



Properties of Graphs Specified by a Regular Language

Volker Diekert¹, Henning Fernau^{2(✉)}, and Petra Wolf²

¹ Formal Methods in Informatics, Universität Stuttgart, Stuttgart, Germany
diekert@fmi.uni-stuttgart.de

² Universität Trier, FB 4 – Informatikwissenschaften, Trier, Germany
{fernau,wolfp}@informatik.uni-trier.de

Abstract. Traditionally, graph algorithms get a single graph as input, and then they should decide if this graph satisfies a certain property Φ . What happens if this question is modified in a way that we get a possibly infinite family of graphs as an input, and the question is if there exists one graph satisfying Φ ? We approach this question by using formal languages for specifying families of graphs. In particular, we show that certain graph properties can be decided by studying the syntactic monoid of the specification language.

Keywords: Graph properties · Regular languages · Periodic semigroups

1 Introduction

When dealing with algorithms on graphs, a graph is often specified by its adjacency matrix, i.e., a graph comes with a linear order on the vertices, and there are no multiple edges. We follow these conventions in our paper. Moreover, we represent graphs by words from a regular set \mathbb{G} over the binary alphabet $\Sigma = \{a, b\}$. Given $w \in \mathbb{G}$, we denote by $\rho(w)$ the corresponding graph. Hence, every subset $L \subseteq \mathbb{G}$ defines a family of graphs $\rho(L)$. Although our results go beyond regular sets L , the focus and the motivation comes from a situation when L is regular. A typical question could be if some (or all) graphs in $\rho(L)$ satisfy a graph property Φ . For example: “are there some planar graphs in $\rho(L)$?” Solving this type of decision problems was the motivation to study *regular realizability problems* in [1, 13] and, independently, calling them *int_{Reg}-problems* (*intersection non-emptiness with regular languages*) in [6, 14, 15].

Typical graph properties ignore the linear vertex orders and the direction of edges. For example, consider the property that the number of vertices is even. The linear order helps describe this property in Monadic Second-Order logic, *MSO* for short. As we will see, we encounter only four different classes $\mathcal{C}_1 \subset \dots \subset \mathcal{C}_4$ of graphs $\rho(L)$.

P. Wolf—Research supported by DFG project FE 560/9-1.

© Springer Nature Switzerland AG 2021

N. Moreira and R. Reis (Eds.): DLT 2021, LNCS 12811, pp. 117–129, 2021.

https://doi.org/10.1007/978-3-030-81508-0_10

1. $\rho(L) \in \mathcal{C}_1$ if and only if the set $\rho(L)$ is finite.
2. $\rho(L) \in \mathcal{C}_2$ implies that $\rho(L)$ has bounded tree-width.
3. $\rho(L) \in \mathcal{C}_3$ implies that every connected finite bipartite graph appears as a connected component of some $G \in \rho(L)$.
4. $\rho(L) \in \mathcal{C}_4$ implies that every connected finite graph appears as a connected component of some $G \in \rho(L)$.

Moreover, if L is regular, then we can compute the smallest ℓ such that $\rho(L) \in \mathcal{C}_\ell$. We use a straightforward encoding of vertices and edges: the i -th vertex u_i of a graph is encoded by $ab^i a$ and the edge (u_i, u_j) is encoded by $ab^i a a b^j a$. Since the syntactic monoid of a regular language is finite, we find some $t, p \in \mathbb{N}$ with $p \geq 1$, *threshold* and *period*, such that for every $n \in \mathbb{N}$ there is some $c \leq t + p - 1$ with $b^c \equiv_L b^n$ where \equiv_L denotes the *syntactic equivalence*. The threshold t tells us that $b^c \equiv_L b^n$ implies $n = c$ for all $0 \leq c < t$ and $b^n \equiv_L b^{n+p} \iff n \geq t$. This is the key observation when proving that we have no more than these four classes above. If $L \subseteq \mathbb{G}$ is not regular, then the syntactic monoid M_L is infinite. We find interesting examples where M_L satisfies the *Burnside condition* that all cyclic submonoids of M_L are finite. If so, then there exist $t, p \in \mathbb{N}$ with $p \geq 1$ such that the syntactic properties stated above hold for the powers of the letter b . In this case, we say that L satisfies the (b, t, p) -*torsion property*. This is a strong restriction, as Theorem 1 shows that for every subset $L \subseteq \mathbb{G}$ satisfying the (b, t, p) -torsion property, there exists a regular set $R \subseteq \mathbb{G}$ such that $\rho(L) = \rho(R)$. This is quite an amazing result. Its proof relies on the fact that $\rho(L)$ is determined once we know the Parikh-image $\pi_C(\text{rf}(L)) \subseteq \mathbb{N}^C$, where for $w \in \mathbb{G}$, the *reduced form* $\text{rf}(w)$ is obtained by replacing every b^n by b^c , where c is the smallest $0 \leq c \leq t + p - 1$ such that $b^c \equiv_L b^n$. Hence, for deciding whether some graph $G \in \rho(L)$ satisfies a property, we can assume that L is regular. We are interested in decidable properties Φ , only. Thus, we assume that the set $\{G \text{ is a finite graph} \mid G \models \Phi\}$ is decidable.

First consider that $\rho(L)$ is finite. Then, we can compute all graphs in $\rho(L)$ and we can output all $G \in \rho(L)$ satisfying Φ . Finiteness of $\rho(L)$ is actually quite interesting and important. It is a case where a representation of L by a DFA or a regular expression can be used for data compression. The minimal size of a regular expression (or the size of a DFA) for L is never worse than listing all graphs in $\rho(L)$, but it might be exponentially better. For a concrete and illustrative case, we refer to Example 2 for a succinct representation of all so-called *crowns with at most n cusps*. The compression rate becomes even better if we use a context-free grammar which produces a finite set L of words in Σ^* , only.

