Decidable properties of extension graphs for substitutive languages

Francesco Dolce*[1], Revekka Kyriakoglou[1] and Julien Leroy†[2]

[1]Université Paris-Est, LIGM, UPEM, 5 Boulevard Descartes, F-77454, Marne-la-Vallée, France
francesco.dolce@u-pem.fr, revekka.kyriakoglou@u-pem.fr
[2]Université de Liège, Institut de mathématique, Allée de la découverte 12 (B37), 4000 Liège, Belgium
J.Leroy@ulg.ac.be

Abstract

Given a set of words S, one can associate with every word $w \in S$ its extension graph which describes the possible left and right extensions of w in S. Families of sets can be defined from the properties of the extension graph of their elements: acyclic sets, tree sets, neutral sets, etc. In this paper we study the specific case of the set of factors of a substitutive language and we show that it is decidable whether these properties are verified or not.

1 Introduction

Given a set S of finite words and an element $w \in S$, the extension graph of w describes the possible left and right extensions of w in S. Some families of sets S are defined by a limitation of the possible extensions graphs of a word in the sets: tree condition, planar tree condition, acyclicity condition, connected condition, etc. (see [BDFD+15a, BDFD+15b, DP16]). The class of tree sets introduced in [BDFD+15a] (i.e. sets satisfying the tree condition) contains the family of Arnoux-Rauzy sets and that of regualar interval exchange sets.

The study of bispecial words is essential in order to understand the properties of a factorial set. By the number of left, right and bi-extension of a word we can compute the factor complexity of a set (see, for example, [Cas97]). In particular, knowing the shape of the extension graph of all bispecial words in a set S allows us to check whether S satisfies one of the above conditions.

For finite sets, all the properties above are trivially decidable. In this paper we consider sets arising as language of a substitutive language. It is known since Cassaigne's work that any long enough bispecial word is the (possibly extended) image under the substitution of a shorter one. Thus, a finite number of bispecial words is enough in order to compute all the bispecial words of a set. Cassaigne's method of describing bispecial words has been formalized by Klouda [Klo12] who studied the two-sided extensions of a bispecial word examining them by pairs. More precisely, he studied bispecial triplets $((l_1, l_2), w, (r_1, r_2))$, where l_1, l_2, r_1, r_2

*This work was supported by grants from Région Île-de-France
†J. Leroy is a FNRS postdoc fellow. This work is supported by a PHC partnership.

are such that $l_1wr_1, l_2wr_2 \in S$. In this paper the same technique is applied but it is formalized in such a way that all extensions are considered together. To acquire this, we make use of the bilateral recognizability introduced by Mossé [Mos92, Mos96] that roughly ensures that any long enough finite word can be uniquely desubstituted, except for a prefix and a suffix.

The paper is organized as follows. In Section 2 we examine the basic notions on sets and words. In particular, we define the extension graph of a word of a set and we introduce the notion of recognizability. In Section 3 we consider the special case of a bispecial word. Specifically, we define the set of initial bispecial words and we show that the extensions of any bispecial word are governed by the extensions of the initial ones.

The main result of this paper is given in Section 4, where we define the graph of extension graphs and we prove that this graph is finite and computable for primitive substitutive languages (Theorem 26). This implies that we can decide whether the language of a primitive morphism satisfies the tree condition (resp. planar tree condition, etc.) or not (Corollary 28).

2 Preliminaries

In this section, we first recall some notions on sets of words including recurrent, uniformly recurrent and tree sets.

2.1 Extension graphs

Let A be a finite alphabet. We denote by A^* the free monoid on A and by ε the empty word. We denote by $A^+ = A^* \setminus \{\varepsilon\}$.

Consider the word $u = u_1 \cdots u_n$ with $u_k \in A$ for all $k \in \{1, \ldots, n\}$. For all i, j such that $1 \leq i \leq j \leq n$, we let $u_{[i:j]}$ denote the factor $u_i \cdots u_j$. We also let $u_{[\,:i]}$ and $u_{[-i:\,]}$ respectively denote the *prefix* $u_1 \cdots u_i$ and the *suffix* $u_{n-i+1} \cdots u_n$. We extend the later notation to sets of words, i.e.,

$$\begin{aligned} S_{[\,:i]} &= \{u_{[\,:i]} \mid u \in S\} \\ S_{[-i:\,]} &= \{u_{[-i:\,]} \mid u \in S\}, \end{aligned}$$

where it is assumed that $i \leq \min_{u \in S} |u|$.

Given a set of words S and a word u, we let Su^{-1} and $u^{-1}S$ denote the sets

$$\begin{aligned} Su^{-1} &= \{v \in A^* \mid vu \in S\} \\ u^{-1}S &= \{v \in A^* \mid uv \in S\} \end{aligned}$$

A set of words on the alphabet A is said to be *factorial* if it contains the alphabet A and all the factors of its elements.

