2004－AL－94　(5)
2004／3／19

Computing Bounded-Degree Phylogenetic Roots of Disconnected Graphs

Zhi-Zhong Chen *　　Tatsuie Tsukiji†

Abstract

The Phylogenetic kth Root Problem (PRk) is the problem of finding a (phylogenetic) tree T from a given graph $G = (V, E)$ such that (1) T has no degree-2 internal nodes, (2) the external nodes (*i.e.* leaves) of T are exactly the elements of V, and (3) $(u, v) \in E$ if and only if the distance between u and v in tree T is at most k, where k is some fixed threshold k. Such a tree T, if exists, is called a *phylogenetic* kth *root* of graph G. The computational complexity of PRk is open, except for $k \leq 4$. Recently, Chen *et al.* investigated PRk under a natural restriction that the maximum degree of the phylogenetic root is bounded from above by a constant. They presented a linear-time algorithm that determines if a given *connected* G has such a phylogenetic kth root, and if so, demonstrates one. In this paper, we supplement their work by presenting a linear-time algorithm for *disconnected* graphs.

1 Introduction

The reconstruction of evolutionary history for a set of species from quantitative biological data has long been a popular problem in computational biology. This evolutionary history is typically modeled by an evolutionary tree or *phylogeny*. A phylogeny is a tree where the leaves are labeled by species and each internal node represents a speciation event whereby a hypothetical ancestral species gives rise to two or more child species. Proximity within a phylogeny in general corresponds to similarity in evolutionary characteristics. Both rooted and unrooted trees have been used to describe phylogenies in the literature, although they are practically equivalent. In this paper, we will consider only unrooted phylogenies for the convenience of presentation. Note that each internal node in a phylogeny has at least 3 neighbors.

Many approaches to phylogenetic reconstruction have been proposed in the literature [8]. In particular, Lin *et al.* [4] recently suggested a graph-theoretic approach for reconstructing phylogenies from similarity data. Specifically, interspecies similarity is represented by a graph G where the vertices are the species and the adjacency relation represents evidence of evolutionary similarity. A phylogeny is then reconstructed from G such that the leaves of the phylogeny are labeled by vertices of G (*i.e.* species) and for any two vertices of G, they are adjacent in G if and only if their corresponding leaves in the phylogeny are at most distance k apart, where k is a predetermined proximity threshold. This approach gives rise to the following algorithmic problem [4]:

> Phylogenetic kth Root Problem (PRk):
> Given a graph $G = (V, E)$, find a phylogeny T with leaves labeled by the elements of V such that for each pair of vertices $u, v \in V$, $(u, v) \in E$ if and only if $d_T(u, v) \leq k$, where $d_T(u, v)$ is the number of edges on the path between u and v in T.

Such a phylogeny T (if exists) is called a *phylogenetic* kth *root*, or a kth *root phylogeny*, of graph G. Graph G is called the kth *phylogenetic power* of T. For convenience, we denote the kth phylogenetic power of any phylogeny T as T^k. That is, $T^k = \{(u, v) \mid u$ and v are leaves of T and $d_T(u, v) \leq k\}$. Thus, PRk asks for a phylogeny T such that $G = T^k$.

*Department of Mathematical Sciences, Tokyo Denki University, Hatoyama, Saitama 350-0394, Japan. Supported in part by the Grant-in-Aid for Scientific Research of the Ministry of Education, Science, Sports and Culture of Japan, under Grant No. 14580390.

†Department of Information Science, Tokyo Denki University, Hatoyama, Saitama 350-0394, Japan

1.1 Previous Results on PRk

PRk was first studied in [4] where linear-time algorithms for PRk with $k \leq 4$ were proposed. At present, the complexity of PRk with $k \geq 5$ is still unknown.

The hardness of PRk for large k seems to come from the unbounded degree of an internal node in the output phylogeny. On the other hand, in the practice of phylogeny reconstruction, most phylogenies considered are trees of degree 3 [8] because speciation events are usually bifurcating events in the evolutionary process. These motivated Chen *et al.* [2] to consider a restricted version of PRk where the output phylogeny is assumed to have degree at most Δ, for some fixed constant $\Delta \geq 3$. We call this restricted version the DEGREE-Δ PRk and denote it for short as ΔPRk.

