SIAM J. DISCRETE MATH.
Vol. 38, No. 2, pp. 1840–1862

ON GRAPHS COVERABLE BY k SHORTEST PATHS*

MAËL DUMAS†, FLORENT FOUCAUD‡, ANTHONY PEREZ†, AND IOAN TODINCA†

Abstract. We show that if the edges or vertices of an undirected graph G can be covered by k shortest paths, then the pathwidth of G is upper-bounded by a single-exponential function of k. As a corollary, we prove that the problem Isometric Path Cover with Terminals (which, given a graph G and a set of k pairs of vertices called *terminals,* asks whether G can be covered by k shortest paths, each joining a pair of terminals) is FPT with respect to the number of terminals. The same holds for the similar problem Strong Geodetic Set with Terminals (which, given a graph G and a set of k terminals, asks whether there exist $\binom{k}{2}$ shortest paths covering G, each joining a distinct pair of terminals). Moreover, this implies that the related problems Isometric Path Cover and Strong Geodetic Set (defined similarly but where the set of terminals is not part of the input) are in XP with respect to parameter k.

Key words. covering problems, shortest paths, graph theory, parameterized algorithms

MSC code. 68W40

DOI. 10.1137/23M1564511

1. Introduction. Path problems such as Hamiltonian Path are among the most fundamental problems in the field of algorithms. Hamiltonian Path can be generalized as the covering problem Path Cover [2], where one asks to cover the vertices of an input graph using a prescribed number of paths. The packing variant is Disjoint Paths where, given a set of k pairs of terminal vertices of a graph G, one asks whether there are k vertex-disjoint paths in G, each joining two paired terminals. Disjoint Paths is a fundamental problem and a precursor to the field of parameterized complexity due to the celebrated fixed-parameter tractable algorithm devised by Robertson and Seymour [25] for the parameter "number of paths". We recall that in the field of parameterized algorithms and complexity, one studies *parameterized problems,* whose input I comes together with a parameter k. A parameterized problem is said to be FPT (fixed-parameter tractable) if it can be solved in time $f(k)\cdot|I|^{O(1)}$ for some computable function f. If the problem can be solved in time $O(|I|^{f(k)})$, it belongs to class XP; see, e.g., [8] for more details.

In this paper, we will not consider arbitrary paths, but *shortest paths,* which are fundamental for many applications. In the problem Disjoint Shortest Paths, given a graph G and k pairs of terminals, one asks whether G contains k vertex-disjoint shortest paths pairwise connecting the k pairs of terminals. This problem was introduced in [13] and recently shown to be polynomial-time solvable for every

*Received by the editors April 7, 2023; accepted for publication (in revised form) December 13, 2023; published electronically June 10, 2024. A preliminary version of this paper appeared in Proceedings of the 33rd International Symposium on Algorithms and Computation, ISAAC 2022, Seoul, Korea, LIPIcs 248, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022, 40.
https://doi.org/10.1137/23M1564511
Funding: The second author was financed by the IFCAM project "Applications of graph homomorphisms" (MA/IFCAM/18/39), by the ANR project GRALMECO (ANR-21-CE48-0004), and by the French government IDEX-ISITE initiative 16-IDEX-0001 (CAP 20-25).
†Université Orléans, INSA Centre Val de Loire, LIFO EA, 4022, F-45067 Orléans Cedex 2, France (mael.dumas@univ-orleans.fr, anthony.perez@univ-orleans.fr, Ioan.Todinca@univ-orleans.fr).
‡Université Clermont-Auvergne, CNRS, Mines de Saint-Étienne, Clermont-Auvergne-INP, LIMOS, 63000, Clermont-Ferrand, France (florent.foucaud@uca.fr).

fixed k by an XP algorithm [5, 20]. The problem Isometric Path Cover with Terminals, which we define as follows, is the covering counterpart of Disjoint Shortest Paths.

Isometric Path Cover with Terminals

Input: A graph G and k pairs of vertices $(s_1, t_1), \dots, (s_k, t_k)$ called *terminals.*
Question: Does there exist a set of k shortest paths, the ith path being an s_i-t_i shortest path, such that each vertex of G belongs to at least one of the paths?

The name Isometric Path Cover with Terminals comes from the related Isometric Path Cover problem where the terminals are not part of the input, which was introduced in [15] in the context of the Cops and Robbers game on graphs (see also [1]).

Isometric Path Cover

Input: A graph G and an integer k.
Question: Does there exist a set of k shortest paths such that each vertex of G belongs to at least one of the paths?

Closely related variants of Isometric Path Cover with Terminals and Isometric Path Cover have been studied, in which there are only k terminals, and one asks to find $\binom{k}{2}$ shortest paths joining each pair of terminals. The version without terminals has been called Strong Geodetic Set in the literature; we call the version with terminals Strong Geodetic Set with Terminals. It was first studied (independently) in [9, 10]; see also [19].

Strong Geodetic Set with Terminals

Input: A graph G and a set of k vertices of G called *terminals.*
Question: Does there exist a set of $\binom{k}{2}$ shortest paths, each path joining a distinct pair of terminals, such that each vertex of G belongs to at least one of the paths?

The variant where the terminals are not given in the input was defined in [4] as follows.

Strong Geodetic Set

Input: A graph G and an integer k.
Question: Does there exist a set of k terminals and a set of $\binom{k}{2}$ shortest paths, each path joining a distinct pair of terminals, such that each vertex of G belongs to at least one of the paths?

The complexity of these problems has been studied in the literature. It was shown in [6] that Isometric Path Cover is NP-complete, even for chordal graphs, a class for which the authors also provide a constant-factor approximation algorithm w.r.t. the number of paths. For general graphs, it was shown in [26] that the problem can be approximated in polynomial time within a factor of $O(\log d)$, where d is the diameter of the input graph. It was proven to be polynomial-time solvable on block graphs [24]. It is shown in [9] that Strong Geodetic Set with Terminals is NP-hard. In [10], it is shown that this holds even for bipartite graphs of maximum degree 4 or diameter 6; however, Strong Geodetic Set with Terminals is polynomial-time solvable on split graphs, graphs of diameter 2, block graphs, and cactus graphs. We prove here (cf. Proposition 5.1) that Isometric Path Cover with Terminals is also NP-complete.

Finally, Strong Geodetic Set is known to be NP-hard [4], even for bipartite graphs, chordal graphs, graphs of diameter 2 and cobipartite graphs [10] as well as for subcubic graphs of arbitrary girth [9]. However, it is polynomial-time solvable on outerplanar graphs [23], cactus graphs, block graphs, and threshold graphs [10].

All of these problems can also be studied in their edge-covering version, where one requires to cover all edges of the input graph by the corresponding shortest paths. For instance, the Strong Edge Geodetic Set problem is studied in [22]. The version of Isometric Path Cover where the solution paths are required to be vertex-disjoint has also been studied under the names of Isometric Path Partition [21] and Shortest Path Partition [14].

Our results. Our main combinatorial theorems are as follows (see section 2 for the definition of pathwidth).

Theorem 1.1. *Let G be a graph whose* edge set *can be covered by at most k shortest paths. Then the pathwidth of G is $O(3^k)$.*

Theorem 1.2. *Let G be a graph whose* vertex set *can be covered by at most k shortest paths. Then the pathwidth of G is $O(k \cdot 3^k)$.*

We actually show that in such a graph G, given an arbitrary vertex a and an integer D, the number of vertices at distance exactly D from a is upper-bounded by a function of k, and the bound does not depend on the size of the input graph. It follows that a very simple linear-time algorithm based on a breadth-first search provides a path decomposition whose width is upper-bounded by twice the aforementioned function of k. The complexity of the algorithm computing the path decomposition does not depend on k.

Besides the combinatorial bounds, we employ the celebrated theorem of Courcelle [7], stating that problems expressible in Monadic Second-Order Logic (MSOL_2) can be solved in linear time for graphs of bounded treewidth (and, thus, of bounded pathwidth). More precisely, we reduce the problem Isometric Path Cover with Terminals to an optimization problem expressible in MSOL_2. The result can also be obtained by dynamic programming but the algorithm would be tedious and not particularly efficient, therefore, we prefer the general logic-based framework for further extensions. The running time is linear in n, the number of vertices of the graph, but superexponential in the parameter k. Together with Theorem 1.2, this implies the following.

Theorem 1.3. Isometric Path Cover with Terminals *and* Strong Geodetic Set with Terminals *are FPT when parameterized by the number of terminals.*

Corollary 1.4. Isometric Path Cover *and* Strong Geodetic Set *are in XP when parameterized by the number of paths, respectively terminals.*

Thanks to the flexibility of MSOL_2 and to Theorem 1.1, our algorithmic results easily extend to the edge-covering versions of our problems, and to variants where we require the paths to be edge-disjoint, or vertex-disjoint like Isometric Path Partition, studied in [14, 21]. The second part of Theorem 1.3 answers positively a question asked in [19].

Outline. After some preliminaries in section 2, we prove Theorems 1.1 and 1.2 in sections 3 and 4, respectively. More specifically, section 3 provides the upper bound on the pathwidth of graphs whose edges are coverable by k shortest paths, then the tools are extended to vertex-coverings in the next section. Algorithmic consequences (Theorem 1.3) are derived in section 5, and we conclude with some open questions.

2. Preliminaries and notations.

Paths and concatenation operators $\oplus$ and $\odot$. We refer to [11] for usual notations on graphs. In this paper we only consider undirected, unweighted graphs. For simplicity, we assume that our input graph $G=(V,E)$ is connected, though all our combinatorial and algorithmic results extend to nonconnected graphs. As usual $N(x)$ denotes the neighborhood of vertex x.

A path P of graph $G=(V,E)$ is a sequence of distinct vertices $(x_1,\ldots,x_l)$ such that for each $i, 1\le i\le l-1$, $\{x_i,x_{i+1}\}$ is an edge of the graph. We also say that P is an x_1-x_l path. Note that our paths are simple as they do not use twice the same vertex. We denote by $V(P)$ the vertices of path P, and by $E(P)$ its edges. Given two vertices $x,y\in V(P)$, we denote by $P[x,y]$ the subpath of P between x and y. Let $|P|$ denote the *length* of path P, that is, its number of edges. The *distance* between two vertices a and b in G is denoted $\operatorname{dist}(a,b)$ and corresponds to the length of a shortest a-b path.

Throughout this paper, we will construct paths by concatenation operations. It is convenient to think of our paths as directed: when we speak of an a-b path, we think of it as being directed from a to b.

Given two vertex disjoint paths $\nu=(x_1,\ldots,x_l)$ and $\eta=(y_1,\ldots,y_t)$ of G such that $\{x_l,y_1\}$ is an edge of G, we define the *concatenation operator* $\oplus$ whose result is $\nu\oplus\eta=(x_1,\ldots,x_l,y_1,\ldots,y_t)$. In particular, $|\nu\oplus\eta|=|\nu|+|\eta|+1$.

We define similarly the *glueing operator* $\odot$ between two paths $\nu=(x_1,\ldots,x_l)$ and $\eta=(x_l,y_1,\ldots,y_t)$ with $V(\nu)\cap V(\mu)=\{x_l\}$ by $\nu\odot\eta=(x_1,\ldots,x_l,y_1,\ldots,y_t)$. Note that in this case $|\nu\odot\eta|=|\nu|+|\eta|$.

Path decompositions through breadth-first search. A path decomposition of $G=(V,E)$ is a sequence $\mathcal{P}=(X_1,X_2,\ldots,X_q)$ of vertex subsets of G, called *bags,* such that for every edge $\{x,y\}\in E$ there is at least one bag containing both endpoints, and for every vertex $x\in V$, the bags containing x form a continuous subsequence of $\mathcal{P}$. The width of $\mathcal{P}$ is $\max\{|X_i|-1\mid 1\le i\le q\}$, and the *pathwidth* $\operatorname{pw}(G)$ of G is the minimum width over all path decompositions of G.

The *treewidth* $\operatorname{tw}(G)$ of graph G is defined similarly (see, e.g., [11]), using a so-called tree decomposition; for our purpose, we only need to know that for any graph G, $\operatorname{tw}(G)\le\operatorname{pw}(G)$ in particular, any path decomposition is also a tree decomposition of the same width. We also need the following folklore lemma on path decompositions.