Let S be a set of words on the alphabet A. For $w \in S$, we denote

$$\begin{aligned} L_S(w) &= \{a \in A \mid aw \in S\} \\ R_S(w) &= \{a \in A \mid wa \in S\} \\ E_S(w) &= \{(a, b) \in A \times A \mid awb \in S\} \end{aligned}$$

and further

$$\ell_S(w) = \mathrm{Card}(L_S(w)), \quad r_S(w) = \mathrm{Card}(R_S(w)), \quad e_S(w) = \mathrm{Card}(E_S(w)).$$

We omit the subscript S when it is clear from the context. A word w is *right-extendable* if $r(w) > 0$, *left-extendable* if $\ell(w) > 0$ and *biextendable* if $e(w) > 0$. A factorial set S is called *right-extendable* (resp. *left-extendable*, resp. *biextendable*) if every word in S is right-extendable (resp. left-extendable, resp. biextendable).

A word w is called *right-special* if $r(w) \geq 2$. It is called *left-special* if $\ell(w) \geq 2$. It is called *bispecial* if it is both left-special and right-special.

For $w \in S$, we denote

$$m_S(w) = e_S(w) - \ell_S(w) - r_S(w) + 1.$$

We omit the subscript S when it is clear from the context. The word w is called *weak* if $m(w) < 0$, *neutral* if $m(w) = 0$ and *strong* if $m(w) > 0$. We say that a factorial set S is *neutral* if every nonempty word in S is neutral.

A set of words $S \neq \{\varepsilon\}$ is *recurrent* if it is factorial and if for any $u, w \in S$ there is a $v \in S$ such that $uvw \in S$. An infinite factorial set is said to be *uniformly recurrent* if for any word $u \in S$ there is an integer $n \geq 1$ such that u is a factor of any word of S of length n. A uniformly recurrent set is recurrent. In [DP16] it is proved that the converse is true for neutral sets.

The *factor complexity* of a factorial set S of words on an alphabet A is the sequence $p_n = \mathrm{Card}(S \cap A^n)$. It can be computed from the set $\{m(w) \mid w \in S\}$ [Cas97].

Let S be a biextendable set of words. For $w \in S$, we consider the set $E(w)$ as an undirected bipartite graph on the set of vertices which is the disjoint union of $L(w)$ and $R(w)$ with edges the pairs $(a, b) \in E(w)$. This graph is called the *extension graph* of w. We note that, since $E(w)$ has $\ell(w)+r(w)$ vertices and $b(w)$ edges, the number $1-m(w)$ is the Euler characteristic of the graph $E(w)$ (see [DP16]).

If the extension graph $E(w)$ is acyclic, then $m(w) \leq 0$. Thus w is weak or neutral. More precisely, one has in this case that $c = 1 - m(w)$ is the number of connected components of the graph $E(w)$.

A biextendable set S is called *acyclic* if for every $w \in S$ the graph $E(w)$ is acyclic. It is *connected* if for every $w \in S$ the graph $E(w)$ is connected.

A biextendable set S is called a *tree set* of characteristic c if for any nonempty $w \in S$, the graph $E(w)$ is a tree and if $E(\varepsilon)$ is a union of c trees (the definition of tree set in [BDFD+15a] corresponds to a tree set of characteristic 1). Note that a tree set of characteristic c is a neutral set of characteristic c.

A *planar tree set* of characteristic c with respect to two orders $<_1$ and $<_2$ on the alphabet A is a tree set of characteristic c compatible with the two orders (see [BDFD+15b]), i.e. for any $w \in S$ one has

$$a <_2 c \implies b \geq_1 d \qquad \text{for any } (a, b), (c, d) \in E(w).$$

For two sets of words X, Y and a word $w \in S$, we denote

$$\begin{aligned} L^X(w) &= \{x \in X \mid xw \in S\}, \\ R^Y(w) &= \{y \in Y \mid wy \in S\}, \\ E^{X,Y}(w) &= \{(x, y) \in X \times Y \mid xwy \in S\}. \end{aligned}$$

As for $E(w)$, we consider $E^{X,Y}(w)$ as an undirected graph on the set of vertices which is the disjoint union of $L^X(w)$ and $R^Y(w)$. Such a graph is called a *generalized extension* *graph*. When $X = A^n$, and S is understood, we write $L_n(w)$ instead of $L^{A^n}(w)$. Similarly we define $R_m(w)$, and $E_{n,m}(w)$. Moreover, when $n = m$, we call the graph $E_{n,n}(w)$ the *uniform generalized extension graph* (of degree n) of w. Note that $E(w) = E_{1,1}(w)$.