The second class \mathcal{C}_2 implies that $\rho(L)$ has bounded tree-width. In this case, by [2, 3, 11] we know that given any property Φ which is expressible in MSO, it is decidable whether there is a graph in $\rho(L)$ satisfying Φ . For languages $L \subseteq \mathbb{G}$ satisfying the (b, t, p) -torsion property, we understand when $\rho(L)$ has finite tree-width. Hence, we have Theorem 2: The satisfiability problem for MSO-sentences is decidable for language in the second class. For the other two classes, the picture changes drastically: the First-Order theory (*FO* for short) becomes undecidable [12]. Conversely, we are not aware of any “natural” graph property Φ where the satisfiability problem for Φ is not trivial for \mathcal{C}_3 and \mathcal{C}_4 .

2 Notation and Preliminaries

We let $\mathbb{N} = \{0, 1, 2, \dots\}$ be the set of natural numbers and $\mathbb{N}_\infty = \mathbb{N} \cup \{\infty\}$. Throughout, if S is a set, then we identify a singleton set $\{x\} \subseteq S$ with the element $x \in S$. The power set of S is identified with 2^S (via characteristic functions). If $E \subseteq X \times Y$ is a relation, then E^{-1} denotes its inverse relation $E^{-1} = \{(y, x) \in Y \times X \mid (x, y) \in E\}$. By id_X , we mean the identity relation $\text{id}_X = \{(x, x) \in X \times X \mid x \in X\}$. Recall that Y^X denotes the set of mappings from a set X to a set Y . If $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ are mappings, then $gf : X \rightarrow Z$ denotes the mapping defined by $gf(x) = g(f(x))$. If convenient, we abbreviate $f(\{x\})$ as $f[x]$. Throughout, Γ denotes a finite alphabet and we fix $\Sigma = \{a, b\}$ with $a \neq b$. Each alphabet is equipped with a linear order on its letters. For Σ , we let $a < b$. The linear order on Γ induces the *short-lex linear order* \leq_{slex} on Γ^* . That is, for $u, v \in \Gamma^*$, we let $u \leq_{\text{slex}} v$ if either $|u| < |v|$ or $|u| = |v|$, $u = pcu'$, and $v = pdv'$ where $c, d \in \Gamma$ with $c < d$. Here, $|u|$ denotes the length of u . Similarly, $|u|_a$ counts the number of occurrences of letter a in u . A language $L \subseteq \Gamma^*$ is a *code* if $c_1 \cdots c_m = d_1 \cdots d_n \in \Gamma^*$ with $c_i, d_j \in L$ implies $m = n$ and $c_i = d_i$ for all $1 \leq i \leq m$. If M is a monoid, then $u \leq v$ means $v \in MuM$, i.e., u is a *factor* of v . This notation is used for the monoids Γ^* and \mathbb{N}^Γ . Elements in \mathbb{N}^Γ are called *vectors*. For $w \in \Gamma^*$, \overleftarrow{w} is the *reversal* of w .

Every subset $R \subseteq \Gamma^*$ has a *syntactic monoid* $M = M_R$, see, e.g., [5]. The elements of M_R are the congruence classes $[u] = \{v \in \Sigma^* \mid v \equiv_R u\}$ w.r.t.the *syntactic congruence* \equiv_R . If R is regular, then M_R is finite. Monoids with a single generator are called *cyclic*. Every finite cyclic monoid M is defined by two numbers $t, p \in \mathbb{N}$ with $p \geq 1$ such that M is isomorphic to the quotient monoid $C_{t,p}$ of $(\mathbb{N}, +, 0)$ with the defining relation $t = t + p$. Hence, the carrier set of $C_{t,p}$ equals $\{0, 1, \dots, t + p - 1\}$. If $t = 0$ and $p = 1$, then $C_{t,p}$ is trivial.

Parikh-Images. If $v, w \in \Gamma^*$, then $|w|_v$ denotes the number how often v appears as a factor in w , i.e., $|w|_v = |\{u \in \Gamma^* \mid \exists s : uvs = w\}|$. If $V \subseteq \Gamma^*$, then the *Parikh-mapping w.r.t.V* is defined by $\pi_V : \Gamma^* \rightarrow \mathbb{N}^V$, mapping a word w to its *Parikh-vector* $(|w|_v)_{v \in V} \in \mathbb{N}^V$. The classical case is $V = \Gamma$; then the Parikh-vector becomes $(|w|_a)_{a \in \Gamma}$ and the Parikh-mapping is the canonical homomorphism from the free monoid Γ^* to the free commutative monoid \mathbb{N}^Γ .

A subset $S \subseteq \mathbb{N}^\Gamma$ is called *positively downward-closed* if, for all $v \in S$, (a) $v(z) \geq 1$ for all $z \in \Gamma$, and (b) $u \leq v$ and $u(z) \geq 1$ for all $z \in \Gamma$ imply $u \in S$. The complement of a positively downward-closed set $S \subseteq \mathbb{N}^\Gamma$ is *upward-closed*, i.e., $u \geq v$ and $v \in S$ imply $u \in S$. An upward-closed set S is determined by its set M_S of minimal elements. By Dickson's Lemma, for every upward-closed subset $S \subseteq \mathbb{N}^\Gamma$, M_S is finite. Hence, every upward-closed subset is semi-linear. As the set of all semi-linear sets in \mathbb{N}^Γ is closed under Boolean operations, every positively downward-closed set $S \subseteq \mathbb{N}^\Gamma$ is also semi-linear, a key for Theorem 1.

Retractions and retracts. Let $\rho : X \rightarrow Y$ and $\gamma : Y \rightarrow X$ be mappings between sets. Then, ρ is called a *retraction* and Y is called a *retract* of X with *section* γ if $\rho(\gamma(y)) = y$ for all $y \in Y$. Then, $\rho^{-1}(y)$ is the *fiber* of $y \in Y$. If $\rho : X \rightarrow Y$ is a homomorphism of groups X and Y and $H = \ker(\rho)$ is the kernel, then ρ is a retraction if and only if X is a semi-direct product of H by Y .

