

FEEDBACK VERTEX SET AND EVEN CYCLE TRANSVERSAL FOR H -FREE GRAPHS: FINDING LARGE BLOCK GRAPHS*

GIACOMO PAESANI[†], DANIËL PAULUSMA[†], AND PAWEŁ RZAŻEWSKI[‡]

Abstract. We prove new complexity results for FEEDBACK VERTEX SET and EVEN CYCLE TRANSVERSAL on H -free graphs, that is, graphs that do not contain some fixed graph H as an induced subgraph. In particular, we prove that for every $s \geq 1$, both problems are polynomial-time solvable for sP_3 -free graphs and $(sP_1 + P_5)$ -free graphs; here, the graph sP_3 denotes the disjoint union of s paths on three vertices and the graph $sP_1 + P_5$ denotes the disjoint union of s isolated vertices and a path on five vertices. Our new results for FEEDBACK VERTEX SET extend all known polynomial-time results for FEEDBACK VERTEX SET on H -free graphs, namely for sP_2 -free graphs [Chiarelli et al., *Theoret. Comput. Sci.*, 705 (2018), pp. 75–83], $(sP_1 + P_3)$ -free graphs [Dabrowski et al., *Algorithmica*, 82 (2020), pp. 2841–2866] and P_5 -free graphs [Abrishami et al., *Induced subgraphs of bounded treewidth and the container method*, in Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA), SIAM, Philadelphia, 2021, pp. 1948–1964]. Together, the new results also show that both problems exhibit the same behavior on H -free graphs (subject to some open cases). This is in part due to a new general algorithm we design for finding in a (sP_3) -free or $(sP_1 + P_5)$ -free graph G a largest induced subgraph whose blocks belong to some finite class \mathcal{C} of graphs. We also compare our results with the state-of-the-art results for the ODD CYCLE TRANSVERSAL problem, which is known to behave differently on H -free graphs.

Key words. feedback vertex set, even cycle transversal, odd cactus, forest, block

MSC codes. 05C85, 05C69

DOI. 10.1137/22M1468864

1. Introduction. For a set of graphs \mathcal{F} , an \mathcal{F} -transversal of a graph G is a set of vertices that intersects the vertex set of every (not necessarily induced) subgraph of G that is isomorphic to some graph of \mathcal{F} . The problem MIN \mathcal{F} -TRANSVERSAL (also called \mathcal{F} -DELETION) is to find an \mathcal{F} -transversal of minimum size (or size at most k , in the decision variant). Graph transversals form a central topic in discrete mathematics and theoretical computer science, both from a structural and an algorithmic point of view.

If \mathcal{F} is the set of all cycles, the set of all even cycles or odd cycles, then we obtain the problems FEEDBACK VERTEX SET, EVEN CYCLE TRANSVERSAL, and ODD CYCLE TRANSVERSAL, respectively. All three problems are NP-complete; hence, they have been studied for special graph classes, in particular *hereditary* graph classes, that is, classes closed under vertex deletion. Such classes can be characterized by a (unique) set \mathcal{H} of minimal forbidden induced subgraphs. Then, in order to initiate a systematic study, it is standard to first consider the case where \mathcal{H} has size 1, say $\mathcal{H} = \{H\}$ for some graph H .

We aim to extend known complexity results for FEEDBACK VERTEX SET for H -

*Received by the editors January 4, 2022; accepted for publication (in revised form) June 1, 2022; published electronically October 4, 2022. An extended abstract containing some of the results in this paper appeared in the proceedings of MFCS 2021 [18].

<https://doi.org/10.1137/22M1468864>

Funding: The first author was supported by the Leverhulme Trust (RPG-2016-258). The third author was supported by Polish National Science Centre grant 2018/31/D/ST6/00062.

[†]Department of Computer Science, Durham University, Durham, DH1 3LE, UK (giacomo.paesani@durham.ac.uk, daniel.paulusma@durham.ac.uk).

[‡]Faculty of Mathematics and Information Science, Warsaw University of Technology, 00-661 Warsaw, Poland, and Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw, 00-927 Warsaw, Poland (p.rzazewski@mini.pw.edu.pl).

free graphs and to perform a new, similar study for EVEN CYCLE TRANSVERSAL (for which, so far, mainly parameterized complexity results exist [2, 3, 14, 16]). To describe the known and new results we need some terminology. The cycle and path on r vertices are denoted C_r and P_r , respectively. The *disjoint union* of two vertex-disjoint graphs G_1 and G_2 is the graph $G_1 + G_2 = (V(G_1) \cup V(G_2), E(G_1) \cup E(G_2))$. We write sG for the disjoint union of s copies of G . For a set $S \subseteq V$, let $G[S]$ be the subgraph of G induced by S . We write $H \subseteq_i G$ (or $G \supseteq_i H$) if H is an induced subgraph of G .

1.1. Known results. By Poljak's construction [19], for every integer $g \geq 3$, FEEDBACK VERTEX SET is NP-complete for graphs of girth at least g (the *girth* of a graph is the length of its shortest cycle). The same holds for ODD CYCLE TRANSVERSAL [8]. It is also known that FEEDBACK VERTEX SET [17] and ODD CYCLE TRANSVERSAL [8] are NP-complete for line graphs and thus for claw-free graphs (the claw is the 4-vertex star). Hence, both problems are NP-complete for the class of H -free graphs whenever H has a cycle or claw. A graph with no cycles and no claws is a forest of maximum degree at most 2. Thus, it remains to consider the case where H is a *linear forest*, that is, a collection of disjoint paths. Both problems are polynomial-time solvable on permutation graphs [6] and thus on P_4 -free graphs [6], on sP_2 -free graphs for every $s \geq 1$ [8] and on $(sP_1 + P_3)$ -free graphs for every $s \geq 0$ [10]. Additionally, FEEDBACK VERTEX SET is polynomial-time solvable on P_5 -free graphs [1], and ODD CYCLE TRANSVERSAL is NP-complete for $(P_2 + P_5, P_6)$ -free graphs [10]. A similar NP-hardness result for FEEDBACK VERTEX SET or EVEN CYCLE TRANSVERSAL is unlikely: for every linear forest H , both problems are quasipolynomial-time solvable on H -free graphs [11] (see section 6 for details).

1.2. New polynomial-time results. We first note that MIN \mathcal{F} -TRANSVERSAL is polynomially equivalent to MAX INDUCED \mathcal{F} -SUBGRAPH, the problem of finding a maximum-size induced subgraph of the input graph G that does not belong to \mathcal{F} (where we assume that G has at least one such subgraph). We say that MAX INDUCED \mathcal{F} -SUBGRAPH is the *complementary* problem of MIN \mathcal{F} -TRANSVERSAL, and vice versa. For example, setting $\mathcal{F} = \{P_2\}$ yields the well-known complementary problems MIN VERTEX COVER and MAX INDEPENDENT SET.

Using the complementary perspective, we now argue that FEEDBACK VERTEX SET and EVEN CYCLE TRANSVERSAL are closely related, in contrast to ODD CYCLE TRANSVERSAL. A graph G is *biconnected* if it has at least two vertices, is connected, and $G - u$ is connected for every $u \in V(G)$. A *block* of a graph G is an inclusionwise maximal biconnected subgraph of G . We now let \mathcal{C} be a set of biconnected graphs. A graph G is a *\mathcal{C} -block graph* if every block of G is isomorphic to some graph in \mathcal{C} . If $\mathcal{C} = \{P_2\}$, then \mathcal{C} -block graphs are precisely forests, and if $\mathcal{C} = \{P_2, C_3, C_5, C_7, \dots\}$, then \mathcal{C} -block graphs are called *odd cacti*. It is well known that a graph is an odd cactus if and only if it does not contain any even cycle as a subgraph. Hence, the complementary problems of EVEN CYCLE TRANSVERSAL and FEEDBACK VERTEX SET are somewhat similar: in particular, both forests and odd cacti have bounded treewidth and their blocks have a very simple structure. This is in stark contrast to ODD CYCLE TRANSVERSAL, whose complementary problem is to find a large induced bipartite subgraph, which might be arbitrarily complicated.

The commonality of complementary problems of EVEN CYCLE TRANSVERSAL and FEEDBACK VERTEX SET leads to the following optimization problem, where \mathcal{C} is some fixed class of biconnected graphs, that is, \mathcal{C} is not part of the input but specified in advance. Note that we consider the more general setting in which every vertex v

of G is equipped with a weight $\mathfrak{w}(v) > 0$, and we must find a solution with maximum total weight.

MAX \mathcal{C} -BLOCK GRAPH

Instance: a graph $G = (V, E)$ with a vertex weight function $\mathfrak{w} : V \rightarrow \mathbb{Q}^+$.

Objective: find a maximum-weight set $X \subseteq V$ such that $G[X]$ is a \mathcal{C} -block graph.

We observe that MAX \mathcal{C} -BLOCK GRAPH is well defined for every set \mathcal{C} , including $\mathcal{C} = \emptyset$, as every independent set in a graph forms a solution. A restriction of the MAX \mathcal{C} -BLOCK GRAPH problem was introduced and studied from a parameterized complexity perspective by Bonnet et al. [4] as BOUNDED \mathcal{C} -BLOCK VERTEX DELETION (so from the complementary perspective) where each block must, in addition, have bounded size.

In section 2 we slightly extend a previously known result, concerning the so-called *blob graphs* [11]. This extended version of the result forms a key ingredient for the proof of our main results, shown in sections 3 and 4, respectively, which are the following two theorems.