2.2 Recognizability and synchronizing delay

Consider a morphism $\sigma : A^* \to A^*$. We are interested in the set S_σ of words that occur as factors of $\sigma^n(a)$ for some $n \in \mathbb{N}$ and some $a \in A$. More precisely, we consider

$$S_\sigma = \{u \in A^* \mid \exists n \in \mathbb{N}, a \in A : \sigma^n(a) \in A^* u A^*\}.$$

Note that we will always assume that S_σ is infinite.

A morphism $\sigma : A^* \to A^*$ is *primitive* if there exists $k \in \mathbb{N}$ such that all letters $a \in A$ occurs in $\sigma^k(b)$ for all $b \in A$. In this case, for all $a \in A$ we have

$$S_\sigma = \{u \in A^* \mid \exists n \in \mathbb{N} : \sigma^n(a) \in A^* u A^*\}.$$

Such a morphism is *aperiodic* if S_σ is not the set of factors of a periodic infinite word.

Example 1. Let $A = \{a, b\}$ and $\varphi : A^* \to A^*$ be the *Fibonacci morphism* defined by $\varphi(a) = ab$, $\varphi(b) = a$. Let us consider the morphism

$$\sigma_F = \varphi^2 : a \mapsto aba, \quad b \mapsto ab.$$

Both morphisms φ and σ_F are primitive and aperiodic. The set $S = S_{\sigma_F}$ is called the *Fibonacci set*. Note that, since σ_F is a power of φ, we have also $S = S_\varphi$.

The extension graph $E(\varepsilon)$ and the generalized uniform extension graph $E_{4,4}(\varepsilon)$ of the empty word are represented in Figure 1.

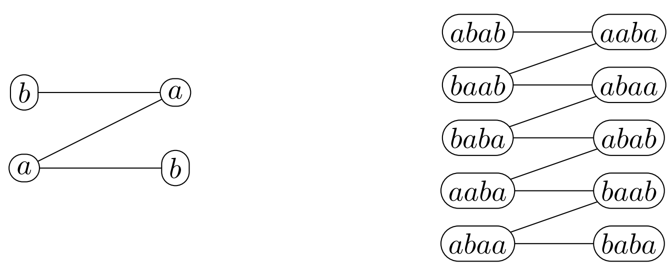

Figure 1: The graphs of $E(\varepsilon)$ (on the left) and $E_{4,4}(\varepsilon)$ (on the right).

Note that identifying in $E_{4,4}(\varepsilon)$ all vertices on the left having the same last letter and all vertices on the right having the same first letter, we find the graph $E(\varepsilon)$.

An important notion when dealing with sets of words associate with morphisms is the notion of *recognizability*. It roughly says that any long enough word in S_σ has a unique pre-image under σ, exept for a prefix and a suffix.

Definition 2. A morphism $\sigma : A^* \to A^*$ is said to be *recognizable* if there exists some constant $C \in \mathbb{N}$ such that for any word $u \in S_\sigma$ of length at least $2C + 1$, there exist $p, v, s \in S_\sigma$ such that

$$u = p\sigma(v)s \qquad \text{with } |s|, |p| \leq C \tag{1}$$

and for all $p', v', s' \in S_\sigma$ satisfying (1), there exists $z, t \in A^*$ such that $v = zv't$, $p' = p\sigma(z)$ and $s' = s\sigma(t)$. The smallest integer C satisying this condition is called the *recognizability constant* of σ.

Theorem 3 ([Mos92, Mos96]). *Any aperiodic primitive morphism is recognizable.*

Example 4. Let σ_F and S as in Example 1. One can check that the recognizability constant of S is 4.

Theorem 3 does not give any information on the constant of recognizablity. When we know it exsits, it can be computed by checking all positive integers one by one until we reach it. The next result gives a theoretical bound on that constant. We set $|\sigma| = \max_{a \in A} |\sigma(a)|$.

Theorem 5 ([DL]). *Let $\sigma : A^* \to A^*$ be an aperiodic primitive substitution. The recognizability constant of σ is bounded by $2K_\sigma^9|\sigma^{8\#AK_\sigma^{24}}| + |\sigma^{\#A}|$, where $K_\sigma = 1 + |\sigma|^{1+3(\#A)^2}$.*

The bound given in Theorem 5 is clearly huge! As explained in [DL], it could be easily improved but this would require much longer statements and heavy notations although the improved bound would probably not be optimal. Using other techniques, the next result provides more reasonable bounds, but for restricted cases. To be able to state it correctly, we need to define the notion of synchronizing point.