Chen *et al.* [2] presented a linear-time algorithm that determines, for any input *connected* graph G and constant $\Delta \geq 3$, if G has a kth root phylogeny with degree at most Δ, and if so, demonstrates one such phylogeny. Unfortunately, their algorithm fails when the input graph G is disconnected. One of their open questions asks for a polynomial-time algorithm for disconnected graphs, because the disconnected case is real in biology.

1.2 Other Problems Related to PRk

A graph G is the kth *power* of a graph H (or equivalently, H is a kth *root* of G), if vertices u and v are adjacent in G if and only if they are at most distance k apart in H. An important special case of graph power/root problems is the TREE kTH ROOT PROBLEM (TRk): Given a graph $G = (V, E)$, we wish to find a tree $T = (V, E_T)$ such that $(u, v) \in E$ if and only if $d_T(u, v) \leq k$. If T exists, then it is called a *tree* kth *root*, or a kth *root tree*, of graph G. There is rich literature on graph roots and powers (see [1, Section 10.6] for an overview), but few results on phylogenetic/tree roots/powers. It is NP-complete to recognize a graph power [6]; nonetheless, we can determine if a graph has a kth root tree, for any fixed k, in cubic time [3]. In particular, determining if a graph has a tree square root can be done in linear time [5]. Moreover, Nishimura *et al.* [7] presented a cubic time algorithm for a variant of PRk with $k \leq 4$, where internal nodes of the output phylogeny are allowed to have degree 2.

1.3 Our Contribution

Our result is a linear-time algorithm that determines, for any input *disconnected* graph G and constant $\Delta \geq 3$, if G has a kth root phylogeny with degree at most Δ, and if so, demonstrates one such phylogeny. This answers an open question in [2]. Combining this algorithm with the algorithm in [2] for connected graphs, we obtain the first linear-time algorithm for ΔPRk for any constants $\Delta \geq 3$ and $k \geq 2$. Our algorithm is complicated and it is based on hidden structures of phylogenetic kth roots of disconnected graphs. Moreover, the algorithm needs a linear-time subroutine for solving a certain optimization problem on each connected component of the input disconnected graph. The subroutine is obtained by nontrivially refining the algorithm in [2].

2 Preliminaries

We employ standard terminologies in graph theory. In particular, the subgraph of a graph G induced by a vertex set U of G is denoted by $G[U]$, the degree of a vertex v in G is denoted by $deg_G(v)$, and the distance between two vertices u and v in G is denoted by $d_G(u, v)$. Moreover, for a set W of vertices in a graph $G = (V, E)$, we write $G - W$ for $G[V - W]$. Furthermore, in a rooted tree, each vertex is both an ancestor and a descendant of itself.

For clarity, if $G = (V, E)$ is a graph and $T = (V_T, E_T)$ is a kth root phylogeny of G for some k, then we call the elements of V *vertices* and call those of V_T *nodes*.

In the remainder of this section, fix a graph $G = (V, E)$ and two integers $k \geq 4$ and $\Delta \geq 3$. A *degree-Δ kth root phylogeny* ((Δ, k)-phylogeny for short) of G is a kth root phylogeny T of G such that the maximum degree of a node in T is at most Δ.

A *degree-Δ kth root quasi-phylogeny* ((Δ, k)-QP for short) of G is a tree Q satisfying the following conditions:

- Each vertex of G is a leaf of Q and appears in Q exactly once. For convenience, we call the leaves of Q that are also vertices of G *true leaves* of Q, and call the other leaves of Q *false leaves* of Q.

- The degree of each node in Q is at most Δ.

- For every two vertices u and v in G, u and v are adjacent in G if and only if $d_Q(u,v) \leq k$.
- For each node x of Q that is a degree-2 node or a false leaf in Q, it holds that $\min_{v \in V} d_Q(x,v) \geq \lfloor \frac{k}{2} \rfloor$.
- If Q has no false leaf, then it has at least one node x such that $2 \leq deg_Q(x) \leq \Delta - 1$ and $\min_{v \in V} d_Q(x,v) \geq \lfloor \frac{k}{2} \rfloor$.

The *cost* of Q is $\max\{1, a + 2b\}$, where a is the number of degree-2 nodes in Q and b is the number of false leaves in Q. Q is an *optimal* (Δ,k)-QP of G if its cost is minimized over all (Δ,k)-QPs of G.