THEOREM 1. *For every integer $s \geq 1$ and every finite class \mathcal{C} of biconnected graphs, MAX \mathcal{C} -BLOCK GRAPH can be solved in polynomial time for sP_3 -free graphs.*

THEOREM 2. *For every integer $s \geq 1$ and every finite class \mathcal{C} of biconnected graphs, MAX \mathcal{C} -BLOCK GRAPH can be solved in polynomial time for $(sP_1 + P_5)$ -free graphs.*

We note that sP_3 -free graphs are the graphs that become a disjoint union of cliques after removing the vertices of any induced $(s-1)P_3$ and their neighbors. The class of $(sP_1 + P_5)$ -free graphs is a natural generalization of the class of $(P_1 + P_5)$ -free graphs. The latter graphs are also known as the *nearly P_5 -free graphs*, that is, graphs in which the subgraph induced by the nonneighborhood of any vertex is P_5 -free. More generally, a graph is *nearly π* for some graph property π if the subgraph induced by the nonneighborhood of any vertex has property π . It is easy to see that MAX INDEPENDENT SET is polynomial-time solvable for graphs that are nearly π if it is so for graphs with property π (see, for example, [5]). However, for many other graph problems, including the problems we study in this paper, such a statement either does not hold, is not known, or could be nontrivial to prove even for graphs that are nearly P_5 -free (such as, for example, CONNECTED VERTEX COVER [13]).

We prove both theorems using the same technique. Essentially we reduce to MAX INDEPENDENT SET for sP_3 -free blob graphs and $(sP_1 + P_5)$ -free blob graphs, respectively. In order to do this, we first perform a structural analysis of sP_3 -free \mathcal{C} -block graphs and $(sP_1 + P_5)$ -free \mathcal{C} -block graphs. These analyses have some common elements, namely they are based on the so-called block-cut forest of the (unknown) maximum-weight solution F . This forest contains as its vertices the cutvertices x and blocks b of F such that xb is an edge if and only if x belongs to b . The precise arguments are different and the resulting polynomial-time algorithms exploit the sP_3 -freeness and $(sP_1 + P_5)$ -freeness in different ways.

1.3. Implications and new hardness results. Theorems 1 and 2 imply corresponding results for FEEDBACK VERTEX SET, as the latter problem is equivalent to MAX $\{P_2\}$ -BLOCK GRAPH. The condition for \mathcal{C} to be finite is critical for our proof technique. Nevertheless, we still have the corresponding result for EVEN CYCLE TRANSVERSAL as well: for sP_3 -free graphs, the cases $\mathcal{C} = \{P_2, C_3, C_5, C_7, \dots\}$ and $\mathcal{C} = \{P_2, C_3, C_5, \dots, C_{4s-3}\}$ are equivalent. Note that we cannot make such an

argument for ODD CYCLE TRANSVERSAL, as arbitrarily large bicliques are $2P_3$ -free.

COROLLARY 3. *For every integer $s \geq 1$, FEEDBACK VERTEX SET and EVEN CYCLE TRANSVERSAL can be solved in polynomial time for sP_3 -free graphs and $(sP_1 + P_5)$ -free graphs.*

Corollary 3 extends the aforementioned results for FEEDBACK VERTEX SET on sP_2 -free graphs and $(sP_1 + P_3)$ -free graphs. In section 5 we prove that EVEN CYCLE TRANSVERSAL is NP-complete for graphs of large girth and for line graphs, and consequently, for H -free graphs where H contains a cycle or a claw. Hence, FEEDBACK VERTEX SET and EVEN CYCLE TRANSVERSAL behave similarly on H -free graphs, subject to a number of open cases, which we listed in Table 1.

TABLE 1

The complexity of FEEDBACK VERTEX SET (FVS), EVEN CYCLE TRANSVERSAL (ECT), and ODD CYCLE TRANSVERSAL (OCT) on H -free graphs for a linear forest H . All three problems are NP-complete for H -free graphs when H is not a linear forest (see also section 5). The bold cases (for FVS and ECT) are the algorithmic contributions of this paper. We write $H \subseteq_i H'$ if H is an induced subgraph of H' . See section 1.1 for references to the known results in the table.

	Polynomial-time	Unresolved	NP-complete
FVS	$H \subseteq_i sP_1 + P_5$ or sP_3 for $s \geq 1$	$H \supseteq_i P_2 + P_4$ or P_6	none
ECT	$H \subseteq_i sP_1 + P_5$ or sP_3 for $s \geq 1$	$H \supseteq_i P_2 + P_4$ or P_6	none
OCT	$H = P_4$ or $H \subseteq_i sP_1 + P_3$ or sP_2 for $s \geq 1$	$H = sP_1 + P_5$ for $s \geq 0$ or $H = sP_1 + tP_2 + uP_3 + vP_4$ for $s, t, u \geq 0, v \geq 1$ with $\min\{s, t, u\} \geq 1$ if $v = 1$, or $H = sP_1 + tP_2 + uP_3$ for $s, t \geq 0, u \geq 1$ with $u \geq 2$ if $t = 0$	$H \supseteq_i P_6$ or $P_2 + P_5$

2. Blob graph of graphs with no large linear forest. Let $G = (V, E)$ be a graph. A (connected) component is a maximal connected subgraph of G . The neighborhood of a vertex $u \in V$ is the set $N_G(u) = \{v \mid uv \in E\}$. For $U \subseteq V$, we let $N_G(U) = \bigcup_{u \in U} N(u) \setminus U$. Two sets $X_1, X_2 \subseteq V(G)$ are adjacent if $X_1 \cap X_2 \neq \emptyset$ or there exists an edge with one endvertex in X_1 and the other in X_2 . The blob graph G° of G is defined as follows:

$$V(G^\circ) := \{X \subseteq V(G) \mid G[X] \text{ is connected}\}$$

$$\text{and } E(G^\circ) := \{X_1 X_2 \mid X_1 \text{ and } X_2 \text{ are adjacent}\}.$$

Gartland et al. [11] showed that for every graph G , the length of a longest induced path in G° is equal to the length of a longest induced path in G . We slightly generalize this result.

THEOREM 4. *For every linear forest H , a graph G contains H as an induced subgraph if and only if G° contains H as an induced subgraph.*

Proof. As G is an induced subgraph of G° , the (\Rightarrow) implication is immediate. We prove the (\Leftarrow) implication by induction on the number k of connected components of H . If $k = 1$, then the claim follows directly from the aforementioned result of Gartland et al. [11]. So assume that $k \geq 2$ and the statement holds for all linear forests H with fewer than k connected components. Let P' be one of the connected components of H , and define $H' := H - P'$.

Suppose that G° contains an induced subgraph isomorphic to H . Let \mathcal{X} be the set of vertices of G° , such that $G^\circ[\mathcal{X}]$ is isomorphic to H . Furthermore, let $\mathcal{Y} \subseteq \mathcal{X}$ be the set of vertices that induce in $G^\circ[\mathcal{X}]$ the component P' of H , that is, $G^\circ[\mathcal{Y}]$ is isomorphic to P' .

Let $Y \subseteq V(G)$ be the union of sets in \mathcal{Y} . Note that $G^\circ[\mathcal{Y}]$ is an induced subgraph of $(G[Y])^\circ$. Thus, by the inductive assumption, $G[Y]$ contains an induced copy of P' .

Let $X \subseteq V(G)$ be the union of sets in $\mathcal{X} \setminus \mathcal{Y}$. Since the copy of H in G° is induced, we know that in G° there are no edges between $\mathcal{X} \setminus \mathcal{Y}$ and \mathcal{Y} . This is equivalent to saying that $X \cap N[Y] = \emptyset$. So we conclude that $G^\circ[\mathcal{X} \setminus \mathcal{Y}]$ is an induced subgraph of $(G - N[Y])^\circ$. Since $G^\circ[\mathcal{X} \setminus \mathcal{Y}]$, and thus $(G - N[Y])^\circ$, contains an induced copy of H' , by the inductive assumption we know that $G - N[Y]$ contains an induced copy of H' . Combining this subgraph with the induced copy of P' in $G[Y]$, we obtain an induced copy of H in G . \square

3. The proof of Theorem 1. We start by analyzing the structure of sP_3 -free \mathcal{C} -block graphs in section 3.1, where \mathcal{C} is any finite class of biconnected graphs. Then, in section 3.2, we present our algorithm for MAX \mathcal{C} -BLOCK GRAPH on sP_3 -free graphs.

3.1. Structural lemmas. From now on, let \mathcal{C} be a finite class of biconnected graphs. For some fixed positive integer s , let $G = (V, E)$ be an sP_3 -free graph with n vertices and vertex weights $\mathbf{w}: V \rightarrow \mathbb{Q}^+$. Let $X \subseteq V$ such that $F = G[X]$ is a \mathcal{C} -block graph. A component of F is *trivial* if it is a single vertex or a single block, otherwise it is *nontrivial*. Let F' be the graph obtained from F by removing all trivial components. Note that F' and F are sP_3 -free, as G is sP_3 -free.

We denote the set of cutvertices of F' and the set of blocks of F' by $\text{Cutvertices}(F')$ and $\text{Blocks}(F')$, respectively. The *block-cut forest* $\text{BCF}(F')$ of F' has vertex set $\text{Cutvertices}(F') \cup \text{Blocks}(F')$ and an edge set that consists of all edges xb such that $x \in \text{Cutvertices}(F')$ and $b \in \text{Blocks}(F')$, and x belongs to b . By definition, each component of F' has a cutvertex; we pick an arbitrary one as the root for the corresponding tree in $\text{BCF}(F')$ to get a parent-child relation. Each leaf of $\text{BCF}(F')$ belongs to $\text{Blocks}(F')$, and we call such blocks *leaf blocks*.

A cutvertex x of F' is a *terminal of type 1* if x has at least two children in $\text{BCF}(F')$ that are leaves, whereas x is a *terminal of type 2* if there exists a leaf block, whose great-grandparent in $\text{BCF}(F')$ is x . In the latter case, there is a three-edge downward path from x to a leaf in $\text{BCF}(F')$; see also Figure 1. Let d be the maximum number of vertices of a graph in \mathcal{C} .

LEMMA 5. *At most $d \cdot (s - 1)$ vertices of F' are terminals of type 1.*

Proof. For the sake of contradiction, suppose that there are at least $d \cdot (s - 1) + 1$ terminals of type 1. We observe that F' is d -colorable. Indeed, each block has at most d vertices, so d colors are sufficient to color each block. Furthermore, we can permute the colors in each block, so that the colorings agree on cutvertices.