Given a word $u = u_1 \cdots u_{|u|} \in S_\sigma$, we say that a triplet (p, v, s) is an *interpretation* of u if $\sigma(v) = puv$. Two interpretations $(p, v, s), (p', v', s')$ are said to be *synchronized at position n* if there exists i, j such that $1 \le i \le |v|$, $1 \le j \le |v'|$ and

$$\sigma(v_1 \cdots v_i) = pu_1 \cdots u_n \quad \text{and} \quad \sigma(v'_1 \cdots v'_j) = p'u_1 \cdots u_n.$$

The word u has a *synchronizing point (at position n)* if all its interpretations are synchronized (at position n). The morphism σ is said to be circular if there is some constant D such that any word of length at least D has a synchronizing point. The smallest such integer D is called the *synchronizing delay* of σ.

A morphism $\sigma : A^* \to A^*$ is *k-uniform* if for all $a \in A$, $|\sigma(a)| = k$.

Theorem 6 ([KM16]). *If $\#A = 2$ and if $\sigma : A^* \to A^*$ is a k-uniform morphism ($k \ge 2$) which is injective on S_σ and recognizable, then its synchronizing delay D is bounded as follows:*

1. *$D \le 8$ if $k = 2$,*
2. *$D \le k^2 + 3k - 4$ if k is an odd prime number,*
3. *$D \le k^2\left(\frac{k}{d} - 1\right) + 5k - 4$ otherwise,*

where d is the least divisor of k greater than 1.

3 Extensions of bispecial words

In the following, we assume that $\sigma : A^* \to A^*$ is an aperiodic primitive morphism and C is its recognizability constant. Since $S_\sigma = S_{\sigma^k}$ for all $k \ge 1$, we can replace σ by a power of itself and assume that $|\sigma(a)| \ge 2$ for all $a \in A$ (as we did in Example 1). Furthermore, the next lemma ensures that the power of σ that is needed is less than $\mathrm{Card}(A)^2$.

Lemma 7 ([HJ90]). *A $d \times d$ matrix M is primitive if and only if there is an integer $k \le d^2 - 2d + 2$ such that M^k contains only positive entries.*

Our goal is to understand the extension graph of all bispecial words in S_σ. It is known that long enough bispecial words are obtained as extended images under σ of shorter bispecial words. We formalize this using the definition of recognizability.

Definition 8. Let u and v be as in Definition 2. We say that v is the *antecedent* of u under σ. We also say that u is an *extended image* of v under σ.

By maximality of v, the antecedent of a word u is obviously unique. The next result states that any long enough bispecial word is an extended image of some shorter bispecial word.

Lemma 9. *If $u \in S_\sigma$ is bispecial and has length at least $2C+1$, then its antecedent v is bispecial and shorter than u.*

Proof. As $\sigma(v)$ is a factor of u and $|\sigma(a)| \geq 2$ for all a, v is shorter than u. Let us prove by contradiction that v is left special, the proof that it is right special is symmetric. Let $p, s \in A^*$ such that $u = p\sigma(v)s$. If v is not left special, there exists a unique letter $a \in A$ such that $av \in S_\sigma$. By maximality of the antecedent, this implies that $|p| < |\sigma(a)|$. This contradicts the fact that u is left special, since $p\sigma(v)$ has a unique left extension which is the letter b such that $\sigma(a) \in A^*bp$. □

Example 10. Let us consider σ_F and S as in Example 1. The recognizability constant of σ_F is 4 (see Example 4). Thus, by Lemma 9, every word of length 9 in S can be uniquely desubstituted. For instance, the word ba is the antecedent of $ba\,ababa\,ab = ba\,\sigma_F(ba)\,ab \in S \cap A^9$.

The next result is a trivial consequence of Lemma 9 and justifies the notion of initial bispecial words.

Corollary 11. *If $u \in S_\sigma$ is bispecial and has length at least $2C+1$, there exists a unique finite sequence $(u_1, u_2, \ldots, u_k)$ of bispecial words in S_σ such that $u_k = u$, $|u_1| \leq 2C$ and for all $i \in \{2, k-1\}$, u_i is the antecedent of u_{i+1} and has length at least $2C+1$.*

Definition 12. A bispecial word is said to be *initial* if it has length at most $2C$.

Example 13. Let σ_F and S be as in Examples 1 and 10. The set of initial bispecial word of S is $\{\varepsilon, a, aba, abaaba\}$.

Let us consider the bispecial word $u = abaababaaba \in S$ of length $11 > 9$. The antecedent $v = aba$ of u is also bispecial, in accord with Lemma 9.

Observe that some initial bispecial word could have a unique antecedent and could therefore not be considered as initial. We could have defined as initial bispecial words the bispecial words that do not have a unique antecedent like it is done in [Klo12]. We decided to proceed in this way because the condition on the length is easier to check.