Lemma 2.1. *Let $G=(V,E)$ be a graph, a vertex $a\in V$, and let K be an upper bound on the number of vertices of G at distance exactly D from a for any integer D.*

Then, $\operatorname{pw}(G)\le 2K-1$. *Moreover, a path decomposition of width $2K-1$ can be computed in linear time by breadth-first search.*

Proof. Let $\operatorname{ecc}(a)$ be the eccentricity of vertex a (i.e., $\max_{x\in V}\operatorname{dist}(a,x)$). For any D with $0\le D\le\operatorname{ecc}(a)$ we denote by $\operatorname{Layer}(D)$ the set of vertices at distance exactly D from a, i.e., the layers of a breadth-first search on G starting at a. Observe that, by taking as bags the unions $\mathrm{Layer}(D) \cup \mathrm{Layer}(D+1)$ of pairs of consecutive layers, $0 \le D < \mathrm{ecc}(a)$, and by ordering them according to D, we obtain a path decomposition of G. Indeed, for each edge $\{x, y\}$ both endpoints are in the same layer or in two consecutive layers, thus will appear in the same bag. For each vertex x, it appears in at most two bags: if $d = \mathrm{dist}(a, x)$, then x is in bags $\mathrm{Layer}(d-1) \cup \mathrm{Layer}(d)$ and $\mathrm{Layer}(d) \cup \mathrm{Layer}(d+1)$ (or one bag if $d = 0$ or $d = \mathrm{ecc}(a)$), and these bags appear consecutively in the decomposition. Since each layer has at most K vertices, the width of this decomposition is at most $2K - 1$. ∎

3. Edge-covering with k shortest paths.

We start by proving Theorem 1.1, upper bounding the pathwidth of graphs $G = (V, E)$ that are edge-coverable by k shortest paths. In this case, there is a simple and elegant encoding of shortest paths leading to a factorial upper bound, given in subsection 3.1. This bound is improved to a single-exponential one in subsection 3.2. We recall that the case of vertex-coverings, which is more technical, will be studied in section 4.

In this section, $G = (V, E)$ denotes a graph whose edge set is coverable by k shortest paths. Let us fix such a set of paths $\mu_1, \ldots, \mu_k$, and call them the *base paths* of G. All constructions in this section are built on this particular set of base paths (without explicitly recalling it for each lemma, in order to ease the notations). We endow each base path μ_c, $1 \le c \le k$ with an arbitrary *direction,* e.g., assuming that the vertices of G are numbered from 1 to n, the direction of path P is from its smallest towards its largest end-vertex. A (directed) subpath $\mu_c[x, y]$ of μ_c is given a positive *sign* $+$ if it follows the direction of μ_c, otherwise it is given a negative sign $-$. For each edge e of G, let $\texttt{Colors}(e)$ be the set of all values $c \in \{1 \ldots, k\}$ such that e is an edge of μ_c.

Good colorings. Let P be an a-b path of G, from vertex a to vertex b. A *coloring* of P is a function $\texttt{col} : E(P) \to \{1, \ldots, k\}$ assigning to each edge e of the path one of its colors $\texttt{col}(e) \in \texttt{Colors}(e)$. The coloring $\texttt{col}$ of P is said to be *good* if, for any color c, the set of edges using this color form a connected subpath $P[x, y]$ of P. (Since our paths are simple, this condition entails that $P[x, y] = \mu_c[x, y]$.) A pair $(P, \texttt{col})$ formed by a path together with a good coloring is called *well-colored path.*

Operator $\odot$ defined in section 2 naturally extends to colored paths. Given a colored path $(P, \texttt{col})$, in a slight abuse of notation we denote by $(P[x, y], \texttt{col})$ its restriction to a subpath $P[x, y]$ of P. Finally, we define for any path P and any color $1 \le c \le k$ the function $\texttt{monochr}_c : E(P) \to \{c\}$. Hence all edges of the colored path $(P, \texttt{monochr}_c)$ have color c.

With these notations, any well-colored a-b path $(P, \texttt{col})$ with colors $(c_1, \ldots, c_l)$ appearing in this order can be written as

$$(\mu_{c_1}[x_1, x_2], \texttt{monochr}_{c_1}) \odot (\mu_{c_2}[x_2, x_3], \texttt{monochr}_{c_2}) \odot \cdots \odot (\mu_{c_l}[x_l, x_{l+1}], \texttt{monochr}_{c_l})$$

for some vertices $a = x_1, x_2, x_3 \ldots, x_l, x_{l+1} = b$. In full words, $P[x_i, x_{i+1}]$ are the monochromatic subpaths of $(P, \texttt{col})$, colored c_i.

LEMMA 3.1 (good coloring lemma). *For any pair of vertices a and b of G, there exists a well-colored a-b path $(P, \texttt{col})$ such that P is a shortest a-b path.*

We will simply call $(P, \texttt{col})$ a well-colored shortest a-b path.

Proof. Among all shortest a-b paths, choose one that admits a coloring with a minimum number of monochromatic subpaths. Let P be this path, and let $\texttt{col}$ be the corresponding coloring. Assume by contradiction that the coloring is not good. Then there exist three edges $e_1 = \{y_1, z_1\}$, $e_2 = \{y_2, z_2\}$, and $e_3 = \{y_3, z_3\}$, appearing in this order, such that $\texttt{col}(e_1) = \texttt{col}(e_3) \neq \texttt{col}(e_2)$. Assume w.l.o.g. that the vertices appear in the order $y_1, z_1, y_2, z_2, y_3, z_3$ from a to b (note that we may have $z_1 = y_2$ or $z_2 = y_3$). Let $c = \texttt{col}(e_1) = \texttt{col}(e_3)$. Therefore, z_1 and y_3 are on the same base path μ_c. Let P' be the path obtained from P by replacing $P[z_1, y_3]$ by $\mu_c[z_1, y_3]$. First, P' is no longer than P, since $\mu_c[z_1, y_3]$ is a shortest possible z_1-y_3 path of graph G (in particular, P' has no repeated vertices). Second, in P' we can color all edges of $P'[z_1, y_3]$ with color c, and keep all other colors unchanged. Hence P' has strictly fewer monochromatic subpaths than P — a contradiction. ∎

Let $(P, \texttt{col})$ be a well-colored a-b path, with colors $(c_1, \ldots, c_l)$ in this order. Recall that each monochromatic subpath $P[x_i, x_{i+1}]$ of P, of color c_i, induces a sign on the corresponding base path μ_{c_i} (positive if $P[x_i, x_{i+1}]$ has the same direction as μ_{c_i}, negative otherwise). Therefore, we can define the *colors-signs word* $\texttt{ColorsSignsWord}(P, \texttt{col}) = ((c_1, s_1), (c_2, s_2), \ldots, (c_l, s_l))$ on the alphabet $\{1, \ldots, k\} \times \{+, -\}$, corresponding to the colors and signs of the monochromatic subpaths of P, according to the ordering in which these subpaths appear from a to b.

3.1. Warm-up: Factorial bound. Observe that colors-signs words have at most k letters on an alphabet of size $2k$. Therefore, the number of different such words is upper bounded by a function of k.

LEMMA 3.2. *The number of possible colors-signs words, over all well-colored paths of G, is upper bounded by $h(k) = \sum_{l=1}^{k} 2^l \frac{k!}{(k-l)!}$.*

Proof. We claim that the number of colors-signs words of l letters is upper bounded by $2^l \frac{k!}{(k-l)!}$. Observe that the colors form a word of length l, on an alphabet of size k, without repetition. The number of such words is $\frac{k!}{(k-l)!}$ (e.g., by choosing l letters among the k possible ones, and applying all possible permutations). Since each letter also has a sign in $\{+, -\}$, we multiply this quantity by 2^l, and the conclusion follows by summing over all possible values of l. ∎

The following crucial lemma implies that, given a start vertex a, a distance D, and a colors-signs word ω, there is at most one vertex b at distance D from a, such that the well-colored shortest a-b path respects word ω. This will allow one to upper bound the number of vertices at distance D from a.

LEMMA 3.3 (colors-signs encoding). *Consider two vertices b and c at the same distance from some vertex a of G. Let $(P, \texttt{col})$ be a well-colored shortest a-b path and $(P', \texttt{col}')$ be a well-colored shortest a-c path. If $\texttt{ColorsSignsWord}(P, \texttt{col}) = \texttt{ColorsSignsWord}(P', \texttt{col}')$, then $b = c$.*

Proof. We proceed by induction on the number of letters of the word:

$$\omega = \texttt{ColorsSignsWord}(P, \texttt{col})$$

Let us denote it by $\omega = ((c_1, s_1), (c_2, s_2), \ldots, (c_l, s_l))$. Let $P[a, x_2]$ (resp., $P'[a, x'_2]$) be the maximal subpath of P (resp., P') of color c_1 starting from a. Assume w.l.o.g. that $P[a, x_2]$ is at least as long as $P'[a, x'_2]$. Since both are subpaths of μ_{c_1}, starting from a and having the same sign s_1 w.r.t. μ_{c_1}, we actually have that $P'[a, x'_2]$ is contained in $P[a, x_2]$, in particular, x'_2 is between a and x_2 in P and in μ_{c_1}.

Observe that if word ω has only one letter, $P[a, x_2] = P$ and $P'[a, x'_2] = P'$, thus they are all of the same length. Since they are of the same sign w.r.t. μ_{c_1}, this implies that $x_2 = x'_2 = b = c$, which proves the base case of our induction.

Assume now that ω has $l \geq 2$ letters and that the lemma is true for words of length $l-1$.

Consider first the case when $P[a, x_2]$ and $P'[a, x_2']$ have the same length. Then $x_2 = x_2'$ is also the first vertex of the subpaths of color c_2 of both P and P'. Then $(P[x_2', b], \texttt{col})$ and $(P'[x_2', c], \texttt{col}')$ are well-colored shortest paths of the same length, and have the same colors-signs word $((c_2, s_2), \dots, (c_l, s_l))$, with $l-1$ letters. Hence the property follows by the induction hypothesis.

We now handle the second and last case, when $P[a, x_2]$ is strictly longer than $P'[a, x_2']$. Let x_3 be the last vertex of the subpath colored c_2 in $(P, \texttt{col})$. In particular, x_2', x_2, and x_3 are all vertices of μ_{c_2}. Let us make an easy but crucial observation: on μ_{c_2}, vertex x_2 is between x_2' and x_3. To prove this claim, note that in path P, vertices a, x_2', and x_2 appear in this order (as observed in the beginning of the proof), and by construction x_2 appears between a and x_3. Therefore, a, x_2', x_2, x_3 appear in this order on P, which is a shortest path. Hence $dist(x_2', x_3) = dist(x_2', x_2) + dist(x_2, x_3)$. Since the three vertices x_2', x_2, x_3 are all on the shortest path μ_{c_2}, they must appear in this order on it. Consequently, $\mu_{c_2}[x_2', x_3]$ induces the same sign s_2 on μ_{c_2} as $P[x_2, x_3] = \mu_{c_2}[x_2, x_3]$. In particular, in path

$$(P[x_2', b], \texttt{col}) = (P[x_2', x_2], \texttt{monochr}_{c_1}) \odot (P[x_2, x_3], \texttt{monochr}_{c_2}) \odot (P[x_3, b], \texttt{col})$$

we can replace the first subpath $P[x_2', x_2]$ colored c_1 by $\mu_{c_2}[x_2', x_2]$, colored c_2, without changing the total length. We obtain the well-colored shortest x_2'-b path

$$(\tilde{P}[x_2', b], \tilde{col}) = (\mu_{c_2}[x_2', x_3], \texttt{monochr}_{c_2}) \odot (P[x_3, b], \texttt{col}).$$

Its colors-signs word is $((c_2, s_2), \dots, (c_l, s_l))$, the same as for the shortest x_2'-c path $(P'[x_2', c], \texttt{col}')$. Moreover, the two paths have the same length, $|P| - |P[a, x_2']|$, hence by the induction hypothesis we have $b = c$, which proves our lemma. ∎

Corollary 3.4. *For any vertex a of G and any integer D, there are at most $h(k) = \sum_{l=1}^{k} 2^l \frac{k!}{(k-l)!}$ vertices at distance exactly D from a.*

Proof. For any fixed vertex a and fixed integer D, thanks to Lemma 3.3 the number of vertices x at distance exactly D from a is upper-bounded by the number of colors-signs words, which is, in turn, upper bounded by $h(k)$ by Lemma 3.2. ∎

Corollary 3.4 together with Lemma 2.1 entail a weaker version of Theorem 1.1: the pathwidth of graphs edge-coverable by k shortest paths is at most $2h(k) - 1$, providing a factorial upper bound.

3.2. Single exponential bound. In this section, we will show that the number of vertices at distance D from a vertex a can actually be bounded by $O(3^k)$. In the previous section we showed that any two shortest well-colored paths of length D starting from a with the same colors-signs words lead to the same vertex. We observe that paths with different colors-signs word may also lead to the same vertex; see Figure 1. In this section, we generalize this idea to a set of shortest paths from a to all vertices at distance D from a.