This implies that there is an independent set X of size at least s , whose every element is a terminal of type 1. For each such terminal v , let its *private* P_3 be a 3-vertex path with v as the central vertex and each endpoint belonging to a different leaf block that is a child of v in $\text{BCF}(F')$. Note that each private P_3 is induced. Furthermore, the private P_3 's of vertices in X are pairwise nonadjacent: this follows from the definition of terminals of type 1 and the fact that X is independent. Thus we have found an induced sP_3 in F , a contradiction. \square

LEMMA 6. *At most $(d + 1) \cdot (s - 1)$ vertices of F' are terminals of type 2.*

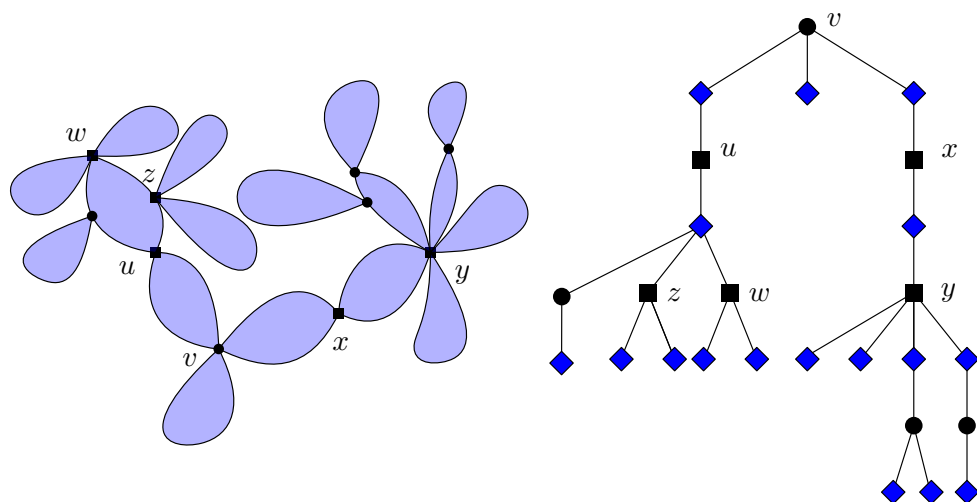


FIG. 1. Left: a graph F' . Blue shapes are blocks, squares are terminals, and dots are non-terminal cutvertices. Right: $\text{BCF}(F')$, rooted in the cutvertex v . Blue diamonds are blocks; w and z are terminals of type 1, u and x are terminals of type 2, and y is a terminal of both types. The remaining cutvertices are not terminals. We will also use this example with this particular $\text{BCF}(F')$ in later figures. (Figure in color online.)

Proof. For the sake of contradiction, suppose that there are at least $(d+1) \cdot (s-1) + 1$ terminals of type 2. Observe that F' has a proper $(d+1)$ -coloring f , satisfying the following two properties:

1. the vertices in each block receive pairwise distinct colors and
2. if b is a block, then any vertex of b receives a color which is different than the color of the cutvertex which is the great-grandparent of b in $\text{BCF}(F')$ (if such a cutvertex exists).

It is easy to find such a coloring of each tree in $\text{BCF}(F')$ by choosing an arbitrary color for the root and proceeding in a top-down fashion. Suppose we want to color the block b and its parent in $\text{BCF}(F')$ is the cutvertex v . Recall that b has at most d vertices and exactly one of them is already colored. Furthermore, we want to avoid the color of the grandparent of v (if such a vertex exists), so we have sufficiently many free colors to color each vertex of $b \setminus \{v\}$ with a different one.

Now, by our assumption, there is a set X of at least s terminals of type 2 that received the same color in f . For each $v \in X$, we define its *private* P_3 as follows. Recall that by the definition of a terminal of type 2, there is a leaf block b , whose great-grandparent in $\text{BCF}(F')$ is v . The private P_3 of v is given by the first three vertices on a shortest path P from v to b . Note that in the extreme case it might happen that both b and its grandparent in $\text{BCF}(F')$ are edges, but P always has at least three vertices.

Clearly, each private P_3 is an induced path. We claim that the private P_3 's associated with any two vertices of X form an induced $2P_3$. For the sake of contradiction, suppose otherwise. Let v_1, v_2 be distinct vertices of X , and let v_i, x_i, y_i be the consecutive vertices of the private P_3 associated with v_i . Let b_i be the block containing v_i and x_i .

First, observe that the sets $\{v_1, x_1, y_1\}$ and $\{v_2, x_2, y_2\}$ are disjoint. Indeed, we know that $v_1 \neq v_2$ by assumption, and because $\text{BCF}(F')$ is a rooted forest, we have

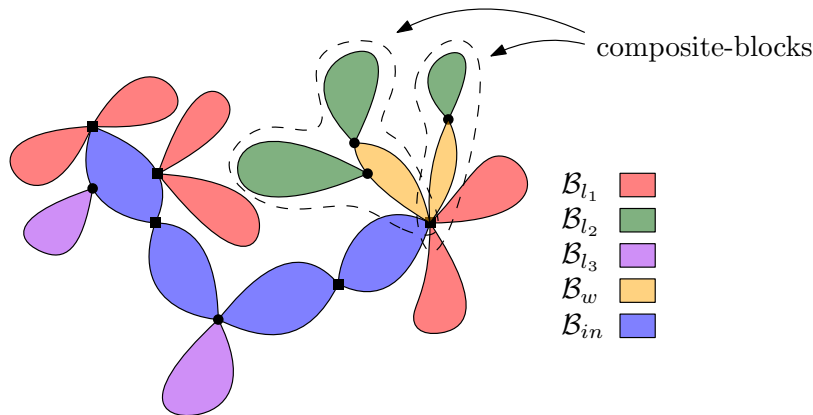


FIG. 2. The classification of blocks of the example of Figure 1 (so this classification is based on the block-cut forest $\text{BCF}(F')$ from Figure 1).

that $\{x_1, y_1\} \cap \{x_2, y_2\} = \emptyset$. Furthermore, recall that $f(v_1) = f(v_2)$ and by the definition of f , we have that the colors of x_i and of y_i are different from the color of v_i .

So now suppose that there is an edge with one endvertex in $\{v_1, x_1, y_1\}$ and the other in $\{v_2, x_2, y_2\}$. Clearly this edge cannot join v_1 and v_2 , as the coloring f is proper. Furthermore, there is no edge between $\{x_1, y_1\}$ and $\{x_2, y_2\}$, as v_1 and v_2 are cutvertices of a rooted tree. Suppose that v_2 is adjacent to x_1 (the case that v_1 is adjacent to x_2 is symmetric). As each vertex of b_1 gets assigned a different color by f and $f(v_1) = f(v_2)$, we observe that v_2 cannot belong to b_1 . Thus x_1 is a cutvertex. However, by the second property of f , we obtain that the color of v_2 must be different from the color of v_1 . As $f(v_1) = f(v_2)$, this is a contradiction.

So finally suppose that v_2 is adjacent to y_1 (the case that v_1 is adjacent to y_2 is symmetric). Note that then y_1 cannot belong to a leaf block, meaning that y_1 belongs to b_1 . Similarly to the previous paragraph, the definition of f implies that the color of v_2 must be different from the color of v_1 , a contradiction.

We conclude that $\{v_1, x_1, y_1, v_2, x_2, y_2\}$ induces a $2P_3$. As v_1 and v_2 were arbitrary vertices of X and $|X| \geq s$, this means we have found an induced sP_3 in F' , a contradiction. \square

Lemmas 5 and 6 imply the following.

LEMMA 7. The number of terminals of F' is at most $(2d + 1) \cdot (s - 1)$.

If v is a terminal of type 2, then by definition there is a cutvertex w that belongs to both a block containing v as well as to some leaf block. We call such w a *witness* of v . We note that a (nonleaf) block may contain multiple witnesses. Some of these witnesses might be terminals (and in that case they are of type 1), whereas others might not be terminals.

We now partition the set of blocks of F' into the following subsets; see also Figure 2:

- B_{l_1} is the set of leaf blocks containing a terminal of type 1;
- B_{l_2} is the set of leaf blocks containing a witness w that is not a terminal of type 1;
- B_{l_3} is the set of remaining leaf blocks, that is, the ones with a cutvertex that

- is neither a terminal nor a witness;
- \mathcal{B}_w is the set of blocks with at least two cutvertices, one of which is a terminal of type 2 and all the other ones are nonterminal witnesses of that type-2 terminal;
- \mathcal{B}_{in} is the set of all remaining blocks.

Note that blocks in \mathcal{B}_{l_2} and \mathcal{B}_w come in groups, that is, for each block B in \mathcal{B}_w , there are at most $d - 1$ blocks B' in \mathcal{B}_{l_2} , such that B and B' share a vertex (note that this common vertex is a nonterminal witness). We will consider this group of blocks as one object. Formally, a *composite-block* is a graph $G[V(B) \cup \bigcup_{1 \leq i \leq r} V(B_i)]$ for some integer $r \leq d - 1$ chosen to be maximum, such that

- $B \in \mathcal{B}_w$, so B has some vertex u that is a terminal of type 2;
- for every $i \in \{1, \dots, r\}$, $B_i \in \mathcal{B}_{l_2}$; and
- for every $i \in \{1, \dots, r\}$, $|V(B) \cap V(B_i)| = 1$ and $u \notin V(B_i)$.

Note that each composite-block has at most $(d-1)d + d = d^2$ vertices and contains exactly one terminal of type 2. Let \mathcal{B}_d be the set of all composite-blocks.

A *backbone* of a component Z of F' is a minimum tree T_Z contained in Z that connects all terminals of F' that belong to Z ; observe that all leaves of T_Z are terminals. The *skeleton* S of F' is the graph obtained from F' by removing all vertices from the blocks in \mathcal{B}_{l_1} except terminals of type 1 and all vertices from the composite-blocks in \mathcal{B}_d except terminals of type 2. Note that every backbone is a subgraph of S . Furthermore, the vertices of the blocks in \mathcal{B}_{l_3} all belong to S .