Lemma 2.1 *Suppose that $G = (V,E)$ is a connected graph. Let Q be an optimal (Δ,k)-QP of G. Then, the following hold:*

1. *Q has no node x with $\min_{v \in V} d_Q(x,v) > \lfloor \frac{k}{2} \rfloor$.*
2. *For each node x with $deg_Q(x) = 2$ or $deg_Q(x) > 3$, each connected component of $Q - \{x\}$ contains at least one true leaf of Q.*

We classify (Δ,k)-QPs Q into four types as follows.

- Q is *helpful* if it has at most one degree-2 node and has no false leaf.
- Q is *moderate* if it has no degree-2 node but has exactly one false leaf.
- Q is *troublesome* if it has at least two degree-2 nodes but has no false leaf.
- Q is *dangerous* if it has at least one false leaf and the total number of false leaves and degree-2 nodes in Q is at least 2.

A (Δ,k)-QP Q is *unhelpful* if it is not helpful.

For a (Δ,k)-QP Q, we define its *port nodes* as follows. If Q is not helpful, then its port nodes are its false leaves and degree-2 nodes. If Q is helpful and has no degree-2 node, then its port nodes are those nodes x with $\min_{v \in V} d_Q(x,v) \geq \lfloor \frac{k}{2} \rfloor$. If Q is helpful and has a degree-2 node, then it has only one port node, namely, its unique degree-2 node.

A *nonport node* of a (Δ,k)-QP Q is a node of Q that is not a port node of Q.

3 Algorithm for Bounded-Degree PRk

Throughout this section, fix two integers $k \geq 4$ and $\Delta \geq 3$. This section presents a linear-time algorithm for solving ΔPRk.

Let $G = (V,E)$ be the input graph. We assume that G is disconnected; otherwise, the linear-time algorithm in [2] solves the problem. Let $G_1, \ldots, G_\ell$ be the connected components of G. For each integer with $1 \leq i \leq \ell$, let V_i be the vertex set of G_i.

The next lemma can be proved by a complicated dynamic programming.

Lemma 3.1 *For every $i \in \{1, \ldots, \ell\}$, we can decide whether G_i has a (Δ,k)-QP, in $O(|V_i|)$ time. Moreover, if G_i has a (Δ,k)-QP, then we can compute an optimal (Δ,k)-QP of G_i in $O(|V_i|)$ time.*

Lemma 3.2 *If for some $i \in \{1, \ldots, \ell\}$, G_i has no (Δ,k)-QP, then G has no (Δ,k)-phylogeny.*

By Lemmas 3.1 and 3.2, we may assume that for each $i \in \{1, \ldots, \ell\}$, G_i has a (Δ,k)-QP. For each $i \in \{1, \ldots, \ell\}$, let Q_i be the optimal (Δ,k)-QP of G_i computed in Lemma 3.1.

Lemma 3.3 *Suppose that G has a (Δ,k)-phylogeny. Then, G has a (Δ,k)-phylogeny T such that $Q_1, \ldots, Q_\ell$ all are subtrees of T.*

In the remainder of this section, a (Δ,k)-phylogeny of G always means one in which $Q_1, \ldots, Q_\ell$ are subtrees. By Lemma 3.3, we lose no generality. For convenience, we call $Q_1, \ldots, Q_\ell$ the *unitary* (Δ,k)-QPs.

Let T be a (Δ,k)-phylogeny T of G. A *junction node* of T is a node x of T such that no unitary (Δ,k)-QP contains x. A node x of T is *overconnected*, if it satisfies one of the following conditions:

(1) $deg_T(x) > 3$ and x is a junction node of T.

(2) $deg_T(x) > 3$ and x is a port node of some unhelpful Q_i $(1 \leq i \leq \ell)$.

(3) x is a nonport node of some unhelpful Q_i $(1 \leq i \leq \ell)$ and $deg_T(x) > deg_{Q_i}(x)$.

A helpful Q_i $(1 \leq i \leq \ell)$ is *mis-connected* in T, if (i) at least one nonport node of Q_i is adjacent to a node outside Q_i in T, or (ii) there are two or

more nodes x outside Q_i such that x is adjacent to a node of Q_i in T.

A (Δ, k)-phylogeny T of G is *canonical*, if it has no over-connected node and no helpful Q_i $(1 \leq i \leq \ell)$ is mis-connected in T.

Lemma 3.4 *If G has a (Δ, k)-phylogeny, then it has a canonical one.*

In the remainder of this section, a (Δ, k)-phylogeny of G always means a canonical one. By Lemma 3.4, we lose no generality.

