.

We must change the Lin algorithm's handling of lemma 15 in [7] such that instead of the representative of the size one internal node being adjacent to two representatives of a leaf critical clique, we add a Steiner point between these two leaf representatives. This change does not affect the correctness of the algorithm as it preserves all adjacencies and degree requirements.

Given a tree chordal graph $T_i \in \mathcal{T}$ decomposed from a graph G, set S corresponding to nodes in T_i of size 1 in $CC(G)$, and a set R corresponding to nodes of $CC(T_i)$ contained in maximal cliques of size three or more in $CC(G)$ produce an S-restricted 3-root Steiner tree as follows. Denote this modified algorithm $ALG(G)$, where G is a tree chordal graph.

- If T_i was part of only large maximal cliques in G, return single representative.
- If T_i was part of a large maximal clique and decomposed into a tree chordal graph of exactly two critical cliques, return tree as in $TrivAlg(T_i)$.
- Produce tree chordal graph T_i' as follows:
  - size two and three external nodes contained in R, change size to four,
  - size one external nodes contained in R adjacent to degree-2 size-2 node in $CC(T_i)$, change size to four,
  - remaining size one external nodes contained in R, change size to two,
  - size one internal nodes contained in R which are not adjacent to an external node not contained in R, change size to two.
- Call the Lin algorithm with modified tree T_i'.
- return no if Lin algorithm fails, or return the S-restricted 3-root Steiner tree.

Lemma 2. *Given a strictly chordal graph G decomposed into a forest of tree chordal graphs $\mathcal{T}$ and set S corresponding to nodes in G of size 1, if $ALG(T)$ fails to produce a valid S-restricted 3-root Steiner tree for any $T_i \in \mathcal{T}$ then no S-restricted 3-root Steiner exists for G.*

Proof. $TrivAlg(T)$ only rejects when $|c_2| = 1$, as $c_2 \in S$ and will never be internal.

To show the correctness of the modified Lin algorithm we prove the contrapositive; assuming G has a S-restricted 3-root Steiner tree T' we will show for any tree chordal graph $T_i \in \mathcal{T}$ produced in the decomposition, the modified algorithm will produce a S-restricted 3-root Steiner tree. Take any $T_i \in \mathcal{T}$ in such a graph G. Denote T_i' as the subtree induced on the tree T' with the representatives of the critical cliques of T_i and the Steiner points on paths between these representatives. We show from T_i' a valid S-restricted 3-root Steiner tree for the T' that $ALG(T_i)$ produces.

First, as T' satisfies all distance requirements, it follows that so will T_i'. All representatives that are not part of large maximal cliques in G will satisfy the size requirements; all Steiner points not adjacent to a representative of a critical clique in large maximal clique will satisfy the minimum degree requirement. Therefore, assume that C is a critical clique in T_i' that fails; it follows either a representative of C is external in T_i' and has size one or a Steiner point adjacent to C has degree two.

We first deal with the case when an external representative is size one. The modifications to $ALG(G)$ set external representatives of size one to size two. Therefore, C must have two representatives, but, the other representative was a leaf in T' implying its size is greater than one. This implies the total size of this critical clique was at least three, but external critical cliques of size three were modified to have size four. Therefore, we can represent these nodes with two representatives of size two each in T_i'.

If the Steiner point adjacent to C has degree two, it must of been adjacent to a Steiner point of the Maximal clique. The representative of the critical clique

C must also have size one, as size greater than two were modified to have size four or more and could have two representatives. This could be the case when the critical clique C is internal as well. By as in the previous case, this implies another representative must exists and we modified critical cliques of size 2 or 3 to size four and, therefore, this critical clique can have two representatives. □

By lemma, if $ALG(T)$ fails for any tree chordal graph, we can return no, as no S-restricted 3-root Steiner tree exits and therefore no 5-root phylogeny will exist. We now enumerate the possibilities of a critical clique returned by $ALG(T)$.

Lemma 3. *Given a tree chordal graph G with at least two critical cliques, $ALG(G)$ leaves the representatives of any critical cliques in the 3-root Steiner tree T in exactly one of the follow states:*

c1: Representatives adjacent to a Steiner point of Steiner degree one; nearest representative of another critical clique is at distance of three with:
  a: a single representative, or
  b: two representatives,
c2: One representative adjacent to a degree two Steiner point, with:
  a: nearest representative of another critical clique is at distance of three, or
  b: nearest representative of another critical clique is at distance of two,
c3: One representative at distance of one to another leaf critical clique and a Steiner point, other critical cliques are at a distance of three,
c4: One representative adjacent to one a representative of another critical clique.
c5: Two adjacent representative; one adjacent to another leaf's representative.