3 Graphs

All graphs are assumed to be (at most) countable, given as a pair $G = (V, E)$ where $E \subseteq V \times V$. An *undirected graph* is the special case where $E = E^{-1}$. If $G = (V, E)$ is a directed graph, then G also defines the *undirected graph* $(V, E \cup E^{-1})$; and it defines the *undirected graph without self-loops* $(V, (E \cup E^{-1}) \setminus \text{id}_V)$. A graph without isolated vertices is called an *edge-graph*; hence, specifying the edge set suffices. If $G' = (V', E')$ and $G = (V, E)$ are graphs such that $V' \subseteq V$ and $E' \subseteq E$, then G' is a *subgraph* of graph G and we denote this fact by $G' \leq G$. If $U \subseteq V$ is any subset, then $G[U] = (U, E \cap U \times U)$ denotes the *induced subgraph* of U in G . A *graph morphism* $\varphi : (V', E') \rightarrow (V, E)$ is given by a mapping $\varphi : V' \rightarrow V$ such that $(u, v) \in E'$ implies $(\varphi(u), \varphi(v)) \in E$. If (V', E') and (V, E) are undirected graphs without self-loops, then $\varphi : (V', E') \rightarrow (V, E)$ is a graph morphism when $(\varphi(u), \varphi(v)) \in E \cup \text{id}_V$. If φ is surjective on vertices and edges, i.e., $\varphi(V') = V$ and $\varphi(E') = E$, φ is a *projection*. We consider graphs up to isomorphism, only. Hence, writing $G = G'$ means that G and G' are isomorphic. A graph $F = (V, E)$ is a *retract* of a graph $F' = (V', E')$ if there are morphisms $\varphi : F' \rightarrow F$ and $\gamma : F \rightarrow F'$ where $\varphi\gamma$ is the identity on vertices and edges of (V, E) , i.e., F appears in F' as the induced subgraph $F'[\gamma(V)]$.

In our paper, every word $w \in \Sigma^*$ represents a directed finite graph $\rho(w) = (V(w), E(w))$ with a linear order on vertices as follows.

$$\begin{aligned} V(w) &= \{ ab^m a \in ab^+ a \mid ab^m a \leq w \} \\ E(w) &= \{ (ab^m a, ab^n a) \in ab^+ a \times ab^+ a \mid ab^m a a a b^n a \leq w \} \end{aligned}$$

The empty word represents the empty graph: there are no vertices and no edges. We extend ρ to 2^{Σ^*} by $\rho(L) = \{\rho(w) \mid w \in L\}$. Vice versa, if $G = (V, E)$ denotes a finite graph with a linear order on its vertices, then, for $1 \leq i, j \in \mathbb{N}$, the i -th vertex is represented by the factor $ab^i a$, and an edge from the i -th vertex to the j -th vertex is represented by the factor $ab^i a a a b^j a$. Thus, vertices are encoded by elements in the set $\mathbb{V} = \{ ab^i a \mid 1 \leq i \in \mathbb{N} \}$ and edges are encoded by elements in the set $\mathbb{E} = \{ ab^i a a a b^j a \mid 1 \leq i, j \in \mathbb{N} \}$. Note that $\mathbb{V} \cap \mathbb{E} = \emptyset$ and $\mathbb{V} \cup \mathbb{E}$ is an infinite regular code. Using these conventions, the regular set $\mathbb{G} = (\mathbb{V} \cup \mathbb{E})^*$ as well as its subset $\mathbb{E}^* \mathbb{V}^*$ represents all finite graphs. The set \mathbb{E}^* represents all edge-graphs, i.e., graphs without isolated vertices. Every nonempty finite graph has infinitely many representations. For example, there are uncountably many subsets $L \subseteq (aba)^+ \subseteq \mathbb{V}^+$ and each $\rho(L)$ represents nothing but the one-point graph without self-loop. In order to choose a unique (and minimal) representation for a finite graph $G = (V, E)$, we choose the minimal word $\gamma(G) = u_1 \cdots u_m v_1 \cdots v_n \in \mathbb{G}$ in the short-lex ordering on Σ^* such that $\rho\gamma(G) = G$, $u_k \in \mathbb{E}$ for $1 \leq k \leq m$ and $v_\ell \in \mathbb{V}$ for $1 \leq \ell \leq n$. Each u_k is of the form $ab^i a a a b^j a$ representing an edge and each v_ℓ is of the form $ab^i a$ representing an isolated vertex. We call $\gamma(G)$ the *short-lex representation* of G . Since $\gamma(G)$ is minimal w.r.t. \leq_{sllex} , we have $m = |E|$ and n is the number of isolated vertices. For a graph without isolated vertices, this means that it is given by its edge list. The set of all $\gamma\rho(\mathbb{G})$ is context-sensitive but not context-free. The *uvwxy*-Theorem does not hold for $\gamma\rho(\mathbb{G})$.

A subset $L \subseteq \mathbb{G}$ is viewed as a description of the set of graphs $\rho(L)$. The mapping $\rho : \rho^{-1}(\mathcal{L}) \rightarrow \mathcal{L}$ is a retraction in the sense of Sect. 2, since $\rho\gamma(G) = G$ for any finite graph G . The main results of the paper are: (1) for $L \subseteq \mathbb{G}$ satisfying the b -torsion property, there is a regular language $R \subseteq \mathbb{G}$ with $\rho(L) = \rho(R)$ and (2) for a context-free language satisfying the b -torsion property (e.g., any regular language) $R \subseteq \mathbb{G}$, we have an effective geometric description of $\rho(R)$. The description is obtained as follows. Using the fact that R is regular, in a first step, we find effectively a semi-linear description of $\rho(R)$. In a second step, we compute a finite set of finite graphs. Each member F in that finite family is a retraction of some possibly infinite graph F^∞ . The description of each $G \in \rho(L)$ is given by selecting some F and the cardinality of every fiber. The precise meaning will become clear later. As a consequence of the description, we are able to show various decidability results. The following example serves as an illustration.

Example 1. In the following, we let $R \subseteq \mathbb{E}^*$ and $t, p \in \mathbb{N}$ with $p \geq 1$, $t > 1$, such that $b^n \equiv_R b^{n+p}$ for all $n \geq t$. Since $t > 1$, we have $[b] = \{b\}$. By a *star*, we denote a graph (V, E) such that there exists a vertex $z \in V$ with the property $E = \{(z, s) \mid s \in V \setminus \{z\}\}$. Thus, a star has a *center* z and the directed edges are the outgoing *rays* of the star.

