Parameterized Algorithms for Finding Square Roots

Manfred Cochefert, Jean-François Couturier, Petr Golovach, Dieter Kratsch, Daniël Paulusma

Parameterized Algorithms for Finding Square Roots[*]

Manfred Cochefert[†]   Jean-François Couturier[†]

Petr A. Golovach[‡]   Dieter Kratsch[†]   Daniël Paulusma[§]

Abstract

We show that the following two problems are fixed-parameter tractable with parameter k: testing whether a connected n-vertex graph with m edges has a square root with at most $n-1+k$ edges and testing whether such a graph has a square root with at least $m-k$ edges. Our first result implies that squares of graphs obtained from trees by adding at most k edges can be recognized in polynomial time for every fixed $k \geq 0$; previously this result was known only for $k = 0$. Our second result is equivalent to stating that deciding whether a graph can be modified into a square root of itself by at most k edge deletions is fixed-parameter tractable with parameter k.

1   Introduction

Squares and square roots are classical concepts in graph theory that are defined as follows. The *square* G^2 of a graph $G = (V_G, E_G)$ is the graph with vertex set V_G such that any two distinct vertices $u, v \in V_G$ are adjacent in G^2 if and only if u and v are of distance at most 2 in G. A graph H is a *square root* of G if $G = H^2$. There exist graphs with no square root, graphs with a unique square root as well as graphs with many square roots.

Mukhopadhyay [18] showed in 1967 that a connected graph G with n vertices $v_1, \ldots, v_n$ has a square root if and only if there exists a set of n

[†]Laboratoire d'Informatique Théorique et Appliquée, Université de Lorraine, 57045 Metz Cedex 01, France, {manfred.cochefert, jean-francois.couturier, dieter.kratsch}@univ-lorraine.fr

[‡]Department of Informatics, University of Bergen, PB 7803, 5020 Bergen, Norway, petr.golovach@ii.uib.no

[§]School of Engineering and Computing Sciences, Durham University, Science Laboratories, South Road, Durham DH1 3LE, UK, daniel.paulusma@durham.ac.uk

complete subgraphs $K^1, \ldots, K^n$ of G with $\bigcup_i V_{K^i} = V_G$ such that K^i contains v_i for all $1 \leq i \leq n$, and K^i contains v_j if and only if K^j contains v_i for all $1 \leq i < j \leq n$. This characterization did not yield a polynomial time algorithm for recognizing squares. In fact, in 1994, Motwani and Sudan [17] showed that the SQUARE ROOT problem, which is that of testing whether a graph has a square root, is NP-complete. This fundamental result triggered a lot of research on the computational complexity of recognizing squares of graphs and computing square roots under the presence of additional structural assumptions. In particular, the following two recognition questions have attracted attention; here $\mathcal{G}$ denotes some fixed graph class.

(1) How hard is it to recognize squares of graphs of $\mathcal{G}$?

(2) How hard is is to recognize graphs of $\mathcal{G}$ that have a square root?

Ross and Harary [20] characterized squares of a tree and proved that if a connected graph has a unique tree square root, then this root is unique up to isomorphism. Lin and Skiena [14] gave linear time algorithms for recognizing squares of trees and planar graphs with a square root. The results for trees [14, 20] were generalized to block graphs by Le and Tuy [12]. Lau [10] gave a polynomial time algorithm for recognizing squares of bipartite graphs. Lau and Corneil [11] gave a polynomial time algorithm for recognizing squares of proper interval graphs and showed that the problems of recognizing squares of chordal graphs, squares of split graphs, and chordal graphs with a square root are all three NP-complete. Le and Tuy [13] gave a quadratic time algorithm for recognizing squares of strongly chordal split graphs. Milanic and Schaudt [15] gave linear time algorithms for recognizing trivially perfect graphs and threshold graphs with a square root. Adamaszek and Adamaszek [1] proved that if a graph has a square root of girth at least 6, then this square root is unique up to isomorphism. Farzad, Lau, Le and Tuy [8] showed that recognizing graphs with a square root of girth at least g is polynomial-time solvable if $g \geq 6$ and NP-complete if $g = 4$. The missing case $g = 5$ was shown to be NP-complete by Farzad and Karimi [7].

1.1   Our Results

The classical SQUARE ROOT problem is a decision problem. We introduce two optimization variants of it in order to be able to take a *parameterized* road to square roots. A problem with input size n and a parameter k is said to be *fixed parameter tractable* (or FPT) if it can be solved in time $f(k) \cdot n^{O(1)}$ for some function f that only depends on k. We consider two natural choices for the parameter k for our optimization variants of the SQUARE ROOT problem and in this way obtain the first FPT algorithms for square root problems.

First, in Section 2, we parameterize the MINIMUM SQUARE ROOT problem, which is that of testing whether a graph has a square root with at most

s edges for some given integer s. Because any square root of a connected n-vertex graph G is a connected spanning subgraph of G, every square root of G has at least $n-1$ edges. Consequently, any instance (G, s) of MINIMUM SQUARE ROOT with $s \leq n-2$ is a no-instance if G is connected, which means that we may assume that $s \geq n-1$. Hence, $k = s - (n-1)$ is the natural choice of parameter. Our main result is that MINIMUM SQUARE ROOT is FPT with parameter k [1]. We prove this result by showing that an instance of MINIMUM SQUARE ROOT can be reduced to an instance of a more general problem, in which we impose additional requirements on some of the edges, namely to be included or excluded from the square root. We prove that the new instance has size quadratic in k. In other words, we show that MINIMUM SQUARE ROOT has a generalized kernel of quadratic size (see Section 1.2 for the definition of this notion). This result is further motivated by the observation that MINIMUM SQUARE ROOT generalizes the problem of recognizing squares of trees (take $s = n-1$). A weaker statement of our FPT result is that of saying that the problem of recognizing squares of graphs of the class

$$\mathcal{G}_k = \{G \mid G \text{ is a graph obtainable from a tree by adding at most } k \text{ edges}\}$$

is polynomial-time solvable for all fixed $k \geq 0$. As such, our result can also be seen as an extension of the aforementioned result of recognizing squares of trees [14].

Second, in Section 3, we parameterize the MAXIMUM SQUARE ROOT problem, which is that of testing whether a given graph G with m edges has a square root with at least s edges for some given integer s. We show that this problem is FPT with parameter $k = m - s$. This choice of parameter is also natural, as G has a square root with at least s edges if and only if G can be modified into a square root (of itself) by at most k edge deletions. Hence, our second FPT result can be added to the growing body of parameterized results for graph editing problems, which form a well studied problem area within algorithmic graph theory. In Section 3 we also present an exact exponential time algorithm for MAXIMUM SQUARE ROOT, which could be seen as an improvement of the algorithm implied by the characterization of Mukhopadhyay [18].

In Section 4 we mention a number of relevant open problems.

---

[1] We restrict ourselves to connected graphs for simplicity. We may do this for the following reason. For disconnected n-vertex graphs with $\ell \geq 2$ connected components the natural parameter is $k = s - (n - \ell)$ instead of $k = s - (n-1)$. Because a square root of a graph is the disjoint union of square roots of its connected components, our FPT result for connected graphs immediately carries over to disconnected graphs if we choose as parameter $k = s - (n - \ell)$ instead.

1.2 Preliminaries

We only consider finite undirected graphs without loops and multiple edges. We refer to the textbook by Diestel [5] for any undefined graph terminology and to the textbooks of Downey and Fellows [6], Flum and Grohe [9], and Niedermeier [19] for detailed introductions to parameterized complexity theory.

Let G be a graph. We denote the vertex set and edge set of G by V_G and E_G, respectively. The subgraph of G induced by a subset $U \subseteq V_G$ is denoted by $G[U]$. The graph $G - U$ is the graph obtained from G by removing all vertices in U. If $U = \{u\}$, we also write $G - u$. The *distance* $\mathrm{dist}_G(u, v)$ between a pair of vertices u and v of G is the number of edges of a shortest path between them. The *open neighborhood* of a vertex $u \in V_G$ is defined as $N_G(u) = \{v \mid uv \in E_G\}$, and its *closed neighborhood* is defined as $N_G[u] = N_G(u) \cup \{u\}$. Two vertices u, v are said to be *true twins* if $N_G[u] = N_G[v]$, and u, v are *false twins* if $N_G(u) = N_G(v)$. A vertex u is *simplicial*, if $N_G(u)$ is a clique. The *degree* of a vertex $u \in V_G$ is denoted $d_G(u) = |N_G(u)|$. The maximum degree of G is denoted $\Delta(G) = \max\{d_G(v) | v \in V_G\}$. A vertex of degree 1 is said to be a *pendant* vertex.

Let G be a connected graph. Let $S \subset V_G$, and let X and Y be two disjoint nonempty vertex subsets of $G - S$. Then S is a *separator* of G if $G - S$ is disconnected, S is an *(X,Y)-separator* if $G - S$ has no path that connects a vertex of X to a vertex of Y, and S is a *minimal (X, Y)-separator* if S is an (X, Y)-separator of G and no proper subset of S is an (X, Y)-separator. Moreover, G is *2-connected* if and only if $|V_G| \geq 3$ and G has no separators of size one.

The *union* of two graphs G_1 and G_2 is the graph $(V_{G_1} \cup V_{G_2}, E_{G_1} \cup E_{G_2})$. The graph K_n denotes the complete graph on n vertices. The graph $K_{1,r}$ denotes the star on $r + 1$ vertices.

A well-known technique to show that a parameterized problem Π is fixed-parameter tractable is to find a *reduction to a problem kernel*. This technique replaces an instance (I, k) of Π with a reduced instance (I', k') of Π called a *(problem) kernel* such that the following three conditions hold:

  i) $k' \leq k$ and $|I'| \leq g(k)$ for some computable function g;

  ii) the reduction from (I, k) to (I', k') is computable in polynomial time;

  iii) (I, k) is a yes-instance of Π if and only if (I', k') is a yes-instance of Π.

If we slightly modify this definition by letting the instance (I', k') belong to a different problem than Π, then (I', k') is called a *generalized* kernel for Π in the literature. This concept has been introduced and named *bikernel* by Alon, Gutin, Kim, Szeider and Yeo [3]; a related notion is compression. An upper bound $g(k)$ on $|I'|$ is called the *kernel size*, and a kernel is called

linear if its size is linear in k and *quadratic* if its size is quadratic in k. It is well known that a parameterized problem is fixed-parameter tractable if and only if it has a kernel (see for example [19]).

2 The Minimum Square Root Problem

As discussed in Section 1.1, we consider connected graphs only and parameterize MINIMUM SQUARE ROOT by $k = s - (n-1)$. From now on we denote this problem as

TREES $+ k$ EDGES SQUARE ROOT
Input: a connected graph G and an integer $k \geq 0$
Parameter: k
Question: has G a square root with at most $n - 1 + k$ edges?

We show the following result.

Theorem 1. *The* TREE $+ k$ EDGES SQUARE ROOT *problem can be solved in time* $2^{O(k^4)} + O(n^4 m)$ *on graphs with n vertices and m edges.*

The remainder of this section is organized as follows. In Section 2.1 we show a number of structural results needed to prove Theorem 1. In Section 2.2 we consider the more general problem

TREE $+ k$ EDGES SQUARE ROOT WITH LABELS
Input: a connected graph G, an integer $k \geq 0$ and two disjoint subsets
            $R, B \subseteq E_G$
Parameter: k.
Question: has G a square root H with at most $n - k + 1$ edges, such that
            $R \subseteq E_H$ and $B \cap E_H = \emptyset$?

Note that the sets R and B in this problem are given sets of *required* edges (that have to be in the square root) and *blocked* edges (that are not allowed to be in the square root), respectively. Also note that TREE $+ k$ EDGES SQUARE ROOT WITH LABELS generalizes TREES $+ k$ EDGE SQUARE ROOT; choose $R = B = \emptyset$. We reduce TREE $+ k$ EDGES SQUARE ROOT to TREE $+ k$ EDGES SQUARE ROOT WITH LABELS where the size of the graph in the obtained instance is $O(k^2)$. In other words, we construct a quadratic generalized kernel for TREE $+ k$ EDGES SQUARE ROOT. This means that to solve an instance of TREES $+ k$ EDGE SQUARE ROOT, we can solve the obtained instance of TREE $+ k$ EDGES SQUARE ROOT WITH LABELS by a brute force algorithm. In Section 2.3 we analyze the corresponding running time and complete the proof of Theorem 1.

2.1 Structural Results

We start with the following observation that we will frequently use.

Observation 1. *Let H be a square root of a connected graph G.*

    i) If u is a pendant vertex of H, then u is a simplicial vertex of G.

    ii) If u, v are pendant vertices of H adjacent to the same vertex, then u, v are true twins in G.

    iii) If u, v are pendant vertices of H adjacent to different vertices, then u and v are not adjacent in G unless $H = K_2$.

