Labelled Well Quasi Ordered Classes of Bounded Linear Clique-Width

Aliaume Lopez ✉ 🆔
University of Warsaw, Poland

Abstract

We construct an algorithm that inputs an MSO-interpretation from finite words to graphs, and decides if there exists a $k \in \mathbb{N}$ such that the class of graphs induced by the interpretation is not well-quasi-ordered by the induced subgraph relation when vertices are freely labelled using $\{1, \ldots, k\}$. In case no such k exists, we also prove that the class of graphs is not well-quasi-ordered by the induced subgraph relation when vertices are freely labelled using any well-quasi-ordered set of labels. As a byproduct of our analysis, we prove that for classes of bounded linear clique-width, a weak version of a conjecture by Pouzet holds.

1 Introduction

A *well-quasi-ordering* (*wqo*) $(X, \leq)$ is a quasi-order that contains no infinite *antichains* (i.e., sets of pairwise incomparable elements) nor infinite *decreasing sequences*. A cornerstone result of structural graph theory is the Graph Minor Theorem of Robertson and Seymour [25], which states that the class of all graphs is well-quasi-ordered by the graph minor relation. This result has profound implications in graph theory, and algorithmic consequences such as the polynomial-time solvability of whether a graph can be embedded into a given surface.

The class of all (finite, undirected) graphs is not well-quasi-ordered by the induced subgraph relation, as witnessed by the antichain composed of finite (chordless) cycles, and there has been considerable interest in characterizing classes $\mathcal{C}$ of finite graphs that are well-quasi-ordered by the induced subgraph relation. It was proven by Ding that all classes of bounded tree-depth are well-quasi-ordered for the induced subgraph relation [12, Theorem 2.2], and it is also the case for *letter graphs* [23]. In general, these theorems can be extended to all classes admitting "well-behaved" tree decompositions, such as classes of *bounded shrub-depth* [14, Corollary 3.9], or *m-partite cographs* [15, Theorem 5.5].

All the theorems mentioned above actually prove a stronger statement: the classes at hand are well-quasi-ordered in the presence of any well-quasi-ordered *labelling of the nodes*. Given a quasi-ordered set $(X, \leq)$ and a class $\mathcal{C}$ of graphs, we call $\mathsf{Label}_X(\mathcal{C})$ the class of graphs obtained by freely labelling vertices using elements of X, that is, graphs $G \in \mathcal{C}$ equipped with a labelling $\mathsf{lbl}_G \colon \mathsf{V}(G) \to X$. The notion of induced subgraph is naturally extended to the one of induced labelled subgraph, and we defer its formal definition to Section 2. For instance, it is shown in [12, Theorem 2.1] that a class of graphs of bounded tree-depth is well-quasi-ordered by the induced labelled subgraph relation, assuming that the labels are themselves well-quasi-ordered. We say that a class $\mathcal{C}$ of graphs is *k-well-quasi-ordered*
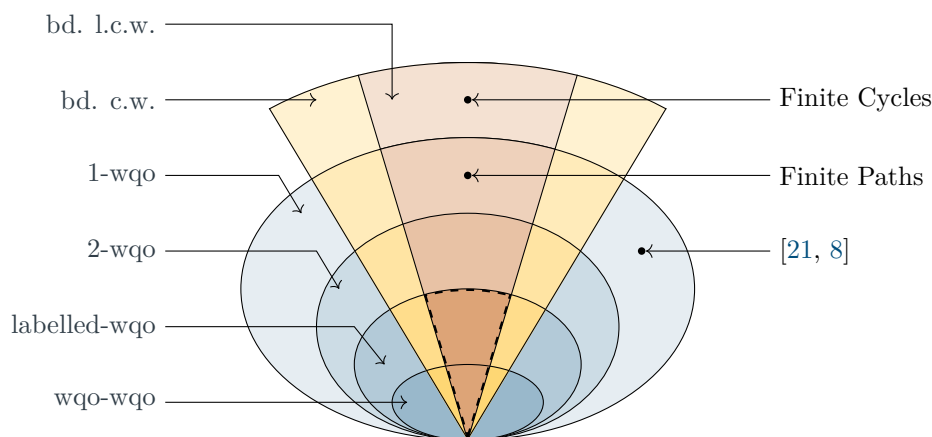
■ **Figure 1** Ellipses represent classes of graphs that are well-quasi-ordered, and the two cones represent respectively the classes of graphs of bounded clique-width and bounded linear clique-width. We chose to not represent the intermediate classes that are k-wqo for all $2 < k < +\infty$ for clarity. The dots represent specific classes of graphs witnessing the non-emptiness of some intersections. Pouzet's conjecture states that all ellipses below 2-wqo collapse for hereditary classes of graphs, and [9, Conjecture 5] states that the region obtained by removing the cone of bounded clique-width from the 2-wqo ellipse is empty. The dashed region represents the classes of graphs that we study in this paper, for which we prove the collapse to wqo-wqo in our Corollary 2.

(k-*wqo*) if $\mathsf{Label}_X(\mathcal{C})$ is well-quasi-ordered for all finite sets X of size at most k, and that it is *labelled-well-quasi-ordered* (*labelled-wqo*) if it is well-quasi-ordered for all finite sets X of labels. Finally, let us say that a class $\mathcal{C}$ is *wqo-well-quasi-ordered* (*wqo-wqo*) if it is well-quasi-ordered for all well-quasi-ordered sets X of labels. Note that wqo-wqo implies labelled-wqo, which implies k-wqo for all $k \in \mathbb{N}$, and that k-wqo implies ℓ-wqo for all $k \geq \ell$.

For hereditary classes of graphs, the notion of 2-well-quasi-ordered is of particular interest, as it implies the existence[1] of a fixed-parameter tractable algorithm to decide the membership of a graph in the class $\mathcal{C}$, such as described in [15, Corollary 5.8], and more generally that the class of graphs is described by finitely many forbidden induced subgraphs as shown by [9, Proposition 3] and [24]. In 2010, Daligault, Rao and Thomassé studied classes of graphs described by *relabel functions*, a generalization of m-partite cographs, and characterized those that are 2-well-quasi-ordered [9, Theorem 3]. In this paper, they also provided a positive answer to the *Pouzet conjecture* [24], which states that 2-wqo, labelled-wqo and wqo-wqo are equivalent for classes described by *relabel functions* [9, p. 4].[2] The classes of graphs studied in [9] are all of bounded clique-width, and it was conjectured that all 2-well-quasi-ordered classes of graphs are of bounded clique-width [9, Conjecture 5]. This started a series of works on the relationship between clique-width and well-quasi-orderings: it was shown that 1-wqo does not imply bounded clique-width [21, 8], that there are classes of graphs that are 1-wqo and described by finitely many forbidden (induced) subgraphs but are not 2-wqo [3], and that there exists (minimal) classes of unbounded clique width described by finitely many forbidden induced subgraphs [2]. We summarize the various properties of classes of graphs in Figure 1.

---

[1] The existence is non-constructive, and building a constructive algorithm requires non-trivial work, see for instance [1] in the case of letter graphs.
[2] The Pouzet conjecture states the same collapse for all classes of graphs.

Contributions. Our first contribution is a decision procedure, that allows us to recognize classes of graphs that are labelled-wqo, provided that they are described using an MSO-interpretation from (finite) words to graphs, the definition of which we defer to Section 2.

▶ **Theorem 1.** *Let I be an* MSO-*interpretation from words to graphs, and let $\mathcal{C}$ be its image. There exists a computable $k \in \mathbb{N}_{\geq 1}$ such that the following are equivalent:*
1. *$\mathcal{C}$ is k-well-quasi-ordered,*
2. *$\mathcal{C}$ is labelled-well-quasi-ordered,*
3. *$\mathcal{C}$ is wqo-well-quasi-ordered.*
Furthermore, these properties are decidable.

We also prove that all hereditary classes of graphs that are 2-wqo and of bounded linear clique-width can be described by such an MSO-interpretation, which justifies their use as an input to our decision procedure. As a corollary, we answer positively to part of Pouzet's conjecture in the case of classes of graphs having bounded linear clique-width, a subset of classes of bounded clique-width, that is incomparable with the classes studied in [9].

▶ **Corollary 2** (Collapse for Bounded Linear Clique-Width). *For all classes of graphs of bounded linear clique-width, the following are equivalent:*
1. *The class is labelled-well-quasi-ordered,*
2. *The class is wqo-well-quasi-ordered.*

The novelty of our approach is to leverage classical tools from automata theory (namely, the factorization forest theorem of Simon [26]), and avoid the use of complex *minimal bad sequence arguments* [22] that were used in [9]. We strongly believe that these results generalize to the case of classes of bounded clique-width. To justify this claim, we derive a self-contained proof of [9, Theorem 3] for classes having bounded linear clique-width. Our algebraic approach provides a new perspective on the statements of [9] by connecting it to the class of totally ordered monoids, that were already investigated in the context of regular languages for completely different reasons [18, 20].

Outline. We provide in Section 2 basic definitions and explain how to deduce Corollary 2 from Theorem 1. In Section 3, we prove our main Theorem 1. Then, in Section 4, we discuss the relationship between our characterization and the previous work of [9], proving in Theorem 24 that their characterization can be seen as a *coarser* version of ours in the case of bounded linear clique-width.