Furthermore, we assume that $R \subseteq (abaaab^n(b^p)^*a)^+$ for some fixed $n \geq t$. This implies $t \leq n < t + p$. Let $w \in R$. We have $w \in (abaaab^n(b^p)^*a)^m$ for $m = |w|_a/5$, i.e., $w = (abaaab^{d_1}a) \cdots (abaaab^{d_m}a)$ where $d_i = n + k_i p$ with $k_i \in \mathbb{N}$ for $1 \leq i \leq m$. The set $\{d_i \mid 1 \leq i \leq m\}$ can have any cardinality s in $\{1, \dots, m\}$. Therefore, $\rho(w)$ is a single star with at least one ray and at most m rays. If R is finite, then $\mathcal{F} = \rho(R)$ is an effective finite collection of stars with at least one ray and at most r rays where $r = \max\{|w|_a/5 \mid w \in R\}$.

Claim: \mathcal{F} is infinite if and only if there is some $M \geq |M_R|$ with $(abaaab^n a)^M \in R$. The claim holds if $\sup\{|w|_a/5 \mid w \in R\} < \infty$, as in this case \mathcal{F} is finite. Thus, let $\sup\{|w|_a/5 \mid w \in R\} = \infty$. Then, there is some $w \in R$ such that $abaaab^n a$ appears at least $|M_R|$ times as a factor. This implies that there is some $M \geq |M_R|$ such that $(abaaab^n a)^M \in R$. The claim follows. Moreover, if \mathcal{F} is infinite, then \mathcal{F} is the set of all finite stars with at least one ray.

One can show that $S = (abaaab^2 b^* a)^*(aba)$ is locally testable and therefore star-free. Hence, the set of all finite stars is specified by a star-free subset of Σ^* .

We study properties of graphs specified by languages $L \subseteq \mathbb{G}$. If L can be arbitrary, then we can specify uncountably many families of graphs. So, we cannot expect any general decidability results. Hence, we restrict our attention to subsets $L \subseteq \mathbb{G}$ where membership for $\rho(L)$ is decidable. In fact, membership for $\rho(L)$ might be decidable although membership for L is undecidable. By Corollary 1, the following definition yields a sufficient condition for decidability of $\rho(L)$.

Definition 1. Let $b \in \Gamma$ be a letter. A subset $L \subseteq \Gamma^*$ satisfies the (b, t, p) -torsion property if we have: $b^t \equiv_L b^{t+p}$. It satisfies the b -torsion property if there are $t, p \in \mathbb{N}$ with $p \geq 1$ such that L satisfies the (b, t, p) -torsion property.

Every regular language $R \subseteq \Gamma^*$ satisfies the b -torsion property because the syntactic monoid M_R is finite. The language $\{wa\overleftarrow{w} \mid w \in \{aba, ab^2a\}^*\}$ is not

regular, but it satisfies the b -torsion property for $t = 3$ and $p = 1$. The b -torsion property is exceptional if R is not regular: even deterministic linear context-free one-counter languages do not satisfy this property, in general. Consider $\{a^n b^n \mid n \in \mathbb{N}\}$. Clearly, $b^k \equiv_R b^m \iff k = m$.

Remark 1. Let $b \in \Gamma$ and $|\Gamma| = m$. Let $L \subseteq \Gamma^*$ with M_L as its syntactic monoid. If all cyclic submonoids of M_L are finite, L satisfies the b -torsion property. In the following, let $1 \leq p \in \mathbb{N}$. Recall that the quotient monoid $\Gamma^*/\{x^p = 1 \mid x \in \Gamma^*\}$ defines the *free Burnside group* $\mathcal{B}(m, p)$. It is a group because every x has the inverse element x^{p-1} as $p \geq 1$. For p large enough, Adjan showed in the 1970s that $\mathcal{B}(2, p)$ is infinite, answering a question of Burnside from 1902. A group is called p -periodic if it is the homomorphic image of some $\mathcal{B}(m, p)$.

Let $\varphi : \Gamma^* \rightarrow G$ be a surjective homomorphism to a group G . Then, the *Word Problem* of G denotes the set $\text{WP}(G) = \{w \in \Gamma^* \mid \varphi(w) = 1\}$. It is a classical fact that the syntactic monoid of $\text{WP}(G)$ is the group G itself. Kharlampovich constructed in [7] a periodic group $B(2, p)$ where the Word Problem is undecidable. Since the $B(2, p)$ is periodic, the b -torsion property holds trivially. Therefore, as we will see, there exists a regular subset R such that $\rho(\text{WP}(B(2, p))) = \rho(R)$.

For the rest of the paper, if $L \subseteq \Sigma^*$ satisfies the b -torsion property, then the cyclic submonoid of M_L generated by the letter b is isomorphic to $C_{t,p}$. That is, we have $t, p \in \mathbb{N}$ with $p \geq 1$, where $t + p$ is minimal such that $\{[b^n] \mid n \in \mathbb{N}\} = \{[b^c] \mid 0 \leq c \leq t + p - 1\}$. Moreover, we assume that L is specified such that on input $n \in \mathbb{N}$, we can compute the value $0 \leq c \leq t + p - 1$ with $b^n \equiv_L b^c$. This assumption is satisfied if L is regular and specified, say, by some NFA. For $L \subseteq \mathbb{G}$, we have $[ab^c a] = a[b^c]a$ and $[ab^c aab^d a] = a[b^c]aaa[b^d]a$.

Definition 2. Let $L \subseteq \mathbb{G}$ satisfy the (b, t, p) -torsion property according to Definition 1. For every $[b^n]$, we define its reduced form by $\text{rf}[b^n] = b^c$ if $[b^c] = [b^n]$ and $0 \leq c \leq t + p - 1$. Given $w \in \mathbb{G}$, we define the reduced form $\text{rf}(w)$ by replacing every factor $ab^m a \leq w$ by $a \text{rf}[b^m]a$. The saturation \widehat{w} of w is defined by replacing every factor $ab^m a \leq w$ by the set $a[b^m]a$. Hence, $\text{rf}(w) \in \widehat{w} \subseteq \mathbb{G}$.

Remark 2. Let $L \subseteq \mathbb{G}$ satisfy the (b, t, p) -torsion property. By possibly decreasing t and/or p , we may assume that for every $1 \leq c \leq t + p - 1$, there is some $w \in L$ such that $ab^c a \leq \text{rf}(w)$. Moreover, we have $[b^c] = \{b^c\}$ if and only if $c < t$.