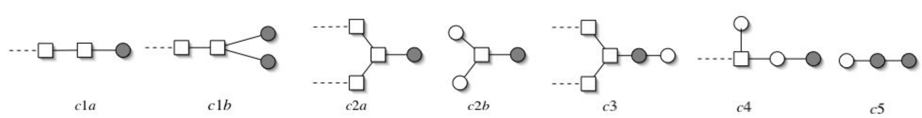


Fig. 3. The seven possible cases for a leaf critical clique in a tree chordal graph as returned by algorithm ALG; representatives are darkened. We denote these cases as $c1a$, $c1b$, $c2a$, etc. Notice only $c2a$ and $c2b$ are unconstrained.

Proof. The following correspondences are for the trivial tree chordal graphs in $TrivAlg(T)$: cases 1a and 2a will correspond to $c4$, case 1b will correspond to $c1b$, case 2b will correspond to $c5$, and the case 2c will correspond to $c2b$.

As the $ALG(G)$ modifies all size one internal nodes to size two, Theorem 3.2 of [7] gives internal nodes of size at least two a single representative ($c2a$), external nodes of size 1, 2 or 3 that were modified to have four representatives ($c1a$) and external nodes of size at least four by two representatives adjacent to a single Steiner point ($c1b$) in a 3-root Steiner tree. Increase the size of all size one external nodes to size two; by Lemma 3.9 in [7] leaves a size two external node as a single representative adjacent to a representative of an internal critical clique ($c4$). The algorithm modifies all size two or more external critical cliques to size at least four; in Theorem 3.2 these will have two representatives adjacent to a Steiner point as in $c2a$. □

2.2 Structure of large maximal cliques

Lemma 4. *[4] Suppose a graph G has a 3rd Steiner root T. Assume there exists in G three maximal cliques K_1, K_2, K_3 such that $K_1 \cap K_2 = I_1 \neq \emptyset$, $K_2 \cap K_3 = I_3 \neq \emptyset$, and $K_1 \cap K_3 = \emptyset$. Let $I_2 = K_2 - I_1 - I_3$. If $I_1 = \{u_1, u'_1\}$, $I_3 = \{u_3, u'_3\}$, and $|I_2| > 0$, then u_1-u'_1-u'_3-u_3 is a path in T and every representative for a critical clique in I_2 is adjacent to either u'_1 or u'_3.*

Lemma 4 is an example of structure that is a potential problem for construction of a S-restricted 3-root Steiner tree. The follow section shows how it becomes unnecessary in the construction of such a Steiner tree.

Lemma 5. *[4] Suppose a graph G has a 3rd Steiner root T, then all maximal cliques with cardinality of 3 or more will have either have exactly two critical cliques each with two representatives as in lemma 4, or will have at most one internal critical clique with two are more representatives in T.*

Lemma 6. *For any maximal clique K represented by the situation of lemma 4 there exists an equivalent representation with a central Steiner point adjacent to the representatives of its critical cliques in K.*

Proof. Given the structure as in lemma 4 identify u_1 and u'_1 into one representative, u^*_1, analogously with u_3 and u'_3, produce u^*_3. Create a new Steiner point s and make u^*_1, u^*_3 and all representatives for I_2 adjacent to s. Notice that since all critical cliques that we adjacent to exactly I_1 we at a distance of 3 from i'_1 are now that distance from s. The same follows for all critical cliques adjacent to i'_3 and critical cliques in I_2. Therefore, the new structure is equivalent to the old tree. □

The following corollary follows easily from Lemmas 5 and 6.

Corollary 1. *Suppose graph G has a 3rd Steiner root tree T, then there exists a representation in which all maximal cliques with cardinality of 3 or more have at most one internal critical clique with two or more representatives.*

3 5PRP on strictly chordal graphs

This section deals with the combination of the Steiner trees returned by $ALG(G)$ and progresses from the most trivial case to the complete case: the solution of the 5-root phylogeny problem over strictly chordal graphs.

3.1 Restriction 1

A *small leaf* is an external critical clique of size 1 in a maximal clique of cardinality at least three. For the following section we will assume that the input graph G contains no small leaves and large maximal cliques K contain at most one critical clique which constrained. Therefore, at most one critical clique in a large maximal clique will be as in $c1a$, $c1b$, $c3$, $c4$, or $c5$. $c4$ is a very restrictive case as the following lemma shows.