3.2. The algorithm. Outline. Our polynomial-time algorithm consists of the following two phases:

1. *Branching Phase*, which consists of the following three steps:
 1. guessing the terminals of F' ;
 2. guessing the backbones of the components of F' ; and
 3. guessing the skeleton of F' ; and
2. *Completion Phase*, where we extend the partial solutions obtained in the Branching Phase to complete ones by finding nonskeleton vertices of F' and trivial components of F ; we do this by
 1. reducing the problem to MAX WEIGHT INDEPENDENT SET for sP_3 -free graphs using the blob graph construction in section 2 and
 2. solving this problem using the polynomial-time algorithm of Brandstädt and Mosca [7].

We now describe our algorithm, prove its correctness, and perform a running time analysis.

Branching Phase. This phase of our algorithm consists of a series of guesses, where we find certain vertices and substructures in G . The total number of vertices to be guessed will be $\mathcal{O}(s^2 d^2)$. Since we guess them exhaustively, this results in a recursion tree with $\mathcal{O}(n^{\mathcal{O}(s^2 d^2)})$ leaves. As both s and d are constants, this bound is polynomial in n . We will ensure that the optimum solution $F = G[X]$ will be found in the call corresponding to at least one of the leaves of the recursion tree. Based on the properties of F , we will expect the guessed vertices to satisfy certain conditions. If, at some point, the guessed vertices do not satisfy these conditions, we just terminate the current call, as it will not lead us to find F . This will be applied implicitly throughout the execution of the algorithm.

The Branching Phase is illustrated in Figures 3–5. We use the convention that gray/black elements are still unknown and blue elements are the ones that we have already guessed.

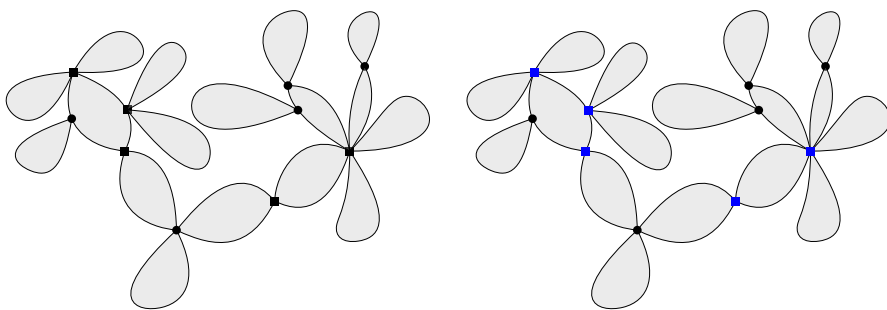


FIG. 3. Step 1 of the Branching Phase. Left: the graph F' . Right: the terminals of F' .

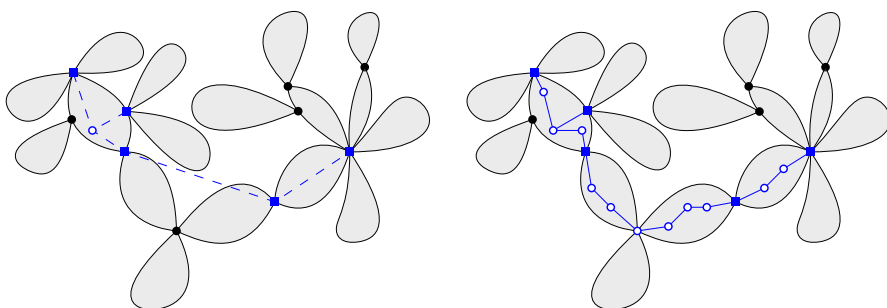


FIG. 4. Step 2 of the Branching Phase. Left: the tree T'_Z . Right: the tree T_Z .

Step 1. Guessing the terminals of \mathbf{F}' . We guess the set $C \subseteq V$ of terminals of F' . By Lemma 7, the total number of terminals is bounded by $(2d + 1) \cdot (s - 1) \leq 3ds$. Hence, we consider $O(n^{3ds})$ options and for each chosen set C we do as follows. For each terminal in C , we guess its type (1, 2, or both). This results in $3^{|C|} \leq 3^{3ds}$ possibilities. We also guess the partition of C , corresponding to the connected components of F . This results in at most $|C|^{|C|} \leq (3ds)^{3ds}$ additional branches. In total, we have $O(n^{O(ds)})$ branches. Step 1 is illustrated in Figure 3.

Step 2. Guessing the backbone of each component of \mathbf{F}' . Let Z be a component of F' . Let $C_Z \subseteq C$ be the subset of terminals that are in Z . Let T_Z be a backbone of Z . Let T'_Z be the tree obtained from T_Z by contracting every path in T_Z whose internal vertices are all nonterminals and of degree 2 to an edge. Note that every nonterminal vertex of T'_Z has degree at least 3. Since T'_Z has at most $|C_Z|$ vertices of degree at most 2, by the handshaking lemma we observe that the total number of vertices of T'_Z is at most $2|C_Z|$. Recall that every edge of T'_Z corresponds to an induced path in T_Z . Since F' is sP_3 -free and thus P_{4s-1} -free, we conclude that T_Z has at most $2|C_Z| \cdot (4s - 2) \leq 8s \cdot |C_Z|$ vertices.

Let T be the forest whose components are the guessed backbones of the components of F' . Note that the total number of vertices of T is at most $\sum_Z 8s \cdot |C_Z| = 8s \cdot |C| \leq 24ds^2$. Thus we may guess the whole forest T , which results in $O(n^{O(ds^2)})$ branches. Step 2 is illustrated in Figure 4.

Step 3. Guessing the skeleton of \mathbf{F}' . Let T be the forest guessed in the previous step; recall that T has at most $24ds^2$ vertices. We guess the partition of $E(T)$ corresponding to *blocks* of F' ; note that a vertex v may be in several blocks: this happens precisely if v is a cutvertex in F' . This results in at most $|E(T)|^{O(|E(T)|)} \leq$

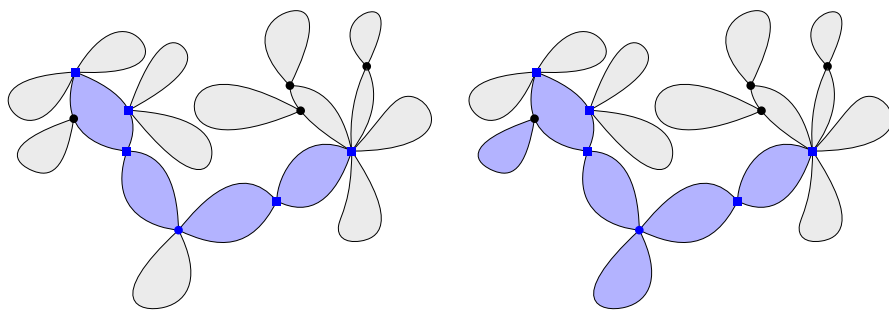


FIG. 5. Step 3 of the Branching Phase. Left: our knowledge about F' after guessing the blocks in \mathcal{B}_{in} . Right: our knowledge about F' after guessing the blocks in \mathcal{B}_{l_3} .

$|V(T)|^{\mathcal{O}(|V(T)|)} \leq (ds)^{\mathcal{O}(ds^2)}$ branches.

We now discuss some properties of the (composite-)blocks. We use the names of vertices as in the definitions introduced in section 3.1; recall also Figure 2. The crucial observation is that now there is a branch, where:

- For each block in \mathcal{B}_{l_1} , we have guessed its cutvertex and no other vertices.
- For each block in \mathcal{B}_{l_2} , we have not guessed any vertices.
- For each block in \mathcal{B}_{l_3} , we guessed no vertices yet except possibly its cutvertex in F' (but in the latter case we have not indicated this yet).
- For each block in \mathcal{B}_w , we have guessed its type-2 terminal vertex and we guessed no other vertices. Thus, for each composite-block in \mathcal{B}_d , we have guessed its cutvertex in F' and no other vertices.
- For each block in \mathcal{B}_{in} , we have guessed at least two vertices.

Now we proceed to the final guessing step; see Figure 5. First, we guess all blocks in \mathcal{B}_{in} . We can do it as

- (i) we know at least two vertices of such a block,
- (ii) the number of these blocks is at most $|E(T)| \leq 24ds^2$, and
- (iii) each block has at most d vertices.

This results in at most $n^{\mathcal{O}(|E(T)| \cdot d)} = n^{\mathcal{O}(d^2s^2)}$ further branches. In each branch, we guessed all vertices of a skeleton S except those that are noncutvertices of F' that belong to the blocks in \mathcal{B}_{l_3} . We will now guess which vertices of S will also belong to exactly one block in \mathcal{B}_{l_3} (so these vertices will be cutvertices in F'). These vertices belong to at least one block of \mathcal{B}_{in} . As the union of the vertices of the blocks in \mathcal{B}_{in} has size at most $24d^2s^2$, this leads to $\mathcal{O}(1)$ extra branches.

Finally, we guess all blocks in \mathcal{B}_{l_3} . Note that we can do it, as

- (i) we know their cutvertices,
- (ii) the number of these cutvertices is at most $24ds^2$,
- (iii) each cutvertex is contained in exactly one block from \mathcal{B}_{l_3} , and
- (iv) each block has at most d vertices.

This results in at most $n^{\mathcal{O}(|V(T)| \cdot d)} = n^{\mathcal{O}(d^2s^2)}$ branches, that is, at most $n^{\mathcal{O}(d^2s^2)}$ sets that are potential skeletons S of F' . Step 3 is illustrated in Figure 5.

The following claim summarizes the outcome of the guessing phase of the algorithm.