Lemma 1. Let $L \subseteq \mathbb{G}$ satisfy the (b, t, p) -torsion property. Then, for every $w \in \mathbb{G}$, $w \in L \iff \widehat{w} \subseteq L \iff \text{rf}(w) \in L$.

The (b, t, p) -torsion property is trivially satisfied if $L \subseteq \mathbb{G}$ is a finite set, an interesting case motivated by data compression. As mentioned in Sect. 1: if L is finite, then the minimal size of a regular expression for L is never worse than listing all graphs in $\rho(L)$, but it might be exponentially better. This type of data compression with formal language methods is also applied in practice [8, 9].

Example 2. Let $G = ([n], E)$ be a connected planar graph with vertices $1, \dots, n$. Then, for every subset $S \subseteq \{n + 1, \dots, 2n\}$, we define a graph G_S by $G_S =$

$([n] \cup S, E \cup \{(s, s - n) \mid s \in S\})$. The family $\mathcal{C}_n = \{G_S \mid S \subseteq \{n + 1, \dots, 2n\}\}$ might contain up to $2^{\Omega(n)}$ connected planar graphs, e.g., if G is a cycle of n nodes. If we embed G in the 2-dimensional sphere where the additional edges are spikes pointing out of the sphere, then G_S can be visualized as a discrete model of a 3-dimensional “crown with at most n cusps”. One can write down a $2n$ -fold concatenation of finite sets describing a finite set $L_n \subseteq \mathbb{G}$ with $\rho(L_n) = \mathcal{C}_n$. The size of the corresponding regular expression is $\mathcal{O}(n^2)$. This leads to a polynomial-size blueprint potentially producing a family of exponentially many “crowns”.

Definition 3. Let $L \subseteq \mathbb{G}$ satisfy the (b, t, p) -torsion property. We introduce two new finite and disjoint alphabets (depending on L)

$$\begin{aligned} A &= \{ \text{rf}(ab^m aaab^n a) \mid m, n \in \mathbb{N}, m, n \geq 1 \} \subseteq \mathbb{E}, \\ B &= \{ \text{rf}(ab^m a) \mid m \in \mathbb{N}, m \geq 1 \} \subseteq \mathbb{V}. \end{aligned}$$

Note that $A \subseteq BaB$. By C , we denote the union of A and B , which is also a finite alphabet with a linear order between letters given by the following definition:

$$x \leq_C y \iff xy \in AB \vee (xy \in (AA \cup BB) \wedge x \leq_{\text{slex}} y). \quad (1)$$

The linear order \leq_C on C defines a short-lex ordering on C^* . Actually, C is a code. Moreover, if $uxv \in C^+$ with $x \in A$ and $u, v \in \Sigma^*$, then $u, v \in C^*$. The analogue for $y \in B$ does not hold, in general. As C is a code, the inclusion $C \subseteq \Sigma^*$ yields an embedding $h_C : C^+ \rightarrow \Sigma^+$. If G is a finite graph, then the minimal element in $\mathbb{H} = h_C^{-1}(\rho^{-1}(G)) \cap A^*B^*$ w.r.t. the short-lex ordering for words in C^* is the same as the minimal element in $h_C(\mathbb{H})$ w.r.t. the ordering $a < b$. We assume henceforth that C only contains factors of words from L .

Lemma 2. Let L, C , and rf as in Definitions 2 and 3. Let $v \in C^*$ and $w \in L$ such that $\pi_C(v) \leq \pi_C(\text{rf}(w))$. If $\pi_C(v)(z) \geq 1$ for all $z \in C$, then we have $\rho(v) \in \rho(L)$.

Theorem 1. Let $L \subseteq \mathbb{G}$ be any language satisfying the b -torsion property. Then, there is a regular set $R \subseteq \mathbb{G}$ such that $\rho(L) = \rho(R)$.

Corollary 1. Let $L \subseteq \mathbb{G}$ satisfy the b -torsion property. Then, given a finite graph $G = (V_G, E_G)$ as an input, it is decidable whether $G \in \rho(L)$.

Corollary 2. Let $L \subseteq \mathbb{G}$ be context-free satisfying the (b, t, p) -torsion property. Then, we can effectively calculate a regular set $R \subseteq \mathbb{G}$ such that $\rho(R) = \rho(L)$.

Let $R \subseteq \mathbb{G}$ be regular. Then, it is well-known that there might be a much more concise representation by some context-free language $K \subseteq \mathbb{G}$ such that $\pi_C(K) = \pi_C(R)$ and hence $\rho(K) = \rho(R)$.

By Theorem 1, we know that regular languages suffice to describe all sets $\rho(L)$ where $L \subseteq \mathbb{G}$ satisfies the b -torsion property. Therefore, we restrict ourselves to regular languages. In the following, $R \subseteq \mathbb{G}$ denotes a regular language. Hence, we can calculate numbers $t \geq 0$ and $p \geq 1$ such that R satisfies the (b, t, p) -torsion

property. As R is regular, the set $L = h_C^{-1}(R) \cap A^*B^*$ is regular; its *Parikh-image* $\pi_C(L) \subseteq \mathbb{N}^C$ is effectively semi-linear. Thus, for some finite set J :

$$\pi_C(L) = \bigcup_{j \in J} (q_j + \sum_{i \in I_j} \mathbb{N}p_i), \tag{2}$$

where $q_j, p_i \in \mathbb{N}^C$ are vectors. Splitting $\pi_C(L)$ into more linear sets by making the index set J larger and the sets I_j smaller (if necessary), we can assume without restriction that for all $j \in J$ and $z \in C$ we have $\sum_{i \in I_j} p_i(z) \leq q_j(z)$. To see this, let $1 \in I_j$. Then, we have

$$q_j + \sum_{i \in I_j} \mathbb{N}p_i = (q_j + \sum_{i \in I_j \setminus \{1\}} \mathbb{N}p_i) \cup (q_j + p_1 + \sum_{i \in I_j} \mathbb{N}p_i).$$