We now state five useful lemmas, the first two of which, Lemmas 1 and 2, can be found implicitly in the paper of Ross and Harary [20]. Ross and Harary [20] consider tree square roots, whereas we are concerned with finding general square roots. As such we give explicit statements of Lemmas 1 and 2. We also give a proof of Lemma 2 (the proof of Lemma 1 is straightforward).

Lemma 1. *Let H be a square root of a graph G. Let $\{u_1, \ldots, u_r\} \subseteq V_H$ for some $r \geq 3$ induce a star in H with central vertex u_1. Let $u_3, \ldots, u_r$ be pendant and $\{u_2\}$ be a $(\{u_1, u_3, \ldots, u_r\}, V_H \setminus \{u_1, \ldots, u_r\})$-separator of H. Then $\{u_1, \ldots, u_r\}$ is a clique of G, and $\{u_1, u_2\}$ is a minimal $(\{u_3, \ldots, u_r\}, V_G \setminus \{u_1 \ldots, u_r\})$-separator of G.*

Lemma 2. *Let G be a connected graph with a square root H. Let $\{u_1, \ldots, u_r\}$, $r \geq 3$ be a clique in G, such that $\{u_1, u_2\}$ is a minimal $(\{u_3, \ldots, u_r\}, V_G \setminus \{u_1, \ldots, u_r\})$-separator of G. Let $\{x_1, \ldots, x_p\} = N_G(u_1) \setminus \{u_1, \ldots, u_r\}$ for some $p \geq 1$ and $\{y_1, \ldots, y_q\} = N_G(u_2) \setminus \{u_1, \ldots, u_r\}$ for some $q \geq 1$, as shown in Figure 1. Then the following three statements hold:*

    i) $u_1 u_2 \in E_H$ and, either $u_3 u_1, .., u_r u_1 \in E_H$, $u_3 u_2, \ldots, u_r u_2 \notin E_H$, $u_1 x_1, \ldots, u_1 x_p \notin E_H$, and $\{u_2\}$ is a minimal $(\{u_1, u_3, \ldots, u_r\}, V_H \setminus \{u_1, \ldots, u_r\})$-separator in H, or $u_3 u_1, \ldots, u_r u_1 \notin E_H$, $u_3 u_2, \ldots, u_r u_2 \in E_H$, $u_2 y_1, \ldots, u_2 y_q \notin E_H$ and $\{u_1\}$ is a minimal $(\{u_2, .., u_r\}, V_H \setminus \{u_1, .., u_r\})$-separator in H (see Figure 2 i)).

    ii) If u_1, u_2 are true twins in G, then either $u_1 x_1, \ldots, u_1 x_p \in E_H$ or $u_2 x_1, \ldots, u_2 x_p \in E_H$. Moreover, in this case, G is the union of two complete graphs with vertex sets $\{u_1, \ldots, u_r\}$ and $\{u_1, u_2, x_1, \ldots, x_p\}$, respectively, and G has two (isomorphic) square roots with edge sets $\{u_1 u_2, \ldots, u_1 u_r\} \cup \{u_2 x_1, \ldots, u_2 x_p\}$ and $\{u_2 u_1, u_2 u_3, \ldots, u_2 u_r\} \cup \{u_1 x_1, \ldots, u_1 x_p\}$, respectively (see Figure 2 ii)).

    iii) If $N_G[u_2] \setminus N_G[u_1] \neq \emptyset$, then $u_2 u_1, \ldots, u_r u_1 \in E_H$, $u_3 u_2, \ldots, u_r u_2 \notin E_H$, $u_1 x_1, \ldots, u_1 x_p \notin E_H$. Moreover, the graph H' obtained from H by deleting all $u_i u_j$ with $3 \leq i < j \leq r$ is a square root of G (in which $\{u_1, \ldots, u_r\}$ induces a star with central vertex u_1 and with leaves $u_2, u_3, \ldots, u_r$ that are pendant vertices except for u_2 (see Figure 2 iii)).
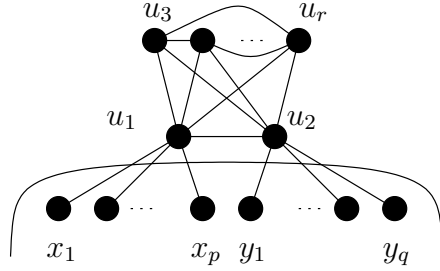
Figure 1: The graph G of Lemma 2. Note that $p \geq 1$ and $q \geq 1$, because $\{u_1, u_2\}$ is a minimal $(\{u_3, \ldots, u_r\}, V_G \setminus \{u_1, \ldots, u_r\})$-separator of G. Also note that $x_i = y_j$ for some $1 \leq i \leq p$ and $1 \leq j \leq q$ is possible.
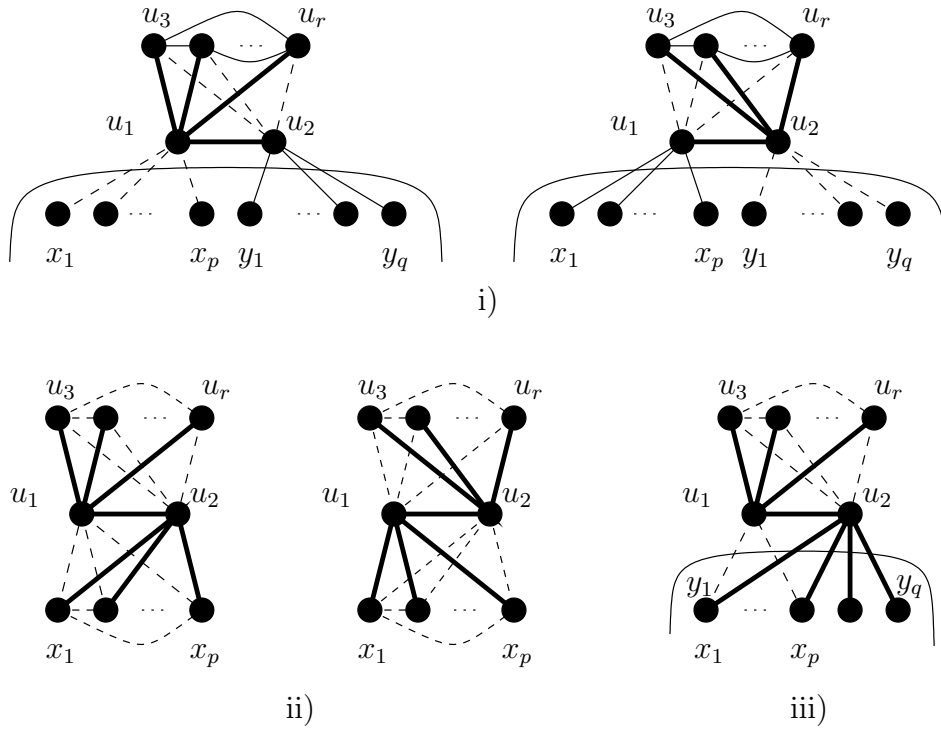
Figure 2: Square roots of G corresponding to statements i)–iii) of Lemma 2, respectively. Edges of G that belong to the square roots are shown by thick lines, whereas edges of G that do not belong to the square roots are shown by dashed lines. In i), edges of G that may be put in a square root of G are shown by thin lines. The square roots in ii) and iii) are the specific square roots defined in statements ii) and iii) of Lemma 2, respectively.

Proof. We first prove i). As $\{u_1, u_2\}$ is a $(\{u_3, \ldots, u_r\}, V_G \setminus \{u_1, \ldots, u_r\})$-separator of G, at least one vertex u_i with $r \geq 3$ is adjacent to one of u_1, u_2 in

H, say to u_1. Then $u_1 x_1, \ldots, u_1 x_p \notin E_H$; otherwise, that is, if u_1 is adjacent to some x_j in H, then $u_i x_j \in E_G$ contradicting the fact that $\{u_1, u_2\}$ is a $(\{u_3, \ldots, u_r\}, V_G \setminus \{u_1, \ldots, u_r\})$-separator of G. Because $u_1 x_1, \ldots, u_1 x_p \notin E_H$, at least one vertex y_h must be adjacent to u_2 in H (as otherwise H is not connected and hence cannot be the square root of G, which is a connected graph). Because $\{u_1, u_2\}$ is a $(\{u_3, \ldots, u_r\}, V_G \setminus \{u_1, \ldots, u_r\})$-separator of G, this means that $u_3 u_2, \ldots, u_r u_2 \notin E_H$. Consequently, $u_1 u_2 \in E_H$ and $\{u_2\}$ is a minimal $(\{u_1, u_3, \ldots, u_r\}, V_H \setminus \{u_1, \ldots, u_r\})$-separator in H. Suppose that there is a vertex u_i, $3 \leq i \leq r$, such that $u_i u_1 \notin E_H$. Since $u_3, \ldots, u_r$ are not adjacent to u_2, it follows that any (u_2, u_i)-path in H has length at least 3, which is not possible as $u_2 u_i \in E_G$. We conclude that $u_3 u_1, \ldots, u_r u_1 \in E_H$. Hence we have shown i).

We now prove ii). Note that $\{x_1, \ldots, x_p\} = \{y_1, \ldots, y_q\}$ with $p = q$. Due to i) either u_1 or u_2 is not adjacent to any x_i. In the first case u_2 must be adjacent to all x_i in H, as otherwise there is no required path of length at most 2 in H between some x_i and u_1. Similarly, in the second case, u_1 must be adjacent to all x_i in H. Hence, $\{u_1, u_2, x_1, \ldots, x_p\}$ is a clique in G. If H has an edge $x_i z$ with $z \notin \{u_1, \ldots, u_r, x_1, \ldots, x_p\}$, then $z u_2 \in E_G$, which is not possible. This means that G is the union of two complete graphs with vertex sets $\{u_1, \ldots, u_r\}$ and $\{u_1, u_2, x_1, \ldots, x_p\}$, respectively. It is readily seen that G has two (isomorphic) square roots with edge sets $\{u_1 u_2, \ldots, u_1 u_r\} \cup \{u_2 x_1, \ldots, u_2 x_p\}$ and $\{u_2 u_1, u_2 u_3, \ldots, u_2 u_r\} \cup \{u_1 x_1, \ldots, u_1 x_p\}$, respectively. Hence we have shown ii).

It remains to prove iii). Let $y_i \in N_G[u_2] \setminus N_G[u_1] \neq \emptyset$. Due to i) we have that $u_1 u_2 \in E_H$, and that either $u_3 u_1, .., u_r u_1 \in E_H$, $u_3 u_2, \ldots, u_r u_2 \notin E_H$, $u_1 x_1, \ldots, u_1 x_p \notin E_H$, or $u_3 u_1, \ldots, u_r u_1 \notin E_H$, $u_3 u_2, \ldots, u_r u_2 \in E_H$, $u_2 y_1, \ldots, u_2 y_q \notin E_H$. If the latter case holds, then any (u_2, y_i)-path in H has length at least 3, which is not possible as $u_2 y_i \in E_G$. Hence the former case must hold. Let H' be a graph obtained from H by deleting all $u_i u_j$ for $i, j \in \{3, \ldots, r\}$. It is readily seen that $H'^2 = H^2 = G$. Hence we have shown iii). $\qquad \square$

Let G be a graph that contains (besides possibly some other vertices) $p + q + r$ distinct vertices $u_1, \ldots, u_r$, $x_1, \ldots, x_p$, $y_1, \ldots y_q$ for some $r \geq 3$, $p \geq 1$ and $q \geq 1$, such that the following conditions hold:

  i) $\{u_1, \ldots, u_r\}$ is a clique in G;

 ii) $\{u_1, u_2, u_3\}$ is a minimal $(\{u_4, \ldots, u_r\}, V_G \setminus \{u_1, \ldots, u_r\})$-separator in G if $r \geq 4$;

iii) $\{u_1, u_3, \ldots, u_r\} \cup \{x_1, \ldots, x_p\} \cup \{y_1, \ldots, y_q\} = N_G(u_2)$;

iv) $\{u_2, u_4, u_5, \ldots, u_r\} = N_G(u_1) \cap N_G(u_3)$;

 v) $\{x_1, \ldots, x_p\} \subseteq N_G(u_1)$ and $\{y_1, \ldots, y_q\} \subseteq N_G(u_3)$;

8

vi) $x_i y_j \notin E_G$ for $i = 1, \ldots, p$ and $y = 1, \ldots, q$.