CLAIM A. In time $\mathcal{O}(n^{\mathcal{O}(s^2d^2)})$ we can enumerate a collection \mathcal{S} of $\mathcal{O}(n^{\mathcal{O}(s^2d^2)})$ triples (S, C_1, C_2) , where $S \subseteq V$ and $C_1, C_2 \subseteq S$ such that \mathcal{S} has the following prop-

erty. Let $X \subseteq V$, such that $F = G[X]$ is a \mathcal{C} -block graph. Let $X' \subseteq X$ be the vertex set of the graph F' obtained from F by removing all trivial components. Then there is at least one triple $(S, C_1, C_2) \in \mathcal{S}$, where

- (a) C_1 is the set of terminals of type 1 in F' ,
- (b) C_2 is the set of terminals of type 2 in F' ,
- (c) $G[S]$ is the skeleton of F' .

Completion Phase. Let \mathcal{S} be the collection from Claim A, and let $(S, C_1, C_2) \in \mathcal{S}$ be a triple that satisfies the properties listed in the statement of Claim A for an optimum solution $F = G[X]$. Let $\mathcal{X} := \mathcal{X}_0 \cup \mathcal{X}_1 \cup \mathcal{X}_2$ be the family of subsets of V with

$$\begin{aligned}\mathcal{X}_0 &:= \{\{v\} \mid v \in V\}, \\ \mathcal{X}_1 &:= \{B \subseteq V \mid G[B] \in \mathcal{C}\}, \text{ and} \\ \mathcal{X}_2 &:= \{B \subseteq V \mid B \text{ is a composite-block whose blocks are in } \mathcal{C}\}.\end{aligned}$$

Let $G^{\mathcal{C}}$ be the graph whose vertex set is \mathcal{X} , and edges join sets that are adjacent in G . Furthermore, we define a weight function $\mathbf{w}^{\mathcal{C}}: \mathcal{X} \rightarrow \mathbb{Q}^+$ as

$$\mathbf{w}^{\mathcal{C}}(A) = \sum_{v \in A} \mathbf{w}(v).$$

Note that in order to complete S to the optimum solution $F = G[X]$, we need to determine

- all blocks in \mathcal{B}_{l_1} ,
- all composite-blocks in \mathcal{B}_d ,
- all trivial components of F .

Note that the vertex sets of all these subgraphs are in the family \mathcal{X} and they form an independent set in $G^{\mathcal{C}}$. Furthermore, since X is of maximum weight, the total weight of selected subsets must be maximized. Thus the idea behind the last step is to reduce the problem to solving MAX WEIGHT INDEPENDENT SET in an appropriately defined subgraph of $G^{\mathcal{C}}$ and weights $\mathbf{w}^{\mathcal{C}}$.

To ensure that the selected subsets are consistent with our guess $(S, C_1, C_2) \in \mathcal{S}$, we will remove certain vertices from $G^{\mathcal{C}}$. In particular, let \mathcal{X}' consist of the sets $A \in \mathcal{X}$, such that

1. $A \in \mathcal{X}_0 \cup \mathcal{X}_1$ and A is nonadjacent to S ; these are the candidates for trivial components of F ,
2. $A \in \mathcal{X}_1$ and A intersects S in exactly one vertex, which is in C_1 ; these are the candidates for blocks in \mathcal{B}_{l_1} ,
3. $A \in \mathcal{X}_2$ and A intersects S in exactly one vertex, which is in C_2 and is not the cutvertex of $G[A]$; these are the candidates for composite-blocks in \mathcal{B}_d .

Now let $\mathcal{I} \subseteq \mathcal{X}'$ be an independent set of $G^{\mathcal{C}}$, and let $S' = \bigcup_{A \in \mathcal{I}} A$. It is straightforward to verify that if $(S, C_1, C_2) \in \mathcal{S}$ satisfies the properties listed in Claim A, then $G[S \cup S']$ is a \mathcal{C} -block graph. Thus, in one of the branches, we will find the optimum solution $F = G[X]$.

Now let us argue that the last step can be performed in polynomial time. First, observe that $|\mathcal{X}| \leq n + n^d + n^{d^2} = n^{\mathcal{O}(d^2)}$ and the family \mathcal{X} can be exhaustively enumerated in time $n^{\mathcal{O}(d^2)}$. Next, \mathcal{X}' can be computed in time polynomial in $|\mathcal{X}|$, and thus in n . This implies that the graph $G^{\mathcal{C}}[\mathcal{X}']$ can be computed in time polynomial in n . We observe that $G^{\mathcal{C}}$, and thus $G^{\mathcal{C}}[\mathcal{X}']$, is an induced subgraph of the blob

graph G° , introduced in section 2. Hence, by Theorem 4, we conclude that $G^{\mathcal{C}}[\mathcal{X}']$ is sP_3 -free.

The final ingredient is the polynomial-time algorithm for MAX WEIGHT INDEPENDENT SET in sP_3 -free graphs by Brandstädt and Mosca [7]. Its running time on an n' -vertex graph is $n'^{\mathcal{O}(s)}$. Since the number of vertices of $G^{\mathcal{C}}[\mathcal{X}']$ is $n^{\mathcal{O}(d^2)}$, we conclude that a maximum-weight independent set in $G^{\mathcal{C}}[\mathcal{X}']$ can be found in time $n^{\mathcal{O}(sd^2)}$.

Summing up, in the guessing phase, in time $n^{\mathcal{O}(s^2d^2)}$ we enumerate the family \mathcal{S} of size $n^{\mathcal{O}(s^2d^2)}$. Then, for each member (S, C_1, C_2) of \mathcal{S} , we try to extend the partial solution to a complete one. This takes time $n^{\mathcal{O}(sd^2)}$ per element of \mathcal{S} . Among all found solutions, we return the one with maximum weight. The total running time of the algorithm is $n^{\mathcal{O}(s^2d^2)}$, which is polynomial in n , since s and d are constants. This completes the proof of Theorem 1.

4. The proof of Theorem 2. In this section we prove that for every integer $s \geq 1$ and every finite class \mathcal{C} of biconnected graphs, MAX \mathcal{C} -BLOCK GRAPH can be solved in polynomial time for $(sP_1 + P_5)$ -free graphs. In section 4.1 we consider two boundary cases, namely the case where $\mathcal{C} = \emptyset$ and the case where $s = 0$. We will use these cases in our algorithm in section 4.3 after first proving some structural lemmas in section 4.2.

4.1. Two boundary cases. First, assume that $\mathcal{C} = \emptyset$. Recall that MAX \emptyset -BLOCK GRAPH is equivalent to MAX INDEPENDENT SET. The latter problem is polynomial-time solvable for P_5 -free graphs (and even for P_6 -free graphs [12]).

THEOREM 8 (see [15]). *MAX INDEPENDENT SET can be solved in polynomial time for P_5 -free graphs.*

We also recall the aforementioned and well-known observation from section 1 on graphs that are nearly π for some graph property π (see, for example, [5]). As a special case, we find that MAX INDEPENDENT SET for $(P_1 + P_5)$ -free graphs is polynomial-time solvable if it is so for P_5 -free graphs. Combining Theorem 8 with s applications of this argument leads to the following (known) extension of Theorem 8, which we will need as a lemma.

LEMMA 9. *For every fixed s , MAX INDEPENDENT SET can be solved in polynomial time in $(sP_1 + P_5)$ -free graphs.*

Now we deal with the case where $s = 0$. That is, we consider MAX \mathcal{C} -BLOCK GRAPH restricted to P_5 -free graphs when \mathcal{C} is a finite class of biconnected graphs. For this case we will use *Monadic Second-Order Logic* (MSO_2) over graphs, which consists of formulas with vertex variables, edge variables, vertex set variables, and edge set variables, quantifiers, and standard logic operators. We also have a predicate $\text{inc}(v, e)$, indicating that the vertex v belongs to the edge e .

Abrihami et al. [1, Theorems 5.3 and 7.3] proved the following result, even for the extension *Counting Monadic Second-Order Logic* (CMSO_2) of MSO_2 , which allows atomic formulas of the form $|X| \equiv p \pmod{q}$, where X is a set variable and $0 \leq p < q$ are integers (however, we do not need this extension for our purposes). We refer the reader to [9] for further information on MSO_2 logic on graphs.

THEOREM 10 (see [1]). *For every fixed CMSO_2 formula Φ and every constant t , it is possible for a P_5 -free graph G with weight function $\mathbf{w} : V(G) \rightarrow \mathbb{Q}^+$, to find in polynomial time a maximum-weight set $X \subseteq V(G)$, such that $G[X]$ is of treewidth at most t and satisfies Φ .*

We use Theorem 10 in our next lemma.

LEMMA 11. *For every finite class \mathcal{C} of biconnected graphs, MAX \mathcal{C} -BLOCK GRAPH can be solved in polynomial time for P_5 -free graphs.*

Proof. Note that every \mathcal{C} -block graph has treewidth at most $\max_{C \in \mathcal{C}} |V(C)|$, which is a constant. In order to use Theorem 10 it remains to show that the property that a set $X \subseteq V(G)$ induces a \mathcal{C} -block graph in G is expressible in CMSO_2 . We show that we can express this property already in MSO_2 .