Before getting into details, let us try to give an informal description of our construction (see Figure 2). Let $\mathcal{P}$ be a family of well-colored shortest paths starting from vertex a, with different endpoints. Initially, this family reaches all vertices at distance D from a. We transform it into another family with the same properties, preserving the endpoints. Let x_n be the furthest vertex from a, contained on all paths of $\mathcal{P}$, such that these paths share the same colored subpath from a to x_n. We focus

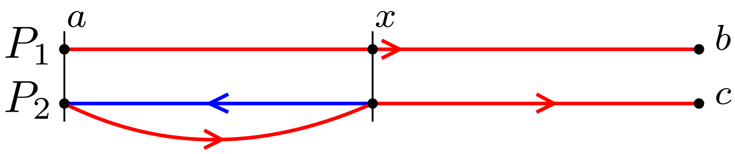

FIG. 1. *$(P_1,\texttt{col}_1)$ and $(P_2,\texttt{col}_2)$ are well-colored shortest paths of same length. The vertex x is shared by the two paths. One can replace $\mu_{blue}[a,x]$ in P_2 by $\mu_{red}[a,x]$, this proves that $b=c$. (Color available online.)*

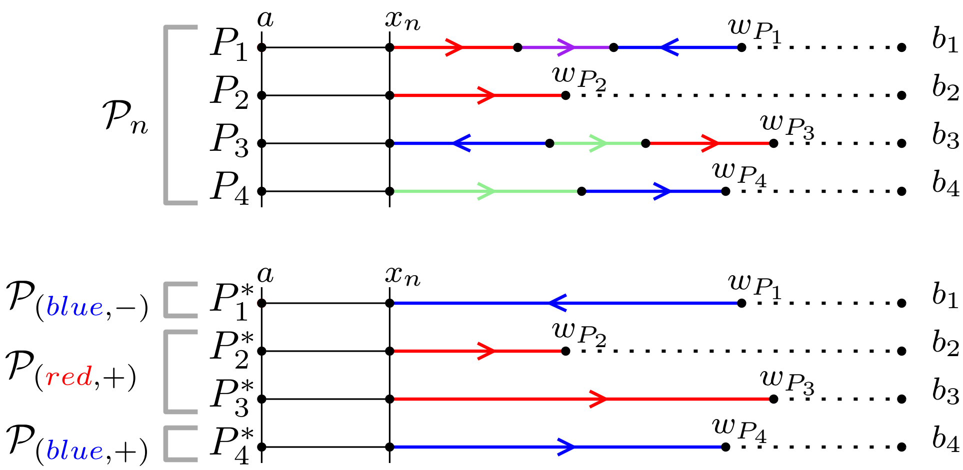

FIG. 2. *Example of the construction in the proof of Lemma 3.6 for a set of paths $\mathcal{P}_n$. In this example, all paths share the same subpath $(P_1[a,x_n],\texttt{col})$, $\mathcal{C}_n=\{red,blue,green\}$ and there are no edges with color in $\mathcal{C}_n$ in the path $(P_i[w_{P_i}:],\texttt{col}_i)$. The sets $\mathcal{P}_{(red,-)},\mathcal{P}_{(green,+)},\mathcal{P}_{(green,-)}$ are empty. (Color available online.)*

on the colors $\mathcal{C}$ appearing on the edges right after x_n. First, we apply the following transformation. For each path $(P,\texttt{col})\in\mathcal{P}$, we consider the furthest edge having its color c in $\mathcal{C}$ (e.g., in Figure 2, for path P_3 this edge is colored red), and we replace the whole subpath of P from x_n to this edge with a subpath of μ_c, of color c. This new family of paths is then split into subfamilies according to the color and sign of the first edge after x_n, and we branch on each subfamily. We eventually prove that this branching process corresponds to a tree (that we call *k-labelled branching tree*), in the sense that the output paths are bijectively mapped on the leaves of the tree. Moreover, we will upper bound the number of leaves of the tree, and thus the number of paths, by $O(3^k)$.

We first define k-labelled branching trees and show in Lemma 3.9 that they have $O(3^k)$ leaves. Then we formally describe the above branching process in Lemma 3.6, deducing the same upper bound on the number of vertices at distance D from a.

Let us define a *rooted labelled tree* as a pair $(T,\texttt{lab})$ where T is a rooted tree and $\texttt{lab}: E(T)\to\mathcal{CS}$ is an edge labelling function with $\mathcal{CS}=\{1,\dots,k\}\times\{+,-\}$. For a node n of T let T_n be the subtree rooted in n. For a node n of T and its children $n_1,\dots,n_t$ we let

$$\mathcal{L}(n)=\{\texttt{lab}(\{n,n_1\}),\dots,\texttt{lab}(\{n,n_t\})\},$$
$$\mathcal{LS}(n)=\{(c,s),(c,\overline{s})\mid(c,s)\in\mathcal{L}(n)\},\ \text{where } \overline{s} \text{ is the opposite of sign } s.$$

Definition 3.5. *A rooted labelled tree* $(T,\texttt{lab})$ *is called a* k-labelled branching tree *if the following properties hold:*

(i) *in every path from the root of* T *to one of its leaves, the edges with the same label form a connected subpath.*

And for every node n *of* T *with children* $n_1,\dots,n_t$*:*

(ii) *the labels on the edges* $\{n,n_i\}, i\in\{1,\dots,t\}$ *are pairwise distinct,*

(iii) *for every* $i\in\{1,\dots,t\}$*,* T_{n_i} *does not contain any edge with labels in* $\mathcal{LS}(n)\setminus\{\texttt{lab}(\{n,n_i\})\}$.

We note that in a general setting where the labelling function lab is defined over some alphabet Σ and the third condition is replaced by "T_{n_i} does not contain any edge with labels in $\mathcal{L}(n)\setminus\{\texttt{lab}(\{n,n_i\})\}$", one can achieve an $O(2^{|\Sigma|})$ bound, thus giving $O(4^k)$ in our case (observe that we simply replaced $\mathcal{LS}$ by $\mathcal{L}$ in the second condition). However by exploiting the specificity of the alphabet $\mathcal{CS}$ we will prove an $O(3^k)$ bound in Lemma 3.9. For the sake of readability we defer this technical proof to the end of the section and first show how k-labelled branching trees can be used to bound the number of vertices at a given distance of any vertex.

In a slight abuse of notation, we define $\texttt{Colors}(P,\texttt{col})$ as the set of colors that appear on any colored path $(P,\texttt{col})$, *i.e.*, $\texttt{Colors}(P,\texttt{col})=\{\texttt{col}(e)\mid e\in E(P)\}$. For a set of colored paths $\mathcal{P}$, $\texttt{Colors}(\mathcal{P})=\bigcup_{(P,\texttt{col})\in\mathcal{P}}\texttt{Colors}(P,\texttt{col})$.

For a coloued a-b path $(P,\texttt{col})$ and a vertex x of P, we let $P[x:]=P[x,b]$, and for a set of colored paths $\mathcal{P}$ that share vertex x we let $\mathcal{P}[x:]=\bigcup_{(P,\texttt{col})\in\mathcal{P}}(P[x:],\texttt{col})$.

Lemma 3.6. *For any vertex* a *of* G *and any integer* D*, there are at most* $c(k)$ *vertices at distance exactly* D *from vertex* a*, where* $c(k)$ *is an upper bound on the number of leaves of any* k*-labelled branching tree.*

Proof. Let $\{b_1\dots,b_q\}$ be the set of vertices at distance D from a in G, and let $\mathcal{P}$ denote a set $\{(P_1,\texttt{col}_1),\dots,(P_q,\texttt{col}_q)\}$ of well-colored shortest a-b_i paths. We aim to construct a k-labelled branching tree $(T,\texttt{lab})$ such that each node n of T is associated with a set of paths $\mathcal{P}_n$. Moreover, for each vertex b_i there is a leaf l of T such that $\mathcal{P}_l$ contains a well-colored shortest a-b_i path and for each leaf l, $\mathcal{P}_l$ contains exactly one path.

We will construct $(T,\texttt{lab})$ recursively starting from the root r of T. Initially, $\mathcal{P}_r=\mathcal{P}$. For every node n of T, $\mathcal{P}_n$ will be a set of well-colored shortest paths from a to a subset of vertices of $\{b_1\dots,b_q\}$. Let n be an already computed node of T. If $|\mathcal{P}_n|=1$, we do nothing: n is a leaf of our tree and no recursive step is performed at n. If $|\mathcal{P}_n|>1$, we will grow the tree from n. We construct the sets of paths that will be associated with the children of n as follows:

1. Let x_n be the vertex at maximum distance from a such that for every pair of colored paths $(P,\texttt{col}),(P',\texttt{col}')\in\mathcal{P}_n$, $(P[a,x_n],\texttt{col})=(P'[a,x_n],\texttt{col}')$. This vertex is guaranteed to exist, notice that x_n can be the vertex a.
2. Let $\mathcal{C}_n=\{\texttt{col}(\{x_n,y_P\})\mid(P,\texttt{col})\in\mathcal{P}_n\}$ where $\{x_n,y_P\}$ denotes the first edge of $P[x_n:]$.
3. For each $(P,\texttt{col})\in\mathcal{P}_n$, let $\{z_P,w_P\}$ be the edge that is farthest from x_n on $P[x_n:]$ whose color is in $\mathcal{C}_n$, and let $c_P=\texttt{col}(\{z_P,w_P\})$. Construct the path
$$(P^*,\texttt{col}^*)=(P[a,x_n],\texttt{col})\odot(\mu_{c_P}[x_n,w_P],\texttt{monochr}_{c_P})\odot(P[w_P:],\texttt{col})$$
4. For each $(c,s)\in\mathcal{CS}_n=(\mathcal{C}_n\times\{+,-\})$ define the set $\mathcal{P}_{(c,s)}$ as $\{(P^*,\texttt{col}^*)\mid(P,\texttt{col})\in\mathcal{P}_n,(c_P,s_P)=(c,s)\}$, where s_P is the sign of $\mu_{c_P}[x_n,w_P]$.

This construction is illustrated in Figure 2. For each nonempty set $\mathcal{P}_{(c,s)}, (c,s) \in \mathcal{CS}_n$, we then create a child n' of n associated with $\mathcal{P}_{n'} = \mathcal{P}_{(c,s)}$, where the edge $\{n, n'\}$ is labelled (c,s). Then, we recursively apply the process to each created node.

CLAIM 3.7. *For any node n of T, $\mathcal{P}_n$ is a set of well-colored shortest paths.*

Proof. This trivially holds for $\mathcal{P}_r$ where r is the root of T. We show that for a node n of T, if $(P, \texttt{col}) \in \mathcal{P}_n$ is a well-colored shortest a-b_i path, then the path $(P^*, \texttt{col}^*) = (P[a, x_n], \texttt{col}) \odot (\mu_{c_P}[x_n, w_P], \texttt{monochr}_c) \odot (P[w_P :], \texttt{col}_i)$ constructed in item 3 of the proof above is a well-colored shortest a-b_i path. It is clear from the construction that P^* is an a-b_i path; moreover, it is also a shortest path since $\mu_{c_P}[x_n, w_P]$ is a shortest x_n-w_P path. Since $(P, \texttt{col})$ is well-colored, so are $(P[a, x_n], \texttt{col})$ and $(P[w_P :], \texttt{col})$. Since $\{z_P, w_P\}$ is the last edge of $(P, \texttt{col})$ colored c_P, this color does not appear in $(P[w_P :], \texttt{col})$. Additionally, if $c_P \in \texttt{Colors}(P[a, x_n])$, it has to be the last color to appear, otherwise $(P, \texttt{col})$ would not be well-colored. It follows that the path $(P^*, \texttt{col}^*)$ is well-colored. Thus, for each $(c,s) \in \mathcal{CS}_n$, the set $\mathcal{P}_{(c,s)}$ associated to a child n' of n is a set of well-colored shortest paths. Hence, by induction, for any node n of T, $\mathcal{P}_n$ is a set of well-colored shortest paths. This concludes the proof of Claim 3.7. ∎

We now show that the construction terminates. In particular, we show that for any nonleaf node n and any of its nonleaf children n_i, we have $\operatorname{dist}(a, x_n) < \operatorname{dist}(a, x_{n_i})$. Since for any node n of T, the paths of sets $\mathcal{P}_n$ are shortest paths of length D, this implies that the construction terminates. By construction (see item 3 of the proof above), for $(P', \texttt{col}') \in \mathcal{P}_{n_i}$, there is a path $(P, \texttt{col}) \in \mathcal{P}_n$ such that $(P', \texttt{col}') = (P^*, \texttt{col}^*)$, implying that $(P[a, x_n], \texttt{col}) = (P'[a, x_n], \texttt{col}')$. Moreover, let $\{x_n, y\}$ be the first edge of $(P'[x_n :], \texttt{col}')$. By definition of $\mathcal{P}_{n_i} = \mathcal{P}_{(c_P, s_P)}$ in item 4 of the proof above, all paths in $\mathcal{P}_{n_i}$ share this edge and hence share the subpath $(P'[a, y], \texttt{col}')$. Hence, $\operatorname{dist}(a, x_n) < \operatorname{dist}(a, y) \le \operatorname{dist}(a, x_{n_i})$ since the paths of $\mathcal{P}_n$ and $\mathcal{P}_{n_i}$ are shortest paths by Claim 3.7. It is possible to prove that the height of T is at most k, but it is not necessary.