3.1 The Case where k is Odd

Throughout this subsection, we assume that k is odd. A *doube* (Δ, k)-QP is a tree $T_{i,j}$ obtained by combining two helpful unitary (Δ, k)-QPs Q_i and Q_j as follows:

1. Select a port node x_i of Q_i, and select a port node x_j of Q_j.
2. Introduce a junction node y, and connect it to both x_i and x_j.

Note that $T_{i,j}$ has exactly one degree-2 node (namely, the junction node y) but has no false leaf. So, $T_{i,j}$ is a helpful (Δ, k)-QP of $G[V_i \cup V_j]$. Moreover, the minimum distance from y to a true leaf in $T_{i,j}$ is exactly $\lfloor \frac{k}{2} \rfloor + 1$ (cf. Statement 1 in Lemma 2.1).

Lemma 3.5 *Suppose that each Q_i $(1 \leq i \leq \ell)$ is helpful or moderate. Then, G has a (Δ, k)-phylogeny if and only if $\ell \geq 2b + 3$, where b is the number of moderate (Δ, k)-QPs among $Q_1, \ldots, Q_\ell$.*

In the sequel, we assume that at least one Q_i $(1 \leq i \leq \ell)$ is troublesome or dangerous (since otherwise Lemma 3.5 solves the problem).

Let T be a (Δ, k)-phylogeny of G. For each dangerous Q_i $(1 \leq i \leq \ell)$, we say that a false leaf x of Q_i is *active* in T, if no connected component of $T - \{x\}$ is a double (Δ, k)-QP. A dangerous Q_i $(1 \leq i \leq \ell)$ is *active* in T if at least one false leaf of Q_i is active in T.

Lemma 3.6 *Suppose that G has a (Δ, k)-phylogeny. Then, G has a (Δ, k)-phylogeny T such that no dangerous Q_i $(1 \leq i \leq \ell)$ is active in T.*

Let I be the set of all $i \in \{1, \ldots, \ell\}$ such that Q_i is dangerous. For each $i \in I$, let t_i be the number of false leaves in Q_i. Let $t = \sum_{i \in I} t_i$. By Lemma 3.6, if G has a (Δ, k)-phylogeny, then there are at least $2t$ helpful unitary (Δ, k)-QPs. So, if there are less than $2t$ helpful unitary (Δ, k)-QPs, then G has no (Δ, k)-phylogeny. In the sequel, we assume that there are at least $2t$ helpful unitary (Δ, k)-QPs. Without loss of generality, we may assume that $Q_1, \ldots, Q_{2t}$ are helpful.

We connect $Q_1, \ldots, Q_{2t}$ to the dangerous unitary (Δ, k)-QPs as follows.

1. Introduce t junction nodes $x_1, \ldots, x_t$, and construct a one-to-one correspondence between them and the t false leaves of the dangerous unitary (Δ, k)-QPs.
2. For each $i \in \{1, \ldots, t\}$, add an edge from x_i to its corresponding false leaf, add an edge from x_i to an (arbitrarily chosen) port node of Q_{2i-1}, and add an edge from x_i to an (arbitrarily chosen) port node of Q_{2i}.

The above modification extends each dangerous unitary (Δ, k)-QP Q_i to a troublesome (Δ, k)-QP R_i. For convenience, let $R_i = Q_i$ for each $i \in \{2t + 1, \ldots, \ell\}$ such that Q_i is not dangerous.

Now, we are left with $R_{2t+1}, \ldots, R_\ell$; none of them is dangerous. Let τ be the number of troublesome (Δ, k)-QPs among $R_{2t+1}, \ldots, R_\ell$. Note that $\tau = |i \in \{1, \ldots, \ell\} \mid Q_i$ is troublesome or dangerous$\}$. So, $\tau \geq 1$. Without loss of generality, we may assume that $R_{2t+1}, \ldots, R_{2t+\tau}$ are troublesome.

By Lemma 3.6, if G has a (Δ, k)-phylogeny, then it has one in which $R_{2t+1}, \ldots, R_\ell$ are subtrees. So, in the remainder of this section, a (Δ, k)-phylogeny of G always means one in which $R_{2t+1}, \ldots, R_\ell$ are subtrees.

A *bridging node* in a (Δ, k)-phylogeny T of G is a node x of T such that no R_i with $2t + 1 \leq i \leq \ell$ contains x. For each (Δ, k)-phylogeny T of G and for each R_i with $2t + 1 \leq i \leq \ell$, each degree-2 node x of R_i is adjacent to exactly one bridging node y in T (by the canonicity of T); we call y the *bridging neighbor* of x in T.