In what follows, x and y are vertex variables, e is an edge variable, while X, X' , and Y denote vertex set variables. We will use some standard shortcuts (see also [9]), for instance:

$$\begin{aligned} \forall(x \in X) : \phi & \quad \text{stands for} & \quad \forall x : (x \in X) \Rightarrow \phi & \quad \text{and} \\ \forall(X' \subseteq X) : \phi & \quad \text{stands for} & \quad \forall X' : (\forall(x \in X') : x \in X) \Rightarrow \phi. \end{aligned}$$

We can now show the required claim. First, we express the property that $G[X]$ is connected in MSO_2 , in the usual way:

$$\text{connected}(X) := \forall(X' \subseteq X) : (\exists(x \in X') \exists(y \in X) \exists e : y \notin X' \wedge \text{inc}(x, e) \wedge \text{inc}(y, e)).$$

We now express the property that $G[X]$ is biconnected in MSO_2 :

$$\text{biconnected}(X) := \text{connected}(X) \wedge \forall(x \in X) : \text{connected}(X \setminus \{x\}).$$

Now $G[X]$ is a block of $G[Y]$ if it is biconnected and maximal with this property:

$$\text{block}(X, Y) := \text{biconnected}(X) \wedge \forall(y \in Y) : (y \notin X \Rightarrow \neg \text{biconnected}(X \cup \{y\})).$$

If C is a fixed graph, then the property that $G[X]$ is isomorphic to C can be easily hard-coded in a formula. We denote this predicate by $\text{is-}C(X)$. This can be extended to checking whether $G[X] \in \mathcal{C}$ (if \mathcal{C} is finite) by setting

$$\text{in-}\mathcal{C}(X) := \bigvee_{C \in \mathcal{C}} \text{is-}C(X).$$

Finally, $G[X]$ is a \mathcal{C} -block graph if and only if X satisfies

$$(4.1) \quad \text{is-}\mathcal{C}\text{-block-graph}(X) := \forall(X' \subseteq X) : \text{block}(X') \Rightarrow \text{in-}\mathcal{C}(X').$$

This completes the proof of the lemma. \square

4.2. Structural lemmas. Let $s \geq 0$, and let G be the $(sP_1 + P_5)$ -free instance graph with weight function \mathbf{w} . Let \mathcal{C} be a finite class of biconnected graphs. Let d be the maximum number of vertices of a graph in \mathcal{C} . Similarly to section 3, we will analyze the structure of an (unknown) maximum-weight solution. Let $X \subseteq V(G)$ be such that $F = G[X]$ is a \mathcal{C} -block graph. Again we consider the block-cut forest $\text{BCF}(F)$ of F . Recall that a leaf block is a block which is a leaf of $\text{BCF}(F)$.

LEMMA 12. *Every cutvertex of F belongs to at most $s + 1$ nonleaf blocks.*

Proof. For the sake of contradiction, let x be a cutvertex that belongs to $s + 2$ blocks b_1, b_2, \dots, b_{s+2} . Consider one such block b_i for $i \in [s + 2]$. As b_i is not a leaf block, there is a cutvertex $y_i \in V(b_i) \setminus \{x_i\}$ and a block $b'_i \neq b_i$ containing y_i ; see Figure 6. Note that $x \notin V(b'_i)$. Let y'_i be any vertex from $V(b'_i) \setminus \{y_i\}$; note that y'_i

is nonadjacent to x . Let Q_i be a shortest x - y'_i -path contained in $V(b_i) \cup V(b'_i)$ and note that Q_i has at least two edges. Furthermore, for $i, j \in [s+2]$, such that $i \neq j$, the paths Q_i and Q_j share one endvertex (namely x) and no other vertices. Thus $G[V(Q_1) \cup V(Q_2)]$ is an induced path with at least five vertices and consequently, $G[V(Q_1) \cup V(Q_2) \cup \bigcup_{i=3}^{s+2} \{y'_i\}]$ contains an induced $sP_1 + P_5$, a contradiction. \square

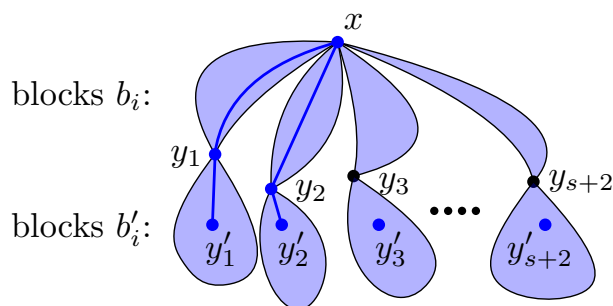


FIG. 6. The induced $sP_1 + P_5$ in the proof of Lemma 12.

A vertex $x \in V(F)$ is called *internal* if it is a cutvertex or belongs to a nonleaf block. All other vertices are *external*; see Figure 7.

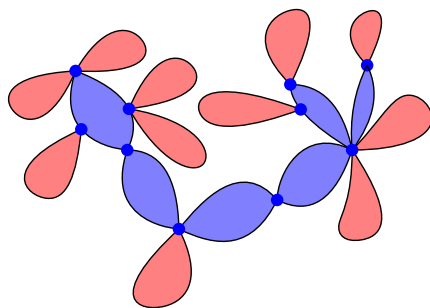


FIG. 7. Internal (blue) and external (red) vertices of F . (Figure in color online.)

LEMMA 13. Every component of F has at most $(5+2s)(d(s+1))^{5+2s}$ internal vertices.

Proof. Let X_{int} be the set of internal vertices of some component of F . Each $v \in X_{int}$ is in at most $s+1$ blocks of $G[X_{int}]$ by Lemma 12, and moreover, it has degree at most $d-1$ in each block (as each block has at most d vertices). Thus the maximum degree in $G[X_{int}]$ is at most $(d-1)(s+1) \leq d(s+1)$. As G is $(sP_1 + P_5)$ -free, $G[X_{int}]$ is $(sP_1 + P_5)$ -free. Hence, $G[X_{int}]$ is also P_{5+2s} -free and as $G[X_{int}]$ is connected, it has a diameter at most $5+2s-1$. Every graph with maximum degree at most $d(s+1)$ and a diameter at most $5+2s-1$ has at most

$$1 + d(s+1) + (d(s+1))^2 + \cdots + (d(s+1))^{5+2s-1} \leq (5+2s)(d(s+1))^{5+2s}$$

vertices.¹ This completes the proof of the lemma. \square

¹An astute reader might notice that this bound can actually be improved to the so-called Moore bound. However, as we do not try to optimize the constants, we kept bounds as simple as possible.

We say that a component F' of F is *big* if $|V(F')| \geq (ds+1) \cdot (5+2s)(d(s+1))^{5+2s}$. Otherwise F' is *small*.

LEMMA 14. *If a component of F is big, then it has a cutvertex belonging to at least s leaf blocks.*

Proof. Let X' be such that $G[X'] = F'$ is a big component of $F = G[X]$. Let X'_{int} and X'_{ext} be the sets of internal and external vertices of X' , respectively. Note that $X' = X'_{int} \cup X'_{ext}$ and $X'_{int} \cap X'_{ext} = \emptyset$. By Lemma 13 we have that $|X'_{int}| \leq (5+2s)(d(s+1))^{5+2s}$. Consequently,

$$\begin{aligned} |X'_{ext}| &= |V(F')| - |X'_{int}| \\ &\geq (ds+1) \cdot (5+2s)(d(s+1))^{5+2s} - |X'_{int}| \\ &\geq ds \cdot (5+2s)(d(s+1))^{5+2s}. \end{aligned}$$

As every block contains at most d vertices, the above implies that $G[X']$ has at least $s \cdot (5+2s)(d(s+1))^{5+2s} \geq s \cdot |X'_{int}|$ leaf blocks. Each leaf block contains exactly one internal vertex, so by the pigeonhole principle we conclude that there must be an internal vertex belonging to at least s leaf blocks. This completes the proof of the lemma. \square

4.3. The algorithm. We are now ready to present our polynomial-time algorithm for $(sP_1 + P_5)$ -free graphs. Let $s \geq 0$, and let G be the $(sP_1 + P_5)$ -free instance graph with weight function \mathbf{w} . Let \mathcal{C} be a finite class of biconnected graphs. Let d be the maximum number of vertices of a graph in \mathcal{C} . Recall that F is the optimum solution we are looking for.

The algorithm consist of three phases, in each of which we look for solutions of a specific type. Afterwards, the algorithm returns the maximum solution found during the whole execution.

Case 1. F has at most three big components. First, suppose that F has exactly three big components $F^1 = G[X^1]$, $F^2 = G[X^2]$, and $F^3 = G[X^3]$. See Figure 8. For $j \in [3]$, let X^j_{int} be the set of internal vertices of $G[X^j]$ (depicted in blue in Figure 8). By Lemma 13 we have that $|X^j_{int}| \leq (5+2s)(d(s+1))^{5+2s}$ and thus the set $X_{int} := \bigcup_{i \in [3]} X^i_{int}$ has at most $3(5+2s)(d(s+1))^{5+2s}$ vertices. We guess the vertices from X_{int} exhaustively; this results in $\mathcal{O}(n^{(5+2s)(d(s+1))^{5+2s}})$ branches. We discard the branches where $G[X_{int}]$ is not a \mathcal{C} -block graph with three components.

For each X_{int} that we have not discarded the only thing left to do is to find

- the small components of F (marked green in Figure 8) and
- the leaf blocks of F^1 , F^2 , and F^3 (marked red in Figure 8).

This task is very similar to the final case of the algorithm in section 3.2. Let \mathcal{X}_s be the family of those subsets of $V(G) \setminus N[X_{int}]$ of size smaller than $(ds+1) \cdot (5+2s)(d(s+1))^{5+2s}$ that induce \mathcal{C} -block graphs. The elements of \mathcal{X}_s are potential candidates for the vertex sets of small components of F . The family \mathcal{X}_s can be enumerated in time $\mathcal{O}(n^{(ds+1) \cdot (5+2s)(d(s+1))^{5+2s}})$.

Let \mathcal{X}_ℓ be the family of the sets $S \subseteq V(G) \setminus X_{int}$, satisfying the following properties:

- (a) there is a unique $x \in X_{int}$ with a neighbor in S ,
- (b) $G[S \cup \{x\}]$ is a graph from \mathcal{C} .