The following observation is verified if n' is a child of n by the previous arguments, hence by induction is also verified for any node of T_n.

Observation 1. Let n be a node of T, and let $n' \neq n$ be a node of T_n. Then, we have $\operatorname{dist}(a, x_n) < \operatorname{dist}(a, x_{n'})$, and for $(P, \texttt{col}) \in \mathcal{P}_n$ and $(P', \texttt{col}') \in \mathcal{P}_{n'}$:

$$(P[a, x_n], \texttt{col}) = (P'[a, x_n], \texttt{col}').$$

Observation 2. Let $\{n, n'\}$ be an edge of T, and let $(P, col) \in \mathcal{P}_{n'}$. Then, the first letter of $\texttt{ColorsSignsWord}(P[x_n, :], \texttt{col})$ is $\texttt{lab}(\{n, n'\})$. If n' is a nonleaf node, there might be more than one letter in $\texttt{ColorsSignsWord}(P[x_n, x_{n'}], \texttt{col})$.

CLAIM 3.8. *$(T, \texttt{lab})$ is a k-labelled branching tree.*

Proof. Let n be a node of T, and let $n_1, \ldots, n_t$ be its children. Item (ii) of Definition 3.5 is verified, since for each $(c,s) \in \mathcal{CS}_n$, at most one edge $\{n, n_i\}$ is labelled (c,s). We now prove item (iii), i.e., T_{n_i} does not contain any edge with labels in $\mathcal{LS}(n) \setminus \{\texttt{lab}(\{n, n_i\})\}$. Recall that $\mathcal{LS}(n) = \{(c,s), (c, \overline{s}) \mid (c,s) \in \mathcal{L}(n)\}$, where $\mathcal{L}(n) = \{\texttt{lab}(\{n, n_1\}), \ldots, \texttt{lab}(\{n, n_t\})\}$ and $\overline{s}$ is the opposite of s. Observe that $\mathcal{LS}(n) \subseteq \mathcal{CS}_n$ (there might be some empty set $\mathcal{P}_{(c,s)}$), and that $\mathcal{CS}_n$ is defined from the set of colors $\mathcal{C}_n$ that appear on the first edge after x_n on paths of $\mathcal{P}_n$.

Fix n_i a child of n, and let $(c_i, s_i) = \texttt{lab}(\{n, n_i\})$. The case when n_i is a leaf is trivial, since T_{n_i} has no edges. Assume now that n_i is not a leaf. Observe that for any $(P, \texttt{col}) \in \mathcal{P}_n$ and by choice of w_P, the only color from $\mathcal{C}_n$ used in $(P^*[x_n :], \texttt{col}^*)$

is c_P. Formally, we have $\texttt{Colors}(P^*[x_n :], \texttt{col}^*) \subseteq \texttt{Colors}(P[x_n :], \texttt{col}) \setminus (\mathcal{C}_n \setminus \{c_P\})$, thus $\texttt{Colors}(\mathcal{P}_{n_i}[x_n :]) \subseteq \texttt{Colors}(\mathcal{P}_n[x_n :]) \setminus (\mathcal{C}_n \setminus \{c_i\})$ for $1 \leq i \leq t$. By induction, this implies that for any node n' of T_{n_i}, $\texttt{Colors}(\mathcal{P}_{n'}[x_{n'} :]) \subseteq (\texttt{Colors}(\mathcal{P}_n[x_n :]) \setminus (\mathcal{C}_n \setminus \{c_i\})$. By Observation 2, the label of an edge $\{n', n''\}$ of T_{n_i} depends on the colors that appear after $x_{n'}$ in paths of $\mathcal{P}_{n''}$, hence no edges of T_{n_i} has a label in $(\mathcal{C}_n \setminus \{c_i\}) \times \{+, -\}$.

We now prove that $(c_i, \overline{s_i})$ is not used as label for edges of T_{n_i}. Let $\{n', n''\}$ be an edge of T_{n_i} and assume by contradiction that $\texttt{lab}(\{n', n''\}) = (c_i, \overline{s_i})$. Moreover, let $(P, \texttt{col}) \in \mathcal{P}_{n_i}$ and $(P', \texttt{col}') \in \mathcal{P}_{n''}$. By Observation 1, $(P[x_n, x_{n_i}], \texttt{col}) = (P'[x_n, x_{n_i}], \texttt{col}')$.

By Observation 2, the first letter of $\texttt{ColorsSignsWord}(P'[x_n :], \texttt{col}')$ is (c_i, s_i). Similarly, the first letter of $\texttt{ColorsSignsWord}(P'[x_{n'} :], \texttt{col}')$ is $(c_i, \overline{s_i})$. This yields a contradiction since the path $(P', \texttt{col}')$ is well-colored (Claim 3.7). Hence, there are no edges of T_{n_i} with labels in $\mathcal{CS}_n \setminus \{(c_i, s_i)\}$, proving item (iii).

It remains to prove item (i), i.e., in every path from the root of T to one of its leaves, the edges with the same label form a connected subpath. Let $r = p_1, p_2, \ldots, p_t = n$ be the path from the root r of T to a nonleaf node n. If there exists a child n' of n such that $\texttt{lab}(\{n, n'\}) = \texttt{lab}(\{p_{i-1}, p_i\}) = (c, s), i \in \{2, \ldots, t\}$, then we have to show that for any $i < j \leq t$, $\texttt{lab}(\{p_{j-1}, p_j\}) = (c, s)$. Suppose that there are such n' and i. For $(P, \texttt{col}) \in \mathcal{P}_{p_i}$ and $(P', \texttt{col}') \in \mathcal{P}_{n'}$, by Observation 1 we have $(P[x_{p_{i-1}}, x_{p_i}], \texttt{col}) = (P'[x_{p_{i-1}}, x_{p_i}], \texttt{col}')$. By Observation 2, the first edge of these subpaths is colored c, and so is the first edge of $(P'[x_n :], \texttt{col}')$. Since $(P', \texttt{col}')$ is well-colored, the whole subpath $(P'[x_{p_{i-1}}, x_n], \texttt{col}')$ is entirely colored with c. For any $i < j \leq t$, $(P''[x_{p_{j-1}}, x_{p_j}], \texttt{col}'') = (P'[x_{p_{j-1}}, x_{p_j}], \texttt{col}')$, with $(P'', \texttt{col}'') \in \mathcal{P}_{p_j}$, and this implies that the first color of $(P''[x_{p_{j-1}}, x_{p_j}], \texttt{col}'')$ is c. Thus $\texttt{lab}(\{p_{j-1}, p_j\}) = (c, s)$ and hence Item (i) is verified. It follows that $(T, \texttt{col})$ is a k-labelled branching tree, which concludes the proof of Claim 3.8. ☐

For a nonleaf node n of T, we can observe that for any a-b_i path in $\mathcal{P}_n$, $b_i \in \{b_1 \ldots, b_q\}$, there is a child n' such that $\mathcal{P}_{n'}$ contains an a-b_i path. This implies that for any b_i there is a leaf l of T such that $\mathcal{P}_l$ contains a well-colored shortest a-b_i path. Since T is a k-labelled branching tree, by Lemma 3.9, it has at most $c(k)$ leaves and for each leaf l of T, $|\mathcal{P}_l| = 1$ (otherwise, our construction would have created a child for node l). It follows that there are at most $c(k)$ vertices at distance D from a. ☐

We now prove the claimed combinatorial bound on the number of leaves contained in a k-labelled branching tree.

LEMMA 3.9. *A k-labelled branching tree $(T, \texttt{lab})$ contains $O(3^k)$ leaves.*

Proof. Let $(T, \texttt{lab})$ be a k-labelled branching tree, with the labelling function $\texttt{lab} : E(T) \to \mathcal{CS}$. Let $\mathcal{B}(n)$ be the set of labels that have not been forbidden in T_n by conditions (i) or (iii) of Definition 3.5 applied to some ancestor of n in the tree T. In particular, if n is the root, then $\mathcal{B}(n) = \mathcal{CS}$, and at any node n, $\mathcal{B}(n)$ is a superset of all labels of $E(T_n)$. If n has exactly one child n', by definition $\mathcal{B}(n') \subseteq \mathcal{B}(n)$ so $|\mathcal{B}(n')| \leq |\mathcal{B}(n)|$.

CLAIM 3.10. *Let n be a node of T with $t \geq 2$ children $n_1, \ldots, n_t$. If n is the root, then $|\mathcal{B}(n_i)| \leq |\mathcal{B}(n)| - (t-1)$, else $|\mathcal{B}(n_i)| \leq |\mathcal{B}(n)| - \max\{2, t-1\}$.*

Proof. Since T is a k-labelled branching tree, by condition (ii) the labels of the edges $\{n, n_j\}$, $1 \leq j \leq t$ are pairwise distinct. In particular, the $t-1$ distinct labels of edges $\{n, n_j\}$, with $j \neq i$, appear in $\mathcal{B}(n)$ but not in $\mathcal{B}(n_i)$, by condition (iii). Therefore, $|\mathcal{B}(n_i)| \leq |\mathcal{B}(n)| - (t-1)$.

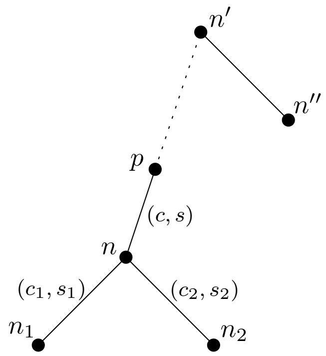

FIG. 3. *Illustration of the configuration one obtains when n is not the root, has at least two children n_1 and n_2 and parent p. Node n' is an ancestor of n while n'' is not.*

It remains to be proven that if n is not the root and has exactly two children n_1, n_2, then $|\mathcal{B}(n_i)| \le |\mathcal{B}(n)| - 2$ for $i \in \{1,2\}$. Let p be the parent node of n in T, we denote $(c,s) = \texttt{lab}(\{p,n\})$ and $(c_i,s_i) = \texttt{lab}(\{n,n_i\})$ for $i \in \{1,2\}$ (see Figure 3). We claim that if $c_i \neq c$, then both (c_i,s_i) and $(c_i,\overline{s_i})$ are in $\mathcal{B}(n)$. First, observe that in this case labels (c_i,s_i) and $(c_i,\overline{s_i})$ do not appear on the path from n to the root of T. Indeed, condition (i) implies that label (c_i,s_i) does not appear on the path from n to the root of T. By condition (iii), nor does label $(c_i,\overline{s_i})$: this would forbid label (c_i,s_i) in the tree below, in particular in T_n. Moreover, the color c_i does not appear on edges incident to the path from n to the root of T. Indeed, by condition (iii), for any edge $\{n',n''\}$ of T such that n' is an ancestor of n and n'' is not an ancestor of n (see Figure 3), c_i cannot appear in the label $\texttt{lab}(\{n',n''\})$ since otherwise (c_i,s_i) and $(c_i,\overline{s_i})$ would be forbidden in the subtree of $T_{n'}$ containing n (which does not contain n''). In particular, (c_i,s_i) could not be used as label of $\{n,n_i\}$. Altogether, both (c_i,s_i) and $(c_i,\overline{s_i})$ are in $\mathcal{B}(n)$ since they could not be forbidden by conditions (i) or (iii).

Assume now w.l.o.g. that $i=1$. If $c_1 = c$, then we claim that $c_2 \neq c$. Indeed, by condition (i), if $c_1 = c_2 = c$, then signs s_1 and s_2 are opposite, so one is the opposite $\overline{s}$ of s, contradicting the fact that label $(c,\overline{s})$ is in $\mathcal{LS}(p)$ and has been forbidden in T_n by condition (iii). Therefore, (c_2,s_2) and $(c_2,\overline{s_2})$ are in $\mathcal{B}(n)$ and by condition (iii) are forbidden in T_{n_1} and thus $|\mathcal{B}(n_1)| \le |\mathcal{B}(n)| - 2$.