Lemma 7. *Given a strictly chordal graph G with a valid S-restricted 3-root Steiner tree T with $S = \emptyset$ and a maximal clique K. If K has an internal critical clique as in $c4$ then the critical clique must be part of at least two maximal cliques with cardinality three or more with other critical cliques unconstrained.*

Proof. The unique representative r of the critical clique in $c4$ is adjacent to another critical clique's representative, therefore, all large maximal cliques that r is a part of must be wide. This implies, that each of these critical cliques is unconstrained and a path of two Steiner point must be between r and these critical cliques. The size of r must be 1, as if had size 2 or more it would have been as in $c1a$ by $ALG(G)$. There exists a representation where all share the Steiner point directly adjacent to r as their path to r, as otherwise, the Steiner point adjacent to r will have degree 2. □

Case $c1a$ leaves has a single representative adjacent to Steiner point with Steiner degree one; as such, if the the single representative corresponding critical clique C has size 1 then we must increase the degree of both this representative and the Steiner point. It follows that C must be part of at least three maximal cliques; the follow operation show how to do this degree increase.

Definition 1 (Operation 1). *Given a critical clique C part of at least three maximal cliques $K_1, K_2, ..., K_n$. If K_1 is as in c1a or c2b and K_2 has all critical cliques unconstrained then assign C a single representative and let the Steiner point adjacent to C in K_1 be adjacent to the Steiner point from K_2 such that all its critical cliques are at distance exactly three from C. For $K_3, ..., K_n$, C now corresponds to c2a and is unconstrained.*

If a critical clique needs to have Operation 1 performed, check that there exists a maximal clique containing it that has all critical cliques unconstrained. If no maximal clique exists, we check if an adjacent critical clique C' is as in $c1a$ or $c1b$, and apply Operation 1. In a similar fashion continue searching for a resolvable path through maximal cliques. Note that such a search is a depth first search through the tree, and in the worst case, and has a linear runtime. Notice that the choice made to change a path by Operation 1 will never effect another path as the search will assign a single representative for a critical clique, and this critical clique will be now be unconstrained. Therefore, we pick the first resolvable path.

Theorem 1. *Given a connected strictly chordal graph G such that G contains no small leaves and large maximal cliques contain at most one constrained critical* *clique, there exists a polynomial time algorithm to recognize whether G has 5-root phylogeny tree T. If T exists, it can be constructed simultaneously.*

Proof. Given G, find $CC(G)$ and create the forest $\mathcal{T}$ by decomposing $CC(G)$, let set S correspond to nodes in G of size 1 in $CC(G)$. For each $T_i \in \mathcal{T}$ find the corresponding 3-root Steiner tree S_i; if one does not exists, by Lemma 1, return no. For each maximal clique K_i create a Steiner point s_i. For each unconstrained critical clique $C \in K_i$, attach its representative to s_i. By Lemma 7, each critical clique as in $c4$ is contained in at least two large maximal cliques, if not, return no; it follows by the precondition, that each of these maximal cliques will have all other critical clique with a single representative. By Lemma 7, create a Steiner point, s for a critical clique as in $c4$, place s adjacent to the critical cliques representative and to all s_i for each K_i. s will have degree at least three as there is at least two maximal cliques K_i. For a critical clique C in K_i as in $c1a$, where $|C| = 1$, apply Operation 1 by searching, if necessary. For a critical clique C in K_i as in $c1a$ with $|C| > 1$, $c1b$, or $c3$ connect s_i to the Steiner point adjacent to C's representative. For critical clique in K_i as in $c5$ connect s_i to the degree one representative.