The elements S of \mathcal{X}_ℓ are potential candidates for the sets of external vertices in the leaf blocks of F^1 , F^2 , and F^3 , where x is the unique neighbor of the block $S \cup \{x\}$

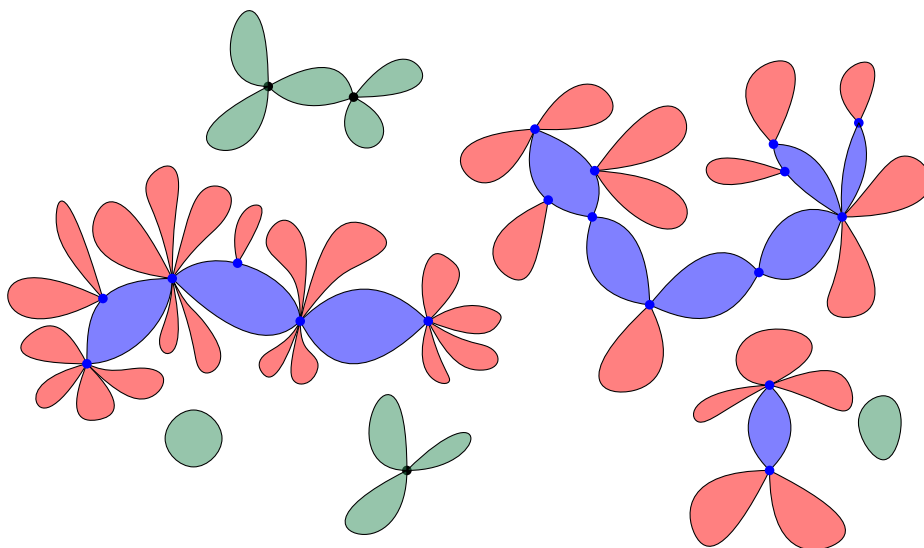


FIG. 8. Case 1 in the algorithm. Internal vertices of the three big components of F are marked blue, while the external ones are red. Small components are marked green. (Figure in color online.)

in $\text{BCF}(F)$. As each block has at most d vertices, each set from \mathcal{X}_ℓ has at most $d - 1$ vertices. Hence, the family \mathcal{X}_ℓ can be enumerated in time $\mathcal{O}(n^{d-1})$.

We now define $\mathcal{X} := \mathcal{X}_S \cup \mathcal{X}_\ell$ and have reduced to MAX INDEPENDENT SET for $(sP_1 + P_5)$ -free graphs. Namely, we build in polynomial time the induced subgraph $G^\circ[\mathcal{X}]$ of G° and the task is to find a maximum independent set in $G^\circ[\mathcal{X}]$. As $G^\circ[\mathcal{X}]$ is $(sP_1 + P_5)$ -free by Theorem 4, we can use the polynomial-time algorithm from Lemma 9 for doing this. Afterwards, we use the solution found, together with \mathcal{X} , to construct a forest F for G . Out of all the forests found in this way, we remember one with maximum weight.

The algorithm also considers the three subcases where F has zero, one, or two big components along the same lines as above but with some straightforward adjustments. In the end it returns a maximum-weight solution amongst the four solutions found. The total running time of Case 1 is polynomial, as there are $\mathcal{O}(n^{(5+2s)(d(s+1))^{5+2s}})$ branches and each of them is processed in time $n^{\mathcal{O}(d-1)}$, i.e., polynomial in n .

Case 2. F has at least four big components. Let X^1, X^2, X^3, X^4 be the vertex sets of pairwise distinct big components of $F = G[X]$. For each $j \in [4]$, there is $x_j \in X^j$ that belongs to at least s leaf blocks of F by Lemma 14. Choose s leaf blocks b_1^j, \dots, b_s^j containing x_j and, let $L^j := (\bigcup_{i=1}^s V(b_i^j)) \setminus \{x_j\}$. Let $L := \bigcup_{j \in [4]} L^j$, and let $G' := G - (N[L] \setminus \{x_1, x_2, x_3, x_4\})$.

Now consider any $X' \subseteq V(G')$, such that $G'[X']$ is a \mathcal{C} -block graph. We observe that $G[X' \cup L]$ is also a \mathcal{C} -block graph.

Due to the above observation we can proceed as follows. We will guess x_1, x_2, x_3, x_4 , and L exhaustively. Note that in the intended solution $\{x_1, x_2, x_3, x_4\} \cup L$ should be a \mathcal{C} -block graph whose block-cut forest is a disjoint union of four stars with x_1, x_2, x_3, x_4 as centers. If this is not the case for some guess, we discard the branch. As $|L| \leq 4ds$, the number of branches is $\mathcal{O}(n^{4+4ds})$. In each of those that we did not discard we will consider the graph $G' = G - (N[L] \setminus \{x_1, x_2, x_3, x_4\})$ and find a maximum-weight set $X' \subseteq V(G')$ such that $G'[X']$ is a \mathcal{C} -block graph. Then, by the above observation,

$X' \cup L$ induces a \mathcal{C} -block graph in G . We will return the maximum-weight solution among all found in the branches.

The only thing left is to show that MAX \mathcal{C} -BLOCK GRAPH can be solved in polynomial time for G' . For this, we make the following combinatorial claim.

LEMMA 15. *The graph G' is P_5 -free.*

Proof. For the sake of contradiction, suppose that G' contains an induced P_5 . As $\{x_1, x_2, x_3, x_4\}$ is an independent set, at least one vertex from this set, say x_1 , does not belong to this path. Thus there exists an induced P_5 in $G' - x_1$.

Note that $G[L^1]$ contains an independent set I of size s : it is sufficient to take one vertex from each block. Furthermore, no vertex from $N[I]$ is in $G' - x_1$. Consequently, the induced P_5 in $G' - x_1$, together with I , induces an $sP_1 + P_5$ in G , a contradiction. \square

Due to Lemma 15, MAX \mathcal{C} -BLOCK GRAPH in G' can be solved in polynomial time by Lemma 11.²

The total running time of Case 2 is polynomial, as there are $\mathcal{O}(n^{4+4ds})$ branches and processing each branch takes polynomial time. This completes the proof of Theorem 2.

5. Hardness results for even cycle transversal on H -free graphs. In this section we prove that subject to a number of unsolved cases, the complexity of EVEN CYCLE TRANSVERSAL for H -free graphs coincides with the one for FEEDBACK VERTEX SET.

An *odd cycle factor* of a graph G is a set of odd cycles such that every vertex of G belongs to exactly one of them. The ODD CYCLE FACTOR problem, which asks if a graph has an odd cycle factor, is known to be NP-complete [20]. The *line graph* $L(G)$ of a graph $G = (V, E)$ has a vertex set E and an edge between two distinct vertices e and f if and only if e and f share an endvertex in G .

The proof of our next result for line graphs is somewhat similar to a proof for ODD CYCLE TRANSVERSAL of [8] but uses some different arguments as well.

THEOREM 16. *EVEN CYCLE TRANSVERSAL is NP-complete for line graphs.*

Proof. Let $G = (V, E)$ be an instance of ODD CYCLE FACTOR with n vertices and m edges. We claim that G has an odd cycle factor if and only if its line graph $L := L(G)$ has an even cycle transversal of size at most $m - n$; see Figure 9.

First, suppose G has an odd cycle factor. Then there is $E' \subseteq E$, such that $|E'| = n$ and $L[E']$ is a disjoint union of odd cycles. Hence, $S := E \setminus E'$ is an even cycle transversal of L of size $|E| - n = m - n$. Now suppose L has an even cycle transversal S with $|S| \leq m - n$. Let $E' := E \setminus S$. As $|E| = m$, we have $|E'| \geq n$.

We prove the following claim.

CLAIM B. *Every component of $L[E']$ is either an odd cycle or the line graph of a tree.*

Proof. Let D be a component of $L[E']$. If D has no cycle, then D is a path, as L is a line graph and thus is claw-free. Hence, D is the line graph of a path, and thus a tree.

²Both the bound on the treewidth of a \mathcal{C} -block graph and the formula (4.1) depend on d , and the dependence of these parameters (especially the CMSO₂ formula) in the work of Abrishami et al. [1] is quite involved. Nevertheless, as d is a constant, the running time of the algorithm in Lemma 11 is polynomial.

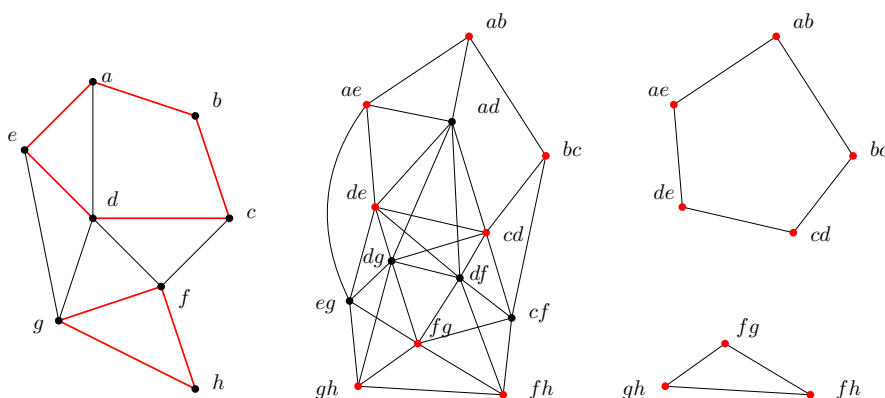


FIG. 9. Left: a graph G with an odd cycle factor. Middle: the graph $L = L(G)$ and the set E' (red). Black vertices form an even cycle factor. Right: the odd cactus $L[E']$.

So suppose D has a cycle C . Then C is odd and induced, as $L[E']$ is an odd cactus. If D has no vertices except for the ones of C , then D is an odd cycle and we are done. Suppose otherwise.

First, assume that C has at least five vertices. Since D has vertices outside C , there is a vertex of C with a neighbor outside C . Hence, D contains either an even cycle or an induced claw, both of which are not possible. So now suppose that C has at most four vertices. Then C is a triangle, as D has no even cycles. Since D is an induced subgraph of L , there exists a subgraph T of G such that $D = L(T)$. As D is a connected graph with at least four vertices, containing a triangle, T is a connected graph with at least four vertices.