If $c_1 \neq c$, then (c,s) is not in $\mathcal{B}(n_1)$ by condition (i) but is in $\mathcal{B}(n)$ and $(c_1,\overline{s_1})$ is not in $\mathcal{B}(n_1)$ by condition (iii) but is in $\mathcal{B}(n)$, hence $|\mathcal{B}(n_1)| \le |\mathcal{B}(n)| - 2$. This concludes the proof of Claim 3.10. ∎

We now prove that T has $O(3^k)$ leaves. For a node n of the tree, we see $|\mathcal{B}(n)|$ as the *budget* of the tree T_n and we estimate the number of leaves w.r.t. this budget. We are reminded that the budget of the root is $2k$. Let $F(b)$ be the maximum number of leaves in any k-labelled branching tree with budget b at its root. Observe that if $b=1$, then $F(b)=1$ and for $b \le b', F(b) \le F(b')$. Let n be a node of T, then $F(|\mathcal{B}(n)|)$ is the maximal number of leaves of T_n. Observe that if n has exactly one child, i.e., it does not branch, then $F(|\mathcal{B}(n)|) = F(|\mathcal{B}(n_1)|)$ since $\mathcal{B}(n_1) \subseteq \mathcal{B}(n)$. Suppose that n has $t \ge 2$ children $n_1, \dots n_t$. If n is the root, then by Claim 3.10:

$$F(|\mathcal{B}(n)|) \le F(|\mathcal{B}(n_1)|) + \cdots + F(|\mathcal{B}(n_t)|) \le t \cdot F(|\mathcal{B}(n)| - (t-1)),$$

else if n is not the root,

$$F(|\mathcal{B}(n)|) \le t \cdot F(|\mathcal{B}(n)| - \max\{2, t-1\}).$$

At the root r of the tree, since $\mathcal{B}(r) = \mathcal{CS}$, we have $|\mathcal{B}(r)| = 2k$. We need to prove that $F(2k) = O(3^k)$. Let us recall some standard techniques for the analysis on the growth of functions described by linear equations (or linear inequalities). The reader can refer to the book of Fomin and Kratsch [16, Chapter 2] for detailed explanations. Consider a function C on positive integers. Assume there are positive integers $\beta_1, \beta_2, \ldots, \beta_q$ such that

$$C(b) \le C(b - \beta_1) + C(b - \beta_2) + \cdots + C(b - \beta_q)$$

for all values b (or at least for all b larger than a threshold b_0). Then $C(b) = O(\alpha^b)$, where α is the (unique) positive real root of equation

$$x^b - x^{b-\beta_1} - x^{b-\beta_2} - \cdots - x^{b-\beta_q} = 0.$$

In [16], $(\beta_1, \beta_2, \ldots, \beta_q)$ is called a branching vector, and α is the branching factor of this vector. Now in our case, for function F, we know that for each value b there exists some $t \ge 3$ such that $F(b) \le t \cdot F(b - (t-1))$, and also (except at the root, if the root has at most two children), $F(b) \le 2 \cdot F(b-2)$ or $F(b) \le F(b-1)$. Each of the inequalities leads to a different branching factor: $\alpha_t = \sqrt[t-1]{t}$ for the first equation, and $\alpha_2 = \sqrt{2}$, $\alpha_1 = 1$ for the latter. Also according to [16], when a function F satisfies one among a family of linear inequalities, the growth of F is $F(b) = O(\alpha^b)$, where α is an upper bound on all branching factors. In our case, observe that the maximum value of $\alpha_1 = 1, \alpha_2 = \sqrt{2}$, and $\alpha_t = \sqrt[t-1]{t}, t \ge 3$ is attained for $t = 3$ with $\alpha_3 = \sqrt{3}$. Therefore, $F(b) = O(3^{b/2})$. Observe that this also holds if the root has one or two children. Indeed, for each child of the root, the number of leaves in the corresponding subtree is $O(3^{b/2})$, and since there are at most two such children, the total number of leaves is $O(3^{b/2})$.

Since at the root $b = 2k$, we conclude that our k-labelled branching tree has $O(3^k)$ leaves. ∎

Combining Lemmas 3.6 and 3.9 we deduce that there are $O(3^k)$ vertices at distance D of any vertex a of graph G. In order to complete the proof of Theorem 1.1 and show that $\mathrm{pw}(G) = O(3^k)$, we hence simply need to apply Lemma 2.1 with $K = O(3^k)$.

4. Vertex-covering with k shortest paths. In this section, $G = (V, E)$ denotes a graph whose *vertices* can be covered by k shortest paths $\mu_1, \ldots, \mu_k$. As before we endow each base path μ_c with a direction, but now colors are assigned to vertices. We can easily adapt the notions of good colorings of the previous section to these vertex-colorings. Again, for any pair of vertices a and b, there is a well-colored shortest path joining them (Lemma 4.1), which defines a colors-signs word. But we shall see that now (unlike in the simpler case of edge-coverings), we may have two distinct vertices b and c at the same distance D from a, and well-colored shortest a-b and a-c paths with the same colors-signs word. More effort will be needed to recover a (slightly larger) upper bound on the number of vertices at distance D from a (Lemma 4.2).

Good colorings. For each vertex v of G, let $\texttt{Colors}(v)$ denote the set of indices (colors) $c \in \{1 \ldots, k\}$ such that v is a vertex of μ_c. Let P be an a-b path of G, from vertex a to vertex b. A *coloring* of P is a function $\texttt{col} : V(P) \to \{1, \ldots, k\}$ assigning to each vertex v of P one of its colors $\texttt{col}(v) \in \texttt{Colors}(v)$. A colored path is a pair $(P, \texttt{col})$. The coloring $\texttt{col}$ of P is said to be *good* if, for any color c, the subgraph induced by the set of *vertices* using this color c forms a connected subpath $P[x, y]$ of

P (which implies that $P[x,y] = \mu_c[x,y]$). A colored path $(P, \mathtt{col})$ where $\mathtt{col}$ is a good coloring is called *well-colored.*

Operators $\oplus$ and $\odot$ naturally extend to (vertex) colored paths, with the precaution that $(\nu, \mathtt{col}) \odot (\eta, \mathtt{col}')$ is defined only when their common vertex x, the last of ν and first of η, satisfies $\mathtt{col}(x) = \mathtt{col}'(x)$. Given a colored path $(P, \mathtt{col})$, we again denote by $(P[x,y], \mathtt{col})$ its restriction to a subpath $P[x,y]$ of P. For each color $1 \leq c \leq k$. Now let $(P, \mathtt{monochr}_c)$ denote the monochromatic coloring of $V(P)$ with color c. With these notations, any well-colored a-b path $(P, \mathtt{col})$ with colors $(c_1, \dots, c_l)$ is of the form

$$(\mu_{c_1}[a_1, b_1], \mathtt{monochr}_{c_1}) \oplus (\mu_{c_2}[a_2, b_2], \mathtt{monochr}_{c_2}) \oplus \dots \oplus (\mu_{c_l}[a_l, b_l], \mathtt{monochr}_{c_l})$$

for some vertices $a = a_1, b_1, a_2, b_2, \dots, a_l, b_l = b$. Like in the previous section, we have the following lemma.

LEMMA 4.1. *For any pair of vertices a and b of G, there exists a well-colored shortest a-b path.*

Proof. Among all shortest a-b paths, choose one that admits a coloring with a minimum number of monochromatic subpaths. Let $(P, \mathtt{col})$ be such a colored path. Assume for a contradiction that the coloring $\mathtt{col}$ is not good. Then there exist three vertices x, y, and z, appearing in this order in P such that $\mathtt{col}(x) = \mathtt{col}(z) \neq \mathtt{col}(y)$. Therefore, x and z are on the same base path μ_c. Let P' be the path obtained from P by replacing $P[x,z]$ by $\mu_c[x,z]$. Notice that P' is no longer than P, since $\mu_c[x,z]$ is a shortest x-z path of graph G. Moreover, in P' we can color all vertices of $P'[x,z]$ with color c, and keep all other colors unchanged. Hence P' has strictly fewer monochromatic subpaths than P—a contradiction. ∎

Colors-signs word. Let $(P, \mathtt{col})$ be a well-colored a-b path; we recall that we see it as being directed from a to b. As in section 3, each monochromatic subpath P' of P, say of color c, induces a sign ($+$ or $-$) depending on its direction w.r.t. μ_c if P' has at least two vertices. If P' has a unique vertex, we assign to it sign $+$. Therefore, we can again define the *colors-signs word* $\mathtt{ColorsSignsWord}(P, \mathtt{col}) = ((c_1, s_1), (c_2, s_2), \dots, (c_l, s_l))$ on the alphabet $\{1, \dots, k\} \times \{+, -\}$, corresponding to the colors and signs of the monochromatic subpaths of P according to the ordering in which these subpaths appear from a to b.

In the case of edge-covering, we had the elegant statement of Lemma 3.3, by which, given a vertex a, a colors-signs word ω, and a distance D, there is a unique vertex b (if any exists) at distance D such that the well-colored shortest a-b path corresponds to this word.

Unfortunately, this does not extend to vertex-covering: Figure 4 presents two distinct vertices b and c located at the same distance D from vertex a, together with a well-colored shortest a-b path $(P, \mathtt{col})$ and a well-colored shortest a-c path

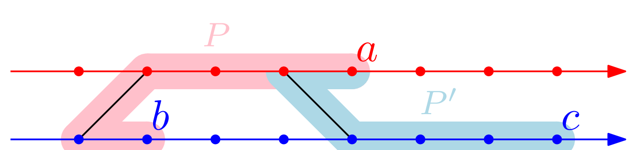

FIG. 4. *Two well-colored paths $(P, \mathtt{col})$ and $(P', \mathtt{col}')$ with same colors-signs word $\omega = ((red, -), (blue, +))$, same length 5 and same start vertex a but different end-vertices (b and c). (Color available online.)*

$(P', \texttt{col}')$. These colored paths starting from a have the same colors-signs word and the same length, but this does not imply that their endpoints are equal. However, we can prove a single-exponential bound for this case as well with similar (but more involved) arguments as the ones in subsection 3.2.

Again, adapting notations to vertex-colored paths, for a colored path $(P, \texttt{col})$ we define $\texttt{Colors}(P, \texttt{col}) = \{\texttt{col}(x) \mid x \in V(P)\}$, the set of colors that appear on $(P, \texttt{col})$. For a set of colored paths $\mathcal{P}$, $\texttt{Colors}(\mathcal{P}) = \bigcup_{(P,\texttt{col})\in\mathcal{P}} \texttt{Colors}(P, \texttt{col})$. For a colored a-b path $(P, \texttt{col})$ and a vertex x of P, we let $P[x:] = P[x, b]$ and for a set of colored paths $\mathcal{P}$ that shares the vertex x we let $\mathcal{P}[x:] = \bigcup_{(P,\texttt{col})\in\mathcal{P}} (P[x:], \texttt{col})$.

LEMMA 4.2. *For any vertex a of G and any integer D, there are $O(k \cdot 3^k)$ vertices at distance exactly D from vertex a.*

Proof. Let $\{b_1 \ldots, b_q\}$ be the set of vertices at distance D from a in G, and let $\mathcal{P}$ denote a set $\{(P_1, \texttt{col}_1), \ldots, (P_q, \texttt{col}_q)\}$ of well-colored shortest a-b_i paths. We aim to construct a k-labelled branching tree $(T, \texttt{lab})$ such that each node n of T is associated with a set of paths $\mathcal{P}_n$ and for each vertex b_i there is a leaf l of T such that $\mathcal{P}_l$ contains a well-colored a-b_i path. However, unlike in the proof of Lemma 3.6 where any leaf node was associated to exactly one path, the set $\mathcal{P}_l$ corresponding to any leaf node l can contain up to $2k + 1$ paths.

We will construct recursively $(T, \texttt{lab})$ starting from the root r of T. Initially, $\mathcal{P}_r = \mathcal{P}$. For every node n of T, $\mathcal{P}_n$ will be a set of well-colored paths from a to a subset of vertices of $\{b_1 \ldots, b_q\}$. These paths may not be shortest paths, but will be of length at most $D + 2d$ where d is the depth of the node n in T (the depth of the root being 0). Let n be an already computed node of T. If there is a path in $\mathcal{P}_n$ such that all the others paths of $\mathcal{P}_n$ are colored subpaths of it, do nothing: n is a leaf of our tree and no recursive step is performed at n. If there is no such path, we will grow the tree from n. We now construct the sets of paths that will be associated with the children of n.

We first consider the case where n is the root r of T. Note that there might exist $(P, \texttt{col})$ and $(P', \texttt{col}')$ in $\mathcal{P}_n$ such that $\texttt{col}(a) \neq \texttt{col}'(a)$. In any case, we have the following:

1.1. Let $\mathcal{C}_n = \{\texttt{col}(a) \mid (P, \texttt{col}) \in \mathcal{P}_n\}$.

1.2. For each $(P, \texttt{col}) \in \mathcal{P}_n$, let w_P be the last vertex of P such that $c_P = \texttt{col}(w_P) \in \mathcal{C}_n$. Construct the path $(P^*, \texttt{col}^*) = (\mu_{c_P}[a, w_P], \texttt{monochr}_{c_P}) \odot (P[w_P:], \texttt{col})$.