For each (Δ, k)-phylogeny T of G, let $\mathcal{M}(T)$ denote the tree obtained by modifying T by merging each R_i with $2t + 1 \leq i \leq \ell$ into a super-node. For convenience, we abuse the notation to let each R_i also denote the super-node of $\mathcal{M}(T)$ corresponding to R_i. Note that each bridging node of T remains

to be an internal node in $\mathcal{M}(T)$ and the leaves of $\mathcal{M}(T)$ one-to-one correspond to the helpful unitary (Δ, k)-QPs among $R_{2t+1}, \ldots, R_\ell$. Moreover, by the canonicity of T and Statement 1 in Lemma 2.1, no two super-nodes can be adjacent in $\mathcal{M}(T)$.

Lemma 3.7 *If G has a (Δ, k)-phylogeny, then it has one T such that there is a path q in $\mathcal{M}(T)$ on which $R_{2t+1}, \ldots, R_{2t+\tau}$ appear.*

Lemma 3.8 *If G has a (Δ, k)-phylogeny, then it has one T such that some path q in $\mathcal{M}(T)$ satisfies the following three conditions:*

1. *$R_{2t+1}, \ldots, R_{2t+\tau}$ and exactly $\tau - 1$ bridging nodes appear on q.*
2. *No two bridging nodes on q are adjacent in T.*
3. *For each bridging node x on q, there is a helpful unitary (Δ, k)-QP R_i such that x is adjacent to a port node of R_i in T.*

In the remainder of this section, a (Δ, k)-phylogeny of G always means one T such that some path q in $\mathcal{M}(T)$ satisfies the three conditions in Lemma 3.8. We call q the *spine* of $\mathcal{M}(T)$. The following corollary shows that it does not matter in which order $R_{2t+1}, \ldots, R_{2t+\tau}$ appear on the spine.

Corollary 3.9 *Let T be a (Δ, k)-phylogeny of G. Then, for every pair (R_i, R_j) of troublesome (Δ, k)-QPs, there is another (Δ, k)-phylogeny T' of G such that the spine of $\mathcal{M}(T')$ can be obtained from that of $\mathcal{M}(T)$ by exchanging the positions of R_i and R_j.*

The following corollary is obvious and shows that it does not matter via which degree-2 nodes each troublesome R_i is connected to the spine.

Corollary 3.10 *Let T be a (Δ, k)-phylogeny of G. Then, for every troublesome R_i and for every pair (x_1, x_2) of degree-2 nodes of R_i, we can obtain another (Δ, k)-phylogeny T' of G by deleting edges (x_1, y_1) and (x_2, y_2) and adding edges (x_1, y_2) and (x_2, y_1), where y_1 (respectively, y_2) is the bridging neighbor of x_1 (respectively, x_2) in T. Moreover, the spines of $\mathcal{M}(T)$ and $\mathcal{M}'(T)$ are the same.*

By Lemma 3.8, if G has a (Δ, k)-phylogeny, then there are at least $\tau - 1$ helpful unitary (Δ, k)-QPs among $R_{2t+\tau+1}, \ldots, R_\ell$. So, if there are less than $\tau - 1$ helpful unitary (Δ, k)-QPs among $R_{2t+\tau+1}, \ldots, R_\ell$, then G has no (Δ, k)-phylogeny. In the sequel, we assume that there are at least $\tau - 1$ helpful unitary (Δ, k)-QPs among $R_{2t+\tau+1}, \ldots, R_\ell$. Without loss of generality, we may assume that $R_{2t+\tau+1}, \ldots, R_{2t+2\tau-1}$ are helpful unitary (Δ, k)-QPs.

If $\tau \geq 2$, then we connect $R_{2t+1}, \ldots, R_{2t+2\tau-1}$ into a single (Δ, k)-QP $\mathcal{R}$ as follows.

1. Introduce $\tau - 1$ bridging nodes $x_1, \ldots, x_{\tau-1}$.
2. Select a degree-2 node y_{2t+1} of R_{2t+1}, and select a degree-2 node $z_{2t+\tau}$ of $R_{2t+\tau}$.
3. For each i with $2t + 2 \leq i \leq 2t + \tau - 1$, select two degree-2 nodes z_i and y_i of R_i.
4. For each i with $1 \leq i \leq \tau - 1$, add edges (x_i, y_{2t+i}) and (x_i, z_{2t+i+1}), and add an edge from x_i to an (arbitrarily chosen) port node of $R_{2t+\tau+i}$.