We call G an *F-graph* and $\{u_1, u_2, u_3\}$ an *F-triple* with *outer vertices* u_1 and u_3, see Figure 3 for an example. Here, F refers to the graph in Figure 4. These notions are further explained by Lemmas 3 and 4.
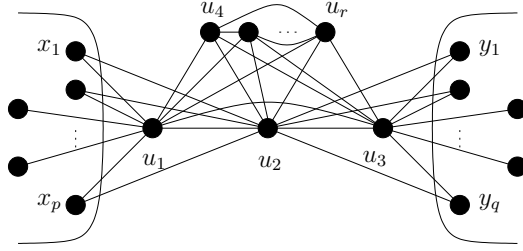

Figure 3: An example of an F-graph with $r \geq 4$. Note that there are no edges between any two vertices x_i and y_j. Also note that the two outer vertices u_1 and u_3 of the F-triple $\{u_1, u_2, u_3\}$ may be adjacent to vertices not adjacent to u_2 (but they may not have any common neighbor in $\{x_1, \ldots, x_p\} \cup \{y_1, \ldots, y_q\}$).
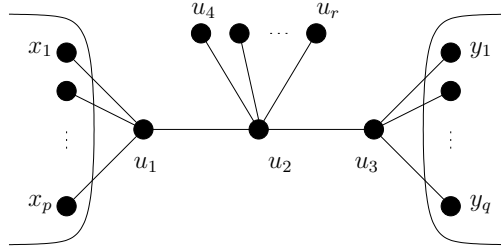
Figure 4: The graph $F = F(p, q, r)$ with $p \geq 1$, $q \geq 1$ and $r \geq 3$; if $r = 3$ then F does not contain any pendant vertices $u_4, \ldots, u_r$. Here, we depicted F as a subgraph of the graph H in Lemma 3. To be more precise, the graph F is exactly the graph with the black edges. In H the vertices $u_1, \ldots, u_r$ have only neighbors that are in F, whereas a vertex x_i or y_j may have one or more neighbors in H that are outside F; however, no x_i and y_j have a common neighbor in H. Moreover, in H, the only edges incident to vertices in F are the black edges depicted (edges of F) and possibly some edges between two vertices x_i, x_j or between two vertices y_i, y_j; such edges have not been depicted in the figure.

Lemma 3. *Let H be a square root of a graph G. Let H contain the graph F of Figure 4 as a subgraph, such that $u_4, \ldots, u_r$ are pendant vertices of H (if $r \geq 4$), $d_H(u_2) = r - 1$, $u_1 u_2 u_3$ is an induced path in H that is not contained in any cycle of length at most 6, $\{x_1, \ldots, x_p\} = N_H(u_1) \setminus \{u_2\}$ and $\{y_1, \ldots, y_q\} = N_H(u_3) \setminus \{u_2\}$. Then G is an F-graph.*

Proof. Conditions (i)-(iii) and (v) are readily seen to hold. Conditions iv) and vi) follow from the condition that the path $u_1u_2u_3$ is not contained in any cycle of length at most 6 in H. $\qquad\square$

Lemma 4. *Let G be a connected F-graph. If H is a square root of G, then the graph F of Figure 4 is a subgraph of H such that $d_H(u_2) = r - 1$, $\{x_1, \ldots, x_p\} = N_H(u_1) \setminus \{u_2\}$ and $\{y_1, \ldots, y_q\} = N_H(u_3) \setminus \{u_2\}$. Moreover, the graph obtained from H by deleting all edges u_iu_j with $4 \leq i < j \leq r$ is a square root of G that contains $u_4, \ldots, u_r$ as pendant vertices (if $r \geq 4$).*

Proof. Let H be a square root of G. We consider the following three cases.

Case 1. $u_1u_2, u_2u_3 \in E_H$. Because $x_1u_3, \ldots, x_pu_3 \notin E_G$, this means that $x_1u_2, \ldots, x_pu_2 \notin E_H$. Symmetrically, $y_1u_2, \ldots, y_qu_2 \notin E_H$. Since each $x_iu_2 \in E_G$ but $x_iu_2 \notin E_H$, H has an (x_i, u_2)-path of length 2. Because $d_G(u_2) = p + q + r - 1$, the middle vertex of this path is in $\{u_1, u_3, \ldots, u_r\}$. Because x_i is not adjacent to $u_3, \ldots, u_r$ in H (as it not so in G), this path goes through u_1. In other words, $x_1u_1, \ldots, x_pu_1 \in E_H$ and, by symmetry, $y_1u_3, \ldots, y_qu_3 \in E_H$. If a vertex $z \notin \{u_2, x_1, \ldots, x_p\}$ is adjacent to u_1 in H, then z is adjacent to both u_2 and x_1 in G. Because $d_G(u_2) = p+q+r-1$, we find that $z \in \{u_3, \ldots, u_r\}$ or $z \in \{y_1, \ldots, y_q\}$. However, none of $\{u_3, \ldots, u_r\}$ is adjacent to x_1, whereas none of $\{y_1, \ldots, y_q\}$ is adjacent to u_2. We conclude that $\{x_1, \ldots, x_p\} = N_H(u_1) \setminus \{u_2\}$ and by using the same arguments that $\{y_1, \ldots, y_q\} = N_H(u_3) \setminus \{u_2\}$.

Now we show that $u_4u_2, \ldots, u_ru_2 \in E_H$. To prove it, assume that some u_i, $4 \leq i \leq r$, is not adjacent to u_2 in H. Then u_1 and u_i are at distance at least 3 in H contradicting $u_1u_i \in E_G$. We already deduced that $x_1u_2, \ldots, x_pu_2 \notin E_H$ and that $y_1u_2, \ldots, y_qu_2 \notin E_H$. By assumption, u_2 is adjacent to both u_1 and u_3. As $d_G(u_2) = p + q + r - 1$, we then find that $d_H(u_2) = r - 1$.

To conclude the proof for this case, it remains to observe that if some u_i, u_j are adjacent in H for $i, j \in \{4, \ldots, r\}$, then the graph H' obtained from H by the removal of these edges is a square root of G.

Case 2. $u_1u_2, u_2u_3 \notin E_H$. Since $u_1u_2 \notin E_H$, $u_1u_2 \in E_G$ and $d_G(u_2) = p + q + r - 1$, there exists a vertex $z \in \{x_1, \ldots, x_p\} \cup \{u_4, \ldots, u_r\}$ such that $u_1z, zu_2 \in E_H$. Because z is not adjacent to $y_1, \ldots, y_q$ in G, we find that $y_1u_2, \ldots, y_qu_2 \notin E_H$. By the same arguments, we obtain $x_1u_2, \ldots, x_pu_2 \notin E_H$. Hence, $z \in \{u_4, \ldots, u_r\}$. By symmetry, some vertex from $\{u_4, \ldots, u_r\}$ is adjacent to u_3 in H. Consequently, each vertex of $\{u_1, u_2, u_3\}$ is adjacent to some vertex in $\{u_4, \ldots, u_r\}$ in H. As $\{u_1, u_2, u_3\}$ separates $\{u_4, \ldots, u_r\}$ from $V_G \setminus \{u_1, \ldots, u_r\}$, this means that H has no edges that join u_1, u_2, u_3 with the vertices of $V_G \setminus \{u_1, \ldots, u_r\}$; a contradiction. Hence, this case is not possible.

By symmetry, it remains to consider the following case.

10

Case 3. $u_1u_2 \in E_H$ and $u_2u_3 \notin E_H$. Because $u_1u_2 \in E_H$ and $y_1u_1, \ldots, y_qu_1 \notin E_G$, we find that $y_1u_2, \ldots, y_qu_2 \notin E_H$. Because $y_1u_2 \in E_G$, this means that H contains a (y_1, u_2)-path of length 2. Because $u_2u_3 \notin E_H$ and $d_G(u_2) = p+q+r-1$, such a path should go through one of the vertices of $\{u_1, u_4, \ldots, u_r\} \cup \{x_1, \ldots, x_p\}$. However, none of these vertices is adjacent to y_1 in G, and consequently not in H either; a contradiction. Therefore, this case is not possible either. □

Lemma 5. *Let u, v be true twins in a connected graph G with at least three vertices. Let G' be the graph obtained from G by deleting v. The following two statements hold:*

i) If H' is a square root of G', then the graph H obtained from H' by adding v with $N_H(v) = N_{H'}(u)$ (that is, by adding a false twin of u) is a square root of G.

ii) If H is a square root of G such that u, v are false twins in H, then the graph H' obtained by deleting v is a square root of G'.

Proof. We first prove i). Let H' be a square root of G', and let H be the graph obtained from H' by adding a false twin v of u. As G is a connected graph with at least three vertices, u is adjacent in H' to some vertex z. Then u and v are adjacent to z in H and thus $d_H(u, v) \leq 2$. Hence, uv is an edge of H^2. Then it is straightforward to see that $G = H^2$. Statement ii) follows from the fact that identifying false twins does not change the distance between any two vertices. □

2.2 Construction of the Generalized Kernel

As discussed, in this section, we reduce TREE + k EDGES SQUARE ROOT to TREE + k EDGES SQUARE ROOT WITH LABELS in such a way that the size of the graph in the obtained instance is $O(k^2)$.

First, we informally sketch the main steps of the reduction. Let G be a connected graph with n vertices, and let k be a positive integer.
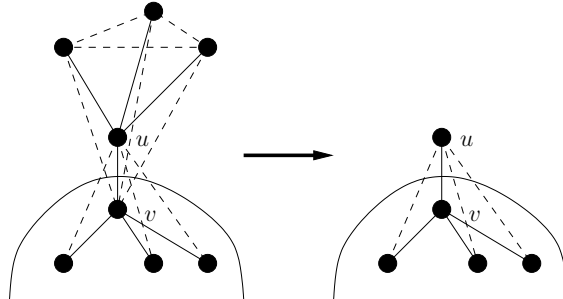

Figure 5: Trimming; the edges of H are shown by the solid lines.

11

Suppose that H is a square root of G with at most $n + k - 1$ edges. If H has a vertex u of degree at least 2 that has exactly one non-pendant neighbor v, then we recognize the corresponding structure in G and delete those vertices of G that are pendant vertices of H adjacent to u as shown in Figure 5, that is, similar to the algorithm of Lin and Skiena [14], we "trim" pendant edges in potential roots. Since the root we are looking for is not a tree, our trimming is more sophisticated and based on Lemmas 1 and 2. We will show that in this way we obtain a graph G' with n' vertices that has the following property: every pendant vertex of any square root H' of G' with at most $n' - 1 + k$ edges is adjacent to a vertex that has at least two non-pendant neighbors in H'.

Suppose that H' has a sufficiently long induced path P, such that every internal vertex of P has exactly two non-pendant neighbors in H'. Let u be an internal vertex of P, and let $x, y \in V_P$ be the two non-pendant neighbors of u. Using Lemmas 3 and 4, we recognize the corresponding structure in G' and modify G' as shown in Figure 6, that is, we delete u in H' and join x any y by an edge. By performing this operation recursively, we obtain a graph G'' with n'' vertices.
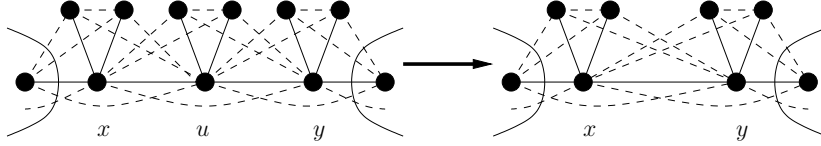

Figure 6: Reduction of paths; the edges of H' are shown by the solid lines.

Suppose that H'' is a square root of G'' with at most $n'' + k - 1$ edges. Let H^* be the graph obtained from H'' by deleting all pendant vertices of H''. Then H^* has no vertices of degree 1, and the length of every path P with internal vertices of degree 2 in H^* is bounded by a constant. This means that the size of H^* is $O(k)$. The vertices of $V_{G''} \setminus V_{H^*}$ are pendant vertices of H''. Consider the set Z of pendant vertices of H'' adjacent to a vertex $u \in V_{H^*}$. Then the vertices of Z are simplicial vertices of G''. Moreover, they are true twins. We use Observation 1 and Lemma 5 to show that we may reduce the number of true twins in G'' if G'' has too many. This results in a graph G''' with n''' vertices such that n''' is $O(k^2)$.

During the reduction from G to G'' we label some edges, that is, we include some edges in sets R or B and, therefore, obtain an instance (G''', k, R, B) of TREE $+ k$ EDGES SQUARE ROOT WITH LABELS.

Before we give a formal description of our reduction, we introduce the following terminology. A square root H of a graph G that has at most $|V_G| - 1 + k$ edges for some $k \geq 0$ is called a *solution* of the instance (G, k) of TREE $+ k$ EDGES SQUARE ROOT. If $R \subseteq E_H$ and $B \cap E_H = \emptyset$ for two disjoint subsets R and B of E_G, then H is also called a *solution* of the

instance (G, k, R, B) of Tree $+$ k Edges Square Root with Labels.

We are now ready to give the exact details of our reduction. Let G be a connected graph with n vertices and m edges, and let k be a positive integer. First we check whether G has a square root that is a tree by using the linear-time algorithm of Lin and Skiena [14]. If we find such a square root, then we stop and return a yes-answer. From now we assume that every square root of G (if there exists one) has at least one cycle.

Because connected graphs that have square roots are 2-connected, we also check whether G is 2-connected. If so, then we stop and return a no-answer. Otherwise we continue as follows. We introduce two sets of edges R and B. Initially, we set $R = B = \emptyset$. Next, we "trim" pendant edges in potential roots, that is, we exhaustively apply the following rule that consists of five steps that must be performed in increasing order.