1.3. For each $(c, s) \in \mathcal{CS}_n = (\mathcal{C}_n \times \{+, -\})$ define the set $\mathcal{P}_{(c,s)}$ as $\{(P^*, \texttt{col}^*) \mid (P, \texttt{col}) \in \mathcal{P}_n, (c_P, s_P) = (c, s)\}$, where s_P is the sign of $\mu_{c_P}[a, w_P]$.

If $n \neq r$, let $(P_0, \texttt{col}_0)$ be a longest path of $\mathcal{P}_n$, and let x_n be the last vertex of P_0 such that for any path $(P, \texttt{col}) \in \mathcal{P}_n$, either $(P_0[a, x_n], \texttt{col}_0) = (P[a, x_n], \texttt{col})$ or $(P, \texttt{col})$ is a colored subpath of $(P_0[a, x_n], \texttt{col}_0)$. This vertex is guaranteed to exist, and notice that x_n can be the vertex a. We proceed as follows:

2.1. Let $\mathcal{P}_n^\circ$ be the set of paths of $\mathcal{P}_n$ that are colored subpaths of $(P_0[a, x_n], \texttt{col}_0)$ and $\hat{\mathcal{P}}_n = \mathcal{P}_n \setminus \mathcal{P}_n^\circ$.

2.2. Let $\mathcal{C}_n = \{\texttt{col}(z_P) \mid (P, \texttt{col}) \in \hat{\mathcal{P}}_n\}$ where z_P denotes the vertex appearing right after x_n on path P.

2.3. For $(P, \texttt{col}) \in \hat{\mathcal{P}}_n$, let w_P be the last vertex of P such that $c_P = \texttt{col}(w_P) \in \mathcal{C}_n$. Let y_P be the second vertex of $\mu_{c_P}[x_n, w_P]$ if $x_n \in V(\mu_{c_P})$; otherwise, y_P is defined as the first vertex of μ_{c_P} adjacent to x_n. We define the following:

$$(P^*, \texttt{col}^*) = (P[a, x_n], \texttt{col}) \oplus (\mu_{c_P}[y_P, w_P], \texttt{monochr}_{c_P}) \odot (P[w_P:], \texttt{col}).$$

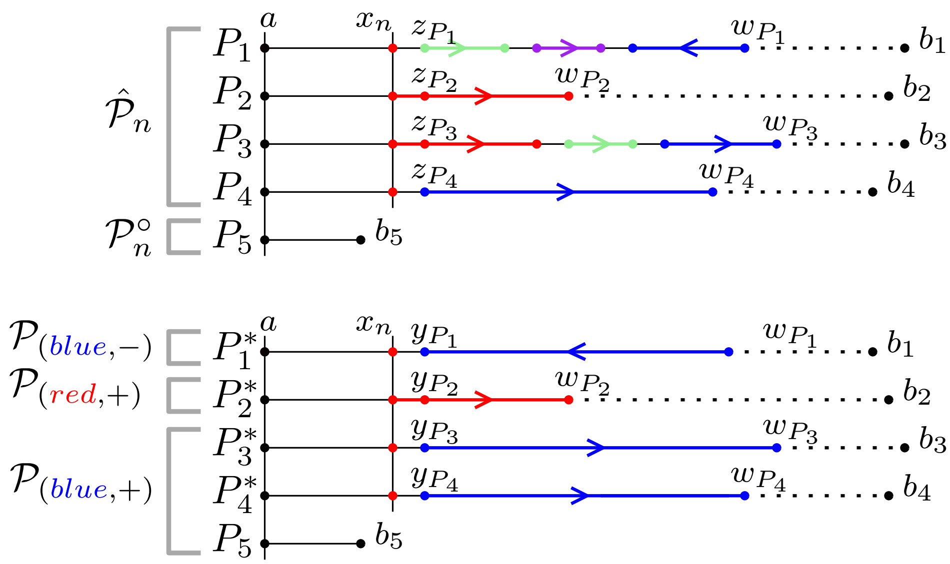

FIG. 5. *Example of the construction in proof of Lemma 4.2 for a set of paths $\mathcal{P}_n$. In this example, all paths share the same subpath $(P_1[a,x_n],\texttt{col})$ except for P_5 that is subpath of the latter. Moreover, these paths may have different length. We have $\mathcal{C}_n = \{red, blue, green\}$. In the path $(P_i[w_{P_i}:],\texttt{col}_i)$ the only vertex whose color is in $\mathcal{C}_n$ is w_{P_i}. The sets $\mathcal{P}_{(red,-)}, \mathcal{P}_{(green,+)}, \mathcal{P}_{(green,-)}$ are empty. The vertices y_{P_1} and $y_{P_3} = y_{P_4}$ are the same if $x_n \notin V(\mu_{blue})$ and distinct else. The path P_2^* has the same length as P_1, the other modified paths might not have the same length than their counterpart, even P_4^* as z_{P_4} might be different from y_{P_4}. The path P_5 was arbitrarily added to $\mathcal{P}_{(blue,+)}$. (Color available online.)*

2.4. For each $(c,s) \in \mathcal{CS}_n = (\mathcal{C}_n \times \{+,-\})$ define the set $\mathcal{P}_{(c,s)}$ as $\{(P^*,\texttt{col}^*) \mid (P,\texttt{col}) \in \hat{\mathcal{P}}_n, (c_P, s_P) = (c,s)\}$. Here s_P is the sign of $\mu_{c_P}[y_P, w_P]$, except when $y_P = w_P$ and $\texttt{col}^*(x_n) = \texttt{col}^*(y_P)$, in which case s_P is the sign of $\mu_{c_P}[x_n, y_P]$.

2.5. Add $\mathcal{P}_n^\circ$ to an arbitrary nonempty set $\mathcal{P}_{(c,s)}$.

This construction is illustrated in Figure 5. For each nonempty set $\mathcal{P}_{(c,s)}, (c,s) \in \mathcal{CS}_n$ we then create a child n' of n associated with $\mathcal{P}_{(c,s)}$ where the edge $\{n,n'\}$ labelled (c,s). Then, we recursively apply the process to each created node.

CLAIM 4.3. *For any node n of depth d of T, $\mathcal{P}_n$ is a set of well-colored paths of length at most $D+2d$.*

Proof. This trivially holds for $\mathcal{P}_r$ where r is the root of T. We show that for a node n of T, if $(P,\texttt{col}) \in \mathcal{P}_n$ is a well-colored a-b path, then the path $(P^*,\texttt{col}^*)$ constructed in either items 1.2 or 2.3 is a well-colored a-b path such that $|P^*| \le |P|+2$.

In both cases it is clear from the construction that P^* is an a-b path. In the case of item 1.2 where $n=r$, since $(P[w_P:],\texttt{col})$ is well-colored and $\texttt{col}(w_P) = c_P$ it follows that $(P^*,\texttt{col}^*)$ is well-colored. Moreover, $(\mu_{c_P}[a,w_P],\texttt{monochr}_{c_P})$ is a shortest a-w_P path, hence $|P^*| = |P|$.

In the case of item 2.3, since $(P,\texttt{col})$ is well-colored so are $(P[a,x_n],\texttt{col})$ and $(P[w_P:],\texttt{col})$. The vertex w_P is colored c_P and if $c_P \in \texttt{Colors}(P[a,x_n])$ it has to be the last color to appear; otherwise, $(P,\texttt{col})$ would not be well-colored. It follows that the path $(P^*,\texttt{col}^*)$ is well-colored.

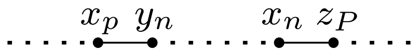

FIG. 6. *Structure of a path* $(P, \texttt{col}) \in \hat{\mathcal{P}}_n$ *with* p *the parent of* n*. Note that* $y_n = x_n$ *is possible.*

We now show that $|P^*| \leq |P| + 2$. This is verified if $x_n \in V(\mu_{c_P})$ since y_P is the second vertex of the shortest x_n-w_P path $\mu_{c_P}[x_n, w_P]$ and hence $P^* = P[a, x_n] \odot \mu_{c_P}[x_n, w_P] \odot P[w_P :]$ and $|P^*| \leq |P|$. Otherwise, in the case when $x_n \notin V(\mu_{c_P})$, y_P is the first vertex of μ_{c_P} adjacent to x_n. The path $\mu_{c_P}[y_P, w_P]$ is a shortest y_P-w_P path, hence $|\mu_{c_P}[y_P, w_P]| \leq |(y_P) \oplus P[x_n, w_P]| \leq 1 + |P[x_n, w_P]|$. This implies the following:

$$\begin{aligned} |P^*| &= |P[a, x_n] \oplus \mu_{c_P}[y_P, w_P] \odot P[w_P :]| \\ &= |P[a, x_n]| + 1 + |\mu_{c_P}[y_P, w_P]| + |P[w_P :]| \\ &\leq |P[a, x_n]| + 2 + |P[x_n, w_P]| + |P[w_P :]| = |P| + 2. \end{aligned}$$

It remains that for each $(c, s) \in \mathcal{CS}_n$, the set $\mathcal{P}_{(c,s)}$ associated to a child n' of n is a set of well-colored paths of length at most $D' + 2$, where D' is the maximum length of paths in $\mathcal{P}_n$. Note that if the paths of $\mathcal{P}_n^\circ$ were added to $\mathcal{P}_{(c,s)}$ in item 2.5, then the property still holds as these paths were not modified. Hence by induction, for any node n of depth d of T, $\mathcal{P}_n$ is a set of well-colored paths of length at most $D + 2d$. This concludes the proof of Claim 4.3. ∎

Let $n \neq r$ be a node of T. Observe that for $(c, s) \in \mathcal{CS}_n$, all paths of $\mathcal{P}_{(c,s)}$ share the same successor of x_n, the vertex y_P defined in item 2.3. Thus for every node $n \neq r$ of T with parent $p \neq r$, we can define a vertex y_n, the successor of x_p in paths of $\hat{\mathcal{P}}_n$; see Figure 6. If $n = r$ or $p = r$, we let $y_n = a$.

Not let $n \neq r$ be a node of T and n' be one of its children. For $(P', \texttt{col}') \in \hat{\mathcal{P}}_{n'}$ there is $(P, \texttt{col}) \in \hat{\mathcal{P}}_n$ such that $(P', \texttt{col}')$ corresponds to the colored path $(P^*, \texttt{col}^*)$ constructed from $(P, \texttt{col})$ at node n; see item 2.3. Therefore, $(P[a, x_n], \texttt{col}) = (P'[a, x_n], \texttt{col}')$. Moreover, the vertex $y_{n'}$ is the successor of x_n shared among all the paths in $\hat{\mathcal{P}}_{n'}$. Hence, if n' is not a leaf, the vertex $x_{n'}$ appears strictly after x_n on paths of $\hat{\mathcal{P}}_{n'}$ since $x_{n'}$ is either $y_{n'}$ or comes after it. Therefore, the following observation is verified by induction.

Observation 3. Let $n \neq r$ be a node of T, $n' \neq n$ a node of T_n, $(P, \texttt{col}) \in \hat{\mathcal{P}}_n$, and $(P', \texttt{col}') \in \hat{\mathcal{P}}_{n'}$. Then $(P[a, x_n], \texttt{col}) = (P'[a, x_n], \texttt{col}')$, and if n' is a not a leaf, x_n appears strictly before $x_{n'}$ on the path P'.

The following observation is a direct consequence of our construction.

Observation 4. Let $\{n, n'\}$ be an edge of T and $(P, col) \in \hat{\mathcal{P}}_{n'}$. Then the letter $\texttt{lab}(\{n, n'\})$ is a letter of $\texttt{ColorsSignsWord}(P[a, x_{n'}], \texttt{col})$. Moreover, the color of $\texttt{lab}(\{n, n'\})$ is the color of the first letter of $\texttt{ColorsSignsWord}(P[y_{n'} :], \texttt{col})$ (the signs can be different, see the special case for the definition of s_P in Item 2.4).

The following claim implies that the construction terminates.