2 Preliminaries

Graphs. In this paper, *graphs* are all finite, simple and undirected unless otherwise explicitly stated. We denote by $\mathsf{V}(G)$ the set of vertices of a graph G, and by $\mathsf{E}(G)$ the set of edges. A map $h \colon G \to H$ between two graphs G and H is an *embedding* if it is injective and satisfies $\{h(u), h(v)\} \in \mathsf{E}(H)$ if and only if $\{u, v\} \in \mathsf{E}(G)$, for all $(u, v) \in \mathsf{V}(G)$. We write $G \subseteq_i H$ if there exists an embedding $h \colon G \to H$, which means that G is an *induced subgraph* of H, or equivalently that H is an *induced extension* of G. A class of graph is *hereditary* when it is closed under taking induced subgraphs. We denote by $\downarrow_{\subseteq_i} \mathcal{C}$ the *hereditary closure* of a class $\mathcal{C}$, obtained by considering all induced subgraphs of graphs in $\mathcal{C}$. The notion of induced subgraph is extended to labelled graphs as follows: a function $f \colon \mathsf{V}(G) \to \mathsf{V}(H)$ is an *embedding* of G into H whenever f is injective, $(u, v) \in \mathsf{E}(G)$ if and only if $(f(u), f(v)) \in \mathsf{E}(H)$, and $\mathsf{lbl}_G(u) \leq \mathsf{lbl}_H(f(u))$ for all $(u, v) \in \mathsf{V}(G)^2$.

▶ **Example 3.** The class of all finite paths is well-quasi-ordered (because a path of length n is an induced subgraph of all paths of length $m \geq n$) but not 2-well-quasi-ordered (which can be seen by coloring the endpoints of the paths). The class of all finite cycles is not well-quasi-ordered (since a cycle of length n is never an induced subgraph of a cycle of length $m \neq n$). The class of all cliques is wqo-wqo, which can be seen by representing colored cliques as finite multisets of colors, and leveraging the fact that these finite multisets are well-quasi-ordered when the colors themselves are well-quasi-ordered (see for instance [10]).

Automata Theory.   We assume that the reader is familiar with the basics of automata theory, in particular the notions of monoid morphisms, idempotents in monoids, monadic second-order (MSO) logic and first-order (FO) logic over finite words (see e.g. [27]). We use symbols Σ, Γ to denote finite *non-empty* alphabets, and letters u, v, w for finite words. The symbol ε is used to denote the empty word, and given a word $w \in \Sigma^*$, we write w_i for the letter at position $1 \leq i \leq |w|$, and $w_{\leq i}$ (resp. $w_{\geq i}$) for the prefix of w of length i (resp. the suffix of w starting at position i), note that these can be empty words. Similarly, we write $w_{(i,j)}$ for the factor of w between positions $i+1$ and $j-1$, which is empty if $j \leq i+1$.

Let us recall that a monoid M is a set endowed with a binary operation that is associative, has an identity element 1_M. Inside a monoid M, an element $e \in M$ is called *idempotent* if $e^2 = e$. A *morphism* of monoids $\mu \colon M \to N$ is a map that preserves the multiplication and the identity element, i.e., $\mu(1_M) = 1_N$ and $\mu(ab) = \mu(a)\mu(b)$ for all $a, b \in M$. The finite words Σ^* is a monoid under concatenation, with the empty word ε as neutral element.

MSO Interpretations.   A (simple) MSO-*interpretation* from finite words (Σ^*) to finite graphs is given by an MSO formula $\varphi_{\mathrm{edge}}(x, y)$ that defines the edge relation of a graph, a domain formula $\varphi_{\mathrm{dom}}(x)$ that defines the domain of the output graph, and a selection formula φ_Δ without free variables, used to restrict the domain of the interpretation.

Let $I := (\varphi_{\mathrm{edge}}, \varphi_{\mathrm{dom}}, \varphi_\Delta)$ be an MSO-interpretation from finite words to graphs. The image of a word $w \in \Sigma^*$ is the graph G with vertices $\{1 \leq i \leq |w| \mid w \models \varphi_{\mathrm{dom}}(i)\}$, and edges $\{i, j\}$ for each pair (i, j) such that $i < j$ and w satisfies $\varphi_{\mathrm{edge}}(i, j)$. The *image* of I is the collection of $I(w)$ where w ranges over the words in Σ^* that satisfy φ_Δ, we write this image $\mathsf{Im}(I)$. We provide examples of interpretations in Examples 4 and 5, the first example is a class that is wqo-wqo, while the second one is not even 2-wqo.

▶ **Example 4.** The class of all cliques can be obtained as the image of the following MSO-interpretation: $\varphi_{\mathrm{dom}}(x) := \top$, $\varphi_{\mathrm{edge}}(x, y) := x \neq y$, and $\varphi_\Delta := \top$.

▶ **Example 5.** The class of all finite paths can be obtained as the image of the following MSO-interpretation: $\varphi_{\mathrm{dom}}(x) := \top$, $\varphi_{\mathrm{edge}}(x, y) := (x \leq y) \wedge \forall z.(x \leq z \leq y) \implies z = x \vee z = y$, and $\varphi_\Delta := \top$.

A class of graphs has *bounded linear clique-width* if there exists a finite alphabet Σ, and an MSO-interpretation I such that $\mathcal{C} \subseteq \mathsf{Im}(I)$ [7, 6]. The notion of *bounded clique-width* is defined similarly, using *trees* instead of words as input of the interpretation. We remark that there are uncountably many classes of graphs that have bounded linear clique-width, and only countably many MSO-interpretations. This discrepancy vanishes when considering hereditary classes of graphs that are 2-well-quasi-ordered (Lemma 6), and we can use this to bridge the gap between the algorithm of Theorem 1 that considers MSO-interpretations and our Corollary 2, by leveraging the interaction between hereditary closure and labelling of a class (Lemma 7).

▶ **Lemma 6** (Sandwich Lemma). *Let $\mathcal{C}$ be a class of graphs that has bounded linear clique-width and is 2-well-quasi-ordered. Then, there exists an MSO-interpretation I such that $\mathcal{C} \subseteq \mathsf{Im}(I) \subseteq \downarrow_{\subseteq_i} \mathcal{C}$. In particular, $\mathcal{C} = \mathsf{Im}(I)$ when $\mathcal{C}$ is hereditary.*

▶ **Lemma 7** (Hereditary Closure Lemma). *Let $k \in \mathbb{N}_{\geq 1}$, and $\mathcal{C}$ be a class of graphs that is $(k+1)$-well-quasi-ordered. Then the hereditary closure of $\mathcal{C}$ is k-well-quasi-ordered.*

▶ **Corollary 2** (Collapse for Bounded Linear Clique-Width). *For all classes of graphs of bounded linear clique-width, the following are equivalent:*

1. *The class is labelled-well-quasi-ordered,*
2. *The class is wqo-well-quasi-ordered.*

Proof. Let $\mathcal{C}$ be a class of graphs that has bounded linear clique-width, and is labelled-well-quasi-ordered. By Lemma 6, there exists an MSO-interpretation I such that $\mathcal{C} \subseteq \mathsf{Im}(I) \subseteq \downarrow_{\subseteq_i} \mathcal{C}$. Since $\mathcal{C}$ is labelled-well-quasi-ordered, so is $\downarrow_{\subseteq_i} \mathcal{C}$ by Lemma 7. As a consequence, $\mathsf{Im}(I)$ is labelled-wqo (as a subset), and using Theorem 1, we conclude that $\mathsf{Im}(I)$ is wqo-wqo. Because $\mathcal{C} \subseteq \mathsf{Im}(I)$, it means that $\mathcal{C}$ is wqo-wqo too. ◀

3 MSO-Interpretations and Monoids

In this section, we will develop an automata theoretic approach to the problem of characterizing images of MSO-interpretations that are labelled-well-quasi-ordered by the induced subgraph relation. The key ingredient is to use the notion of *factorization forests* [26] to provide an alternative syntax of such classes of graphs where *words* (linear trees of unbounded depth) are replaced by *forests* (branching trees of bounded depth). Let us first justify that in the upcoming analysis, the domain formula φ_{dom} and the universe formula φ_Δ can be safely ignored (i.e., set to be constantly true). Beware that, contrary to Lemma 6, the following construction is *effective* which is essential for the algorithm of Theorem 1.

▶ **Lemma 8** (Simple Interpretations). *Let I be an MSO-interpretation. There exists an MSO-interpretation I' that does not use φ_Δ and φ_{dom}, and such that for all sets $(Q, \leq)$, $\mathsf{Label}_Q(\mathsf{Im}(I))$ is well-quasi-ordered if and only if $\mathsf{Label}_Q(\mathsf{Im}(I'))$ is.*
Furthermore, I' is effectively computable from I.

Monoid Interpretations. The equivalence between regular languages, finite automata, languages recognized by finite monoids and MSO formulas ([13]) provides the existence (and computability), for every formula $\varphi(x, y) \in \mathsf{MSO}$, of a finite monoid M, together with a surjective morphism $\mu \colon \Sigma^* \twoheadrightarrow M$, and a subset $P \subseteq M^3$, such that, for all words $w \in \Sigma^*$, for all couples $i < j$ of positions in w, we have:

$$w \models \varphi(i, j) \iff (\mu(w_{\leq i}), \mu(w_{(i,j)}), \mu(w_{\geq j})) \in P \quad .$$

As a consequence, given a finite alphabet Σ and an edge formula $\varphi_{\mathrm{edge}}(x, y)$ defining an MSO-interpretation, one can construct a *monoid interpretation*: it is described by a finite monoid M, a morphism $\mu \colon \Sigma^* \to M$, and a subset $P \subseteq M^3$ that we call an *edge selector*. To a word $w \in \Sigma^*$, it associates a graph G with vertices $\mathsf{V}(G)$ defined as the set of positions of w, and creates an edge $(i, j) \in \mathsf{E}(G)$ if $(\mu(w_{\leq i}), \mu(w_{(i,j)}), \mu(w_{\geq j})) \in P$. We provide hereafter two examples of monoid interpretations that will be used to illustrate our constructions.