Trimming Rule

1. Find a pair $S = \{u_1, u_2\}$ of two adjacent vertices such that one connected component of $G - S$ consists of $r \geq 3$ vertices $u_3, \ldots, u_r$ that together with u_1, u_2 form a clique in G.

2. If $N_G[u_1] = N_G[u_2]$ then stop and return a no-answer.

3. If $N_G[u_1] \setminus N_G[u_2] \neq \emptyset$ and $N_G[u_2] \setminus N_G[u_1] \neq \emptyset$, then stop and return a no-answer.

4. If $N_G[u_1] \setminus N_G[u_2] \neq \emptyset$, then rename u_1 by u_2 and u_2 by u_1 (this step is for notational convenience only and has no further meaning).

5. Define sets $R' = \{u_1 u_2, \ldots, u_1 u_r\}$ and $B' = \{u_i u_j \mid 2 \leq i < j \leq r\} \cup \{u_1 x \mid x \in N_G(u_1) \setminus \{u_2, \ldots, u_r\}\}$.

6. If $R \cap B' \neq \emptyset$ or $R' \cap B \neq \emptyset$, then stop and return a no-answer. Otherwise, set $R = R \cup R'$, $B = B \cup B'$, delete $u_3, \ldots, u_r$ from G and also delete all edges incident to $u_3, \ldots, u_r$ from R and B.

Exhaustively applying the trimming rule yields a sequence of instances $(G_0, k, R_0, B_0), \ldots, (G_\ell, k, R_\ell, B_\ell)$ of Tree $+$ k Edges Square Root with Labels for some integer $\ell \geq 0$, where $(G_0, k, R_0, B_0) = (G, k, \emptyset, \emptyset)$ and where $(G_\ell, k, R_\ell, B_\ell)$ is an instance for which we have either returned a no-answer (in steps 2, 3 or 6) or for which there does not exist a set S as specified in step 1. For $0 \leq i \leq \ell - 1$ we denote the sets R' and B' constructed in the $(i + 1)$th call of the trimming rule by R'_i and B'_i, respectively. We need the following lemma.

Lemma 6. *The instance $(G_\ell, k, B_\ell, R_\ell)$ has no solution that is a tree, and G_ℓ is 2-connected. Moreover, $(G_\ell, k, R_\ell, B_\ell)$ has a solution if and only if (G_0, k, R_0, B_0) has a solution. If the trimming rule returned a no-answer for $(G_\ell, k, R_\ell, B_\ell)$, then (G_0, k, R_0, B_0) has no solution.*

Proof. For $0 \leq i \leq \ell$, we use induction to show that the graph G_i is 2-connected and that (G_i, k, B_i, R_i) has no solution that is a tree. Moreover, for all $1 \leq i \leq \ell$, we show that (G_i, k, R_i, B_i) has a solution if and only if $(G_{i-1}, k, R_{i-1}, B_{i-1})$ has a solution. Finally, we prove that if the trimming rule returned a no-answer for $(G_\ell, k, R_\ell, B_\ell)$, then (G_0, k, R_0, B_0) has no solution.

If $i = 0$, then G_i is 2-connected and (G_0, k, B_0, R_0) has no solution that is a tree by our initial assumption (as we had preprocessed G with respect to these two properties). Now suppose that $1 \leq i \leq \ell$. By our induction hypothesis, we may assume that G_{i-1} is 2-connected and that $(G_{i-1}, k, B_{i-1}, R_{i-1})$ has no solution that is a tree.

Because the trimming rule applied on $(G_{i-1}, k, R_{i-1}, B_{i-1})$ yielded a new instance (G_i, k, R_i, B_i), the graph G_{i-1} has a pair $S = \{u_1, u_2\}$ of adjacent vertices such that one connected component of $G - S$ consists of vertices $u_3, \ldots, u_r$ that together with u_1, u_2 form a clique in G_{i-1}. Step 6 implies that $G_i = G_{i-1} - \{u_3, \ldots, u_r\}$. Because we did not return a no-answer for $(G_{i-1}, k, R_{i-1}, B_{i-1})$, we find that $N_{G_{i-1}}[u_1] \subset N_{G_{i-1}}[u_2]$. Hence, G_{i-1} is not a complete graph. Because G_{i-1} is 2-connected, this means that G_i is 2-connected. We now show that any solution for $(G_{i-1}, k, B_{i-1}, R_{i-1})$ corresponds to a solution for (G_i, k, B_i, R_i), and vice versa.

First suppose that H_{i-1} is an arbitrary solution for $(G_{i-1}, k, B_{i-1}, R_{i-1})$. Let $N_{G_{i-1}}(u_1) \setminus \{u_2, \ldots, u_r\} = \{x_1, \ldots, x_p\}$. Because $N_{G_{i-1}}[u_2] \setminus N_{G_{i-1}}[u_1] \neq \emptyset$, we find that $G_{i-1} - \{u_1, u_2\}$ contains at least two connected components. As G_{i-1} is 2-connected, this means that $\{u_1, u_2\}$ is a minimal $\{u_3, \ldots, u_r\}, V_{G_{i-1}} \setminus \{u_1, \ldots, u_r\}$-separator of G_{i-1}. Hence we may apply Lemma 2 iii), which tells us that $u_2 u_1, \ldots, u_r u_1 \in E_{H_{i-1}}$, $u_3 u_2, \ldots, u_r u_2 \notin E_{H_{i-1}}$, and $u_1 x_1, \ldots, u_1 x_p \notin E_{H_{i-1}}$. As $R_i \subseteq R_{i-1} \cup \{u_1 u_2\}$ and $B_i \subseteq B \cup \{u_1 x_1, \ldots, u_1 x_p\}$, this means that the graph obtained from H_{i-1} by deleting $u_3, \ldots, u_r$ is a solution for (G_i, k, R_i, B_i); in particular note that $|E_{H_i}| \leq |E_{H_{i-1}}| - (r - 3) \leq |V_{G_{i-1}}| - 1 + k - (r - 3) = |V_{G_i}| - 1 + k$, as required.

Now suppose that H_i is an arbitrary solution for (G_i, k, R_i, B_i). Then adding the edges $u_1 u_3, \ldots, u_1 u_r$ to H_i yields a graph H that is a square root of G_{i-1}. The edges $u_1 u_3, \ldots, u_1 u_r$ are not in B_{i-1}, as they are in the set R'_{i-1} constructed in step 5 and $R'_{i-1} \cap B_{i-1} = \emptyset$ (otherwise the trimming rule would have stopped when processing $(G_{i-1}, k, R_{i-1}, B_{i-1})$ in step 6). Now suppose that R_{i-1} contains an edge not in H. By definition of R_i, this edge must be between some u_s and u_t with $3 \leq s < t \leq r$. Then $u_s u_t$ belongs to R_i, because it was placed in the set R_h for some $h \leq i - 1$. In step 4 of the corresponding call of the trimming rule, also one of the edges $u_s u_1$ or $u_t u_1$ was placed in B_h. Hence either $u_s u_1$ or $u_t u_1$ belongs to B_{i-1}. This yields a contraction as both $u_s u_1$ and $u_t u_1$ belong to R'_{i-1} and $R'_{i-1} \cap B_{i-1} = \emptyset$ (otherwise the trimming rule would have stopped when processing $(G_{i-1}, k, R_{i-1}, B_{i-1})$ in step 6). Hence, after observing

that $|E_H| = |E_{H_i}| + (r-3) \leq |V_{G_i}| - 1 + k + (r-3) = |V_{G_{i-1}}| - 1 + k$, we conclude that H is a solution for $(G_{i-1}, k, R_{i-1}, B_{i-1})$. We observe that H_i cannot be a tree, as this would imply that H is a tree, which is not possible as $(G_{i-1}, k, R_{i-1}, B_{i-1})$ does not have such a solution.

We are left to show that if the trimming rule returned a no-answer for $(G_\ell, k, R_\ell, B_\ell)$, then (G_0, k, R_0, B_0) has no solution. Due to the above, this comes down to showing that $(G_\ell, k, R_\ell, B_\ell)$ has no solution.

Suppose that the trimming rule returned a no-answer for $(G_\ell, k, R_\ell, B_\ell)$. Then this must have happened in step 2, 3 or 6, thus after step 1. Hence, there exists a pair of adjacent vertices $S = \{u_1, u_2\}$ in G_ℓ, such that one connected component of $G_\ell - S$ has vertex set $\{u_3, \ldots, u_r\}$ and $\{u_1, \ldots, u_r\}$ is a clique.

First assume that S is not a separator of G_ℓ, that is, G_ℓ is a complete graph with vertex set $\{u_1, \ldots, u_r\}$. Then $N_G[u_1] = N_G[u_2]$ (and the no-answer given by the trimming rule happens in step 2). In order to obtain a contradiction, assume that $(G_\ell, k, R_\ell, B_\ell)$ has a solution H. Any star on $|V_{G_\ell}|$ vertices is a square root of G_ℓ with at most $|V_{G_\ell}| - 1 + k$ edges. However, H cannot be such a star, as $(G_\ell, k, R_\ell, B_\ell)$ has no solution that is a tree. Hence, $R_\ell \neq \emptyset$ or $B_\ell \neq \emptyset$. Recall that $B_0 = R_0 = \emptyset$. Hence, $\ell \geq 1$, and non-emptiness of R_ℓ or B_ℓ must have been obtained in a previous call of the trimming rule, say in the $(h+1)$th call of the trimming rule for some $0 \leq h \leq \ell - 1$. By definition of steps 5 and 6, we find that $B_h \neq \emptyset$ implies that $R_h \neq \emptyset$. Hence, $R_h \neq \emptyset$. Let $u_i u_j \in R_h$. By steps 5 and 6, this edge has an end-vertex, say u_i, such that $u_i u_s \in B_\ell$ for all $s \in \{1, \ldots, r\} \setminus \{i, j\}$. Consequently, $u_j u_s \in E_H$ for all $s \in \{1, \ldots, r\} \setminus \{j\}$. Because the star with central vertex u_j and leaves $V_{G_\ell} \setminus \{u_j\}$ is not a solution for $(G_\ell, k, R_\ell, B_\ell)$, there must be an edge $u_s u_t \in R_\ell$ with $s, t \in \{1, \ldots, r\} \setminus \{j\}$. However then, due to steps 5 and 6, $u_j u_s \in B_\ell$ or $u_j u_t \in B_\ell$, that is, at least one of these edges cannot be in H; a contradiction.

Now assume that S is a separator of G_ℓ. Because G_ℓ is 2-connected, both u_1 and u_2 have at least one neighbor in $V_{G_\ell} \setminus \{u_1, \ldots, u_r\}$. Hence $\{u_1, u_2\}$ is a minimal separator (and we may apply Lemma 2 in the remainder). Recall that the trimming rule only returns a no-answer in steps 2, 3, or 6. We consider each of these three cases separately.

Case 1. The no-answer is given in step 2. Then $N_G[u_1] = N_G[u_2]$. By Lemma 2 i) and ii), G_ℓ is the union of two cliques $\{u_1, \ldots, u_r\}$ and $\{u_1, u_2, x_1, \ldots x_p\}$ where $\{x_1, \ldots, x_p\} = N_G(u_1) \setminus \{u_2, \ldots, u_r\}$. In order to obtain a contradiction, suppose that $(G_\ell, k, R_\ell, B_\ell)$ has a solution H. By Lemma 2 i) and ii), we may assume without loss of generality that $u_1 u_2, \ldots, u_1 u_r \in E_H$, $u_2 u_3, \ldots, u_2 u_r \notin E_H$, $u_1 x_1, \ldots, u_1 x_p \notin E_H$ and $u_2 x_1, \ldots, u_2 x_p \in E_H$. Recall that $(G_\ell, k, R_\ell, B_\ell)$ has no solution that is a tree. Hence, there exists an edge $u_i u_j \in R_\ell$ for some $i, j \in \{2, \ldots, r\}$ or an edge $x_i x_j \in R_\ell$ for some $i, j \in \{x_1, \ldots, x_p\}$. By symmetry, we only need to

15

consider the case $u_i u_j \in R_\ell$. This edge was placed in R_ℓ in some previous call of the trimming rule. However, due to steps 5 and 6 performed in that call, we find that $u_i u_1 \in B_\ell$ or $u_j u_1 \in B_\ell$, that is, at least one of these two edges cannot be in H; a contradiction.

Case 2. The no-answer is given in step 3. Then we have $N_G[u_1] \setminus N_G[u_2] \neq \emptyset$ and $N_G[u_2] \setminus N_G[u_1] \neq \emptyset$. Due to Lemma 2 i) and iii), $(G_\ell, k, R_\ell, B_\ell)$ has no solution.