Our built 3-root Steiner tree T' is now connected as G was connected and we have connected all the tree chordal graphs by their maximal cliques. As each critical clique had at most one constrained critical clique, each maximal clique will have diameter at most 3 in T'; this satisfies Lemma 2. As the minimum diameter of a maximal clique in T' is 2 and as $c4$ is the only case where a representative is adjacent to another representative, it follows that all nonadjacent critical clique's representatives are at distance at least 4. All size one representatives in each T_i will all be internal now as we assumed no small leaves exist. Therefore the algorithm, produces a S-restricted 3-root Steiner tree T'. To produce the 5-root phylogeny tree T, we replace each representative with a Steiner point and place the representatives adjacent to this Steiner point. By Lemma 1, we have a valid 5-root phylogeny tree T. This construction is polynomial, as it calls ALG(G) at most once for each critical clique and performs a linear amount of work for each of these cliques. □

3.2 Restriction 2

In following section, we assume that the input graph G contains no small leaves. A strictly chordal graph may have a large maximal clique having more than one constrained critical clique; if all except one cannot be modified to be unconstrained, then the algorithm returns no, by Corollary 1. If a critical clique in a 3-root Steiner tree is as $c4$, Lemma 7 forces the structure for all maximal cliques it is contained in; $c3$ and $c5$ are similarly restrictive.

Lemma 8. *Given a strictly chordal graph G with a valid S-restricted 3-root Steiner tree T with $S = \emptyset$ and a maximal clique K. If K has critical clique C as in $c3$ or $c5$ of Lemma 3 then any other maximal cliques with cardinality three or more containing C will have all critical cliques as unconstrained.*

Proof. As a representative of the critical clique C is adjacent to a critical clique not in K; any other critical clique that is adjacent to C must be at distance of exactly three from this representative. Therefore, similar to Lemma 7 assign these critical cliques a single representative. □

Thus, given a representative as in cases $c3$, $c4$, or $c5$, we can immediately decide if the maximal clique can be recombined. As $c2a$ and $c2b$ both have a single representative adjacent to a Steiner point of degree at least three, we now deal with the cases $c1a$ and $c1b$.

Lemma 9. *Given a strictly chordal graph G with a corresponding 3-root Steiner tree T and a critical clique C, where C is part of maximal cliques $K_1, K_2, ..., K_n$ and K_1 is as in* c1a *or* c1b*, then at least one maximal clique must have all critical cliques other than C unconstrained.*

Proof. If all maximal cliques have two constrained critical clique, then at least one maximal clique will have diameter of four. □

Theorem 2. *Given a connected strictly chordal graph G, G contains no small leaves, there exists a polynomial time algorithm to recognize whether G has 5-root phylogeny tree T. If T exists, it can be constructed simultaneously.*

Proof. We proceed as Theorem 1 until we recombine large maximal cliques. For a maximal clique that contains a critical clique as in $c3$, $c4$ or $c5$, by Lemmas 7 and 8 we know that all other critical cliques must have a single representative; if not, no 5-root phylogeny will exit. For a critical clique C in K_i as in $c1a$, where $|C| = 1$, apply Operation 1 by searching, if necessary. For each critical clique C as in case $c1a$ with $|C| > 1$ or $c1b$, if it is part of exactly two maximal cliques, then the maximal cliques corresponding to the large maximal clique must have all its critical cliques unconstrained, by Lemma 9; if yes, create a Steiner point and set the Steiner point of the $c1b$ critical clique and each of the representatives for each critical clique adjacent to it.

Let the set M consist of all maximal cliques which contain at least two critical cliques as in case $c1b$ or $c1a$ with $|C| > 1$. Let the set N consist of all maximal cliques which contain a single critical clique as in $c1b$ or $c1a$ with $|C| > 1$ and has all other critical cliques unconstrained. By Lemma 9, find a critical clique $C \in N \cap M$; perform Operation 1 with searching on this maximal clique and, if possible, remove C from M. Continue until either M is empty, the algorithm has resolved all maximal cliques in M, or $N \cap M$ contains no such C, and therefore, no phylogeny tree will exist. The algorithm removes one C from the list each time, and we will have at most $O(|V|)$ searches of the maximal cliques, therefore, this will run in polynomial time.

When $M = \emptyset$ every maximal clique will only contain critical cliques that are unconstrained, therefore create a Steiner point for each maximal clique and connect each representative to it. We will now have S-restricted 3-root Steiner tree and, thus, the 5-root phylogeny tree. The construction is polynomial as we use the polynomial construction from Theorem 1 and the searching of M takes polynomial time. □

3.3 No restrictions

In any maximal clique there exists at most one leaf critical clique; otherwise, these multiple critical clique would have the same set of neighbors, and therefore, would be a larger critical clique. [7] shows a size one leaf critical clique could never exist in a maximal clique of cardinality two. For the remainder of the paper let S contain all size one critical cliques in G.