CLAIM 4.4. *The height of* T *is at most* k*.*

Proof. Let $n \neq r$ be a node of T and $(P, \texttt{col}) \in \hat{\mathcal{P}}_n$. We show that $\texttt{Colors}(P^*[y_P :])$ is strictly included in $\texttt{Colors}(P[y_n :])$. From the choice of w_P in item 2.3, we have the following:

$$\texttt{Colors}(P^*[y_P :]) \subseteq \texttt{Colors}(P[y_n :]) \setminus (\mathcal{C}_n \setminus \{c_P\}).$$

If $|\mathcal{C}_n| \geq 2$, it follows that $\texttt{Colors}(P^*[y_P :]) \subsetneq \texttt{Colors}(P[y_n :])$. If $|\mathcal{C}_n| = 1$, then $c = \texttt{col}(y_n) \notin \mathcal{C}_n$. Indeed, if c was the only color in $\mathcal{C}_n$, since paths in $\hat{\mathcal{P}}_n$ are well-colored and $y_n \neq y_P$, they all share the subpath $(\mu_c[y_n, y_P], \texttt{monochr}_c)$, a contradiction with the choice of x_n. Hence $\texttt{Colors}(P^*[y_P :]) \subseteq \texttt{Colors}(P[y_n :]) \setminus \{c\}$. It follows that for any child n' of n, $\texttt{Colors}(\hat{\mathcal{P}}_{n'}[y_{n'} :]) \subsetneq \texttt{Colors}(\hat{\mathcal{P}}_n[y_n :])$. Hence for a node $n \neq r$ of T, $|\texttt{Colors}(\hat{\mathcal{P}}_n[y_n :])| \geq 1$ and for a child n' of n, $|\texttt{Colors}(\hat{\mathcal{P}}_{n'}[y_{n'} :])| < |\texttt{Colors}(\hat{\mathcal{P}}_n[y_n :])|$. It follows that the height of T is at most k since $|\texttt{Colors}(\mathcal{P}_r)| \leq k$ and that for any child n' of r $\texttt{Colors}(\mathcal{P}_{n'}) \subseteq \texttt{Colors}(\mathcal{P}_r)$. This concludes the proof of Claim 4.4. ☐

CLAIM 4.5. $(T, \texttt{lab})$ *is a k-labelled branching tree.*

Proof. Let n be a node of T and $n_1, \dots, n_t$ its children. Item (ii) of Definition 3.5 is verified since for each $(c, s) \in \mathcal{CS}_n$ at most one edge $\{n, n_i\}$ is labelled (c, s).

We now prove item (iii), i.e., T_{n_i} does not contain any edge with labels in $\mathcal{LS}(n) \setminus \{\texttt{lab}(\{n, n_i\})\}$. Recall that $\mathcal{LS}(n) = \{(c, s), (c, \overline{s}) \mid (c, s) \in \mathcal{L}(n)\}$, where $\mathcal{L}(n) = \{\texttt{lab}(\{n, n_1\}), \dots, \texttt{lab}(\{n, n_t\})\}$ and $\overline{s}$ is the opposite of s. Observe that $\mathcal{LS}(n) \subseteq \mathcal{CS}_n$ (there might be some empty set $\mathcal{P}_{(c,s)}$), and that $\mathcal{CS}_n$ is defined by the set of colors $\mathcal{C}_n$ that appear on the first vertices after x_n on paths of $\hat{\mathcal{P}}_n$.

Fix n_i a child of n and let $(c_i, s_i) = \texttt{lab}(\{n, n_i\})$. The case when n_i is a leaf is trivial, since T_{n_i} has no edges, so let us assume that n_i is not a leaf. For any $(P, \texttt{col}) \in \hat{\mathcal{P}}_n$ by choice of w_P, $\texttt{Colors}(P^*[y_P :], \texttt{col}^*) \subseteq (\texttt{Colors}(P[y_n :], \texttt{col}) \setminus (\mathcal{C}_n \setminus \{c_P\})$ (replace y_P by a if n is the root). Thus $\texttt{Colors}(\hat{\mathcal{P}}_{n_i}[y_{n_i} :]) \subseteq (\texttt{Colors}(\hat{\mathcal{P}}_n[y_n :]) \setminus (\mathcal{C}_n \setminus \{c_i\}))$. By induction, this implies that for any node n' of T_{n_i}, $\texttt{Colors}(\hat{\mathcal{P}}_{n'}[y_{n'} :]) \subseteq (\texttt{Colors}(\hat{\mathcal{P}}_n[y_n :]) \setminus (\mathcal{C}_n \setminus \{c_i\}))$. By Observation 4, the label of an edge $\{n', n''\}$ of T_{n_i} contains the color of $y_{n''}$ in paths of $\hat{\mathcal{P}}_{n''}$, hence no edges of T_{n_i} have a label in $(\mathcal{C}_n \setminus \{c_i\}) \times \{+, -\}$. We now prove that $(c_i, \overline{s_i})$ is not used as label for edges of T_{n_i}. Let $\{n', n''\}$ be an edge of T_{n_i} and assume that $\texttt{lab}(\{n', n''\}) = (c_i, \overline{s_i})$. Let $(P, \texttt{col}) \in \hat{\mathcal{P}}_{n_i}$ and $(P', \texttt{col}') \in \hat{\mathcal{P}}_{n''}$. By Observation 3, $(P[a, x_{n_i}], \texttt{col}) = (P'[a, x_{n_i}], \texttt{col}')$, hence Observation 4 implies that $\texttt{lab}(\{n, n_i\}) = (c_i, s_i)$ is a letter of $\texttt{ColorsSignsWord}(P'[a, x_{n_i}], \texttt{col}')$. Similarly, we have that $(c_i, \overline{s_i})$ is a letter of $\texttt{ColorsSignsWord}(P', \texttt{col}')$. This is a contradiction since the path $(P', \texttt{col}')$ is well-colored (Claim 4.3). Hence there are no edges of T_{n_i} with labels in $\mathcal{CS}_n \setminus \{(c_i, s_i)\}$, proving item (iii).

It remains to prove item (i), i.e., in every path from the root of T to one of its leaves, the edges with the same label form a connected subpath. Let $r = p_1, p_2, \dots, p_t = n$ be the path from the root r of T to a nonleaf node n. If there exists a child n' of n such that $\texttt{lab}(\{n, n'\}) = \texttt{lab}(\{p_{i-1}, p_i\}) = (c, s), i \in \{2, \dots, t\}$, then we have to show that for any $i < j \leq t$, $\texttt{lab}(\{p_{j-1}, p_j\}) = (c, s)$. Suppose that there exist such n' and i. For $(P, \texttt{col}) \in \hat{\mathcal{P}}_{p_i}$ and $(P', \texttt{col}') \in \hat{\mathcal{P}}_{n'}$, by Observation 3 $(P[a, x_{p_i}], \texttt{col}) = (P'[a, x_{p_i}], \texttt{col}')$. By Observation 4, the vertex y_{p_i} (that appears before x_{p_i}) is colored c in both of these subpaths, and so is $y_{n'}$ in $(P', \texttt{col}')$. Since $(P', \texttt{col}')$ is well-colored, the subpath $(P'[y_{p_i}, y_{n'}], \texttt{col}')$ is entirely colored with c. For any $i < j \leq t$, $(P''[a, x_{p_j}], \texttt{col}'') = (P'[a, x_{p_j}], \texttt{col}')$, with $(P'', \texttt{col}'') \in \hat{\mathcal{P}}_{p_j}$. This implies that the first color of $(P''[y_{p_j} :], \texttt{col}'')$ is c. Since there is no edge of T_{n_i} labelled $(c, \overline{s})$, it follows that $\texttt{lab}(\{p_{j-1}, p_j\}) = (c, s)$ and hence item (i) is verified. It remains that $(T, \texttt{col})$ is a k-labelled branching tree, which concludes the proof of Claim 3.8. This concludes the proof of Claim 4.5. ☐

For a nonleaf node n of T, we can observe that for any a-b_i path in $\mathcal{P}_n$, $b_i \in \{b_1 \dots, b_q\}$, there is a child n' such that $\mathcal{P}_{n'}$ contains a a-b_i path (notice that it may be added via the set $\mathcal{P}^\circ_n$). This implies that for any b_i there is a leaf l of T such that $\mathcal{P}_l$ contains a well-colored a-b_i path. For a leaf l of T, by construction, $\mathcal{P}_l$ contains a path $(P, \texttt{col})$ including any other path of $\mathcal{P}_l$. Moreover, the length of P is at most $D+2k$ by Claims 4.3 and 4.4. This implies that $|\mathcal{P}_l| \leq 2k+1$, because $\mathcal{P}_l$ only contains subpaths of P, of length between D and $D+2k$. Since T is a k-labelled branching tree, by Lemma 3.9, it has $O(3^k)$ leaves, hence there are $O(k \cdot 3^k)$ vertices at distance D from a. ∎

In order to complete the proof of Theorem 1.2 and show that $\text{pw}(G) = O(k \cdot 3^k)$, we simply apply Lemma 2.1 with $K = O(k \cdot 3^k)$.

5. Algorithmic consequences.

Problem Strong Geodetic Set with Terminals is known to be NP-complete by [9]. By a simple reduction, so is Isometric Path Cover with Terminals.

Proposition 5.1. Isometric Path Cover with Terminals *is NP-complete.*

Proof. We provide a straightforward reduction from Strong Geodetic Set with Terminals. Let $(G = (V, E), k)$ be an instance of Strong Geodetic Set with Terminals with terminals $v_1, \ldots, v_k$. We build an instance of Isometric Path Cover with Terminals by considering all $\binom{k}{2}$ possible pairs of terminals i.e., $(G = (V, E), \bigcup_{1 \leq i < j \leq k} \{(v_i, v_j)\})$. By definition, (G, k) is a Yes-instance of Strong Geodetic Set with Terminals, i.e., there exists a set of $\binom{k}{2}$ shortest paths covering $V(G)$ if and only if (G, k') is a Yes-instance of Isometric Path Cover with Terminals, i.e., a set of k' shortest v_i-v_j paths, $1 \leq i < j \leq k$ covering $V(G)$. ∎

We now prove the algorithmic consequences stated Theorem 1.3 and Corollary 1.4. We recall these theorems for the sake of readability.

Theorem 1.3. Isometric Path Cover with Terminals *and* Strong Geodetic Set with Terminals *are FPT when parameterized by the number of terminals.*

Corollary 1.4. Isometric Path Cover *and* Strong Geodetic Set *are in XP when parameterized by the number of paths, respectively terminals.*

Proof of Theorem 1.3 and Corollary 1.4. Recall that, for simplicity, we assume that our input graph is connected, but all results easily extend to disconnected graphs. We first show that problem Isometric Path Cover with Terminals is FPT when parameterized by k, the number of pairs of terminals. As a first consequence, so is problem Strong Geodetic Set with Terminals, a special case of Isometric Path Cover with Terminals with $\binom{k}{2}$ pairs of terminals. Corollary 1.4 follows immediately, since for both Isometric Path Cover and Strong Geodetic Set it suffices to try all possible sets of terminals and use the FPT algorithms for the versions with terminals.

Let us focus on Isometric Path Cover with Terminals, with parameter k, input G, and the k pairs of terminals $(s_1, t_1), \ldots, (s_k, t_k)$. We can assume that the pathwidth (and hence treewidth) of the input graph is upper bounded by a function of k, as stated in Theorem 1.2, and that we have in the input a tree decomposition of such width. Indeed, recall that Theorem 1.2 does not only provide a combinatorial bound on the pathwidth of YES-instances, but also a simple, BFS-algorithm for computing the suitable path decompositions (which is also, as stated in section 2, a tree decomposition of the same width). If the algorithm fails to find a path decomposition of small width, we can directly conclude that our input graph is a NO-instance.

Therefore, we can use the classical Monadic Second-Order Logic of graphs (henceforth called MSOL_2) tools on bounded treewidth graphs. MSOL_2 includes the logical connectives $\vee$, $\wedge$, $\neg$, $\Leftrightarrow$, $\Rightarrow$, variables for vertices, edges, sets of vertices, and sets of edges, the quantifiers $\forall$ and $\exists$ that can be applied to these variables, and five binary relations: $\mathrm{adj}(u,v)$, where u and v are vertex variables and the interpretation is that u and v are adjacent; $\mathrm{inc}(v,e)$, where v is a vertex variable and e is an edge variable and the interpretation is that v is incident to e; $v \in V'$, where v is a vertex variable and V' is a vertex set variable; the similar $e \in E'$ on edge variable e and edge set variable E', and eventually equality of two variables of the same nature.

By a celebrated theorem of Courcelle [7], any problem expressible in MSOL_2 can be solved in time $f(\mathrm{tw}) \cdot n$ time on bounded treewidth graphs if a tree decomposition of the input graph is also given. Function f depends on the formula (hence, on the problem). Courcelle's theorem extends in several ways to optimization problems, and slightly larger classes of formulae, e.g., allowing one to identify a fixed number of terminal vertices, as we shall detail later. Here we will refer to [3], one of the (alternative) proofs of Courcelle's theorem, with some extensions. As noted in [3], MSOL_2 allows one to express properties as $\mathrm{Connected}(V', E')$ where V' is a vertex set variable and E' is an edge set variable and the property is true if and only if (V', E') is a connected subgraph of G. Also let $\mathrm{Cover}(V_1, \ldots, V_k)$ express the fact that vertex subsets $V_1, \ldots, V_k$ cover all vertices of the graph, by simply stating that $\forall x(x \in V_1 \vee x \in V_2 \vee \ldots x \in V_k)$.