Case 3. The no-answer is given in step 6. Then $R_\ell \cap B'_\ell \neq \emptyset$ or $R'_\ell \cap B_\ell \neq \emptyset$. By step 4, we may assume that $N_G[u_1] \setminus N_G[u_2] = \emptyset$ and that $N_G[u_2] \setminus N_G[u_1] \neq \emptyset$. In order to obtain a contradiction, suppose that $(G_\ell, k, R_\ell, B_\ell)$ has a solution H. By Lemma 2 iii), $R'_\ell = \{u_2 u_1, \ldots, u_r u_1\} \subseteq E_H$. Hence $R'_\ell \cap B_\ell = \emptyset$, which means that $R_\ell \cap B'_\ell \neq \emptyset$.

Let $\{x_1, \ldots, x_p\} = N_G(u_1) \setminus \{u_1, \ldots, u_r\}$. Then we have that $B'_\ell = \{u_i u_j \mid 2 \leq i < j \leq r\} \cup \{u_1 x_1, \ldots, u_1 x_p\}$. By the same arguments as used in Case 1, we find that $u_i u_j \notin R_\ell$ for all $2 \leq i < j \leq r$. By Lemma 2 iii), we find that E_H, and hence R_ℓ, does not contain the edges $u_1 x_1, \ldots, u_1 x_p$. We conclude that $R_\ell \cap B'_\ell = \emptyset$; a contradiction. $\qquad\square$

Lemma 6 shows that the trimming rule is safe, that is, we either found that $(G, k, \emptyset, \emptyset)$ has no solution, or that we may continue with the instance $(G_\ell, k, R_\ell, B_\ell)$ instead. Suppose the latter case holds. Recall that $(G_\ell, k, R_\ell, B_\ell)$ has no set S as specified in step 1, as otherwise we would have applied the trimming rule once more.

To simplify notation, we write $(G, k, R, B) = (G_\ell, k, R_\ell, B_\ell)$. We need the following properties that hold for every solution of (G, k, R, B) (should (G, k, R, B) have a solution).

Lemma 7. *Any solution H of (G, k, R, B) satisfies the following properties:*

  i) *the neighbor of every pendant vertex of H has at least two non-pendant neighbors in H;*

  ii) *only edges of G incident to pendant vertices of H can be in R or B;*

  iii) *if a pendant vertex v of H is incident to an edge of R in G, then all other edges of G that are incident to v are in B.*

Proof. In order to show i), suppose that H is a solution of an instance (G, k, R, B), such that H contains a pendant vertex u adjacent to a vertex v. If $d_H(v) = 1$, then H is isomorphic to K_2, which is not possible as (G, k, R, B) has no solution that is a tree. Hence $d_H(v) \geq 2$ and v has at least one neighbor other than u. If all neighbors of v are pendant, then H is a tree; a contradiction. Hence, v has at least one non-pendant neighbor. If v has a unique non-pendant neighbor w, then by Lemma 1, $G - \{v, w\}$ contains a connected component induced by the pendant neighbors of v whose vertices

together with v and w form a clique in G. Hence, we can apply the trimming rule on $S = \{v, w\}$, which is a contradiction. Properties ii) and iii) follow from the construction of R and B in steps 4 and 5 of the trimming rule. $\quad\square$

We now exhaustively apply the following rule on (G, k, R, B). This rule consists of four steps that must be performed in increasing order.

Path Reduction Rule

1. Find an F-triple $S = \{u_1, u_2, u_3\}$.

2. Set $R' = \{u_2u_1, u_2u_3, \ldots, u_2u_r\}$ and $B' = \{x_1u_2, \ldots, x_pu_2\} \cup \{y_1u_2, .., y_qu_2\} \cup \{u_1u_3, \ldots, u_1u_r\} \cup \{u_3u_4, \ldots, u_3u_r\}$ (note that the set $\{u_3u_4, \ldots, u_3u_r\} = \emptyset$ if $r = 3$).

3. If $R \cap B' \neq \emptyset$ or $R' \cap B \neq \emptyset$, then stop and return a no-answer.

4. Delete $u_2, u_4, \ldots, u_r$ from G. Delete all edges incident to $u_2, u_4, \ldots, u_r$ from R and B. If $u_1u_3 \in B$, then delete u_1u_3 from B. Add u_1u_3 to R. Add $x_1u_3, \ldots, x_pu_3$ and $y_1u_1, \ldots, y_qu_1$ in G. Put these edges in B.

Exhaustively applying the path reduction rule yields a sequence of instances $(G_0, k, R_0, B_0), \ldots, (G_\ell, k, R_\ell, B_\ell)$ of TREE + k EDGES SQUARE ROOT WITH LABELS for some integer $\ell \geq 0$, where $(G_0, k, R_0, B_0) = (G, k, R, B)$ and where $(G_\ell, k, R_\ell, B_\ell)$ is an instance for which we have either returned a no-answer (in step 3) or for which there does not exist an F-triple S. For $0 \leq i \leq \ell$ we denote the sets R' and B' constructed in the $(i + 1)$th call of the path reduction rule by R'_i and B'_i, respectively.

We need the following lemma, which we will use at several places.

Lemma 8. *Let $1 \leq i \leq \ell$ and $\{u_1, u_2, u_3\}$ be the F-triple that yielded instance (G_i, k, R_i, B_i). If H_i is a solution for (G_i, k, R_i, B_i), then $u_1u_3 \in E_{H_i}$ and the graph H_{i-1} obtained from H_i by removing the edge u_1u_3 and by adding u_2 and vertices $u_4, \ldots, u_r$ (if $r \geq 4$) together with edges $u_2u_1, u_2u_3, \ldots, u_2u_r$ is a solution for $(G_{i-1}, k, R_{i-1}, B_{i-1})$.*

Proof. We find that u_1u_3 is an edge in H_i, because $u_1u_3 \in R_i$ due to step 4 of the last call of the path reduction rule. The graph H_{i-1} is not only a square root of G_{i-1} but even a solution for $(G_{i-1}, k, R_{i-1}, B_{i-1})$ for the following reasons. First, H_{i-1} has at most $|V_{G_{i-1}}| - 1 + k$ edges. Second, H_{i-1} contains no edge of B_{i-1} as the added edges $u_2u_1, u_2u_3, \ldots, u_2u_r$ are all in R'_{i-1} and $R'_{i-1} \cap B_{i-1} = \emptyset$. Third, H_{i-1} contains all the edges of R_{i-1}, which can be seen as follows. Suppose that H_{i-1} misses an edge of R_{i-1}. Then this edge must be in $\{x_1u_2, \ldots, x_pu_2\} \cup \{y_1u_2, .., y_qu_2\} \cup \{u_1u_3, \ldots, u_1u_r\} \cup \{u_3u_4, \ldots, u_3u_r\}$. However, this set is equal to B'_{i-1} and $R_{i-1} \cap B'_{i-1} = \emptyset$. $\quad\square$

We also need the following lemma about true twins in $G_0, \ldots, G_\ell$ that we will use later as well.

Lemma 9. *Let* $1 \leq i \leq \ell$ *and* $\{u_1, u_2, u_3\}$ *be the F-triple that yielded instance* (G_i, k, R_i, B_i). *Then any true twins* $v, w \in V_{G_i} \setminus \{u_1, u_3\}$ *in* G_i *are true twins in* G_{i-1}.

Proof. Suppose that G_i has true twins $v, w \in V_{G_i} \setminus \{u_1, u_3\}$ that are not true twins in G_{i-1}. Consider the corresponding F-graph that yielded the instance (G_i, k, R_i, B_i). Because v, w are not true twins in G_{i-1}, the neighborhood of v or w is modified by the path reduction rule. We may assume without loss of generality that the neighborhood of v is changed. Note that neither $v = u_2$ nor $v \in \{u_4, \ldots, u_r\}$ if $r \geq 3$, because these vertices have been removed in step 4 of the path reduction rule when G_i was constructed. As $v \notin \{u_1, u_3\}$ either, we find that $v \in \{x_1, \ldots, x_p\} \cup \{y_1, \ldots, y_q\}$. By symmetry we may assume that $v \in \{x_1, \ldots, x_p\}$. We observe that v is adjacent to both u_1 and u_3 in G_i. Because the neighborhood of each x_i is modified in the same way (namely by the removal of u_2 and the addition of u_3), we find that $w \notin \{x_1, \ldots, x_p\}$. Because u and v are true twins, they are adjacent. Because no two vertices x_i and y_j are adjacent in G_i, we then obtain that $w \notin \{y_1, \ldots, y_q\}$. We conclude that the neighborhood of w is not modified by the application of the path reduction rule. Because v is adjacent to u_1 and u_3 in G_i and v, w are true twins in G_i, this means that w is adjacent to u_1 and u_3 in G_{i-1} already. However, by definition of an F-graph, $N_{G_{i-1}}(u_1) \cup N_{G_{i-1}}(u_3) = \{u_2, u_4, \ldots, u_r\}$, and $u_2, u_4, \ldots, u_r$ are not in G_i as they were removed by the path reduction rule; a contradiction. $\square$

The next lemma is the analog of Lemma 6 for the path reduction rule.

Lemma 10. *The instance* $(G_\ell, k, B_\ell, R_\ell)$ *has no solution that is a tree, and* G_ℓ *is 2-connected. Moreover,* $(G_\ell, k, R_\ell, B_\ell)$ *has a solution if and only if* (G_0, k, R_0, B_0) *has a solution. If the path reduction rule returned a no-answer for* $(G_\ell, k, R_\ell, B_\ell)$, *then* (G_0, k, R_0, B_0) *has no solution.*

Proof. For $0 \leq i \leq \ell$, we use induction to show that the graph G_i is 2-connected and that (G_i, k, B_i, R_i) has no solution that is a tree. Moreover, for all $1 \leq i \leq \ell$, we show that (G_i, k, R_i, B_i) has a solution if and only if $(G_{i-1}, k, R_{i-1}, B_{i-1})$ has a solution. Finally, we prove that if the path reduction rule returned a no-answer for $(G_\ell, k, R_\ell, B_\ell)$, then (G_0, k, R_0, B_0) has no solution.

If $i = 0$, then G_i is 2-connected and (G_0, k, B_0, R_0) has no solution that is a tree by Lemma 6. Now suppose that $1 \leq i \leq \ell$. By our induction hypothesis, we may assume that G_{i-1} is 2-connected and that $(G_{i-1}, k, B_{i-1}, R_{i-1})$ has no solution that is a tree.

Because the path reduction rule applied on $(G_{i-1}, k, R_{i-1}, B_{i-1})$ yielded a new instance (G_i, k, R_i, B_i), the graph G_{i-1} has an F-triple $S =$

$\{u_1, u_2, u_3\}$. Because G_{i-1} is 2-connected, G_i is 2-connected; in particular note that $p \geq 1$ and $q \geq 1$ by definition of an F-triple.

First suppose that H_{i-1} is a solution for $(G_{i-1}, k, R_{i-1}, B_{i-1})$. We claim that H_{i-1} contains no edge $u_s u_t \in R_{i-1}$ with $4 \leq s < t \leq r$. We prove this claim by contradiction: let $u_s u_t \in E_{H_{i-1}} \cap R_{i-1}$ for some $4 \leq s < t \leq r$.

Suppose that $u_s u_t \in R_0$. We may apply Lemma 7 as (G_0, k, R_0, B_0) has a solution H_0; if $i \geq 1$ this fact follows from the induction hypothesis. By Lemma 7 we find that either u_s is a pendant vertex in H_0 with u_t as its (unique) neighbor, or the other way around. We may assume without loss of generality that the first case holds, that is, u_s is pendant in H_0 and has u_t as its neighbor. Note that $N_{G_0}[u_s] \subseteq N_{G_0}[u_t]$. We claim that $N_{G_h}[u_s] \subseteq N_{G_h}[u_t]$ for all $0 \leq h \leq i - 1$. To obtain a contradiction, suppose not. Then at some point u_s will be made adjacent to a vertex v not adjacent to u_t for the first time in step 4 of some call of the path reduction rule. Let $S = \{u'_1, u'_2, u'_3\}$ be the corresponding F-triple. Then we may assume without loss of generality that either $u_s \notin \{u'_1, u'_2, u'_3\}$ is adjacent to u'_1 and u'_2 but not to $u'_3 = v$, or that $v \notin \{u'_1, u'_2, u'_3\}$ is adjacent to u'_1, u'_2 but not to $u'_3 = u_s$. In the first case, u_t is not in $\{u'_1, u'_2\}$, but must be adjacent to u'_1 and u'_2 by our assumption, and hence, the edge $u_t u'_3 = u_t v$ will be added in the same step; a contradiction. In the second case, as u_s is adjacent to u'_1 and u'_2, also u_t is adjacent to u'_1 and u'_2 (again by our assumption). Because u_t does not get removed in this step (as u_t belongs to G_{i-1}), this violates the definition of an F-triple. We conclude that $N_{G_h}[u_s] \subseteq N_{G_h}[u_t]$ for all $0 \leq h \leq i - 1$.