We now show that the extensions of a non-initial bispecial word are completely determined by the extensions of its antecedent.

Lemma 14. *Let $u \in S_\sigma$ with length at least $2C+1$ and v be its antecedent. If $p, s \in A^*$ are such that $u = p\sigma(v)s$, then for all $m \geq |p|$ and all $n \geq |s|$, we have*

$$\begin{aligned}
\mathrm{L}_m(u) &= A^m \cap \sigma(\mathrm{L}_m(v))p^{-1} \\
\mathrm{R}_n(u) &= A^n \cap s^{-1}\sigma(\mathrm{R}_n(v)) \\
\mathrm{B}_{m,n}(u) &= \left\{(x,y) \in A^m \times A^n \mid \exists\, (x',y') \in \mathrm{B}_{m,n}(v) : \sigma(x') \in A^*xp,\ \sigma(y') \in syA^*\right\}.
\end{aligned}$$

Proof. For each equality, the inclusion $\supseteq$ is obvious. Let us show that

$$\mathrm{L}_m(u) \subseteq A^m \cap \sigma(\mathrm{L}_m(v))p^{-1} = \left\{x \in A^m \mid \exists\, x' \in \mathrm{L}_m(v) : \sigma(x') \in A^*xp\right\}, \qquad (2)$$

the other two cases are similar. Let $x \in A^m$ such that $xu \in S_\sigma$. By uniqueness of v, there exists $x' \in A^*$ such that $x'v \in S_\sigma$ and $\sigma(x') \in A^*xp$. It suffices to show that one can choose x' with length m. If there exists such a x', with $|x'| < m$, then the result follows from the fact that S_σ is biextendable: one can extend x' to the left until we reach the wanted length. On the other hand, if one can find x' with $|x'| > m$, then its suffix z of length m also satisfies $\sigma(z) \in A^*xp$. Indeed, by the assumption made on σ, we have $|\sigma(z)| \geq 2|z| = 2m \geq |xp|$, which concludes with the inclusion (2). ☐

Definition 15. Let u, v be words in A^+. We set

1. $f_L(u,v)$ the longest common suffix of $\sigma(u)$ and $\sigma(v)$;
2. $f_R(u,v)$ the longest common prefix of $\sigma(u)$ and $\sigma(v)$.

Lemma 16. *Let $w \in S_\sigma$ be a word of length at least $C/2$. If $u, v \in \mathrm{L}_C(w)$ have distinct last letter, we have $|f_L(u,v)| \leq C$. In particular, $|\sigma(u)f_L(u,v)^{-1}| \geq C$ and $|\sigma(v)f_L(u,v)^{-1}| \geq C$.*

Symmetrically, if $u, v \in \mathrm{R}_C(w)$ have distinct first letter, we have $|f_R(u,v)| \leq C$. In particular, $|f_R(u,v)^{-1}\sigma(u)| \geq C$ and $|f_R(u,v)^{-1}\sigma(v)| \geq C$.

Proof. Let us prove the first part of the result, the second one is symmetric. Proceed by contradiction and suppose that $|f_L(u,v)| \geq C+1$. By the assumption made on σ, we have $|\sigma(w)| \geq C$. Thus, by recognizability, the word $f_L(u,v)\sigma(w)$ has an antecedent x and we have

$$f_L(u,v)\sigma(w) = p\sigma(x)s \quad \text{with } |p|, |s| \leq C.$$

As $f_L(u,v)$ is a suffix of $\sigma(u)$, the antecedent x can be writen as $x = yz$, where y is a suffix of u and z is a prefix of w. Similarly the antecedent x can be writen as $x = y'z$, where y' is a suffix of v. As u and v do not have any common proper suffix by assumtion, we have $y = y' = \varepsilon$. We thus get

$$f_L(u,v)\sigma(w) = p\sigma(z)s,$$

with $\sigma(w) = \sigma(z)s$. We get a contradiction since $|f_L(u,v)| > |p|$.

The fact that $|\sigma(u)f_L(u,v)^{-1}| \geq C$ directly follows from the assumption made on σ: $|\sigma(u)| \geq 2|u| = 2C$. ☐

In what follows, our aim is to describe the extensions of a long enough word knowing the extensions of its antecedent. We thus introduce the following notation.

Definition 17. Let u, v be words in A^+. We set

1. $g_L(u,v) = \left(\sigma(u)(f_L(u,v))^{-1}\right)_{[-C:]}$;
2. $g_R(u,v) = \left((f_R(u,v))^{-1}\sigma(u)\right)_{[:C]}$.