Lemma 10. *Given a strictly chordal graph G and a corresponding S-restricted 3-root Steiner tree T, if there exists a small leaf l in a maximal clique K, then:*

1. *l is internal in T,*
2. *each critical clique $C \in K \setminus l$ has all adjacent critical cliques not in K at a distance of at least 2 in T,*
3. *at least one critical clique $C \in K \setminus l$ has all adjacent critical cliques not in K at a distance of 3 in T, and*
4. *every critical clique in K has a single or 2 adjacent representatives.*

Proof. Assume G has such a maximal clique l, by definition of S-restricted 3-root Steiner, must be internal. As the maximum diameter of a maximal clique is 4 and l is internal, any critical clique not in K adjacent to a critical clique K would be adjacent to l, thus claim two holds. The third claim follows as l is internal and therefore is adjacent to at least one other critical clique C; critical cliques adjacent to C must be at distance of three from C in T, otherwise, adjacent to l. If K has an internal critical clique, C, with at least two nonadjacent representatives in T, l will be adjacent to the same center Steiner point or representative of the critical clique and will be adjacent to all C's neighbors. A critical clique not in K can be adjacent to a critical clique with a single representatives or two adjacent representatives and not to l, thus, the fourth claim holds. □

By this lemma, a critical clique C in a maximal clique containing a small leaf can be as in $c1a$, $c1b$, $c2a$ or $c2b$. $c4$ is impossible as the critical clique is adjacent to another critical clique failing to satisfy condition 2. $c3$ is impossible as the small leaf would have to be at a distance of exactly three from single representative, but then it would be a leaf in T, failing to satisfy condition 1. Similarly, $c5$ a small leaf would be distance three from the degree two representative but then a leaf in T.

We now introduce two operations to change a critical clique to satisfy condition 3. The algorithm applies these operations if no suitable critical clique exists to satisfy condition 3 of Lemma 10. Notice that only one of these operations can apply to a set of maximal cliques. In addition, the search as done for Operation 1 can be applied to these operations.

Definition 2 (Operation 2a). *Given a critical clique, C, which is part of at least three maximal cliques, $K_1, K_2, ..., K_n$. If at most one of $K_2, ..., K_n$ was part of a decomposed tree chordal graph with C as in $c1a$, $c2b$ $c2a$, or $c2b$ and all critical cliques in the remainder are unconstrained. Then give all critical cliques C a single representative and let the Steiner point adjacent to C be adjacent to* *the Steiner points from $K_2, ..., K_n$ such that each remaining critical clique is at distance exactly three from C.*

Definition 3 (Operation 2b). *Given a critical clique, $|C| \geq 2$, which is part of exactly two large maximal cliques K_1 and K_2. If all critical cliques in K_1 and K_2 other than C are unconstrained then create two Steiner points, p_1 and p_2; let all critical cliques in K_1 other than C be adjacent to p_1, all critical cliques in K_2 other than C be adjacent to p_2, and give C two adjacent representatives where one is adjacent to p_1 and the other to p_2.*

Lemma 11. *Given a graph with a S-restricted 3-root Steiner tree and a large maximal clique K containing a small leaf l, then one critical clique $C \in K \setminus l$ can have Operation 2a applied, Operation 2b applied, or is as $c2a$.*

Proof. Lemma 10 case 3 shows that at least one critical clique must all adjacent critical cliques not in K at distance exactly 3, the above enumerates the possibilities. To see that there does not exists any other situations to consider, we note that cases $c1a$, $c1b$, $c2a$, $c2b$ are the only cases for a maximal clique with a small leaf. Operation 2a and 2b show how to construct such a distance 3 situation. For $c2b$, representing the critical clique by two adjacent representatives will force all critical cliques in K to be wide in their maximal cliques; thus, all will have to be as in the lemmas three cases. □

Lemma 12. *Given a strictly chordal graph G and a corresponding S-restricted 3-root Steiner tree T, if there exists a small leaf l in a maximal clique K, then:*

1. *if $card(K) = 3$ and there exists exactly one critical clique $C \in K \setminus l$ with two adjacent representative then all other critical cliques have all adjacent critical cliques not in K at a distance of exactly 3 in T,*
2. *if $card(K) = 3$ and no critical clique $C \in K \setminus l$ has two adjacent representative then all critical cliques having all adjacent critical cliques not in K at a distance of exactly 3 in T, and*
3. *if $card(K) \geq 4$ then there exists a critical clique $C \in K \setminus l$ with Operation 2a applicable or C is as $c2a$.*