▶ **Example 9.** Let $\Sigma := \{a\}$, $M := (\{0, 1, 2\}, +_2)$ where $a +_2 b := \min(2, a + b)$, $\mu(a) := 1$, and $P := \{(x, y, z) \in M^3 \mid y = 0\}$. Then, the image of the monoid interpretation I_{paths} is the class of finite paths.

▶ **Example 10.** Let $\Sigma := \{a\}$, $M := (\mathbb{Z}/2\mathbb{Z}, +)$, $\mu(a) := 1$, and $P := \{(x, y, z) \in M^3 \mid y = 0\}$. Then, the image of the monoid interpretation I_{bp} is included in the class of bipartite cliques.

Although Examples 9 and 10 share the same alphabet, similar morphisms and similar edge selectors, their behavior is drastically different. The image of I_{paths} is 1-wqo but not 2-wqo, while the image of I_{bp} is wqo-wqo.

3.1 Factorisation Forests

Let us now introduce one of the key ingredients of our proof, which is the notion of factorization forest of a finite word over a finite monoid M. Initially introduced by Simon in [26], they provide a way to efficiently compute the product of *factors* of a word in M^*, by grouping computations using a tree of bounded depth, as restated in Theorem 11. The main idea of this paper is to leverage the factorization forest theorem of Simon to provide a (bounded depth) *tree-decomposition* of the graphs in the image of a monoid interpretation. We will equip these tree-decompositions with a well-quasi-ordering (Lemma 12), which will induce a (new) well-quasi-ordered relation on the image of our monoid interpretation. At the level of graphs, this new relation will be a restriction of the induced subgraph relation, obtained by only considering embeddings that are *compatible with the tree-decompositions* of the graphs.

▶ **Theorem 11** ([26]). *Let Σ be a finite alphabet, and $\mu\colon \Sigma^* \to M$ be a morphism to a finite monoid. There exists[3] a computable depth $d \in \mathbb{N}$ such that every word $w \in \Sigma^*$ can be factorized as a tree of depth at most d built using:*

- Leaves*: $a \in \Sigma$, that evaluate to $\mu(a)$,*
- Binary nodes*: $t_1 \curlywedge t_2$, that evaluate to $a \cdot b$ if t_1 evaluates to a and t_2 evaluates to b,*
- Idempotent nodes*: $\curlywedge\!\!\!\!\!\curlywedge(t_1, \ldots, t_n)$, if $n \geq 2$ and $t_1, \ldots, t_n$ are subtrees that evaluate to the same idempotent element $e \in M$. The idempotent nodes evaluate to their corresponding idempotent e.*

We call such a tree a factorization forest of w. We write $\mathcal{F}_d(M)$ for the set of factorization forests of depth at most d for the monoid M, leaving Σ implicit.

As the name suggests, the factorization theorem provides a *factorization* of a word w into a tree, and its subtrees correspond to factors of w. Let us illustrate in Figure 2 a factorization of a word for the monoid used in Example 9. For the interpretation I_{bp} of Example 10, a similar picture is obtained by replacing all 2's from Figure 2 by 0's.

Ordering Factorisation Forests. By definition, factorization forests are *nested words* over the alphabet Σ^*. It has been proven by Higman that if a set $(X, \leq)$ is well-quasi-ordered, then the set of finite words $(X^*, \leq^*)$ is also well-quasi-ordered [17], where $u \leq^* v$ if there exists an increasing map $f\colon \{1, \ldots, |u|\} \to \{1, \ldots, |v|\}$ such that $u_i \leq v_{f(i)}$ for all i. When $u \leq^* v$ we say that u is a *(scattered) subword* of v.

Given a finite alphabet Σ, a finite monoid M and a surjective morphism $\mu\colon \Sigma^* \twoheadrightarrow M$, we define by induction on $d \in \mathbb{N}$ a well-quasi-ordering $\preceq_d$ on the set of factorization forests of depth at most d. For $d = 0$, the forests are all leaves, and we equip them with the equality quasi-order. For the inductive case, we let $t \preceq_{d+1} t'$ if one of the following holds:

- $t \preceq_d t'$ (hence, both are of depth at most d),
- $t = t_1 \curlywedge t_2$, $t' = t'_1 \curlywedge t'_2$, $t_1 \preceq_d t'_1$ and $t_2 \preceq_d t'_2$,
- $t = \curlywedge\!\!\!\!\!\curlywedge(t_1, \ldots, t_n)$ and $t' = \curlywedge\!\!\!\!\!\curlywedge(t'_1, \ldots, t'_n)$ and $t_1 \cdots t_n \preceq_d^* t'_1 \cdots t'_n$ for the subword embedding relation over $(\mathcal{F}_d(M))^*$.

---

[3] It can be proven that this depth is at most $3|M|$ [5], but we will not use this fact.
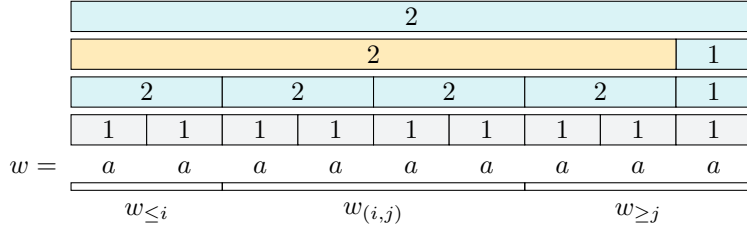
Figure 2 We represent a factorization forest for the word $w = a^9$ using the monoid of the interpretation I_{paths} of Example 9. Gray nodes are leaves, blue nodes are binary (or unary), and yellow nodes are idempotent nodes. For every node, we used as a label the evaluation of the subtree rooted at this node. We depicted the factors $w_{\leq i}$, $w_{(i,j)}$, and $w_{\geq j}$, where $i = 2$ and $j = 7$.

Let us briefly remark that this definition of ordering using "nested subword embeddings" was used to verify *priority channel systems* in [16], and is a specialization of the notion of the gap-embedding relation on trees of [11]. Furthermore, the definition of the well-quasi-ordering $\preceq_d$ can be extended to the case where nodes of the forests are *labelled by* elements from a quasi-ordered set $(Q, \leq)$ as follows: whenever we compare two nodes t and t', we also check if their labels are comparable. The following lemma states that $(\mathcal{F}_d(M), \preceq_d)$ is a well-quasi-order for every d, and follows immediately from Higman's lemma [17].

▶ **Lemma 12** (Factorization Forests are Well-Quasi-Ordered). *Let M be a finite monoid, Σ be a finite alphabet, $\mu\colon \Sigma^* \to M$ be a morphism, $(Q, \leq)$ be a well-quasi-ordered set, and $d \in \mathbb{N}$. The set of Q-labelled factorization forests of depth at most d is a well-quasi-order for $\preceq_d$.*

Let us fix for the rest of this section a monoid interpretation $I = (\Sigma, M, \mu, P)$, and a constant d such that every word in $w \in \Sigma^*$ has a factorization forest of depth at most d (using Theorem 11). Because factorization forests are an additional structure over a finite word w, the interpretation I can be extended to a map from factorization forests to graphs (ignoring the factorization of the underlying word). If this map $I\colon \mathcal{F}_d(M) \to I(\Sigma^*)$ was *order-preserving*, i.e., if $t \preceq_d t'$ implies $I(t) \subseteq_i I(t')$ for every $t, t' \in \mathcal{F}_d(M)$, then we could immediately conclude that $I(\Sigma^*)$ is well-quasi-ordered: the image of a well-quasi-ordered set by an order-preserving map is well-quasi-ordered [10]. Let us argue that it is *almost* the case.

By construction, if $t \preceq_d t'$, then t can be obtained from t' by inserting new children to idempotent nodes. In particular, it induces a map f from nodes of t to nodes of t' that respects: the least common ancestor relation, the evaluation of subtrees, the type of nodes (binary node, leaf, idempotent node). Now, if t and t' are factorization forests of words u and v respectively, then f can be seen as a map from positions of u to positions of v, hence a map from vertices of $I(u)$ to vertices of $I(v)$, which is *almost* an embedding.

▶ **Lemma 13.** *Let $u \in \Sigma^*$ be a word, $t_u \in \mathcal{F}_d(M)$ be a factorization of u, e be an idempotent element of M, x be an idempotent node that evaluates to e in t_u, and t_w be a factorization of a word w that evaluates to e. Let t_v be obtained by inserting t_w as a new child of the node x in u, and $f\colon \{1, \ldots, |u|\} \to \{1, \ldots, |v|\}$ be the corresponding map at the level of words. Then, for every pair of positions $i < j$ that are not in the factor of u described by x, we have: $\mu(u_{\leq i}) = \mu(v_{\leq f(i)})$, $\mu(u_{\geq j}) = \mu(v_{\geq f(j)})$, and $\mu(u_{(i,j)}) = \mu(v_{(f(i),f(j))})$.*

As a consequence, for such pairs, $(i,j) \in \mathsf{E}(I(u))$ if and only if $(f(i), f(j)) \in \mathsf{E}(I(v))$.