If $\tau = 1$, we let $\mathcal{R} = R_{2t+1}$.

Note that $\mathcal{R}$ is a troublesome (Δ, k)-QP. By Lemma 3.8 and Corollaries 3.9 and 3.10, if G has a (Δ, k)-phylogeny, then G has one T such that $\mathcal{R}, R_{2t+2\tau}, \ldots, R_\ell$ are subtrees of T. In the remainder of this section, a (Δ, k)-phylogeny of G always means such a tree T. Let h be the number of degree-2 nodes in $\mathcal{R}$. Let $x_1, \ldots, x_h$ be the degree-2 nodes of $\mathcal{R}$.

Lemma 3.11 *If G has a (Δ, k)-phylogeny, then it has one T such that for all but one $x_i \in \{x_1, \ldots, x_h\}$, the connected component of $T - \{x_i\}$ containing no node of $\mathcal{R}$ is a double (Δ, k)-QP.*

By Lemma 3.11, if G has a (Δ, k)-phylogeny, then there are at least $2h - 2$ helpful unitary (Δ, k)-QPs among $R_{2t+2\tau}, \ldots, R_\ell$. So, if there are less than $2h - 2$ helpful unitary (Δ, k)-QPs among $R_{2t+2\tau}, \ldots, R_\ell$, then G has no (Δ, k)-phylogeny. In the sequel, we assume that there are at least $2h - 2$ helpful unitary (Δ, k)-QPs among $R_{2t+2\tau}, \ldots, R_\ell$. We may further assume that $R_{2t+2\tau}, \ldots, R_{2t+2\tau+2h-3}$ are helpful unitary (Δ, k)-QPs. For each $i \in \{2t + 2\tau, \ldots, 2t + 2\tau + 2h - 3\}$, let z_i be an (arbitrarily chosen) port node of R_i.

We connect $\mathcal{R}, R_{2t+2\tau}, \ldots, R_{2t+2\tau+2h-3}$ into a single (helpful) (Δ, k)-QP $\mathcal{R}'$ by performing the following steps:

1. Introduce $h-1$ bridging nodes $s_1, \ldots, s_{h-1}$.

2. For each $i \in \{1, \ldots, h-1\}$, add edges $(s_i, z_{2t+2\tau+2i-2})$, $(s_i, z_{2t+2\tau+2i-1})$, and (s_i, x_i).

Now, we are left with $\mathcal{R}', R_{2t+2\tau+2h-2}, \ldots, R_\ell$ each of which is helpful or moderate. Moreover, by Lemma 3.11, if G has a (Δ, k)-phylogeny, then it has one in which $\mathcal{R}', R_{2t+2\tau+2h-2}, \ldots, R_\ell$ are subtrees. So, we can modify the proof of Lemma 3.5 to show that G has a (Δ, k)-phylogeny if and only if $a' \geq b'+3$, where a' (respectively, b') is the number of helpful (respectively, moderate) (Δ, k)-QPs among $\mathcal{R}', R_{2t+2\tau+2h-2}, \ldots, R_\ell$.

In summary, we have the following:

Theorem 3.12 *Suppose that k is odd. Then, we can decide if G has a (Δ, k)-phylogeny, and construct one if so, in linear time.*

3.2 The Case where k is Even

Throughout this subsection, we assume that k is even. The contents in this subsection are very similar to those in the last subsection. In particular, the lemmas in this subsection one-to-one correspond to the lemmas in the last subsection. Moreover, the proof of each lemma in this subsection is very similar to (indeed a bit simpler than) its corresponding lemma in the last subsection.

Lemma 3.13 *Suppose that each Q_i ($1 \leq i \leq \ell$) is helpful or moderate. Then, G has a (Δ, k)-phylogeny if and only if $a \geq 2$, where a is the number of helpful (Δ, k)-QPs among $Q_1, \ldots, Q_\ell$.*

In the sequel, we assume that at least one Q_i ($1 \leq i \leq \ell$) is troublesome or dangerous (since otherwise Lemma 3.13 solves the problem).

Let T be a (Δ, k)-phylogeny of G. For each dangerous Q_i ($1 \leq i \leq \ell$), we say that a false leaf x of Q_i is *active* in T, if no connected components of $T - \{x\}$ is a helpful unitary (Δ, k)-QP. A dangerous Q_i ($1 \leq i \leq \ell$) is *active* in T if at least one false leaf of Q_i is active in T.