We first assume that $u_s u_2$ is an edge in G_0. Step 4 of the path reduction rule only moves an edge $u'_1 u'_3$ from a B-set to an R-set if u'_1 and u'_3 are outer vertices of an F-triple. In that case all their common neighbors will be removed from the graph by the definition of an F-triple. Because $N_{G_h}[u_s] \subseteq N_{G_h}[u_t]$ for all $0 \leq h \leq i - 1$, we find that u_t is a common neighbor of u_2 and u_s in G_h for all $0 \leq h \leq i - 1$; in particular u_t belongs to G_{i-1}. Hence, the edge $u_s u_2$ will never be moved from B_h to R_h in step 4 of the $(h+1)$th call of the path reduction rule for some $0 \leq h \leq i - 1$. If $u_s u_2$ is not an edge in G_0, then at some point it will be an edge due to step 4 of some call of the path reduction rule, say the $(h^* + 1)$th call for some $0 \leq h^* \leq i - 1$. In the same step, $u_s u_2$ will be placed in the set B_{h^*}. Then, again because $N_{G_h}[u_s] \subseteq N_{G_h}[u_t]$ for all $0 \leq h \leq i - 1$, the edge $u_s u_2$ will never be moved from B_{h^*} to a set R_h for some $h^* < h \leq i - 1$. Hence, in both cases, we find that $u_s u_2 \in B_{i-1}$ even if $i \geq 1$. As $u_s u_2 \in R'_{i-1}$ (due to step 2 in the ith call), we find that $R'_{i-1} \cap B_{i-1} \neq \emptyset$. Hence, the path reduction rule would return a no-answer for $(G_{i-1}, k, R_{i-1}, B_{i-1})$ in step 3, and consequently the instance (G_i, k, R_i, B_i) would not exist; a contradiction.

Now suppose that $u_s u_t$ was placed in some set R_h for some $1 \leq h \leq i - 1$. Properties ii) and iii) of an F-graph together with step 4 of the path reduction rule imply the following: if u_s and u_t form a triangle with

some vertex z, then $u_s z \in B_h$ or $u_t z \in B_h$. Moreover, in the case in which $z \in V_{G_{i-1}}$, this property is not violated by any subsequent intermediate calls of the path reduction rule. Hence, if $u_s u_t \in R_{i-1}$, then $u_s u_2 \in B_{i-1}$ or $u_t u_2 \in B_{i-1}$, and as $\{u_s u_2, u_t u_2\} \subseteq R'_{i-1}$ as well, we derive the same contradiction as before. We conclude that H_{i-1} contains no edge $u_s u_t \in R_{i-1}$ with $4 \le s < t \le r$. Also, by Lemma 4, we may assume without loss of generality that H_{i-1} contains no edge $u_s u_t \notin R_{i-1}$ with $4 \le s < t \le r$; otherwise we could remove such an edge from H_{i-1}, and the resulting graph would still be a solution for $(G_{i-1}, k, R_{i-1}, B_{i-1})$. Consequently, $u_4, \ldots, u_r$ are pendant vertices of H_{i-1}. This means that the graph H obtained from H_{i-1} by deleting vertices $u_2, u_4, \ldots, u_r$ and adding the edge $u_1 u_3$ is not only a square root of G_i with at most $|V_{G_i}| - 1 + k$ edges but even a solution for (G_i, k, R_i, B_i).

Now suppose that H_i is a solution for (G_i, k, R_i, B_i). By Lemma 8, the graph H obtained from H_i by removing the edge $u_1 u_3$ and by adding u_2 and vertices $u_4, \ldots, u_r$ (if $r \ge 4$) together with edges $u_2 u_1, u_2 u_3, \ldots, u_2 u_r$ is a solution for $(G_{i-1}, k, R_{i-1}, B_{i-1})$. We observe that H_i cannot be a tree, as this would imply that H is a tree, which is not possible as $(G_{i-1}, k, R_{i-1}, B_{i-1})$ does not have such a solution by the induction hypothesis.

Finally, suppose that the path reduction rule returned a no-answer for $(G_\ell, k, R_\ell, B_\ell)$. We must show that (G_0, k, R_0, B_0) has no solution. Due to the above this comes down to showing that $(G_\ell, k, R_\ell, B_\ell)$ has no solution. The only step in which the path reduction rule can return a no-answer is in step 3, meaning that G_ℓ has an F-triple $S = \{u_1, u_2, u_3\}$ such that $R_\ell \cap B'_\ell \ne \emptyset$ or $R'_\ell \cap B_\ell \ne \emptyset$.

In order to obtain a contradiction, suppose that $(G_\ell, k, R_\ell, B_\ell)$ has a solution H. By Lemma 4, the graph F shown in Figure 4 is a subgraph of H such that $d_H(u_2) = r - 1$, $\{x_1, \ldots, x_p\} = N_H(u_1) \setminus \{u_2\}$ and $\{y_1, \ldots, y_q\} = N_H(u_3) \setminus \{u_2\}$. Consequently, $R'_\ell = \{u_2 u_1, u_2 u_3, \ldots, u_2 u_r\} \subseteq E_H$, and hence $R'_\ell \cap B_\ell = \emptyset$, and moreover, $E_H \cap B'_\ell = E_H \cap (\{x_1 u_2, \ldots, x_p u_2\} \cup \{y_1 u_2, .., y_q u_2\} \cup \{u_1 u_3, \ldots, u_1 u_r\} \cup \{u_3 u_4, \ldots, u_3 u_r\}) = \emptyset$, and hence $R_\ell \cap B'_\ell = \emptyset$; a contradiction. $\qquad \square$

Lemma 10 shows that the path reduction rule is safe, that is, we either found that (G_0, k, R_0, B_0) has no solution, or that we may continue with the instance $(G_\ell, k, R_\ell, B_\ell)$ instead. Suppose the latter case holds. Recall that $(G_\ell, k, R_\ell, B_\ell)$ has no F-triple, as otherwise we would have applied the path reduction rule once more. Also recall that R_0 is the set of vertices in the set R immediately after the trimming rule. We write $R^1 = R_0 \cap R_\ell$ and $R^2 = R_\ell \setminus R_0$. To simplify notation, from now on, we also write $(G, k, R, B) = (G_\ell, k, R_\ell, B_\ell)$; note that $R = R^1 \cup R^2$. We need the following properties that hold for every solution of (G, k, R, B) (should (G, k, R, B) have a solution).

We call an induced cycle C in a graph H *semi-pendant* if all but at most one of the vertices of C are only adjacent to pendant vertices of H and their

neighbors on C. Similarly, we call an induced path P in a graph H *semi-pendant* if all internal vertices of P are only adjacent to pendant vertices of H and their neighbors on P.

Lemma 11. *Any solution H of (G, k, R, B) has the following properties:*

i) *the neighbor of every pendant vertex of H has at least two non-pendant neighbors in H;*

ii) *only edges of G incident to pendant vertices of H can be in R^1, and if a pendant vertex v of H is incident to an edge of R, then all other edges of G that are incident to v are in B;*

iii) *no edge of R^2 is incident to a pendant vertex of H;*

iv) *the length of every semi-pendant path in H is at most 5;*

v) *the length of every semi-pendant cycle in H is at most 6.*

Proof. We prove that property i) holds by contradiction. Suppose that H contains a vertex v that is the (unique) neighbor of a pendant vertex u, such that v has at most one non-pendant neighbor in H. If all neighbors of v in H are pendant, then H is a tree. However, this would contradict Lemma 10. Hence, v has a unique non-pendant neighbor in H. Recall that H is a solution for $(G_\ell, k, R_\ell, B_\ell)$. Note that if v is an outer vertex of the corresponding F-triple, then Lemma 8 tells us that $(G_{\ell-1}, k, R_{\ell-1}, B_{\ell-1})$ has a solution $H_{\ell-1}$ in which v is a non-pendant vertex that has at least one pendant neighbor and that has a unique non-pendant neighbor. Hence, by applying Lemma 8 inductively, we obtain that (G_0, k, R_0, B_0) has a solution H_0 containing a vertex with exactly the same property. This contradicts Lemma 7 i). We conclude that property i) holds.

We now show property ii). By Lemma 7, every edge of G_0 that is in R_0 is incident to a pendant vertex u of any solution for (G_0, k, R_0, B_0) such that all the other edges of u belong to B_0. We observe that, when applying the path reduction rule, u will neither be in an F-triple nor removed from the graph, but u could be a vertex of x-type or y-type. Hence, the path reduction rule may change the neighbors of u but if so any new edges incident to it will be placed in B (and stay in B afterward). Consequently, u must be a pendant vertex in any solution for $(G, k, R, B)(= (G_\ell, k, R_\ell, B_\ell))$ as well. We conclude that ii) holds.

We now prove prove property iii). Recall that we applied the path reduction rule only after first applying the trimming rule exhaustively. When we apply the path reduction rule on an F-triple $\{u_1, u_2, u_3\}$, then afterward u_1 and u_3 have degree at least 2 in any solution for the resulting instance, which can be seen as follows. The edge $u_1 u_3$ is added to $R^2 \subseteq R$, and hence belongs to any solution. We also have that u_1 is adjacent to x_1 in G, whereas

the edge u_3x_1 belongs to B. This means that u_1 cannot be made adjacent to x_1 via the path $u_1u_3x_1$ in H, and as such must have at least one other neighbor in H. For the same reason u_3, which is adjacent to y_1 in G whereas $u_1y_1 \in B$, must have another neighbor in H besides u_1. As a consequence, any edge in R^2 cannot be incident to a pendant vertex of H, that is, we have shown property iii).

We now prove property iv). Let P be a semi-pendant path of length at least 6 in H. By definition, P is an induced path. Hence, we can take any three consecutive vertices of P as the three vertices u_1, u_2, u_3 in Lemma 3. By applying this lemma, we find that G is an F-graph implying that we could have applied the path reduction rule once more; a contradiction. Property v) can be proven by using the same arguments. $\square$

We need the following lemma that holds in case a solution exists for (G, k, R, B).

Lemma 12. *The number of non-pendant vertices of any solution for (G, k, R, B) is at most $15k - 14$.*

Proof. Suppose (G, k, R, B) has a solution H. Let Z be the set of pendant vertices of H, and let $H^* = H - Z$. We need to show that V_{H^*} has at most $15k - 14$ vertices. Let V' be the set of vertices that have degree at least 3 in H^*, and let V'' be the set of vertices of degree 2 in H^*. By Lemma 11 i) every vertex of H that is adjacent to a pendant vertex of H has degree at least 2 in H^*. Hence, H^* has no vertices of degree at most 1, that is, $V_{H^*} = V' \cup V''$. Because H is a solution for (G, k, R, B), we have that $|E_H| \leq |V_G| - 1 + k = |V_H| - 1 + k$. This means that

$$
\begin{aligned}
|V'| + |V''| - 1 + k &= |V_H| - |Z| - 1 + k \\
&\geq |E_H| - |Z| \\
&= |E_{H^*}| \\
&= \tfrac{1}{2} \sum_v d_{H^*}(v) \\
&\geq \tfrac{1}{2}(3|V'| + 2|V''|).
\end{aligned}
$$

Hence, $|V'| \leq 2k - 2$. Let α be the number of paths in H^* that only have internal vertices of degree 2; note that by Lemma 11 iv) the length of such paths is at most 5. Let β be the number of cycles in H^* that have exactly one vertex of degree at least 3; note that by Lemma 11 v) the length of such cycles is at most 6. Because $|E_{H^*}| \leq |V'| + |V''| - 1 + k$, we find that $\alpha + \beta \leq 2k - 2 - 1 + k = 3k - 3$ and that $\beta \leq k$. Hence, $|V''| \leq 5k + 4((3k - 3) - k) = 13k - 12$. Consequently, H^* has at most $2k - 2 + 13k - 12 = 15k - 14$ vertices. $\square$

We are now ready to state our final reduction rule. The goal of this rule is to apply it once in order to deduce either that (G, k, R, B) has no solution

or to derive a new instance of bounded size. A *true twin partition* of a set of vertices S of a graph G is a partition $S_1, \ldots, S_t$ of S such that for all $u, v \in S$ and all $1 \le i \le t$ we have that u and v are in S_i if and only if u and v are true twins in G. If S consists of simplicial vertices only we observe that there is no edge between any two vertices that belong to different sets S_i and S_j.

Simplicial Vertex Reduction Rule

1. Find the set S of all simplicial vertices of G that are not incident to the edges of R^2, and moreover, that have all but one of their incident edges in B should they be incident to an edge of R^1.

2. If $|V_G \setminus S| > 15k - 14$, then stop and return a no-answer.

3. Construct the true twin partition $S_1, \ldots, S_t$ of S. Let $X_1, \ldots, X_t$ be the sets of vertices incident to an edge of R^1 in $S_1, \ldots, S_t$, respectively.

4. If $t > 15k - 14$, then stop and return a no-answer.

5. If there exist a set X_i such that the edges of R^1 incident to a vertex of X_i have no common end-vertex, then stop and return a no-answer.