Proof. Let us consider two positions $i < j$ of the word u. Assume that the insertion of t_w is done between i and j (the other cases are similar). Then, one can write $u = u_{\leq i} u_{(i,j)}^{\text{left}} x^{\text{left}} x^{\text{right}} u_{(i,j)}^{\text{right}} u_{\geq j}$, where $x^{\text{left}} x^{\text{right}}$ is the factor of u corresponding to the idempotent node x, split where the insertion of the word w happens. By construction, $v =$
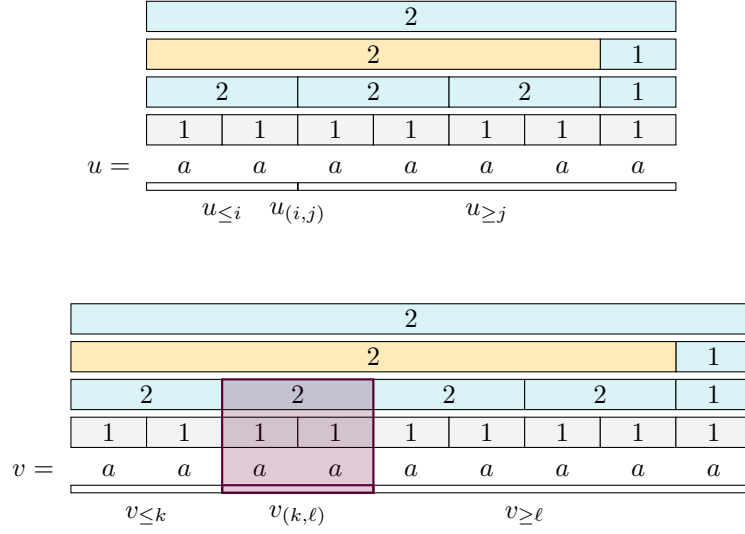
Figure 3 We represent factorization forest for the words $u = a^7$ and $v = a^9$, using the monoid of the interpretation I_{paths} of Example 9, following the same coloring conventions as in Figure 2. The forest for v is obtained by inserting the purple subtree in the forest for u. The map $f \colon \{1, \ldots, 7\} \to \{1, \ldots, 9\}$ derived from this embedding is such that $f(x) = x$ if $x \le 2$ and $f(x) = x + 2$ otherwise. In this drawing, $k = f(i) = 2$, $j = 3$, and $\ell = f(j) = 5$. The pair (i, j) belongs to $\mathsf{E}(I(u))$ because $\mu(u_{(i,j)}) = \mu(\varepsilon) = 0$, but the pair $(f(i), f(j)) = (k, \ell)$ does not belong to $\mathsf{E}(I(v))$ because $\mu(v_{(k,\ell)}) = \mu(a^2) = 2$.

$u_{\le i} u_{(i,j)}^{\text{left}} x^{\text{left}} w x^{\text{right}} u_{(i,j)}^{\text{right}} u_{\ge j}$. Because $u_{\le i} = v_{\le f(i)}$ and $u_{\ge j} = v_{\ge f(j)}$, we immediately obtain the first and third desired equalities.

$$
\begin{aligned}
\mu(v_{(f(i),f(j))}) &= \mu(u_{(i,j)}^{\text{left}})\mu(x^{\text{left}})\mu(w)\mu(x^{\text{right}})\mu(u_{(i,j)}^{\text{right}}) \\
&= \mu(u_{(i,j)}^{\text{left}})e^2\mu(u_{(i,j)}^{\text{right}}) && x \text{ is an idempotent node} \\
&= \mu(u_{(i,j)}^{\text{left}})e\mu(u_{(i,j)}^{\text{right}}) && e \text{ is idempotent} \\
&= \mu(u_{(i,j)}^{\text{left}})\mu(x^{\text{left}}x^{\text{right}})\mu(u_{(i,j)}^{\text{right}}) && x \text{ is an idempotent node} \\
&= \mu(u_{(i,j)})
\end{aligned}
$$

Which concludes the proof. ◀

Let us illustrate why considering positions that are "too close" to the inserted forest is problematic in Figure 3. This is not surprising as the class of graphs described by the interpretation I_{paths} of Figure 3 is not 2-wqo, hence it is impossible for the interpretation I_{paths} to transport embeddings between forests to embeddings between their images.

3.2   Characterizing Labelled Well-Quasi-Ordered Classes

Generalizing the example of Figure 3, we can now define what are the formal *obstructions* inside the factorization forests for the map I to be order-preserving. We will call them bad forest paths. To that end, let us first generalize the notion of evaluation of a monoid interpretation from factorization forests to account for a *context* $(m_l, m_r) \in M^2$: $I(m_l \cdot t \cdot m_r)$ is the graph obtained by evaluating the factorization forest t with the morphism μ and the edge selector P_{m_l, m_r}, where $(x, y, z) \in P_{m_l, m_r} \iff (m_l \cdot x, y, z \cdot m_r) \in P$. This will allow

us to reason about the image of a subtree t' of a factorization forest t, which corresponds to an induced subgraph $I(m_l \cdot t' \cdot m_r)$ of the generated graph $I(t)$, where m_l and m_r are the monoid elements obtained by evaluating the word to the left and to the right of the selected subtree. Let us also define the *type of a leaf* i in a factorization tree $t \in \mathcal{F}_d(M)$ of a word $u \in \Sigma^*$, as the triple $(\mu(u_{<i}), \mu(u_i), \mu(u_{>i}) \in M^3$, that we denote as $\mathsf{tp}_t(i)$.

▶ **Definition 14.** *Let $(t_i)_{1 \leq i \leq k}$ be a sequence of elements in $\mathcal{F}_d(M)$, all evaluating to the same idempotent element $e \in M$ and $t = \sqcap (t_i)_{1 \leq i \leq k}$. The sequence is a good forest path if for all $(m_l, m_r) \in M^2$, there exists another sequence $(h_i)_{1 \leq i \leq \ell}$ of elements of $\mathcal{F}_d(M)$ evaluating to e, a tree $h = \sqcap (h_i)_{1 \leq i \leq \ell}$, and an embedding $f \colon \mathsf{V}(T) \to \mathsf{V}(H)$ where $T = I(m_l \cdot t \cdot m_r)$ and $H = I(m_l \cdot h \cdot m_r)$, such that:*

1. *for every element $x \in \mathsf{V}(T)$, originating from a leaf of some t_i, that is sent to a vertex originating from h_j via f, we have $\mathsf{tp}_{t_i}(x) = \mathsf{tp}_{h_j}(f(x))$,*
2. *the vertices originating from t_1 are mapped to vertices originating from h_1,*
3. *the vertices originating from t_k are mapped to vertices originating from h_ℓ,*
4. *there exists $1 \leq i \leq \ell$ such that the image of f does not intersect the vertices originating from h_i.*

We say that a good forest path is split by the sequence $(h_i)_{1 \leq i \leq \ell}$ and embedding f in the context (m_l, m_r). A bad forest path is a sequence of factorization forests $\sqcap (t_i)_{1 \leq i \leq k}$, that is not a good forest path.

By definition, good forest paths can be split into smaller independent parts. Beware that the split can reorder the nodes in some arbitrary fashion, and may not respect the order of the leaves in the original forest. Our first result is to prove that the existence of arbitrarily long bad forest paths is sufficient to conclude that the class $I(\Sigma^*)$ is not labelled-wqo. In fact, the number of colors used will be related to the size of the monoid M.

▶ **Lemma 15.** *Assume that bad forest paths of arbitrarily large length in $\mathcal{F}_d(M)$. Then, $I(\Sigma^*)$ is not k-wqo, where $k = |M|^2 \times 3$.*

Proof. Assume that there exists arbitrarily long bad forest paths in $\mathcal{F}_d(M)$. Then, there exists a sequence of forests $(t_i)_{i \in \mathbb{N}}$ such that for all i, $t_i = \sqcap (h_{i,k})_{1 \leq j \leq n_i}$, where $(h_{i,k})_{1 \leq k \leq n_i}$ is a bad forest path, and n_i is increasing with respect to i. Since these are bad forest paths, there exists $(a_i, b_i) \in M^2$ such that t_i cannot be split in the context (a_i, b_i).

Because the monoid M is finite, we can assume that all the forests t_i evaluate to the same idempotent element $e \in M$, and that there exists $(a, b) \in M^2$ such that $a_i = a$ and $b_i = b$ for all $i \in \mathbb{N}$. Extracting a subsequence also allows us to assume that for all $i \in \mathbb{N}$, n_{i+1} is greater than the number of leaves in t_i.

Let us now consider the sequence of graphs $G_i := I(a \cdot t_i \cdot b)$, that we color as follows: every vertex is colored by its leaf type in its original factorization forest $h_{i,k}$, $1 \leq k \leq n_i$. We furthermore add three new colors, distinguishing vertices originating from the first and last factorization forest of the bad forest path $h_{i,1}$ and h_{i,n_i} from the rest of the vertices. The final coloring corresponds to $|M|^3 \times 3$ labels. We now prove that the sequence $(G_i)_{i \in \mathbb{N}}$ is an infinite antichain with the provided coloring, which concludes the proof.