Lemma 3.14 *Suppose that G has a (Δ, k)-phylogeny. Then, G has a (Δ, k)-phylogeny T such that no dangerous unitary (Δ, k)-QP is active in T.*

Let I be the set of all $i \in \{1, \ldots, \ell\}$ such that Q_i is dangerous. For each $i \in I$, let t_i be the number of false leaves in Q_i. Let $t = \sum_{i \in I} t_i$. By Lemma 3.14, if G has a (Δ, k)-phylogeny, then there are at least t helpful unitary (Δ, k)-QPs. So, if there are less than t helpful unitary (Δ, k)-QPs, then G has no (Δ, k)-phylogeny. In the sequel, we assume that there are at least t helpful unitary (Δ, k)-QPs. Without loss of generality, we may assume that $Q_1, \ldots, Q_t$ are helpful.

We connect $Q_1, \ldots, Q_t$ to the dangerous unitary (Δ, k)-QPs as follows.

1. Construct a one-to-one correspondence between $Q_1, \ldots, Q_t$ and the t false leaves of the dangerous unitary (Δ, k)-QPs.

2. For each $i \in \{1, \ldots, t\}$, add an edge from an (arbitrarily chosen) port node of Q_i to the false leaf corresponding to Q_i.

The above modification extends each dangerous unitary (Δ, k)-QP Q_i to a troublesome (Δ, k)-QP R_i. For convenience, let $R_i = Q_i$ for each $i \in \{t+1, \ldots, \ell\}$ such that Q_i is not dangerous.

Now, we are left with $R_{t+1}, \ldots, R_\ell$; none of them is dangerous. Let τ be the number of troublesome (Δ, k)-QPs among $R_{t+1}, \ldots, R_\ell$. Note that $\tau = |i \in \{1, \ldots, \ell\} \mid Q_i$ is troublesome or dangerous$\}$. So, $\tau \geq 1$. Without loss of generality, we may assume that $R_{t+1}, \ldots, R_{t+\tau}$ are troublesome.

By Lemma 3.14, if G has a (Δ, k)-phylogeny, then it has one in which $R_{t+1}, \ldots, R_\ell$ are subtrees. So, in the remainder of this section, a (Δ, k)-phylogeny of G always means one in which $R_{t+1}, \ldots, R_\ell$ are subtrees.

For each (Δ, k)-phylogeny T of G, let $\mathcal{M}(T)$ denote the tree obtained by modifying T by merging each R_i with $t+1 \leq i \leq \ell$ into a super-node. For convenience, we abuse the notation to let each R_i also denote the super-node corresponding to R_i in $\mathcal{M}(T)$.

Lemma 3.15 *If G has a (Δ, k)-phylogeny, then it has one T such that there is a path in $\mathcal{M}(T)$ on which $R_{t+1}, \ldots, R_{t+\tau}$ appear.*

Lemma 3.16 *If G has a (Δ, k)-phylogeny, then it has one T such that there is a path in $\mathcal{M}(T)$ whose nodes are exactly $R_{t+1}, \ldots, R_{t+\tau}$.*

In the remainder of this section, a (Δ, k)-phylogeny of G always means one T such that

there is a path q in $\mathcal{M}(T)$ whose nodes are exactly $R_{t+1}, \ldots, R_{t+\tau}$. We call q the *spine* of $\mathcal{M}(T)$. Obviously, Corollaries 3.9 and 3.10 still hold even if k is even.

If $\tau \geq 2$, then we connect $R_{t+1}, \ldots, R_{t+\tau}$ into a single (Δ, k)-QP $\mathcal{R}$ as follows.

1. Select a degree-2 node y_{t+1} of R_{t+1}, and select a degree-2 node $z_{t+\tau}$ of $R_{t+\tau}$.

2. For each i with $t+2 \leq i \leq t+\tau-1$, select two degree-2 nodes z_i and y_i of R_i.

3. For each i with $t+1 \leq i \leq t+\tau-1$, add edge (y_i, z_{i+1}).

If $\tau = 1$, we let $\mathcal{R} = R_{t+1}$.