6. If there exist a set S_i such that $|S_i \setminus X_i| \ge 15k - 13$ and such that there are three vertices $u \in X_i$, $v \in N_G(u)$ and $x \in S_i \setminus X_i$ with $uv \in R^1$ and $xv \in B$, then stop and return a no-answer.

7. For $i = 1, \ldots, t$, if $|X_i| > 1$, then take $|X_i| - 1$ arbitrary vertices of X_i and delete them both from G and from S_i, also delete the edges of R and B that are incident to these vertices.

8. For $i = 1, \ldots, t$, if $|S_i| > 15k - 13$, then delete $|S_i| - 15k + 13$ arbitrary vertices of $S_i \setminus X_i$ from G, also delete the edges of R and B that are incident to these vertices.

Applying the simplicial vertex reduction rule on (G, k, R, B) either yields a no-answer (in step 2, 4, 5 or 6) or a new instance $(\hat{G}, k, \hat{R}, \hat{B})$ of TREE + k EDGES SQUARE ROOT WITH LABELS. We will show that if $\hat{G}$ exists, then its size is bounded by a quadratic function of k. For doing so we first need the following two lemmas.

Lemma 13. *For $i = 1, \ldots, t$, no vertex of $S_i \setminus X_i$ is incident to an edge in R.*

Proof. By definition of S_i, no vertex of S_i, and hence no vertex of $S_i \setminus X_i$, is incident to an edge in R^2. By definition of X_i, no vertex in $S_i \setminus X_i$ is incident to an edge in R^1. Because $R = R^1 \cup R^2$, we have proven Lemma 13. $\qquad \square$

For $x \in V_G$, we let $B(x)$ denote the set of edges of B incident to x.

Lemma 14. $B(x) = B(y)$ *for all* $x, y \in S_i \setminus X_i$.

Proof. Let $x, y \in S_i \setminus X_i$ and let $xz \in B$ for some $z \in V_G$. We first show that $y \neq z$ and we then prove that $yz \in B$.

In order to obtain a contradiction, assume that $y = z$. Then xy was included in B either by an application of the trimming rule or by an application of the path reduction rule. In both cases, xy was also made adjacent to an edge of R. This edge may be deleted later on. Deleting an edge e from R happens either in step 6 of the trimming rule or in step 4 of the path reduction rule. However, both rules add a new edge e' to R that is adjacent to all the edges that were previously adjacent to e and that were not deleted by the two rules. Hence, xy is still adjacent to an edge of R in G. In other words, x or y is incident to an edge of R in G. Because x and y belong to $S_i \setminus X_i$, this is not possible due to Lemma 13. Hence, $y \neq z$.

In order to show that $yz \in B$, we again use the observation that whenever the trimming or path reduction rule deletes an edge $e \in R$, the rule adds a new edge e' in R such that e' is adjacent to all the edges uv that were previously adjacent to e and that were not deleted by the rules. In this case we make the extra observation that if a vertex u is an end-vertex of e that is not deleted by the rule, then u is an end-vertex of e'. Because the vertices in $S_i \setminus X_i$ are not incident to any edges in R by Lemma 13, we find that z was incident to an edge of R after applying the trimming rule or path reduction rule that added the edge xz to B. We also observe that an edge in B is only deleted from B if one of its end-vertices is deleted unless it is added to R by the path reduction rule. This means that we can argue as follows.

First suppose that xz was added to B due to an application of the trimming rule. If y was adjacent to z when the rule was applied, then yz was included in B as well by the definition of this rule. If y was made adjacent to z by the path reduction rule afterwards, then $yz \in B$ by the definition of the path reduction rule.

Now suppose that xz was added to B due to an application of the path reduction rule. By definition of this rule, x and z were not adjacent to each other before. Suppose that $yz \notin B$. Then xy, yz are edges of the original input graph of the TREE $+ k$ EDGES SQUARE ROOT problem. Because xz was not such an edge, x and y only became true twins due to an application of the path reduction rule. Then, by Lemma 9, x or y must be an outer vertex of some F-triple, that is, at least one of these two vertices must be incident to an edge of R. Then there is an edge of R incident to at least one of these two vertices after the exhaustive application of the path reduction rule. Because x and y are in $S_i \setminus X_i$, this is a contradiction to Lemma 13. Hence, $yz \in B$. This completes the proof of Lemma 14. $\qquad \square$

We prove the following lemma, which is our final lemma; in particular note that if $\hat{G}$ exists then its size is bounded by a quadratic function of k.

Lemma 15. *If the simplicial vertex reduction rule returned a no-answer for (G, k, R, B), then (G, k, R, B) has no solution. Otherwise, the new instance $(\hat{G}, k, \hat{R}, \hat{B})$ has a solution if and only if (G, k, R, B) has a solution. Moreover, $\hat{G}$ has at most $(15k - 14)(15k - 12)$ vertices.*

Proof. We start by showing that (G, k, R, B) has no solution if the simplicial vertex reduction rule returned a no-answer for (G, k, R, B). This can happen in step 2, 4, 5 or 6, each of which we discuss in a separate case.

Case 1. The no-answer is given in step 2. Suppose (G, k, R, B) has a solution H. We will prove that $|V_G \setminus S| \leq 15k - 14$, which means that returning a no-answer is correct if $|V_G \setminus S| > 15k - 14$.

Let Z be the set of pendant vertices of H, and let $H^* = H - Z$. By Observation 1 i), vertices in Z are simplicial vertices of G. Then, by Lemma 11 ii) and iii), we find that $Z \subseteq S$. Hence, $|V_G \setminus S| = |V_G| - |S| = |V_H| - |S| \leq |V_H| - |Z| = |V_{H^*}| \leq 15k - 14$, where the last inequality follows from Lemma 12.

Case 2. The no-answer is given in step 4. Suppose (G, k, R, B) has a solution H. We will prove that $t \leq 15k - 14$, which means that returning a no-answer is correct if $t > 15k - 14$.

Let H^* be the graph obtained from H after removing all pendant vertices of H. Then $|V_{H^*}| \leq 15k - 14$ by Lemma 12. If a set S_i contains a pendant vertex u of H, then u is adjacent to a vertex v of H^*. Then, by Observation 1 ii), v is not adjacent to pendant vertices of H in any S_j with $j \neq i$. Otherwise S_i consists of non-pendant vertices of H, that is, vertices of H^*; being nonempty S_i contains at least one vertex of H^*. We conclude that every set in the true twin partition of S corresponds to at least one unique vertex of H^*. If their total number $t > 15k - 14$, this means that $|V_{H^*}| > 15k - 14$; a contradiction. Hence, $t \leq 15k - 14$, as we had to show.

Case 3. The no-answer is given in step 5. Suppose that (G, k, R, B) has a solution H. We will prove that the edges of R^1 incident to a set X_i have a common end-vertex for $i = 1, \ldots, t$, which means that returning a no-answer is correct should this not be the case.

In order to obtain a contradiction, suppose that some set X_i contains two vertices u and v that are incident to edges $uu', vv' \in R^1$ with $u' \neq v'$. By Lemma 11 ii), we find that uu' and vv' are incident to pendant vertices of H. By Observation 1 iii), these pendant vertices are not adjacent in G. However, from the definition of S_i we deduce that u, v, u', v' are mutually adjacent; a contradiction. This completes Case 3.

Case 4. The no-answer is given in step 6. Then there exists a set S_i such that $|S_i \setminus X_i| \geq 15k - 13$ and such that there are three vertices $u \in X_i$, $v \in N_G(u)$ and $x \in S_i \setminus X_i$ with $uv \in R^1$ and $xv \in B$. In order to obtain a contradiction, assume that (G, k, R, B) has a solution H.

25

By Lemma 12, H has at most $15k - 14$ non-pendant vertices. Because $|S_i \setminus X_i| \geq 15k - 13$, this means that at least one vertex $y \in S_i \setminus X_i$ is a pendant vertex of H. Also, $u \in X_i$ is a pendant vertex of H that has v as its unique neighbor, because $uv \in R^1$ and all other edges incident to u belong to B by definition of S. If $y = x$, then v is not adjacent to y in H, because $xv \in B$. If $y \neq x$, then v is not adjacent to y in H either, because $xv \in B$ and $B(x) = B(y)$ (due to Lemma 14) imply $yv \in B$. We conclude that u and y are pendant vertices of H adjacent to different vertices. However, from Observation 1 iii) we derive that u and y are not adjacent in G. This is a contradiction, because u and y are true twins in G by definition of S_i. This completes Case 4.

From now on assume that the simplicial vertex reduction rule did not return a no-answer after performing step 6. Let (G', k, R', B') be the instance created after applying step 7 to some set $X_i = \{x_1, \ldots, x_\ell\}$ with $\ell \geq 2$, that is, G' is the graph obtained from G after deleting $x_2, \ldots, x_\ell$, whereas R' and B' are the sets obtained from R and B, respectively, after deleting edges incident to $x_2, \ldots, x_\ell$ from them. We claim that (G', k, R', B') has a solution if and only if (G, k, R, B) has a solution. Before we prove this claim, we first observe that in any solution H for (G, k, R, B) the vertices $x_1, \ldots, x_\ell$ are pendant vertices in H. This is because $x_1, \ldots, x_\ell$ are incident to exactly one edge in R^1, whereas all the other edges incident to them belong to B. Moreover, $x_1, \ldots, x_\ell$ have a (unique) common neighbor in H, as otherwise a no-answer would have been returned in step 5. We let v denote this common neighbor. Similarly, x_1 is a pendant vertex that has v as its (unique) neighbor in any solution H' for (G', k, R', B').

First suppose that (G', k, R', B') has a solution H'. Then the graph obtained from H' by adding the vertices $x_2, \ldots, x_\ell$ and the edges $x_2 v, \ldots, x_\ell v$ is a square root of G by Lemma 5 i). By definition of R', B' and the set X_i (all of whose vertices are incident to one edge of $R^1 \subseteq R$ and to edges in B) it is a solution for (G, k, R, B) as well.

Now suppose that (G, k, R, B) has a solution H. Then the graph obtained from H after deleting $x_2, \ldots, x_\ell$ is a square root of G' by Lemma 5 ii). By definition of R' and B', it is a solution for (G', k, R', B') as well.

We denote the instance resulting from step 7 by (G, k, R, B) again and observe that every X_i now contains at most one vertex. It remains to consider what happens at step 8. We let (G', k, R', B') be the instance created after applying step 8 to some set S_i with $|S_i| > 15k - 13$, that is, G' is the graph obtained from G after deleting a set T of $|S_i| - 15k + 13 \geq 1$ arbitrary vertices from $S_i \setminus X_i$ (note that this is possible as $|X_i| \leq 1$), whereas R' and B' are the sets obtained from R and B, respectively, after deleting the edges that are incident to vertices of T. We claim that (G', k, R', B') has a solution if and only if (G, k, R, B) has a solution.

First suppose that (G', k, R', B') has a solution H'. Because we could not

apply the trimming and path reduction rules for (G, k, R, B), we cannot apply these rules for (G', k, R', B') either. Then, by using the same arguments that we applied for (G, k, R, B) in the proof of Lemma 12, we find that H' contains at most $15k - 14$ non-pendant vertices. Note that H' contains at least $15k - 13$ vertices, which are all in S_i. Hence, H' has at least one pendant vertex x that belongs to S_i. Let v be the (unique) vertex adjacent to x in H'. Then the graph H obtained from H' by adding the vertices of T and their edges incident to v is a square root of G by Lemma 5 i). We argue that H is a solution for (G, k, R, B) as well. Because the vertices of $T \subseteq S_i \setminus X_i$ are not incident to the edges of R due to Lemma 13, we have to show that none of the $|T|$ edges that we added in order to obtain H belong to B. If $x \in S_i \setminus X_i$, then $xv \notin B$ and because $B(x) = B(y)$ for all $y \in S_i \setminus X_i$, we have that $yv \notin B$ for all $y \in T$. Assume that $x \in X_i$. Recall that $|X_i| \leq 1$ after step 7. Because $|S_i| > 15k - 13$ after step 7, $|S_i \setminus X_i| \geq 15k - 13$. Then $yv \notin B$ for all $y \in S_i \setminus X_i$ as otherwise the algorithm would have produced a no-answer at step 6.

Now suppose that (G, k, R, B) has a solution H. By Lemma 12, the graph H contains at most $15k - 14$ non-pendant vertices. Hence, H has at least $|S_i| - 15k + 14 \geq 15k - 12 - 15k + 14 = 2$ pendant vertices. Because vertices in $S_i \setminus X_i$ are true twins not incident to edges of R and $B(x) = B(y)$ for any $x, y \in S_i \setminus X_i$, we may assume without loss of generality that the vertices of T are amongst these pendant vertices of H. If $X_i = \{x\} \neq \emptyset$, then x is a pendant vertex in H incident to a unique edge $xv \in R^1$. By Observation 1, all pendant vertices of H that are in S_i are adjacent to v in H. Then the graph obtained from H after deleting the vertices of T is a square root of G' by Lemma 5 ii). By definition of R' and B', it is a solution for (G', k, R', B') as well. If $X_i = \emptyset$, then all pendant vertices of H that are in S_i are adjacent to some v in H by Observation 1. Then, by Lemma 5 (ii), the graph obtained from H by deleting the vertices of T is a square root of G'. By definition of R' and B', it is a solution for (G', k, R', B') as well.