Corollary 18. *Let $w \in S_\sigma$ be a bispecial word of length at least $C/2$. For all $(u_1,v_1),(u_2,v_2) \in \mathrm{B}_{C,C}(w)$ such that u_1 and u_2 have distinct last letter and v_1 and v_2 have distinct first letter, the word $w' = f_L(u_1,u_2)\sigma(w)f_R(v_1,v_2)$ is a bisepcial word in S_σ.*

Furthermore, if $|w'| \geq 2C+1$, then w is the antecedent of w' and we have

$$\begin{aligned}
\mathrm{L}_C(w') &= (\sigma(\mathrm{L}_C(w))f_L(u_1,u_2)^{-1})_{[-C:]} \\
\mathrm{R}_C(w') &= (f_R(u_1,u_2)^{-1}\sigma(\mathrm{R}_C(w)))_{[:C]} \\
\mathrm{B}_{C,C}(w') &= \big\{(x,y) \in A^C \times A^C \mid \exists\,(x',y') \in \mathrm{B}_{C,C}(w) : \\
&\qquad \sigma(x') \in A^* x f_L(u_1,u_2), \sigma(y') \in f_R(v_1,v_2) y A^*\big\}\,.
\end{aligned}$$

Proof. Since $(u_1,v_1),(u_2,v_2) \in \mathrm{B}_{C,C}(w)$, we have $\sigma(u_1wv_1), \sigma(u_2wv_2) \in S_\sigma$, with

$$\sigma(u_1wv_1) = g_L(u_1,u_2)w'g_R(v_1,v_2) \quad \text{and} \quad \sigma(u_2wv_2) = g_L(u_2,u_1)w'g_R(v_2,v_1).$$

Furthermore, $g_L(u_1,u_2)$ and $g_L(u_2,u_1)$ have distinct last letter by definition of $f_L(u_1,u_2)$ and $g_R(v_1,v_2)$ and $g_R(v_2,v_1)$ have distinct first letter by definition of $f_R(u_1,u_2)$. The word w' is thus bispecial. The second part of the result is a direct consequence of Lemma 14 and Lemma 16. □

Definition 19. Let w and w' as in Corollary 18. We say that w' is a *bispecial extended image* of w.

Example 20. Let us consider again σ_F and S as in Examples 1 and 4.

The word $w = aba$ is a bispecial word of S of length more than $4/2$. The uniform generalized extension graph of degree 4 of w is represented on the left of Figure 2. The only bispecial extended image of w is the word $w' = abaababa\,aba$. Indeed one has, for example, $(baab, abaa), (aaba, baab) \in E_{4,4}(aba)$ with $f_L(baab, aaba) = \varepsilon$ and $f_R(abaa, baab) = aba$. The uniform generalized extension graph of degree 4 of w' is represented on the center of Figure 2.

Figure 2: The graphs of $E_{4,4}(aba)$ on the left and $G = E_{4,4}(w')$ on the right.

4 Graph of extension graphs

In this section we build a finite graph $\mathcal{G}(S_\sigma)$ whose set of vertices is the set of extension graphs of the bispecial words in S_σ. For this we first consider a larger graph and define $\mathcal{G}(S_\sigma)$ as a subgraph of it.

Remark 21. In the second part of Corollary 18, the equalities between the sets of (left-, right- and bi-) extensions depend only on the sets themselves, not on the bispecial words w and w'. Roughly speaking, if $w, w' \in S_\sigma$ are long enough and such that $E_{C,C}(w) = E_{C,C}(w')$, then they give rise to the "same" bispecial extended images. More precisely, for all pairs $(u_1,v_1),(u_2,v_2) \in E_{C,C}(w)$ such that u_1 and u_2 have distinct last letter and v_1 and v_2 have distinct first letter, the bispecial extended images $f_L(u_1,u_2)\sigma(w)f_R(v_1,v_2)$ and $f_L(u_1,u_2)\sigma(w')f_R(v_1,v_2)$ have the same uniform generalized extension graph of degree C.

The previous remark allows us to define a relation $\mathcal{R}$ on $\mathcal{E}(S_\sigma) \times \mathcal{E}(S_\sigma)$, where $\mathcal{E}(S_\sigma)$ is the set of possible uniform generalized extension graph of degree C of a bispecial word in S_σ. The relation associates such an extension graphs with the extension graph of any of the long enough bispecial extended image that can occur. The following definition is rather long but is just a formalization of this relation, using notation of Corollary 18.