We aim to show that T is a tree. For the sake of contradiction, suppose that T contains a cycle C_T . Then C_T must be a triangle, as otherwise D would contain an even cycle or an odd cycle with at least five vertices. Let a, b, c be the vertices of C_T . As T is connected and has at least four vertices, at least one of $\{a, b, c\}$, say a , must have a neighbor $d \notin \{b, c\}$. However, the edges $ad - ab - bc - ac$ form a C_4 in D , a contradiction with D being an odd cactus. So we conclude that T contains no cycles and thus T is a tree. \square

Each component of $L[E']$ that is an odd cycle corresponds to an odd cycle in G . By Claim B, each component D of $L[E']$ that is not an odd cycle is the line graph of some subtree T of G . So, if D has r vertices, then T has $r + 1$ vertices. Furthermore, the vertex sets of G corresponding to distinct components of $L[E']$ are pairwise disjoint. Suppose that $L[E']$ has $p \geq 0$ components that are not odd cycles. Let Q be the set of vertices incident to at least one edge of E' . Then $n = |V(G)| \geq |Q| = |E'| + p \geq n + p$. Hence, $p = 0$ and $|Q| = n$. So, the components of $L[E']$ correspond to an odd cycle factor of G . This completes the proof. \square

We make a straightforward observation similar to an observation for FEEDBACK VERTEX SET [8, 19], except that we must subdivide edges of a graph an even number of times.

THEOREM 17. *For every $p \geq 3$, EVEN CYCLE TRANSVERSAL is NP-complete for graphs of girth at least p .*

Proof. We reduce from EVEN CYCLE TRANSVERSAL for general graphs by noting the following. Namely, the size of a minimum even cycle transversal in G is equal to

the size of a minimum even cycle transversal in the graph G' obtained from G by subdividing every edge $2p$ times, and the girth of G' is at least p . \square

The next theorem is analogous to the one for FEEDBACK VERTEX SET; see also Table 1.

THEOREM 18. *Let H be a graph. Then EVEN CYCLE TRANSVERSAL for H -free graphs is polynomial-time solvable if $H \subseteq_i sP_3$ or $H \subseteq_i sP_1 + P_5$ for some $s \geq 0$, and it is NP-complete if H is not a linear forest.*

Proof. If $H \subseteq_i sP_3$ or $H \subseteq_i sP_1 + P_5$ for some integer $s \geq 0$, then we use Corollary 3. If H is not a linear forest, then it has a cycle or a claw. If H has a cycle, then we apply Theorem 17 for $p = |V(H)| + 1$. Otherwise, H has an induced claw and we apply Theorem 16. \square

6. Conclusions. We have proven that the MAX \mathcal{C} -BLOCK GRAPH problem is polynomial-time solvable on sP_3 -free graphs and $(sP_1 + P_5)$ -free graphs (for every $s \geq 1$). Hence, we have shown that for a large family of graphs \mathcal{F} , the MIN \mathcal{F} -TRANSVERSAL problem is polynomial-time solvable on these graph classes. The two best-known problems in this framework are FEEDBACK VERTEX SET and EVEN CYCLE TRANSVERSAL. Our results for FEEDBACK VERTEX SET extend all the known polynomial-time results for FEEDBACK VERTEX SET on H -free graphs, namely for sP_2 -free graphs [8], $(sP_1 + P_3)$ -free graphs [10], and P_5 -free graphs [1]. By proving some new hardness results we have also shown that in contrast to the situation for ODD CYCLE TRANSVERSAL, all other known complexity results for FEEDBACK VERTEX SET on H -free graphs hold for EVEN CYCLE TRANSVERSAL as well. Hence, so far both problems behave the same on special graph classes.

Due to the above, it would be interesting to prove polynomial equivalency of the two problems more generally. Table 1 still shows some missing cases for each of the three problems. In particular, we highlight the following borderline cases, namely the cases $H = P_2 + P_4$ and $H = P_6$ for FEEDBACK VERTEX SET and EVEN CYCLE TRANSVERSAL and the case $H = P_1 + P_4$ for ODD CYCLE TRANSVERSAL.

We recall that in section 4.1 we showed that the MAX \mathcal{C} -BLOCK GRAPH problem is a special case of finding a maximum-weight subset of vertices that induces a bounded-treewidth graph which satisfies a given CMSO₂ formula. The latter problem can be solved in *quasipolynomial time* for P_r -free graphs for any fixed r [11]. Thus we immediately obtain the following.

COROLLARY 19. *For every linear forest H and every finite class \mathcal{C} of biconnected graphs, MAX \mathcal{C} -BLOCK GRAPH can be solved in quasipolynomial time for H -free graphs.*

In particular, this implies quasipolynomial-time algorithms for FEEDBACK VERTEX SET and EVEN CYCLE TRANSVERSAL for H -free graphs if H is a linear forest, whereas ODD CYCLE TRANSVERSAL is NP-complete even for P_6 -free graphs [10]. Hence, a polynomial-time algorithm for FEEDBACK VERTEX SET and EVEN CYCLE TRANSVERSAL on P_6 -free graphs would show that these two problems, restricted to H -free graphs, differ in their complexity from ODD CYCLE TRANSVERSAL.

Acknowledgments. The first author thanks Carl Feghali for an inspiring initial discussion. The third author thanks Marcin Pilipczuk for some fruitful discussions including an alternative polynomial-time algorithm for FEEDBACK VERTEX SET on $(P_1 + P_5)$ -free graphs.

REFERENCES

- [1] T. ABRISHAMI, M. CHUDNOVSKY, M. PILIPCZUK, P. RZAŻEWSKI, AND P. SEYMOUR, *Induced subgraphs of bounded treewidth and the container method*, in Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA), SIAM, Philadelphia, 2021, pp. 1948–1964, <https://doi.org/10.1137/1.9781611976465.116>.
- [2] Y. AOIKE, T. GIMA, T. HANAKA, M. KIYOMI, Y. KOBAYASHI, Y. KOBAYASHI, K. KURITA, AND Y. OTACHI, *An improved deterministic parameterized algorithm for cactus vertex deletion*, Theory Comput. Syst., 66 (2022), pp. 502–515.
- [3] B. BERGOUNOUX, É. BONNET, N. BRETTELL, AND O-J. KWON, *Close relatives of Feedback Vertex Set without single-exponential algorithms parameterized by treewidth*, in Proceedings of the 15th International Symposium on Parameterized and Exact Computation, LIPIcs. Leibniz Int. Proc. Inform. 180, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2020, 3.
- [4] É. BONNET, N. BRETTELL, O-J. KWON, AND D. MARX, *Parameterized vertex deletion problems for hereditary graph classes with a block property*, in Graph-Theoretic Concepts in Computer Science, WG 2016, Lecture Notes in Comput. Sci. 9941, Springer, Berlin, 2016, pp. 233–244.
- [5] A. BRANDSTÄDT AND C. T. HOÀNG, *On clique separators, nearly chordal graphs, and the maximum weight stable set problem*, Theoret. Comput. Sci., 389 (2007), pp. 295–306.
- [6] A. BRANDSTÄDT AND D. KRATSCHE, *On the restriction of some NP-complete graph problems to permutation graphs*, in Fundamentals of Computation Theory (Cottbus, 1985), Lecture Notes in Comput. Sci. 199, Springer, Berlin, 1985, pp. 53–62.
- [7] A. BRANDSTÄDT AND R. MOSCA, *Maximum weight independent set for l -claw-free graphs in polynomial time*, Discrete Appl. Math., 237 (2018), pp. 57–64.
- [8] N. CHIARELLI, T. R. HARTINGER, M. JOHNSON, M. MILANIĆ, AND D. PAULUSMA, *Minimum connected transversals in graphs: New hardness results and tractable cases using the price of connectivity*, Theoret. Comput. Sci., 705 (2018), pp. 75–83.
- [9] B. COURCELLE AND J. ENGELFRIET, *Graph Structure and Monadic Second-Order Logic - A Language-Theoretic Approach*, Encyclopedia Math. Appl. 138, Cambridge University Press, Cambridge, 2012.
- [10] K. K. DABROWSKI, C. FEGHALI, M. JOHNSON, G. PAESANI, D. PAULUSMA, AND P. RZAŻEWSKI, *On cycle transversals and their connected variants in the absence of a small linear forest*, Algorithmica, 82 (2020), pp. 2841–2866.
- [11] P. GARTLAND, D. LOKSHTANOV, M. PILIPCZUK, M. PILIPCZUK, AND P. RZAŻEWSKI, *Finding large induced sparse subgraphs in $C_{>t}$ -free graphs in quasipolynomial time*, in Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, ACM, New York, pp. 330–341.
- [12] A. GRZESIK, T. KLIMOSOVÁ, M. PILIPCZUK, AND M. PILIPCZUK, *Polynomial-time algorithm for maximum weight independent set on P_6 -free graphs*, in Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, 2019, pp. 1257–1271.
- [13] M. JOHNSON, G. PAESANI, AND D. PAULUSMA, *Connected vertex cover for $(sP_1 + P_5)$ -free graphs*, Algorithmica, 82 (2020), pp. 20–40.
- [14] S. KOLAY, D. LOKSHTANOV, F. PANOLAN, AND S. SAURABH, *Quick but odd growth of cacti*, Algorithmica, 79 (2017), pp. 271–290.
- [15] D. LOKSHANTOV, M. VATSHELLE, AND Y. VILLANGER, *Independent set in P_5 -free graphs in polynomial time*, in Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, ACM, New York, 2014, pp. 570–581.
- [16] P. MISRA, V. RAMAN, M. S. RAMANUJAN, AND S. SAURABH, *Parameterized algorithms for even cycle transversal*, in Graph-theoretic Concepts in Computer Science, Lecture Notes in Comput. Sci. 7551, Springer, Heidelberg, pp. 172–183.
- [17] A. MUNARO, *On line graphs of subcubic triangle-free graphs*, Discrete Math., 340 (2017), pp. 1210–1226.
- [18] G. PAESANI, D. PAULUSMA, AND P. RZAŻEWSKI, *Feedback vertex set and even cycle transversal for H -free graphs: Finding large block graphs*, in Proceedings of the 46th International Symposium on Mathematical Foundations of Computer Science, LIPIcs. Leibniz Int. Proc. Inform. 202, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021, 82.
- [19] S. POLJAK, *A note on stable sets and colorings of graphs*, Comment. Math. Univ. Carolinae, 15 (1974), pp. 307–309.
- [20] O. VORNBERGER, *Komplexität von Wegeproblemen in Graphen*. Reihe Theoretische Informatik 5, 1979.