Assume towards a contradiction that there exists $i < j$ and a labelled embedding $f \colon \mathsf{V}(G_i) \to \mathsf{V}(G_j)$. Because of the coloring chosen on G_i and G_j, the map f respects the types of the leaves (Item 1), sends the vertices originating from $h_{i,1}$ to the vertices originating from $h_{j,1}$ (Item 2), and the vertices originating from h_{i,n_i} to the vertices originating from h_{j,n_j} (Item 3). Furthermore, the image of f must not intersect the vertices originating from some $h_{j,k}$, $1 \leq k \leq n_j$ (Item 4), because n_j is greater than the number of leaves in t_i, hence the number of vertices in G_i. We conclude that $(h_{i,k})_{1 \leq k \leq n_i}$ is a good forest path, which can be split in the context (a, b), contradicting our initial assumption. ◀

Let us now prove that the absence of arbitrarily long bad forest paths is sufficient to conclude that the class $I(\Sigma^*)$ is wqo-well-quasi-ordered. The key ingredient of the proof is to split the good forest paths, which has the effect of adding *dummy idempotent nodes* separating the leaves of our factorization forests. Adding these nodes will prevent the issues encountered in generalizing Lemma 13 to nodes that are "too close" to the inserted forest, ensuring that the case of Figure 3 never happens. A first observation is that to split a good forest path, one only needs to repeat it three times.

▶ **Lemma 16.** *Let $(t_i)_{1 \le i \le n}$ be a good forest path. Then, for all contexts $(a, b) \in M^2$, there exists an embedding $f \colon \mathsf{V}(T) \to \mathsf{V}(T^3)$ satisfying Items 1–4 of Definition 14, where $T = I(a \cdot \text{⑂}(t_1, \ldots, t_n) \cdot b)$, and $T^3 = I(a \cdot \text{⑂}(t_1, \ldots, t_n, t_1, \ldots, t_n, t_1, \ldots, t_n) \cdot b)$. Furthermore, the embedding f can be chosen such that for all x, $f(x)$ is either x in the first copy of the sequence, or x in the third copy of the sequence, i.e., f is uniquely determined by a map from the vertices of T to $\{\mathsf{first}, \mathsf{third}\}$.*

Proof. Let $(h_i)_{1 \le i \le \ell}$ be a sequence of factorization forests that splits the good forest path in the context (a, b) using an embedding $g \colon \mathsf{V}(T) \to \mathsf{V}(H)$, where $H = I(a \cdot \text{⑂}(h_1, \ldots, h_\ell) \cdot b)$. Let us write $1 < N_0 < \ell$ for the smallest index such that the image of g does not intersect the vertices originating from h_{N_0}. Let us also write $\rho \colon \mathsf{V}(T) \to \{1, \ldots, \ell\}$ the map that assigns to every vertex x of T the index of the factorization forest from which $g(x)$ originates in H.

Recall that every vertex of T exists in three copies in T^3 (one for each repetition of the sequence of forests). We define the map $f \colon \mathsf{V}(T) \to \mathsf{V}(T^3)$ as follows, given a vertex $i \in \mathsf{V}(T)$:

- If i originates from t_1, then it is mapped to its first copy in T^3,
- If i originates from t_n, then it is mapped to its last copy in T^3,
- If i originates from t_j, $1 < j < n$, and $\rho(i) < N_0$, then it is mapped to its first copy in T^3,
- If i originates from t_j, $1 < j < n$, and $\rho(i) > N_0$, then it is mapped to its last copy in T^3.

It is clear that f satisfies Items 1–4. What remains to be shown is that f is an embedding of T into T^3, which will follow from the fact that g was itself an embedding. Given two vertices $x, y \in \mathsf{V}(T)$ four cases can be considered:

- If x and y both originate from t_1. Then, $(x, y) \in \mathsf{E}(T)$ if and only if $(x, y) \in \mathsf{E}(I(a \cdot \text{⑂}(t_1, \ldots, t_n) \cdot b))$, if and only if $(x, y) \in \mathsf{E}(I(a \cdot t_1 \cdot eb))$, because $n \ge 2$ for every idempotent nodes. Therefore, $(x, y) \in \mathsf{E}(T)$ if and only if $(f(x), f(y)) \in \mathsf{E}(T^3)$.
- If x and y both originate from t_n, the same argument applies but with the last copy of t.
- If x and y originate from a tree t_j, $1 < j < n$, and are sent to the same copy of t by f. Then a similar argument applies too.
- If x and y originate from a tree t_j, $1 < j < n$, and are sent to different copies of t by f. Then, one idempotent element e is inserted when computing the edge relation between $f(x)$ and $f(y)$, which may change the result (as in Figure 3). However, $(x, y) \in \mathsf{E}(T)$ if and only if $(g(x), g(y)) \in \mathsf{E}(H)$ because g is an embedding. It follows from the same proof scheme as for Lemma 13 that $(g(x), g(y)) \in \mathsf{E}(H)$ if and only if $(f(x), f(y)) \in \mathsf{E}(T^3)$, as they share the same leaf types, and are both separated by the same idempotent e.

We have proven that f is an embedding of T into T^3. ◀

We are now ready to upgrade the forests in $\mathcal{F}_d(M)$, assuming the existence of a bound N_0 on the length of bad forest paths. Let us overload notations for the binary nodes so that $\text{⑁}(t_1, \ldots, t_k) := \text{⑁}(t_1, \text{⑁}(t_2, \ldots, t_k))$, $\text{⑁}(t_1, t_2) := t_1 \text{⑁} t_2$, and $\text{⑁}(t_1) := t_1$. Note that the depth of $\text{⑁}(t_1, \ldots, t_k)$ is bounded by $k + d_{\max}$, where $d_{\max}$ is the maximal depth of the trees t_i.

Let us illustrate our construction on an example. Let $t = \text{⑂}(t_1, t_2, t_3, t_4)$, where t_i are factorization forests evaluating to some idempotent element $e \in M$, and assuming that $N_0 = 1$. The first step is to group the children of t into *buckets* of size $N_0 + 1$ (that is,

2): $b_1 := (t_1, t_2)$ and $b_2 := (t_3, t_4)$. Now, because b_1 and b_2 are of length $N_0 + 1$, they are good forest paths, and can be split by repeating their sequence three times (Lemma 16). We obtain new buckets $b'_1 := (t_1, t_2, t_1, t_2, t_1, t_2)$ and $b'_2 := (t_3, t_4, t_3, t_4, t_3, t_4)$. We label the leaves of the trees in the sequences b'_1 and b'_2 as valid if they are in the image of the split, and dummy otherwise. The resulting forest is obtained as follows:

$$t' := \underbrace{(\vdash(t_1, t_2))}_{\text{Leftmost}} \vdash (\vdash(\underbrace{t_1, t_2}_{\text{padding}} , \underbrace{\vdash(t_1, t_2, t_3, t_4)}_{\text{Group}}, \underbrace{t_3, t_4}_{\text{padding}})) \vdash \underbrace{(\vdash(t_3, t_4))}_{\text{Rightmost}} \quad . \tag{1}$$

Where the *padding* elements have all their leaves labelled with dummy. The induced subgraph of $I(t')$ obtained by removing all the vertices labelled by dummy is isomorphic to $I(t)$. Furthermore, every pair of leaves (x, y) of $I(t')$ that are labelled by valid have the following property: either x and y belong to the same child of t' (for instance, $\vdash (t_1, t_2)$, or $\vdash (t_1, t_2, t_3, t_4)$), or they are separated by a child of t' whose leaves are all labelled by dummy. In particular, for the tree t', inserting new children to the idempotent node does not modify the induced subgraph of the leaves labelled by valid.

The following technical lemma states that this construction can be applied inductively to obtain a function $\tau : \mathcal{F}_d(M) \to \mathcal{F}_{d'}(M)$, for some bounded $d' \in \mathbb{N}$ that depends on N_0 and d.

▶ **Lemma 17.** *Assume that there exists a bound $N_0 \in \mathbb{N}$ on the length of bad forest paths in $\mathcal{F}_d(M)$. Then, there exists a $d' \geq d$ and a function $\tau : \mathcal{F}_d(M) \to \mathcal{F}_{d'}(M)$ that maps a factorization forest t to a forest $\tau(t)$ that is labelled using $\{\mathsf{valid}, \mathsf{dummy}\}$, and such that:*
- *Every graph $G \in I(\Sigma^*)$ is obtained as the restriction of a graph $I(t)$ to its valid vertices for some $t \in F$,*
- *For every $t, t' \in \mathcal{F}_d(M)$, for every function f witnessing that $\tau(t) \preceq_{d'} \tau(t')$ (as labelled forests), the restriction of f to the induced subgraph of valid vertices of $I(t)$ is an embedding into the induced subgraph of valid vertices of $I(t')$.*

From Lemma 17, we immediately conclude that the class $I(\Sigma^*)$ is wqo-well-quasi-ordered. To simplify the proof, let us recall an equivalent characterization of well-quasi-ordered sets $(X, \leq)$. A sequence $(x_i)_{i \in \mathbb{N}}$ is *good* if there exists $i < j$ such that $x_i \leq x_j$. A set $(X, \leq)$ is well-quasi-ordered if and only if every infinite sequence of elements of X is good.

▶ **Corollary 18.** *Assume that there exists a bound $N_0 \in \mathbb{N}$ on the length of bad forest paths in $\mathcal{F}_d(M)$. Then, $I(\Sigma^*)$ is wqo-well-quasi-ordered.*

Proof. Consider an infinite sequence of graphs $G_i \in I(\Sigma^*)$, with labellings $\mathsf{lbl}_{G_i} : \mathsf{V}(G_i) \to Q$, where $(Q, \leq)$ is a well-quasi-ordered set. Let us equip the set $\{\mathsf{valid}, \mathsf{dummy}\}$ with the equality relation, and remark that $Q \times \{\mathsf{valid}, \mathsf{dummy}\}$ is also well-quasi-ordered when endowed with the product ordering.