Splitting L into even more but finitely many cases, we can assume without restriction (for simplifying the notation) that the set J is a singleton. Thus, $\pi_C(L) = q + \sum_{i \in I} \mathbb{N}p_i$ for some $q, p_i \in \mathbb{N}^C$ such that $\sum_{i \in I} p_i(z) \leq q(z)$. By possibly reducing A, B, C , we may assume that $q(z) \geq 1$ for all $z \in C$ and $C = A \cup B$. In order to understand the set of graphs in $\rho(R)$, it suffices to understand the set of finite graphs defined by linear sets of the form $S = q + \sum_{i \in I} \mathbb{N}p_i \subseteq \mathbb{N}^C$, where $q(z) \geq 1$ for all $z \in C$ and $\sum_{i \in I} p_i \leq q$. For that purpose, we let $r = \sum_{i \in I} p_i \leq q$ and we define a function $\alpha : C \rightarrow \mathbb{N}_\infty$ as follows.

$$\alpha(z) = \begin{cases} q(z) & \text{if } r(z) = 0 \wedge \exists m \in \mathbb{N} : t \leq m \wedge ab^m a \leq z \\ \infty & \text{if } r(z) \geq 1 \wedge \exists m \in \mathbb{N} : t \leq m \wedge ab^m a \leq z \\ 1 & \text{otherwise. That is: } \forall m \in \mathbb{N} : ab^m a \leq z \implies m < t. \end{cases} \tag{3}$$

For all $z \in C$, let $L_z \subseteq \Sigma^*$. Then, we write $\prod_{z \in C} L_z = L_{z_1} \cdots L_{z_{|C|}}$, where $z_i \leq z_j$ for all $i \leq j$ according to the linear order defined in Eq. (1). Observe that $\prod_{z \in C} L_z$ is regular if all L_z are regular. With this notation, we define:

$$R_\alpha = \prod_{z \in C} z^{\alpha(z)} \quad \text{and} \quad L_\alpha = \prod_{z \in C} [z]^{\alpha(z)} \tag{4}$$

Notice that L^∞ is just another notation for L^+ if L is any set of words.

Lemma 3. *The sets R_α, L_α of Eq. (4) are regular with $R_\alpha \subseteq L_\alpha, \rho(L_\alpha) = \rho(R)$.*

Now, we define for α a finite family of finite graphs \mathcal{F}_α and then, for each $F \in \mathcal{F}_\alpha$, we define a possibly infinite graph F^∞ , using the notion of marked graphs.

Definition 4. *For $z \in C$, let $\alpha'(z) = \alpha(z)$ if $\alpha(z) < \infty$ and $\alpha'(z) = 1$, otherwise. We let $R'_\alpha = \prod_{z \in C} z^{\alpha'(z)}$, and we define $\mathcal{F}_\alpha = \rho(R'_\alpha)$.*

Since R'_α is a finite set of words, \mathcal{F}_α is a finite set of finite graphs. Now, we define the crucial notion of a marked graph, with some vertices and edges marked.

Definition 5. *A marked graph is a tuple $F = (V_F, E_F, \mu)$, where (V_F, E_F) is a finite graph and $\mu \subseteq V_F \cup E_F$ denotes the set of marked vertices and edges. Isolated vertices may appear, but if an isolated vertex is marked, then there is exactly*

one isolated vertex. We also require that whenever an edge (u, v) is marked, then at least one of its endpoints is marked, too. A marked edge-graph is a marked graph without isolated vertices.

In the following, each graph $(V_F, E_F) \in \mathcal{F}_\alpha$ as in Definition 4 defines a marked graph $F = (V_F, E_F, \mu)$ as follows, where μ denotes a marking (as in Definition 5) that we call the *canonical marking*. We begin by marking those vertices and edges $z \in C$ where $\alpha(z) = \infty$. In particular, if $z \in C$ is marked, then $[z] \neq \{z\}$ and $[z]$ is an infinite set. In the second step, we mark also all vertices u which satisfy $[u] \neq \{u\}$ and which appear as an endpoint in some marked edge. Thereafter, every marked edge contains at least one marked endpoint. In the third step, if an isolated vertex is marked, then remove all isolated marked vertices $y \in B$ except one isolated vertex which is marked. In particular, after that procedure, if a marked isolated vertex y appears, then $\alpha(y) = \infty$.

Now, we switch to a more abstract viewpoint. We let \mathcal{F} be any finite family of marked graphs. For each $F = (V_F, E_F, \mu) \in \mathcal{F}$, we define a possibly infinite graph F^∞ where (V_F, E_F) appears as an induced subgraph, and we define a family of finite graphs \mathcal{G}_F . We consider finitely many \mathcal{F}_α , and then we study $\bigcup \{\mathcal{G}_F \mid (V_F, E_F) \in \mathcal{F}_\alpha\}$, where $F = (V_F, E_F, \mu)$ is the marked graph obtained by the canonical marking procedure above (which might have removed isolated marked vertices). For understanding $\rho(R)$, we need to describe sets \mathcal{G}_F for marked graphs $F = (V_F, E_F, \mu)$. This requires to define F^∞ as follows.

Definition 6. Let $F = (V_F, E_F, \mu)$ be a marked graph as in Definition 5. Then, the graph $F^\infty = (V_F^\infty, E_F^\infty)$ is defined as follows.

$$V_F^\infty = V_F \times \{0\} \cup \bigcup_{u \in V_F} \{(u, k) \mid u \text{ is marked} \wedge k \in \mathbb{N}\},$$

$$E_F^\infty = E_F \times \{0\} \cup \{(u, k), (v, \ell) \in V_F^\infty \times V_F^\infty \mid (u, v) \in E_F \wedge (u, v) \text{ is marked}\},$$

with $E_F \times \{0\} = \{(u, 0), (v, 0) \mid (u, v) \in E_F\}$. The family \mathcal{G}_F is the set of finite subgraphs of F^∞ containing $(V_F \times \{0\}, E_F \times \{0\})$ as an induced subgraph.