Proof. When $card(K) = 3$ and l is internal, at least one of the critical clique must be adjacent to l, as it is internal. If the other critical clique is not adjacent and does not have two adjacent representatives then the Steiner point adjacent to l will have degree two, a contradiction. Thus, the first claim holds. In case 2, as every critical clique has two adjacent representative or a single representative, the two non-leaf critical cliques have single representatives that are adjacent to the small leaf, otherwise, the maximal clique has width four. It follows from Lemma 3, that Operation 2a and $c2a$ represent the only cases when the critical clique C in its other maximal cliques. When $card(K) \geq 4$, at least one critical clique representative r will have a single representative adjacent to l, otherwise, not internal. All critical cliques adjacent to r which are not in not in K must be at distance of three, otherwise, adjacent to r. Therefore, the third case holds. □

Theorem 3. *Given a connected strictly chordal graph G, there exists a polynomial time algorithm to recognize whether G has 5-root phylogeny tree T. Moreover, such a T can be constructed at the same time, if it exists.*

Proof. Proceed as in Theorem 1 until the recombination of large maximal cliques. Lemmas 7 and 8 handles tree chordal graphs returned as in $c3$, $c4$, and $c5$. Define the set L as all maximal cliques K which contain a small leaf l. All critical cliques as in $c1a$ and $c1b$ which are in a maximal clique in L will need to be changed using either Operation 1 or 2a with searching. For each $K \in L$ such that $card(K) = 3$ if one critical clique C in K is as $c2b$ and in no other maximal clique, then if $|C| > 1$ then given C two adjacent representatives such that one is adjacent to l. All other critical cliques must be one of the choices in Lemma 11, otherwise return no. If $|C| = 1$ then no S-restricted 3-root Steiner tree will exist by Lemma 12. Otherwise by Lemma 12 both critical cliques must be one of the choices in Lemma 11, otherwise return no. Perform operation if needed and set representatives adjacent to l.

All maximal cliques in L now have cardinality at least 4. If any critical clique is as in $c2a$ and contained in exactly two maximal cliques, then set the leaf adjacent to it and all remaining critical cliques single representative adjacent to a Steiner point adjacent to the leaf. For a maximal clique containing multiple critical cliques as in $c1a$ or $c1b$, first apply Operation 2a, if possible, and then, apply Operation 1 if possible. If neither operation is applicable, then no tree exists by Lemma 10. Every maximal clique in L must have a critical clique changed by Operation 2a, otherwise by Lemma 12 no S-restricted 3-root Steiner tree exists; combine these maximal cliques by setting l adjacent to this critical clique. Set the other critical cliques, which are now all necessarily unconstrained, adjacent to a Steiner point adjacent to l. All maximal cliques in L will now be recombined, by Lemma 12.

Finally continue as in Theorem 2 by combining large maximal cliques with more than one constrained critical clique. We produce the 5-root phylogeny tree as in Theorem 1. This construction adds a linear amount of work for each maximal clique containing a small leaf. Therefore, the algorithms overall runtime is still polynomial. □

References

1. A. Brandstadt, V. B. Le, and J. P. Spinrad. *Graph Classes: A Survey.* SIAM, Monographs on Discrete Mathematics and Applications, SIAM, Philadelphia, 1999.
2. R. Haenni and N. Lehmann. Efficient hypertree construction. Technical Report 99-3, Institute of Informatics, University of Fribourg, 1999.
3. P. E. Kearney and D. G. Corneil. Tree powers. *Journal of Algorithms*, 29(1):111 – 131, 1998.
4. W. Kennedy, G. Lin, and G. Yan. Strictly chordal graphs are leaf powers. *Journal of Discrete Algorithms*, 2005.
5. H. Kong and G. Yan. Algorithm for phylogenic 5-root problem. 2003.

6. G. Lin, T. Jiang, and P. E. Kearney. Phylogenetic k-root and steiner k-root. In *The 11th Annual International Symposium on Algorithms and Computation (ISAAC 2000)*, volume 1969, pages 539–551, 2000.
7. G. Lin, W. Kennedy, H. Kong, and G. Yan. The 5-root phylogeny construction for tree chordal graphs. 2005.
8. Y. L. Lin and S. S. Skiena. Algorithms for square roots of graphs. *SIAM Journal on Discrete Mathematics*, 8:99 – 118, 1995.
9. R. Motwani and M. Sudan. Computing roots of graphs is hard. *Discrete Applied Mathematics*, 54:81 – 88, 1994.