By definition, every graph G_i is the image of a word $u_i \in \Sigma^*$ through I. Because we chose the value of d according to Simon's factorization theorem (Theorem 11), every word u_i has a factorization forest $t_i \in \mathcal{F}_d(M)$. We can construct build the sequence $h_i := \tau(t_i)$, where τ is the function provided by Lemma 17.

Because the set of $(Q \times \{\mathsf{valid}, \mathsf{dummy}\})$-labelled factorization forests of depth at most d' is well-quasi-ordered for $\preceq_{d'}$ (Lemma 12), there exists an increasing pair $i < j$ of indices such that $h_i \preceq_{d'} h_j$. This defines a function $f : \mathsf{V}(I(h_i)) \to \mathsf{V}(I(h_j))$ that is an embedding when restricted to the induced subgraph of valid vertices, and respects the labelling $Q \times \{\mathsf{valid}, \mathsf{dummy}\}$. Because the induced subgraph of valid vertices of $I(h_i)$ is precisely $I(t_i) = I(u_i) = G_i$, and the same holds for $I(h_j) = G_j$, this provides a function $f : \mathsf{V}(G_i) \to \mathsf{V}(G_j)$ that is a labelled embedding of G_i into G_j. ◀

The sequence $(G_i)_{i \in \mathbb{N}}$ is therefore a good sequence. and we have proven that $I(\Sigma^*)$ is wqo-well-quasi-ordered. ◀

We have proven that the existence of a bound on the length of bad forest paths charcterizes whether $I(\Sigma^*)$ is wqo-well-quasi-ordered. To finalize the proof of our Theorem 1, we only need to show that the existence of such a bound is decidable. This is because the language of bad forest paths of depth at most d is a regular language, and one can decide whether a regular language is unbounded.

▶ **Lemma 19.** *Let $d \in \mathbb{N}$. It is decidable whether $\mathcal{F}_d(M)$ contains bad forest paths of unbounded length.*

Proof of Theorem 1 page 3. Let I be an MSO-interpretation from Σ^* to graphs. Using Lemma 8, we can compute an MSO-interpretation I' that only uses the formula φ_{edge}, such that $I'(\Sigma^*)$ is k-wqo if and only if $I(\Sigma^*)$ is k-wqo for all $k \in \mathbb{N}$. From I', we can compute a monoid interpretation $I'' = (\Sigma, M, \mu, P)$ by standard techniques. Applying Theorem 11, we compute a number d such that every word $u \in \Sigma^*$ has a factorization forest $t \in \mathcal{F}_d(M)$. We then use Lemma 19 to decide whether $\mathcal{F}_d(M)$ has bad forest paths of unbounded length. If it does, then $I''(\Sigma^*)$ is not k-well-quasi-ordered, where $k = |M|^3 \times 3$ (Lemma 15), hence $I(\Sigma^*)$ is not k-wqo. If it does not, then $I''(\Sigma^*)$ is wqo-well-quasi-ordered (Corollary 18), and therefore so is $I(\Sigma^*)$.

In particular, we have computed a number $k = |M|^3 \times 3$ such that $I(\Sigma^*)$ is k-well-quasi-ordered if and only if $I'(\Sigma^*)$ is wqo-well-quasi-ordered. ◀

4   A Semigroup Approach to Well-Quasi-Orders of Relabel Functions

Classes that are labelled-well-quasi-ordered and of bounded clique-width had already been studied by [9], which leveraged the notion of k-node labelled controlled graphs, a notion introduced by [28]. The class $\mathsf{NLC}_k^{\mathcal{F}}$ of *k-node labelled controlled graphs* with *relabel functions* $\mathcal{F}$ from $\{1, \dots, k\}$ to $\{1, \dots, k\}$ is defined inductively on k-labelled graphs as follows: one can create a single vertex graph with label $i \in \{1, \dots, k\}$, one can relabel a k-labelled graph G with a function $f \in \mathcal{F}$ (applying f to every vertex label of G), and one can combine two k-labelled graphs G and H using a set $S \subseteq \{1, \dots, k\}^2$ by creating the disjoint union of G and H, and adding edges between vertices x and y in G and H respectively if $(\mathsf{lbl}_G(x), \mathsf{lbl}_H(y)) \in S$.

It is a folklore result that every class of graphs $\mathcal{C}$ that has bounded linear clique-width is a subset of some $\mathsf{NLC}_k^{\mathcal{F}}$ for some k and some $\mathcal{F}$, and [9] completely characterized the pairs $(k, \mathcal{F})$ such that $\mathsf{NLC}_k^{\mathcal{F}}$ is 2-well-quasi-ordered or labelled-well-quasi-ordered, showing the equivalence of these two properties [9, Theorem 3]. The theorem relies on the notion of *totally ordered sets of relabel functions*: given two functions $f, g \colon \{1, \dots, k\} \to \{1, \dots k\}$, they define a relation $f \preccurlyeq g$ as $\mathrm{Img}(f) \subseteq \mathrm{Img}(f \circ g)$, and study families of functions $\mathcal{F}$ that are closed under composition, contain the identity function, and *totally ordered* for this relation [9, Section 2].

▶ **Theorem 20** ([9, Theorem 3, Corollary 1, Corollary 2]). *For all k, for all monoids $\mathcal{F}$ of endofunctions of $\{1, \dots, k\}$, the following are equivalent and decidable properties:*
1. *$\mathcal{F}$ is totally ordered by the relation $\preccurlyeq$,*
2. *The class of graphs $\mathsf{NLC}_k^{\mathcal{F}}$ is 2-well-quasi-ordered.*
3. *The class of graphs $\mathsf{NLC}_k^{\mathcal{F}}$ is labelled-well-quasi-ordered.*

Let us mention that Theorem 20 does not allow us to decide whether a class of bounded linear clique-width is 2-well-quasi-ordered or not, nor does it solve the Pouzet Conjecture in that case. While it is true that every class of graphs of bounded linear clique-width is a **subset** of $\mathsf{NLC}_k^{\mathcal{F}}$ for some k and some $\mathcal{F}$ [7], no analogues of Lemma 6 exist for this representation of classes of graphs: it may be that the class of graphs is 2-well-quasi-ordered while the class $\mathsf{NLC}_k^{\mathcal{F}}$ that contains it is not. This mismatch was conjectured to never happen [9, Conjecture 4], but it remains an open question. Informally, one can state that the difference between Theorem 20 and the results of the present paper is akin to the difference between analyzing properties of *transition monoids* and properties of *regular languages*: in our case, the set of accepting states (the edge selector) is fixed, which leads to more precise results, and in particular allows us to transfer our results from images of monoid interpretations to all classes of bounded linear clique-width. We will make this intuition precise in our Theorem 24 and Lemma 25, by showing that the condition of being ordered is equivalent to the fact that there are no long bad forest paths *for every choice of edge selector*.

4.1 Totally Ordered Monoids

In this section, we argue that the seemingly arbitrary notion of ordering between relabel functions $\preccurlyeq$ can be naturally understood in the language of semigroup theory, namely that they are related to the classical *Green Relations* [5]. The connection can be made because the sets $\mathcal{F}$ of functions studied by [9] are closed under compositions and contain the identity function, and therefore form a (finite) monoid M.

Let M be a finite monoid and $m \in M$ be an element. We denote by $(m)_{\mathfrak{J}}$ the *bilateral ideal* of m, i.e., the set of elements xmy where x, y ranges in M. Similarly, we define the *left ideal* $(m)_{\mathfrak{L}}$ and the *right ideal* $(m)_{\mathfrak{R}}$ respectively as the sets of elements xm and my where x, y ranges in M.

▶ **Lemma 21** (Green Formulations of Total Orderings). *Let M be a finite monoid. The following are equivalent:*

1. *M is totally ordered by the relation defined by $x \leq y$ if and only if $(x)_{\mathfrak{R}} \subseteq (xy)_{\mathfrak{R}}$,*
2. *M is totally ordered by the relation defined by $x \leq y$ if and only if $(x)_{\mathfrak{L}} \subseteq (yx)_{\mathfrak{L}}$,*
3. *M is totally ordered by the relation defined by $x \leq y$ if and only if $(x)_{\mathfrak{J}} = (xy)_{\mathfrak{J}} = (yx)_{\mathfrak{J}}$,*
4. *The bilateral ideals of M are totally ordered for inclusion, and for all $x, y \in M$, $(xy)_{\mathfrak{J}} = (x)_{\mathfrak{J}} \cap (y)_{\mathfrak{J}}$.*

Whenever one of these conditions is satisfied, all the defined preorders coincide. Furthermore, if M is a monoid of relabel functions $\mathcal{F}$, then the above conditions are equivalent to the condition that $\mathcal{F}$ is totally ordered by the relation $\preccurlyeq$.

Let us now introduce a definition for so-called "totally ordered monoids" that is based on Item 3. Examples of such totally ordered monoids are: all groups, and all *band monoids*, i.e., the monoids M such that for all $a, b \in M$, there exists $c \in M$ such that $a = cbc$.