Observe that $F^\infty = F$ if and only if there is no marking, i.e., if $\mu = \emptyset$. We embed F into F^∞ by a graph morphism γ which maps each vertex $u \in V_F$ to the pair $\gamma(u) = (u, 0) \in V_F^\infty$. The projection onto the first component $\varphi(u, k) = (u)$ yields a retraction for every $G \in \mathcal{G}_F$ with retract F . If no isolated vertex is marked, then F^∞ has at most $|V_F|$ isolated vertices, but if there are marked vertices, then for every sufficiently large k , there is some graph in \mathcal{G}_F which has exactly k isolated vertices. In order to understand the graphs in \mathcal{G}_F (which is our goal), it is enough to understand the graphs G satisfying $F \leq G \leq F^\infty$. For $F = F^\infty$, we know everything about that set. Let us hence consider $F \neq F^\infty$. Proposition 1 shows that $\rho(R)$ is rather rich as soon as some $F \in \mathcal{F}_\alpha$ satisfies $F \neq F^\infty$. Confer the next result with the classification $\mathcal{C}_1 \subset \dots \subset \mathcal{C}_4$ of graphs from Sect. 1.

Proposition 1. *Let $F = (V_F, E_F, \mu)$ be any marked graph.*

1. *If F contains a marked edge (u, v) where v is marked, then every finite star with center $(u, 0)$ appears as an induced subgraph of some $G \in \mathcal{G}_F$.*
2. *Suppose we represent a bipartite graph as a triple (U, V, E) where $U \cap V = \emptyset$ and $E \subseteq U \times V$. Let H be any finite bipartite edge-graph. If F contains a marked edge (u, v) where u and v are marked, then a disjoint union of F and H appears in \mathcal{G}_F .*
3. *Let H be any finite graph. If F contains a marked self-loop (u, u) , then the disjoint union of F and H belongs to \mathcal{G}_F .*
4. *Let F be any marked graph such that one or two vertices are marked. Then, the following holds. A disjoint union of F and any non-bipartite graph appears in \mathcal{G}_F if and only if there is some marked self-loop in F .*

By Schützenberger’s classical theorem [10] characterizing star-freeness via finite and aperiodic syntactic monoids, this case distinction entails:

Corollary 3. *Let $L \subseteq \mathbb{G}$ be any language satisfying the b -torsion property. If there is a star-free language R such that $\rho(L) = \rho(R)$, then there is no $F \in \mathcal{F}_\alpha$ such that a disjoint union of F and a triangle appears in \mathcal{G}_F .*

4 Graph Properties

Throughout this section, F denotes a marked graph and \mathcal{G}_F denotes the family of graphs defined in Definition 6. A *graph property* is a **decidable** subset $\Phi \subseteq \mathbb{G}$. For a finite graph G , we write $G \models \Phi$ if the short-lex representation $\gamma(F)$ belongs to Φ . Given a word $w \in \mathbb{G}$, we can compute $\gamma\rho(w)$. Hence, we can assume $\rho^{-1}(\rho(\Phi)) = \Phi$. As $\rho(w)$ is realized as a graph with a natural linear order on the vertices, we have $ab^c a \leq ab^d a \iff c \leq d$. We consider properties of undirected finite graphs, only: if $u \in \mathbb{G}$ represents the graph $\rho(u) = (V, E)$, then $\rho(u) \models \Phi$ if and only if $(V, E \cup E^{-1}) \models \Phi$. We focus on the satisfiability problem $\text{Sat}(\mathcal{G}_F, \Phi)$:

- Input: A marked graph F .
- Question: “ $\exists G \in \mathcal{G}_F : G \models \Phi$?”

For various well-studied graph properties, $\text{Sat}(\mathcal{G}_F, \Phi)$ is decidable. For example, when Φ states that a graph is planar, or k -colorable, etc. This follows from:

Proposition 2. *Let either \mathcal{G}_F be finite or Φ be any graph property which is closed under taking induced subgraphs (or both). Then, $\text{Sat}(\mathcal{G}_F, \Phi)$ is decidable.*

In many cases, graph properties are expressible either in MSO or even in FO. MSO is a rich and versatile class to define graph properties¹. Since $w \in \mathbb{G}$ defines graphs with a linear order, we can express in MSO, for example, that the number of vertices is even. We use the following well-known results as black boxes. First, (Trakhtenbrot’s Theorem) [12]: given an FO-sentence Φ , it is undecidable whether there exists a graph (resp. bipartite graph) satisfying Φ . Second, given an MSO-sentence Φ and $k \in \mathbb{N}$, it is decidable whether there exists a graph of tree-width at most k satisfying Φ , see, e.g., [2, 3, 11].

¹ For our purposes, we allow quantification over both sets of vertices and sets of edges.

Theorem 2. *Let Φ be an MSO-sentence. Then, $\text{Sat}(\mathcal{G}_F, \Phi)$ is decidable for marked graphs $F = (V_F, E_F, \mu)$ if at most one endpoint of each edge is marked.*

Theorem 3. *Let Φ be an FO-sentence. Then, $\text{Sat}(\mathcal{G}_F, \Phi)$ is undecidable for marked graphs $F = (V_F, E_F, \mu)$ where both endpoints of some edge are marked.*

Some graph properties where the problem $\text{Sat}(\mathcal{G}_F, \Phi)$ is trivially decidable are covered by the next theorem, including the problem whether \mathcal{G}_F contains a non-planar graph, and various parametrized problems like: “Is there some $(V_G, E_G) \in \mathcal{G}_F$ with a clique bigger than $\sqrt{|V_G|}$?”.

Theorem 4. *Let F be any marked graph and Φ be a non-trivial graph property such that $G \models \Phi$ if and only if there is a connected component G' of G such that $G' \models \Phi$. Then, the answer to $\text{Sat}(\mathcal{G}_F, \Phi)$ is “Yes” in the following two cases. [3] (a) The property Φ is true for some bipartite edge-graph and there is some marked edge where both endpoints are marked. (b) There is some marked self-loop.*

Example 3 lists a few graph properties which are not covered by the results above, but nevertheless the satisfiability problem is decidable.

Example 3. Let $F = (V_F, E_F, \mu)$ denote a marked graph as input. Then, the following problems are decidable. Is there some $G \in \mathcal{G}_F$ (a) with a Hamiltonian cycle (See [4]), or (b) with a perfect matching, or (c) with a dominating set of size at most $\sqrt{|V_G|}$?