From the above it follows that the instance $(\hat{G}, k, \hat{R}, \hat{B})$ obtained after step 8 has a solution if and only if (G, k, R, B) has a solution. In order to complete the proof, we must show that $\hat{G}$ has at most $(15k - 14)(15k - 12)$ vertices. Each S_i has at most $15k - 13$ vertices due to step 8, and we also have $t \leq 15k - 14$ due to step 4. Hence $|S| \leq (15k - 14)(15k - 13)$. As the number of vertices in $V_G \setminus S$ is at most $15k - 14$ due to step 2, we obtain that $|V_{\hat{G}}| \leq (15k - 14)(15k - 13) + 15k - 14 = (15k - 14)(15k - 12)$, as required. $\qquad\square$

2.3 Solving the Labeled Variant and Running Time Analysis

Let n and m denote the number of vertices and edges of the graph G of the original instance (G, k) of TREE $+ k$ EDGES SQUARE ROOT. In order to complete the proof of Theorem 1, we first note that the trimming and

path reduction rules are applied at most n times to construct the instance $(\hat{G}, k, \hat{R}, \hat{B})$. Each application of the trimming rule can be done in time $O(n^2 m)$ and each application of the path reduction rule takes time $O(n^3 m)$. Finally, the simplicial vertex reduction rule can be done in time $O(nm)$. Hence, our kernelization algorithm runs in time $O(n^4 m)$, and it remains to solve the obtained reduced instance $(\hat{G}, k, \hat{R}, \hat{B})$. Because $\hat{G}$ has at most $(15k - 14)(15k - 12)$ vertices, $\hat{G}$ has at most $\frac{1}{2}(15k - 14)(15k - 12)((15k - 14)(15k - 12) - 1) = O(k^4)$ edges. Therefore, we can solve TREE $+ k$ EDGES SQUARE ROOT WITH LABELS for instance $(\hat{G}, k, \hat{R}, \hat{B})$ in time $2^{O(k^4)}$; we consider all edge subsets of $\hat{G}$ that have size at most $|V_{\hat{G}}| - 1 + k$ and use brute force. We conclude that the total running time of our algorithm is $2^{O(k^4)} + O(n^4 m)$, as required.

We finish this section with the following remarks. First, recall that our quadratic kernel is a generalized kernel for the TREE $+ k$ EDGES SQUARE ROOT problem. We believe that a quadratic kernel exists for this problem as well by using a similar reduction. However, proving this seemed to be more technical and also to yield a graph with more than $(15k - 14)(15k - 12)$ vertices. We therefore chose to prove our FPT result by using a reduction leading to a generalized kernel. Second, it should also be noted that our generalized kernel for TREE $+ k$ EDGES SQUARE ROOT does not imply a kernel for TREE $+ k$ EDGES SQUARE ROOT WITH LABELS, because our reduction rules require that the original instance is unlabeled. We do not know whether the (more general) problem TREE $+ k$ EDGES SQUARE ROOT WITH LABELS is FPT as well.

3   The Maximum Square Root Problem

Recall that the MAXIMUM SQUARE ROOT problem is that of testing whether a given graph G with m edges has a square root with at least s edges for some given integer s. In this section we give an FPT algorithm for this problem with parameter $k = m - s$. In other words, we show that the problem of deciding whether a graph G has a square root that can be obtained by removing at most k edges of G is fixed-parameter tractable when parameterized by k. We also present an exact algorithm for the MAXIMUM SQUARE ROOT problem. Both algorithms are based on the observation that in order to construct a square root H from a given graph G, we must delete at least one of every pair of adjacent edges that do not belong to a triangle in G. We therefore construct an auxiliary graph $\mathcal{P}(G)$ that has vertex set E_G and an edge between two vertices e_1 and e_2 if and only if $e_1 = xy$ and $e_2 = yz$ for three distinct vertices $x, y, z \in V_G$ with $xz \notin E_G$. Observe that $\mathcal{P}(G)$ is a spanning subgraph of the line graph of G. We need the following lemma.

Lemma 16. *Let H be a spanning subgraph of a graph G. Then H is a square root of G if and only if E_H is an independent set of $\mathcal{P}(G)$ and every two adjacent vertices in G are at distance at most 2 in H.*

Proof. First suppose that H is a square root of G. By definition, every two adjacent vertices in G are of distance at most 2 in H. In order to show that E_H is an independent set in $\mathcal{P}(G)$, assume that two edges $e_1, e_2 \in E_H$ are adjacent vertices in $\mathcal{P}(G)$. Then $e_1 = xy$ and $e_2 = yz$ for three distinct vertices $x, y, z \in V_G$ with $xz \notin E_G$. This means that x and z are of distance 2 in H implying that $xz \in E_G$, which is a contradiction.

Now suppose that E_H is an independent set of $\mathcal{P}(G)$ and that every two adjacent vertices in G are at distance at most 2 in H. In order to show that H is a square root of G, it suffices to show that every two non-adjacent vertices in G have distance at least 3 in H. Let u and v be two non-adjacent vertices in G that have distance at most 2 in H. Then there exists a vertex $z \notin \{u, v\}$ such that $uz, vz \in E_H$. Then $e_1 = uz$ and $e_2 = vz$ are adjacent in $\mathcal{P}(G)$ contradicting the independence of E_H in $\mathcal{P}(G)$. $\square$

We use Lemma 16 to prove Propositions 1 and 2. Here, we use the O^*-notation to suppress any polynomial factors. A *vertex cover* is a subset $U \subseteq V$ such that every edge is incident with at least one vertex of U. The VERTEX COVER problem is that of testing whether a given graph has a vertex cover of size at most p for a given integer p.

In Proposition 1 we prove that there is a $O^*(2^k)$ time algorithm to decide whether a given graph G has square root H such that $|E_G \setminus E_H| \leq k$.

Proposition 1. MAXIMUM SQUARE ROOT *can be solved in time $O^*(2^k)$.*

Proof. Let G be a graph with n vertices and m edges, and let $k \geq 0$ be an integer. By Lemma 16 it suffices to check whether $\mathcal{P}(G)$ has a vertex cover U of size at most k such that $H_U = (V_G, E_G \setminus U)$ is a square root of G. All vertex covers of size at most k of a graph can be enumerated by adapting the standard $O^*(2^k)$ branching algorithm for the VERTEX COVER problem (see for example [6]). It requires $O(m^2)$ time to compute $\mathcal{P}(G)$ and $O(nm)$ time to check whether a graph H_U is a square root of G. Hence the overall running time of our algorithm is $O^*(2^k)$. $\square$

We observe that MAXIMUM SQUARE ROOT has a linear kernel for connected graphs. This immediately follows from a result of Aingworth, Motwani and Harary [2], who proved that if H is a square root of a connected n-vertex graph $G \neq K_n$, then $|E_G \setminus E_H| \geq n - 2$. Hence, $n \leq k + 2$ for every yes-instance (G, k) of MAXIMUM SQUARE ROOT with $G \neq K_n$ (trivially, K_n is its own square root). Note that this kernel does not lead to a faster running time than $O^*(2^k)$.

In Proposition 2 we present our exact algorithm, which does not only solve the decision problem but in fact determines a square root of a given graph that has maximum number of edges.

Proposition 2. MAXIMUM SQUARE ROOT *can be solved in time* $O^*(3^{m/3})$ *on graphs with m vertices.*

Proof. Let G be a graph with n vertices and m edges, and let $k \geq 0$ be an integer. We compute the graph $\mathcal{P}(G)$, enumerate all maximal independent sets I of $\mathcal{P}(G)$, and verify for each $I \subseteq E_G$ whether G is the square of the graph $H_I = (V_G, I)$. Out of those graphs H_I that are square roots of G, return the one with maximum number edges; if no such graph H_I has been found, then G has no square roots. Correctness follows from Lemma 16. Recall that $\mathcal{P}(G)$ can be computed in time $O(m^2)$. All the maximal independent sets of the m-vertex graph $\mathcal{P}(G)$ can be enumerated in time $O^*(3^{m/3})$ using the polynomial delay algorithm of Tsukiyama et al. [21], since $\mathcal{P}(G)$ has at most $3^{m/3}$ maximal independent sets [16]. Finally, recall that for each maximal independent set I, we can check in time $O(nm)$ whether $(H_I)^2 = G$. Hence the overall running time of our algorithm is $O^*(3^{m/3})$. □

4 Open Problems

We conclude our paper with two open problems. First, is it also possible to construct an exact algorithm for MINIMUM SQUARE ROOT that is better than the trivial exact algorithm?

Second, recall that if H is a square root of a connected n-vertex graph $G \neq K_n$, then $|E_G \setminus E_H| \geq n-2$ [2]. Is it FPT to decide whether a connected n-vertex graph $G \neq K_n$ has a square root that can be obtained by removing at most $n-2+k$ edges, or equivalently, whether a connected n-vertex graph $G \neq K_n$ has a square root with at least $|E_G| - n + 2 - k$ edges, when parameterized by k? In particular, can it be decided in polynomial time whether a connected graph G has a square root with *exactly* $|E_G| - |V_G| + 2$ edges?

References

[1] A. ADAMASZEK AND M. ADAMASZEK, *Uniqueness of graph square roots of girth six*, Electr. J. Comb., 18 (2011).

[2] D. AINGWORTH, R. MOTWANI, AND F. HARARY, *The difference between a graph and its square*, Util. Math., 54 (1998), pp. 223–228.

[3] N. ALON, G. GUTIN, E. J. KIM, S. SZEIDER, AND A. YEO, *Solving max-r-sat above a tight lower bound*, Algorithmica, 61 (2011), pp. 638–655.

[4] M. COCHEFERT, J.-F. COUTURIER, P. A. GOLOVACH, D. KRATSCH, AND D. PAULUSMA, *Sparse square roots*, in WG 2013, vol. 8165 of Lecture Notes Comp. Sci., Springer, 2013, pp. 177–188.

[5] R. DIESTEL, *Graph theory*, vol. 173 of Graduate Texts in Mathematics, Springer, Heidelberg, fourth ed., 2010.

[6] R. G. DOWNEY AND M. R. FELLOWS, *Parameterized complexity*, Monographs in Computer Science, Springer-Verlag, New York, 1999.

[7] B. FARZAD AND M. KARIMI, *Square-root finding problem in graphs, a complete dichotomy theorem*, CoRR, abs/1210.7684 (2012).

[8] B. FARZAD, L. C. LAU, V. B. LE, AND N. N. TUY, *Complexity of finding graph roots with girth conditions*, Algorithmica, 62 (2012), pp. 38–53.

[9] J. FLUM AND M. GROHE, *Parameterized Complexity Theory*, Texts in Theoretical Computer Science. An EATCS Series, Springer-Verlag, Berlin, 2006.

[10] L. C. LAU, *Bipartite roots of graphs*, ACM Transactions on Algorithms, 2 (2006), pp. 178–208.

[11] L. C. LAU AND D. G. CORNEIL, *Recognizing powers of proper interval, split, and chordal graph*, SIAM J. Discrete Math., 18 (2004), pp. 83–102.

[12] V. B. LE AND N. N. TUY, *The square of a block graph*, Discrete Mathematics, 310 (2010), pp. 734–741.

[13] ——, *A good characterization of squares of strongly chordal split graphs*, Inf. Process. Lett., 111 (2011), pp. 120–123.

[14] Y.-L. LIN AND S. SKIENA, *Algorithms for square roots of graphs*, SIAM J. Discrete Math., 8 (1995), pp. 99–118.

[15] M. MILANIC AND O. SCHAUDT, *Computing square roots of trivially perfect and threshold graphs*, Discrete Applied Mathematics, in press.

[16] J. W. MOON AND L. MOSER, *On cliques in graphs*, Israel J. Math., 3 (1965), pp. 23–28.

[17] R. MOTWANI AND M. SUDAN, *Computing roots of graphs is hard*, Discrete Applied Mathematics, 54 (1994), pp. 81–88.

[18] A. MUKHOPADHYAY, *The square root of a graph*, J. Combinatorial Theory, 2 (1967), pp. 290–295.

[19] R. NIEDERMEIER, *Invitation to fixed-parameter algorithms*, vol. 31 of Oxford Lecture Series in Mathematics and its Applications, Oxford University Press, Oxford, 2006.

[20] I. C. ROSS AND F. HARARY, *The square of a tree*, Bell System Tech. J, 39 (1960), pp. 641–647.

[21] S. TSUKIYAMA, M. IDE, H. ARIYOSHI, AND I. SHIRAKAWA, *A new algorithm for generating all the maximal independent sets*, SIAM J. Comput., 6 (1977), pp. 505–517.