This allows us to express Isometric Path Cover with Terminals as an optimization MSOL_2 problem, called an EMS-problem in [3]. Let $\varphi(E_1, \ldots, E_k)$ be the formula on edge sets $E_1, \ldots, E_k$ expressing the property that there exist k connected subgraphs $(V_1, E_1), \ldots, (V_k, E_k)$ of G such that the sets $V_1, \ldots, V_k$ cover all vertices of G, and graph (V_i, E_i) contains terminals s_i, t_i, for all $1 \le i \le k$. More formally,

$$\begin{aligned}\varphi(E_1, \ldots, E_k) = \exists\, V_1, \ldots V_k\ [&(s_1 \in V_1) \wedge (t_1 \in V_1) \wedge \cdots \wedge (s_k \in V_k) \wedge (t_k \in V_k) \\ &\wedge \mathrm{Cover}(V_1, \ldots, V_k) \wedge \mathrm{Connected}(V_1, E_1) \wedge \cdots \wedge \mathrm{Connected}(V_k, E_k)].\end{aligned}$$

Consider now the optimization version of this problem, where the goal is to find edge sets $E_1, \ldots, E_k$ satisfying $\varphi(E_1, \ldots, E_k)$ and minimizing $|E_1| + \cdots + |E_k|$. Let OptCover denote this optimum. By Theorem 5.6 in [3], this problem can be solved in linear time on bounded treewidth graphs. More precisely, Arnborg, Lagergren, and Seese [3] call such problems "EMS linear extremum problems", in the sense that they correspond to linear optimization functions over the sizes of set variables, when these variables satisfy an MSOL_2 formula over labelled graphs with a fixed number of labels (here we consider each terminal vertex labelled with a different label). In contemporary terms, the problem is FPT parameterized by treewidth plus the number of terminals, and the running time is linear in n.

We now observe that our input is a YES-instance of Isometric Path Cover with Terminals if and only if $\mathrm{OptCover} = \mathrm{dist}(s_1, t_1) + \mathrm{dist}(s_2, t_2) + \cdots + \mathrm{dist}(s_k, t_k)$. Indeed, if there exist the required shortest s_i-t_i paths $P_1, \ldots, P_k$ covering the vertex set of the whole graph, then their edge sets $E_1, \ldots, E_k$ provide a solution for our optimization problem whose objective is the sum of the lengths of the paths. Conversely, for any of the k connected subgraphs (V_i, E_i) of G such that $s_i, t_i \in V_i$, we have $|E_i| \ge \mathrm{dist}(s_i, t_i)$. Therefore, by simply checking if OptCover corresponds to the sum of the distances between pairs of terminals, we decide whether the input satisfies Isometric Path Cover with Terminals. Altogether, this problem is FPT parameterized by k, which concludes the proof of Theorem 1.3.

As mentioned in the introduction, the same techniques extend to variants where the covering paths are required to be edge-disjoint or vertex-disjoint, by simply adding disjointness conditions in the MSOL_2 formula φ.

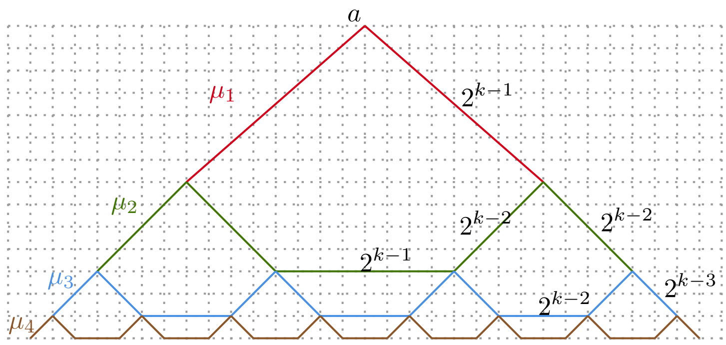

Fig. 7. *A graph that can be edge-covered with k shortest paths and with 2^k vertices at distance 2^k-1 from a. Therefore, a path decomposition obtained by a BFS from vertex a has width exponential in k. Nevertheless, one can easily prove that this graph has pathwidth at most k. (Color available online.)*

6. Conclusion. We have shown that graphs that can be covered by k shortest paths have their pathwidth upper-bounded by a function of k. Our bound is exponential, and the first natural open question is whether it can be improved to a polynomial bound. Such an improvement cannot rely on path decompositions based on the layers of an arbitrary BFS, since we have examples (see Figure 7) where the same layer contains 2^k vertices. Nevertheless, we leave as an open question whether graphs whose vertices (or edges) can be covered by k shortest paths have treewidth at most a polynomial in k.

Observe that the approach does not generalize to coverings with few *induced paths,* since grids have arbitrarily large treewidth but are edge-coverable by four induced paths.

On the algorithmic side, we have proved that the Isometric Path Cover with Terminals and Strong Geodetic Set with Terminals problems are FPT parameterized by the number of terminals. This directly entails that problems Isometric Path Cover and Strong Geodetic Set are in XP with respect to the same parameter, by simply enumerating all possible pairs (respectively, sets) of terminals. An exciting open question is whether these two problems are FPT. By Theorem 1.2, this is equivalent to asking if the problems are FPT when parameterized by the solution size (i.e., number of paths/terminals) + pathwidth. (Indeed, if k is the number of terminals, Theorem 1.2 ensures that either the pathwidth pw of the input graph is upper bounded by a function $f(k)$, or we can directly reject the input for being a NO-instance. Therefore, if one of the problems is FPT parameterized by $k + \mathrm{pw}$, we obtain an FPT algorithm parameterized by k as follows. The algorithm checks that $\mathrm{pw} \leq f(k)$ as in Theorem 1.2, by a simple breadth-first search from an arbitrary vertex. If the assertion is false, the algorithm rejects. Otherwise, it simply remains to apply the algorithm parameterized by $k+\mathrm{pw}$ on parameter $k+f(k)$. Nevertheless, the answer to the question whether these problems are FPT for parameter $k + \mathrm{pw}$ seems nontrivial. At least, while many optimization problems are FPT when parameterized by treewidth/pathwidth, several problems including constraints on distances remain $W[1]$-hard even when parameterized by such structural parameters, plus solution size. We can cite recent hardness results for d-Scattered Set [17], whose goal is to find a large set of vertices at pairwise distance at least d or, even closer to our problems, Geodetic set [18], where one aims to find a small set of terminals of the input graph such that the set of all shortest paths between every pair of terminals covers the graph.

Another natural question is the study of these problems on directed graphs. Note that Isometric Path Partition, the partition version of Isometric Path Cover, was proved to be W[1]-hard on DAGs for the parameter solution size in [14]. Moreover, there are tournaments (orientations of cliques) whose vertices can be covered by a unique shortest path. For instance, those created from a directed Hamiltonian path with all remaining arcs going in the reverse direction. The treewidth of the underlying undirected graph, which is a clique, is not bounded. A DAG that can be covered by two directed paths but whose underlying undirected graph has large treewidth is the following: consider two disjoint directed paths, and add all arcs from the vertices of the first path to all vertices of the second one. Thus, our results cannot be directly extended to digraphs, not even to DAGs.

REFERENCES

[1] M. Aigner and M. Fromme, *A game of cops and robbers*, Discrete Appl. Math., 8 (1984), pp. 1–11.

[2] G. Andreatta and F. Mason, *Path covering problems and testing of printed circuits*, Discrete Appl. Math., 62 (1995), pp. 5–13.

[3] S. Arnborg, J. Lagergren, and D. Seese, *Easy problems for tree-decomposable graphs*, J. Algorithms, 12 (1991), pp. 308–340.

[4] A. Arokiaraj, S. Klavžar, P. D. Manuel, E. Thomas, and A. Xavier, *Strong geodetic problems in networks*, Discuss. Math. Graph Theory, 40 (2020), pp. 307–321.

[5] M. Bentert, A. Nichterlein, M. Renken, and P. Zschoche, *Using a geometric lens to find k disjoint shortest paths*, in Proceedings of the 48th International Colloquium on Automata, Languages, and Programming (ICALP 2021), Leibniz International Proceedings in Informatics (LIPIcs) 198, N. Bansal, E. Merelli, and J. Worrell, eds., Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2021, 26.

[6] D. Chakraborty, A. Dailly, S. Das, F. Foucaud, H. Gahlawat, and S. K. Ghosh, *Complexity and algorithms for isometric path cover on chordal graphs and beyond*, in Proceedings of the 33rd International Symposium on Algorithms and Computation, ISAAC 2022, LIPIcs 248, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022, 12.

[7] B. Courcelle, *The monadic second-order logic of graphs.* I. *Recognizable sets of finite graphs*, Inform. and Comput., 85 (1990), pp. 12–75.

[8] M. Cygan, F. V. Fomin, L. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, and S. Saurabh, *Parameterized Algorithms*, Springer, Cham, 2015.

[9] T. Davot, L. Isenmann, and J. Thiebaut, *On the approximation hardness of geodetic set and its variants*, in Proceedings of the 27th International Computing and Combinatorics Conference, COCOON 2021, Lecture Notes in Comput. Sci. 13025, Springer, 2021, pp. 76–88.

[10] J. H. G. de Sousa, *Exact Algorithms and Computational Complexity for the Strong Geodetic Set Problem*, Master's thesis, Universidade Federal de Minas Gerais, Brazil, 2018.

[11] R. Diestel, *Graph Theory*, 4th ed., Grad. Texts in Math. 173, Springer, 2012.

[12] M. Dumas, F. Foucaud, A. Perez, and I. Todinca, *On graphs coverable by k shortest paths, Seoul*, in Proceedings of the 33rd International Symposium on Algorithms and Computation, ISAAC 2022, 2022, Seoul, Korea, LIPIcs 248, S. W. Bae and H. Park, eds., Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022, 40.

[13] T. Eilam-Tzoreff, *The disjoint shortest paths problem*, Discrete Appl. Math., 85 (1998), pp. 113–138.

[14] H. Fernau, F. Foucaud, K. Mann, R. R. K. N., and U. Paradiya, *Parameterizing path partitions*, in .Proceedings of the 13th International Symposium on Algorithms and Complexity, CIAC 2023, Lecture Notes in Comput. Sci, Springer, 2023.

[15] D. C. Fisher and S. L. Fitzpatrick, *The isometric number of a graph*, J. Combin. Math. Combin. Comput., 38 (2001), pp. 97–110.

[16] F. V. Fomin and D. Kratsch, *Exact Exponential Algorithms*, Texts in Theoretical Computer Science, An EATCS Series, Springer, Heidelberg, 2010.

[17] I. Katsikarelis, M. Lampis, and V. T. Paschos, *Structurally parameterized d-scattered set*, Discret. Appl. Math., 308 (2022), pp. 168–186.

[18] L. Kellerhals and T. Koana, *Parameterized complexity of geodetic set*, in Proceedings of the 15th International Symposium on Parameterized and Exact Computation, IPEC 2020, LIPIcs 180, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, 20.

[19] C. V. G. C. Lima, V. F. d. Santos, J. H. G. Sousa, and S. A. Urrutia, *On the computational complexity of the strong geodetic recognition problem*, preprint, https://arxiv.org/abs/2208.01796, 2022.

[20] W. Lochet, *A polynomial time algorithm for the k-disjoint shortest paths problem*, in Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, SIAM, 2021, pp. 169–178, https://doi.org/10.1137/1.9781611976465.12.

[21] P. Manuel, *On the isometric path partition problem*, Discuss. Math. Graph Theory, 41 (2021), pp. 1077–1089.

[22] P. Manuel, S. Klavžar, A. Xavier, A. Arokiaraj, and E. Thomas, *Strong edge geodetic problem in networks*, Open Math., 15 (2017), pp. 1225–1235.

[23] M. Mezzini, *An $O(mn^2)$ algorithm for computing the strong geodetic number in outerplanar graphs*, Discuss. Math. Graph Theory, 42 (2022), pp. 591–599.

[24] J. Pan and G. J. Chang, *Isometric-path numbers of block graphs*, Inform. Process. Lett., 93 (2005), pp. 99–102.

[25] N. Robertson and P. D. Seymour, *Graph minors.* XIII. *The disjoint paths problem*, J. Combin. Theory Ser. B, 63 (1995), pp. 65–110.

[26] M. Thiessen and T. Gaertner, *Active learning of convex halfspaces on graphs*, in Proceedings of the 35th Conference on Neural Information Processing Systems, NeurIPS 2021, Vol. 34, Curran Associates, 2021, pp. 23413–23425.