▶ **Definition 22.** *A monoid M is* totally ordered *if for all $a, b \in M$, either $(ab)_{\mathfrak{J}} = (a)_{\mathfrak{J}}$ or $(ab)_{\mathfrak{J}} = (b)_{\mathfrak{J}}$. We write $a \leq_{\mathfrak{J}} b$ if $(a)_{\mathfrak{J}} \subseteq (b)_{\mathfrak{J}}$, and $a \equiv_{\mathfrak{J}} b$ if $(a)_{\mathfrak{J}} = (b)_{\mathfrak{J}}$.*

Let us mention that totally ordered monoids are exactly those that only recognize languages L having the *finite power property*, i.e., such that languages such that there exists $n \in \mathbb{N}$ such that $L^* = \bigcup_{i \leq n} L^i$ [18, 20]. Furthermore, totally ordered monoids enjoy the following *cancellation property*, that was noticed by [9], and powers their combinatorial results. We will leverage this property to prove the upcoming Lemma 25, that these cancellations prevent the existence of long bad forest paths.

▶ **Lemma 23** (Cancellation Property). *Let M be a totally ordered finite monoid. Then for all $a, b, c \in M$, if $b \leq_{\mathfrak{J}} a$ and $abc = ab$ then $bc = b$. Similarly, if $cba = ba$ then $cb = b$.*

4.2   Bad Forest Paths in Totally Ordered Monoids

Let us now relate our characterization of classes of bounded linear clique-width that are labelled-well-quasi-ordered to the one introduced by [9]. This is done by showing that totally ordered monoids (i.e., the families $\mathcal{F}$ studied in [9]), are precisely those for which the edge selector of our monoid interpretations (as defined in Section 3) does not matter: every choice leads to a wqo-well-quasi-ordered class of graphs.

▶ **Theorem 24.** *Let Σ be a finite alphabet, M be a finite monoid, and $\mu \colon \Sigma^* \twoheadrightarrow M$ be a surjective morphism. Let $k := |M|$ be the size of M, and $\mathcal{F}$ be the set of functions from $\{1, \dots, k\}$ to itself obtained by considering the action of M on itself. Then, the following are equivalent:*

1. *The monoid M is totally ordered,*
2. *The union $\bigcup_{P \subseteq M^3} \mathsf{Im}(I_P)$ is 2-well-quasi-ordered, where $I_P := (\Sigma, M, \mu, P)$.*
3. *The union $\bigcup_{P \subseteq M^3} \mathsf{Im}(I_P)$ is wqo-well-quasi-ordered, where $I_P := (\Sigma, M, \mu, P)$.*
4. *The class $\mathsf{NLC}_k^{\mathcal{F}}$ is wqo-well-quasi-ordered.*
5. *The class $\mathsf{NLC}_k^{\mathcal{F}}$ is 2-well-quasi-ordered.*

Notice that Theorem 24 also proves that not considering the edge selector from our interpretations (by taking the union over all possible choices) smooths the combinatorial properties of Theorem 1: from a computable k such that the class is k-well-quasi-ordered if and only if it is wqo-well-quasi-ordered, we drop to the constant 2, answering positively to the Pouzet Conjecture in this case. The proof of Theorem 24 follows straightforwardly from the following lemma that characterizes the absence of bad forest paths in the case of totally ordered monoids, regardless of the choice of the edge selector. The lemma itself uses the cancellation properties of totally ordered monoids (Lemma 23).

▶ **Lemma 25** (Bad Forest Paths in Totally Ordered Monoids). *Let Σ be a finite alphabet, let M be a totally ordered finite monoid, $P \subseteq M^3$ be an edge selector, $\mu \colon \Sigma \twoheadrightarrow M$ be a surjective morphism, and $I = (\Sigma, M, \mu, P)$ be a monoid interpretation. Then, there are no bad forest paths in $\mathcal{F}_d(M)$, for all $d \in \mathbb{N}$.*

Proof. Let $t \in \mathcal{F}_d(M)$ be factorization forest of a word $u \in \Sigma^*$. Assume that t evaluates to an idempotent element $e \in M$. Because $\mu(u) = e$, and M is totally ordered, there must exist a factorization $u = vaw$ such that $\mu(v) = m_1$, $\mu(a) = m_2$, and $\mu(w) = m_3$ and $m_2 \equiv_{\mathfrak{J}} e$. This is proven by induction on the size of u, using the fact that $(xy)_{\mathfrak{J}} = (x)_{\mathfrak{J}}$ or $(xy)_{\mathfrak{J}} = (y)_{\mathfrak{J}}$.

Let us now construct a function from the leaves of t into the leaves of $t^3 := \Vdash(t, t, t)$. To that end, let us remark that the leaves of t^3 are the positions of the word u^3, and let us define the function f as follows:

- For every position i that belongs to va, we $f(i) = i$, i.e., its corresponding position in the first copy of u in u^3.
- For every position i that belongs to w, we $f(i) = i + 2|u|$, i.e., its corresponding position in the third copy of u in u^3.

This function satisfies all requirements of the definition of a split (Definition 14), and we only have to prove that it is an embedding of $I(t)$ into $I(t^3)$. To that end, it suffices to prove that for every decomposition $u = u_{\leq i} u_{(i,j)} u_{\geq j}$, we have $\mu(u_{(i,j)}) = \mu((uuu)_{(f(i), f(j))})$, $\mu(u_{\leq i}) = \mu((uuu)_{\leq f(i)})$, and $\mu(u_{\geq j}) = \mu((uuu)_{\geq f(j)})$.[4] Let us only deal with the case of $u_{(i,j)}$, as the other cases are similar.

---

[4] This works regardless of the choice of P, that can only use the values of these factors to determine the presence of an edge.

If the factor $u_{(i,j)}$ is contained in va or in w, then the result is immediate. Let us now assume that the range (i,j) intersects both va and w. We can write $u = v_1 v_2 a w_1 w_2$, such that $u_{(i,j)} = v_2 a w_1$, $v = v_1 v_2$, and $w = w_1 w_2$. One can then compute $(uuu)_{(f(i),f(j))} = (v_2 a w)(vaw)(vaw_1)$. Our goal is to prove that $\mu(v_2 a w_1) = \mu((v_2 a w)(vaw)(vaw_1))$. To that end, let us remark that $\mu(u) = e = e^3 = \mu(uuu)$. Let us also remark that $e \equiv_{\mathfrak{J}} \mu(v_2 a w_1) \leq_{\mathfrak{J}} \mu(v_1)$, and $e \equiv_{\mathfrak{J}} \mu(v_2 a w_1) \leq_{\mathfrak{J}} \mu(w_2)$, and that one can therefore apply the cancellation property of Lemma 23 to obtain the following equalities:

$$\mu(u) = \mu(uuu)$$
$$\mu(v_1)\mu(v_2 a w_1)\mu(w_2) = \mu(v_1)\mu((v_2 a w)(vaw)(vaw_1))\mu(w_2)$$
$$\mu(v_2 a w_1)\mu(w_2) = \mu((v_2 a w)(vaw)(vaw_1))\mu(w_2) \qquad \text{left cancellation}$$
$$\mu(v_2 a w_1) = \mu((v_2 a w)(vaw)(vaw_1)) \qquad \text{right cancellation}$$

We have proven that there exists no bad forest path in $\mathcal{F}_d(M)$. ◀

Proof of Theorem 24 page 14. First, the equivalence between Item 1 and Items 4 and 5 follows from Theorem 20. Furthermore, it is clear that Item 3 implies Item 2.

To prove that Item 1 implies Item 3, let us assume that M is totally ordered. Then by Lemma 25, we know that for all $P \subseteq M^3$, there are no bad forest paths in $\mathcal{F}_d(M)$ for all $d \in \mathbb{N}$, and using Corollary 18 we conclude that $\mathsf{Im}(I_P)$ is wqo-well-quasi-ordered. As a consequence, the finite union of these classes is also wqo-well-quasi-ordered.

To prove that Item 2 implies Item 1, let us assume that that the monoid M is not totally ordered, then there exists $m_1, m_2 \in M$ such that $(m_1 m_2)_{\mathfrak{J}} \neq (m_1)_{\mathfrak{J}}$ and $(m_1 m_2)_{\mathfrak{J}} \neq (m_2)_{\mathfrak{J}}$. Because the morphism $\mu \colon \Sigma^* \twoheadrightarrow M$ is surjective, there exists words $u, v \in \Sigma^*$ such that $\mu(u) = m_1$ and $\mu(v) = m_2$. In particular, both u and v must be non-empty words. Let us define the edge selector $P := M \times (M \setminus (m_1 m_2)_{\mathfrak{J}}) \times M$, and consider the sequence of words $w_n := a(uv)^n a$, for some letter $a \in \Sigma$ and for all $n \in \mathbb{N}$. If there is an edge between the i-th and j-th letter of w_n, then $|i - j| \leq 2(|u| + |v|)$. As a consequence, the distance between the first and last positions of w_n in the graph $I_P(w_n)$ is at least $n/2(|u| + |v|)$, which tends to $+\infty$ as n increases. Furthermore, there exists a path from the first to the last position of w_n, because for every factor u' of u, $\mu(u') \notin (m_1 m_2)_{\mathfrak{J}}$ (and similarly for v). Hence, the distance between the vertices originating from the first and last positions in w_n is at most $|w_n|$. By coloring the vertices of the graphs $I_P(w_n)$ to distinguish the first and last positions of w_n, and extracting a subsequence of w_n such that the distance between the first and last positions in w_{n+1} is greater than $|w_n|$, we obtain an infinite antichain of 2-labelled graphs. Therefore, $\mathsf{Im}(I_P)$ is not 2-well-quasi-ordered, which completes the proof. ◀

5 Outlook

The automata based approach. We believe that the results presented in this paper strongly advocate for an automata based approach to the characterization of graph classes that are labelled-well-quasi-ordered. In particular, there are two main directions that naturally emerge from our work. The first one is to investigate the case of classes of bounded clique-width, that is, moving from *words* to *trees* as input of our MSO-interpretations. The second direction is to investigate whether the bound k of our main Theorem 1 can be improved to 2, as it is the case in Theorem 24.