Perfect Matching. Let $V_F = \{x_1, \dots, x_k\}$ and suppose some $G = (V_G, E_G) \in \mathcal{G}_F$ has a perfect matching. We have $V_F \subseteq V_G$. Hence, all $x_i \in V_F$ are matched by vertices $y_i \in V_G$. The induced subgraph $G[V_F \cup \{y_1, \dots, y_k\}]$ has a perfect matching with at most $2|V_F|$ vertices. All such small $G \in \mathcal{G}_F$ can be enumerated.

Dominating Set. If F contains no marked edge, then decide if the property holds for (V_F, E_F) . Otherwise, there is a marked edge $(u, v) \in E_F$. Let $V_G = (V_F \times \{0\}) \cup (\{v\} \times \{1, \dots, |V_F|^2\})$ and $E_G = E_F \cup \{((u, 0), (v, i)) \mid 1 \leq i \leq |V_F|^2\}$. Then, $G = (V_G, E_G) \in \mathcal{G}_F$. Also, V_F is a sufficiently small dominating set of G .

With the results above, we have a meta-theorem for graph properties Φ with a decidable satisfiability problem, covering all cases where we have positive results.

Theorem 5. *Let $r : \mathbb{N} \rightarrow \mathbb{N}$ be a non-decreasing computable function and let Φ be a graph property such that, for each marked graph $F = (V_F, E_F, \mu)$, if some graph in \mathcal{G}_F satisfies Φ , then there is a graph $G = (V, E) \in \mathcal{G}_F$ such that $G \models \Phi$ and $|V| \leq r(|V_F|)$. Then, given as input a context-free language $L \subseteq \mathbb{G}$ satisfying the (b, t, p) -torsion property, $\text{Sat}(\rho(L), \Phi)$ is decidable.*

5 Conclusion and Open Problems

The starting point of our paper was the following idea: Decide a graph property Φ not for a single instance as in traditional algorithmic graph theory, but generalize

this question to a set of graphs specified by a regular language. We chose a natural representation of graphs by words over a binary alphabet Σ , but other choices would work equally well. Next, pick your favorite graph property Φ . For example, Φ says that the number of vertices is a prime number. The property does not look very regular, there is no way to express the property, say, in MSO. Still, given a context-free language $L \subseteq \Sigma^*$ which satisfies the b -torsion property and which encodes sets of graphs, we can answer the question if there exists a graph represented by L and which satisfies Φ . This is a consequence of Theorem 5 and Bertrand's postulate that for all $n \geq 1$, there is a prime between n and $2n$.

Various problems remain open. For instance, given a graph property Φ , we can define $\mathcal{G}(\Phi) = \{G \text{ is a finite graph} \mid G \models \Phi\}$. Suppose that $\rho^{-1}(\mathcal{G}(\Phi))$ is regular. Given a regular language $R \subseteq \Sigma^*$, can we decide whether $\mathcal{G}(\Phi) \subseteq \rho(R)$? What about the equality $\mathcal{G}(\Phi) = \rho(R)$? We can ask the same two questions if R is context-free. Future research should address complexity issues. For example, given a typical NP-complete graph property Φ and ask how complex it is to decide the satisfiability for \mathcal{G}_F if the input is a marked graph F .

Note: Missing proofs can be found in [4].

References

1. Anderson, T., Loftus, J., Rampersad, N., Santean, N., Shallit, J.: Detecting palindromes, patterns and borders in regular languages. *Inf. Comp.* **207**, 1096–1118 (2009)
2. Courcelle, B.: The expression of graph properties and graph transformations in Monadic Second-Order Logic. In: Rozenberg, G. (ed.) *Handbook of Graph Grammars and Computing by Graph Transformations*, Vol. 1: Foundations, pp. 313–400. World Scientific (1997)
3. Courcelle, B., Engelfriet, J.: *Graph Structure and Monadic Second-Order Logic - A Language-Theoretic Approach*, *Encyclopedia of Mathematics and Its Applications*, vol. 138. Cambridge University Press, Cambridge (2012)
4. Diekert, V., Fernau, H., Wolf, P.: Properties of graphs specified by a regular language. arXiv eprints [arXiv:2105.00436](https://arxiv.org/abs/2105.00436) (2021)
5. Eilenberg, S.: *Automata, Languages, and Machines*, vol. A. Academic Press, Cambridge (1974)
6. Güler, D., Krebs, A., Lange, K.-J., Wolf, P.: Deciding regular intersection emptiness of complete problems for PSPACE and the polynomial hierarchy. In: Klein, S.T., Martín-Vide, C., Shapira, D. (eds.) *LATA 2018*. LNCS, vol. 10792, pp. 156–168. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-77313-1_12
7. Kharlampovich, O.: The word problem for the burnside varieties. *J. Algebra* **173**, 613–621 (1995)
8. Larsson, N.J., Moffat, A.: Off-line dictionary-based compression. *Proc. IEEE* **88**, 1722–1732 (2000)
9. Lohrey, M., Maneth, S., Mennicke, R.: XML tree structure compression using RePair. *Inf. Syst.* **38**(8), 1150–1167 (2013)
10. Schützenberger, M.P.: On finite monoids having only trivial subgroups. *Inf. Control* **8**, 190–194 (1965)
11. Seese, D.: The structure of the models of decidable monadic theories of graphs. *Ann. Pure Appl. Logic* **53**, 169–195 (1991)

12. Trahtenbrot, B.A.: The impossibility of an algorithm for the decision problem for finite domains. In: *Doklady Akademii Nauk SSSR, New Series*, vol. 70, pp. 569–572 (1950)
13. Vyalyi, M.N., Rubtsov, A.A.: On regular realizability problems for context-free languages. *Probl. Inf. Transm.* **51**(4), 349–360 (2015). <https://doi.org/10.1134/S0032946015040043>
14. Wolf, P.: On the decidability of finding a positive ILP-instance in a regular set of ILP-instances. In: Hospodár, M., Jirásková, G., Konstantinidis, S. (eds.) *DCFS 2019*. LNCS, vol. 11612, pp. 272–284. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23247-4_21
15. Wolf, P.: From decidability to undecidability by considering regular sets of instances. In: Cordasco, G., et al. (eds.) *ICTCS Proceedings of the CEUR*, vol. 2756, pp. 33–46. CEUR-WS.org (2020)