Definition 22. We consider the set $\mathcal{E}(S_\sigma)$ of undirected bipartite graphs $G = (V(G), E(G))$, or simply $G = (V, E)$ when it is clear from the context, such that

1. V is a disjoint union of V_{Left} and V_{Right}, where
   (a) $V_{\text{Left}} \subseteq A^C$ and $V_{\text{Right}} \subseteq A^C$;
   (b) there exists $u_1, u_2 \in V_{\text{Left}}$ with distinct last letter and there exist $v_1, v_2 \in V_{\text{Right}}$ with distinct first letter;
2. $E \subseteq V_{\text{Left}} \times V_{\text{Right}}$ is such that all vertices have positive degree;

We define the relation $\mathcal{R} \subseteq \mathcal{E}(S_\sigma) \times \mathcal{E}(S_\sigma)$ by $(G, H) \in \mathcal{R}$, whenever there exist two edges $(u_1, v_1), (u_2, v_2)$ in $E(G)$ such that

1. u_1 and u_2 have distinct last letter and v_1 and v_2 have distinct first letter;
2. the vertices $V_{\text{Left}}(H)$ and $V_{\text{Right}}(H)$ of H are defined by

$$\begin{aligned} V_{\text{Left}}(H) &= (\sigma(V_{\text{Left}}(G)) f_L(u_1, u_2)^{-1})_{[-C:]} \\ V_{\text{Right}}(H) &= (f_R(u_1, u_2)^{-1} \sigma(V_{\text{Right}}(G)))_{[:C]} \end{aligned}$$

3. the edges $E(H)$ of H are defined by

$$\begin{aligned} E(H) = \big\{ &(x, y) \in A^C \times A^C \mid \exists (x', y') \in E(G) : \\ &\sigma(x') \in A^* x f_L(u_1, u_2),\ \sigma(y') \in f_R(v_1, v_2) y A^* \big\}. \end{aligned}$$

Lemma 23. *We have $(G, H) \in \mathcal{R}$ if and only if there exist two bispecial words w and w' such that $|w'| \geq 2C + 1$, w' is a bispecial extended image of w, $G = E_{C,C}(w)$ and $H = E_{C,C}(w')$.*

Proof. This follows from Corollary 18 and from Remark 21. □

Let us now define the graph $\mathcal{G}(S_\sigma)$.

Definition 24. Let $\mathcal{K}(S_\sigma)$ be the graph whose set of vertices is $\mathcal{E}(S_\sigma) \cup \mathcal{IE}(S_\sigma)$, where

$$\mathcal{IE}(S_\sigma) = \{(w, E_{C,C}(w)) \mid w \in S_\sigma \text{ is an initial bispecial word}\}$$

and where the edges are defined as follows.

1. There is an edge from $(u, E_{C,C}(u)) \in \mathcal{IE}(S_\sigma)$ to $(v, E_{C,C}(v)) \in \mathcal{IE}(S_\sigma)$ whenever v is a bispecial extended image of u.
2. There is an edge from $(u, E_{C,C}(u)) \in \mathcal{IE}(S_\sigma)$ to $G \in \mathcal{E}(S_\sigma)$ whenever there exists $v \in S_\sigma$ such that v is a bispecial extended image of u, $|v| \geq 2C + 1$ and $G = E_{C,C}(v)$. This edge is labeled by (p, s) if $p, s \in A^*$ are such that $v = p\sigma(u)s$.

3. There is an edge from $G \in \mathcal{E}(S_\sigma)$ to $H \in \mathcal{E}(S_\sigma)$ whenever $(G, H) \in \mathcal{R}$. This edge is labeled by $(f_L(u_1, u_2), f_R(v_1, v_2))$, where (u_1, v_1) and (u_2, v_2) are as in Definition 22.

The graph $\mathcal{G}(S_\sigma)$ is the subgraph of $\mathcal{K}(S_\sigma)$ whose set of vertices if the set of vertices of $\mathcal{K}$ that are accessible from a vertex in $\mathcal{IE}(S_\sigma)$.

In other words, vertices in the graph $\mathcal{G}(S_\sigma)$ (as well as in $\mathcal{K}(S_\sigma)$) could have two different forms: either they are pairs of an initial bispecial word w together with its generalized extension graph $E_{C,C}(w)$, or they are just generalized extension graphs of the form $E_{C,C}(u)$ without the information of the word u.

Example 25. Let σ_F and S be as in Example 1. From Example 13 it follows that

$$\mathcal{IE}(S) = \left\{ \left(\varepsilon, E_{4,4}(\varepsilon)\right), \left(a, E_{4,4}(a)\right), \left(aba, E_{4,4}(aba)\right), \left(abaaba, E_{4,4}(abaaba)\right) \right\},$$

whith $E_{4,4}(\varepsilon)$ represented in Figure 1 on the right, $E_{4,4}(aba)$ in Figure 2 on the left and $E_{4,4}(a)$, $E_{4,4}(abaaba)$ are represented in Figure 3.