We strongly believe that our proof technique can be adapted to trees, as it is fundamentally based on two key theorems for words that are already known to be adaptable to trees: the fact that finite words are equipped with a well-quasi-ordering by Higman's lemma [17], for

which the generalization to trees is known as Kruskal's tree embedding theorem [19]; and the fact that finite words over a finite monoids can be factored into bounded height trees, i.e. Simon's factorization theorem [26], which has been adapted to trees by Colcombet [4].

As to the bound k in Theorem 1, we conjecture that it can be improved to 2, based on a finer analysis of bad forest paths (Lemma 15), which would bridge the gap between our analysis and the previous results from [9], and prove that the Pouzet conjecture holds for classes of graphs of bounded linear clique width.

New decompositions. Let us also remark that the proof of Theorem 1 actually provides a tree-decomposition of graphs that is, in some sense, *successor-free*: adding new nodes between two nodes in the tree does not change the semantics of the tree. We believe it is worth investigating whether these decompositions could be exploited in other contexts than labelled-well-quasi-orderings.

Interpreting paths. Finally, we conjecture that *paths* are the only obstruction to being labelled-well-quasi-ordered, as all (currently known by the authors) proofs that some classes are not labelled-wqo end up extracting finite paths from the graph classes we consider (see for instance Lemma 15 and Theorem 24, [9, Theorem 3], [12]). This leads us to formulate the following Conjecture 26.

▶ **Conjecture 26.** *Let $\mathcal{C}$ be a class of graphs. Then, $\mathcal{C}$ is not labelled-well-quasi-ordered if and only if there exists a finite set Q of labels, and an order-preserving MSO-interpretation from $\mathsf{Label}_Q(\mathcal{C})$ to the class of all finite paths.*

—— **References** ——

1 Bogdan Alecu, Mamadou Moustapha Kanté, Vadim Lozin, and Viktor Zamaraev. Lettericity of graphs: an fpt algorithm and a bound on the size of obstructions, 2024. `doi:10.48550/arXiv.2402.12559`.

2 Aistis Atminas, Robert Brignall, Vadim V. Lozin, and Juraj Stacho. Minimal classes of graphs of unbounded clique-width defined by finitely many forbidden induced subgraphs. *Discrete Applied Mathematics*, 295:57–69, May 2021. `doi:10.1016/j.dam.2021.02.007`.

3 Robert Brignall, Michael Engen, and Vincent Vatter. A counterexample regarding labelled well-quasi-ordering. *Graphs and Combinatorics*, pages 1395–1409, 2018. `doi:10.1007/s00373-018-1962-0`.

4 Thomas Colcombet. A combinatorial theorem for trees. In Lars Arge, Christian Cachin, Tomasz Jurdziński, and Andrzej Tarlecki, editors, *Automata, Languages and Programming*, pages 901–912, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. `doi:10.1007/978-3-540-73420-8_77`.

5 Thomas Colcombet. Green's relations and their use in automata theory. In Adrian-Horia Dediu, Shunsuke Inenaga, and Carlos Martín-Vide, editors, *Language and Automata Theory and Applications*, pages 1–21, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. `doi:10.1007/978-3-642-21254-3_1`.

6 Bruno Courcelle. Monadic second-order definable graph transductions: a survey. *Theoretical Computer Science*, 126(1):53–75, 1994. `doi:10.1016/0304-3975(94)90268-2`.

7 Bruno Courcelle, Joost Engelfriet, and Grzegorz Rozenberg. Handle-rewriting hypergraph grammars. *Journal of Computer and System Sciences*, 46(2):218–270, 1993. `doi:10.1016/0022-0000(93)90004-G`.

8 Konrad K. Dabrowski, Vadim V. Lozin, and Daniël Paulusma. Well-quasi-ordering versus clique-width: New results on bigenic classes. *Journal of Combinatorial Theory, Series B*, 130:1–18, 2018. `doi:10.1016/j.jctb.2017.09.012`.

9   Jean Daligault, Michael Rao, and Stéphan Thomassé. Well-quasi-order of relabel functions. *Order*, 27(3):301–315, September 2010. `doi:10.1007/s11083-010-9174-0`.

10   Stéphane Demeri, Alain Finkel, Jean Goubault-Larrecq, Sylvain Schmitz, and Philippe Schnoebelen. Algorithmic aspects of wqo theory. Course notes for the MPRI Master's program, Paris, 2012. URL: `https://cel.archives-ouvertes.fr/cel-00727025`.

11   Nachum Derhowitz and Iddo Tzameret. Gap embedding for well-quasi-orderings1. *Electronic Notes in Theoretical Computer Science*, 84:80–90, 2003. WoLLIC'2003, 10th Workshop on Logic, Language, Information and Computation. `doi:10.1016/S1571-0661(04)80846-6`.

12   Guoli Ding. Subgraphs and well-quasi-ordering. *Journal of Graph Theory*, 16(5):489–502, November 1992. `doi:10.1002/jgt.3190160509`.

13   Samuel Eilenberg. *Automata, Languages, and Machines*, volume A of *Automata, Languages, and Machines*. Academic Press, 1974. `doi:10.5555/540244`.

14   Robert Ganian, Petr Hliněný, Jaroslav Nešetřil, Jan Obdržálek, and Patrice Ossona de Mendez. Shrub-depth: Capturing height of dense graphs. *Logical Methods in Computer Science*, Volume 15, Issue 1, January 2019. `doi:10.23638/LMCS-15(1:7)2019`.

15   Robert Ganian, Petr Hliněný, Jaroslav Nešetřil, Jan Obdržálek, Patrice Ossona de Mendez, and Reshma Ramadurai. *When Trees Grow Low: Shrubs and Fast MSO1*, pages 419–430. Springer Berlin Heidelberg, 2012. `doi:10.1007/978-3-642-32589-2_38`.

16   Christoph Haase, Sylvain Schmitz, and Philippe Schnoebelen. The power of priority channel systems. In Pedro R. D'Argenio and Hernán Melgratti, editors, *CONCUR 2013 – Concurrency Theory*, pages 319–333, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. `doi:10.1007/978-3-642-40184-8_23`.

17   Graham Higman. Ordering by divisibility in abstract algebras. *Proceedings of the London Mathematical Society*, 3:326–336, 1952. `doi:10.1112/plms/s3-2.1.326`.

18   Daniel Kirsten. The finite power problem revisited. *Information Processing Letters*, 84(6):291–294, 2002. `doi:10.1016/S0020-0190(02)00319-8`.

19   Joseph B. Kruskal. The theory of well-quasi-ordering: A frequently discovered concept. *Journal of Combinatorial Theory, Series A*, 13:297–305, 1972. `doi:10.1016/0097-3165(72)90063-5`.

20   Michal Kunc. Regular solutions of language inequalities and well quasi-orders. *Theoretical Computer Science*, 348(2):277–293, 2005. Automata, Languages and Programming: Algorithms and Complexity (ICALP-A 2004). `doi:10.1016/j.tcs.2005.09.018`.

21   Vadim V. Lozin, Igor Razgon, and Viktor Zamaraev. Well-quasi-ordering does not imply bounded clique-width. In *Graph-Theoretic Concepts in Computer Science*, pages 351–359, Berlin, Heidelberg, 2016. Springer Berlin Heidelberg. `doi:10.1007/978-3-662-53174-7_25`.

22   Crispin St. John Alvah Nash-Williams. On well-quasi-ordering transfinite sequences. *Mathematical Proceedings of the Cambridge Philosophical Society*, 61:33–39, 1965. `doi:10.1017/S0305004100038603`.

23   Marko Petkovšek. Letter graphs and well-quasi-order by induced subgraphs. *Discrete Mathematics*, 244(1–3):375–388, February 2002. `doi:10.1016/s0012-365x(01)00094-2`.

24   Maurice Pouzet. Un bel ordre d'abritement et ses rapports avec les bornes d'une multirelation. *CR Acad. Sci. Paris Sér. AB*, 274:A1677–A1680, 1972.

25   Neil Robertson and Paul D. Seymour. Graph minors. xx. wagner's conjecture. *Journal of Combinatorial Theory, Series B*, 92(2):325–357, 2004. Special Issue Dedicated to Professor W.T. Tutte. `doi:10.1016/j.jctb.2004.08.001`.

26   Imre Simon. Factorization forests of finite height. *Theoretical Computer Science*, 72(1):65–94, 1990. `doi:10.1016/0304-3975(90)90047-L`.

27   Wolfgang Thomas. Languages, automata, and logic. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of formal languages*, pages 389–455. Springer, 1997. `doi:10.1007/978-3-642-59136-5`.

28   Egon Wanke. k-nlc graphs and polynomial algorithms. *Discrete Applied Mathematics*, 54(2):251–266, 1994. `doi:10.1016/0166-218X(94)90026-4`.