Note that $\mathcal{R}$ is a troublesome (Δ, k)-QP. By Lemma 3.16 and Corollaries 3.9 and 3.10, if G has a (Δ, k)-phylogeny, then G has one T such that $\mathcal{R}, R_{t+\tau+1}, \ldots, R_\ell$ are subtrees of T. In the remainder of this section, a (Δ, k)-phylogeny of G always means such a tree T. Let h be the number of degree-2 nodes in $\mathcal{R}$. Let $x_1, \ldots, x_h$ be the degree-2 nodes of $\mathcal{R}$.

Lemma 3.17 *If G has a (Δ, k)-phylogeny, then it has one T such that for all but one $x_i \in \{x_1, \ldots, x_h\}$, the connected component of $T - \{x_i\}$ containing no node of $\mathcal{R}$ is a helpful unitary (Δ, k)-QP.*

By Lemma 3.17, if G has a (Δ, k)-phylogeny, then there are at least $h-1$ helpful unitary (Δ, k)-QPs among $R_{t+\tau+1}, \ldots, R_\ell$. So, if there are less than $h-1$ helpful unitary (Δ, k)-QPs among $R_{t+\tau+1}, \ldots, R_\ell$, then G has no (Δ, k)-phylogeny. In the sequel, we assume that there are at least $h-1$ helpful unitary (Δ, k)-QPs among $R_{t+\tau+1}, \ldots, R_\ell$. We may further assume that $R_{t+\tau+1}, \ldots, R_{t+\tau+h-1}$ are helpful unitary (Δ, k)-QPs. For each $i \in \{t+\tau+1, \ldots, t+\tau+h-1\}$, let z_i be an (arbitrarily chosen) port node of R_i.

We connect $\mathcal{R}, R_{t+\tau+1}, \ldots, R_{t+\tau+h-1}$ into a single (helpful) (Δ, k)-QP $\mathcal{R}'$ by adding edges $(x_1, z_{t+\tau+1})$, ..., $(x_{h-1}, z_{t+\tau+h-1})$.

Now, we are left with $\mathcal{R}', R_{t+\tau+h}, \ldots, R_\ell$ each of which is helpful or moderate. Moreover, by Lemma 3.17, if G has a (Δ, k)-phylogeny, then it has one in which $\mathcal{R}', R_{t+\tau+h}, \ldots, R_\ell$ are subtrees. So, we can modify the proof of Lemma 3.13 to show that G has a (Δ, k)-phylogeny if and only if $a' \geq 2$, where a' is the number of helpful (Δ, k)-QPs among $\mathcal{R}', R_{t+\tau+h}, \ldots, R_\ell$.

In summary, we have the following:

Theorem 3.18 *Suppose that k is even. Then, we can decide if G has a (Δ, k)-phylogeny, and construct one if so, in linear time.*

References

[1] A. Brandstädt, V. B. Le, and J. P. Spinrad, *Graph Classes: a Survey*, SIAM Monographs on Discrete Mathematics and Applications, SIAM, Philadelphia, 1999.

[2] Z.-Z. Chen, T. Jiang, and G.-H. Lin, *Computing phylogenetic roots with bounded degrees and errors*, SIAM Journal on Computing, 32 (2003) 864–879.

[3] P. E. Kearney and D. G. Corneil, *Tree powers*, Journal of Algorithms, 29 (1998) 111–131.

[4] G.-H. Lin, P. E. Kearney, and T. Jiang, *Phylogenetic k-root and Steiner k-root*, in: The 11th Annual International Symposium on Algorithms and Computation (ISAAC 2000), Lecture Notes in Computer Science, 1969 (2000) 539–551.

[5] Y.-L. Lin and S. S. Skiena, *Algorithms for square roots of graphs*, SIAM Journal on Discrete Mathematics, 8 (1995) 99–118.

[6] R. Motwani and M. Sudan, *Computing roots of graphs is hard*, Discrete Applied Mathematics, 54 (1994) 81–88.

[7] N. Nishimura, P. Ragde, and D. M. Thilikos, *On graph powers for leaf-labeled trees*, in: Proceedings of the 7th Scandinavian Workshop on Algorithm Theory (SWAT 2000), Lecture Notes in Computer Science, 1851 (2000) 125–138.

[8] D. L. Swofford, G. J. Olsen, P. J. Waddell, and D. M. Hillis, *Phylogenetic inference*, in: D. M. Hillis, C. Moritz, and B. K. Mable (Ed.), Molecular Systematics (2nd Edition), Sinauer Associates, Sunderland, Massachusetts, 1996, pp. 407–514.