The graph H represented on the right of Figure 3 (that has the same shape of $E_{4,4}(abaaba)$) is an element of $\mathcal{E}(S)$. Indeed, $H = \mathcal{E}_{4,4}(w'')$, with w'' the only bispecial extended image of $abaaba$.

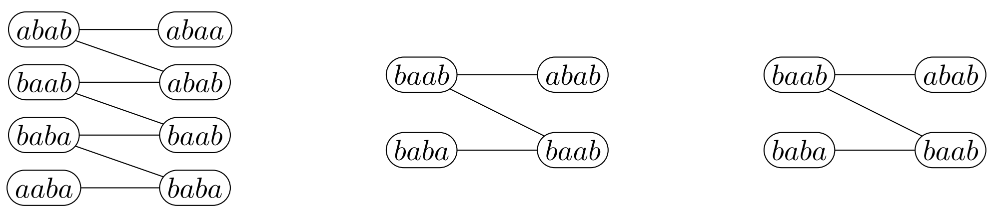

Figure 3: The graphs of $\mathcal{E}_{4,4}(a)$ (on the left), $\mathcal{E}_{4,4}(abaaba)$ (on the center) and H (on the right).

Theorem 26. *The graph $\mathcal{G}(S_\sigma)$ is finite and computable. Moreover,*

1. *For any bispecial word $w \in S_\sigma$, there exists a vertex $\mathcal{V} \in \mathcal{G}(S_\sigma)$ such that $\mathcal{V} = E_{C,C}(w)$ or $\mathcal{V} = (w, E_{C,C}(w))$.*

2. *For any vertex $\mathcal{V}$ of $\mathcal{G}(S_\sigma)$, there exists a bispecial word $w \in S_\sigma$ such that $\mathcal{V} = E_{C,C}(w)$ or $\mathcal{V} = (w, E_{C,C}(w))$.*

Proof. The graph $\mathcal{G}(S_\sigma)$ is finite since its set of vertices is a subset of $\mathcal{IE}(S_\sigma) \cup \mathcal{E}(S_\sigma)$ which is finite. The subgraph of $\mathcal{G}(S_\sigma)$ involving vertices in $\mathcal{IE}(S_\sigma)$ is computable since it only involves words of bounded length and this bound is computable by Theorem 5. For the same reason, the set $\mathcal{E}(S_\sigma)$ and the relation $\mathcal{R} \subset \mathcal{E}(S_\sigma) \times \mathcal{E}(S_\sigma)$ are computable. The graph $\mathcal{K}(S_\sigma)$ is thus computable, hence so is $\mathcal{G}(S_\sigma)$.

Let w be a bispecial word in S_σ. If w is initial, then $(w, E_{C,C}(w))$ is a vertex in $\mathcal{IE}(S_\sigma)$ and we are done. If w is not initial, then $E_{C,C}(w)$ is a graph in $\mathcal{E}(S_\sigma)$, hence a vertex in $\mathcal{K}(S_\sigma)$. This vertex is accessible from a vertex in $\mathcal{IE}(S_\sigma)$ by Corollary 11.

Let $\mathcal{V}$ be a vertex of $\mathcal{G}(S_\sigma)$. If $\mathcal{V} \in \mathcal{IE}(S_\sigma)$, then $\mathcal{V}$ is associated with an initial bispecial word in S_σ and we are done. If $\mathcal{V} \notin \mathcal{IE}(S_\sigma)$, then $\mathcal{V}$ is a vertex in $\mathcal{K}(S_\sigma)$ that is accessible from a vertex in $\mathcal{IE}(S_\sigma)$. In other words, and using the definition of the edges of $\mathcal{K}(S_\sigma)$, there exists a path $(\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_k)$ in $\mathcal{G}(S_\sigma)$ such that

1. $\mathcal{V}_1 = (u, E_{C,C}(u)) \in \mathcal{IE}$, $\mathcal{V}_k = \mathcal{V}$ and $\mathcal{V}_i \notin \mathcal{IE}(S_\sigma)$ for all $i > 1$;
2. there exists a bispecial extended image w of u such that $|w| \geq 2C+1$ and $E_{C,C}(w) = \mathcal{V}_2$;
3. $(\mathcal{V}_i, \mathcal{V}_{i+1}) \in \mathcal{R}$ for all $2 \leq i \leq k-1$.

The result then follows from Lemma 23. $\square$

Example 27. Let σ_F and S be as in Example 1. The graph of graphs $\mathcal{K}(S)$ is represented in Figure 4. The set $\mathcal{IE}(S)$ is defined in Example 25 and the initial vertices are colored in blue in Figure 4. The other two vertices of $\mathcal{K}(S)$ are the graphs G of Example 20 and H of Example 25.

$$\begin{tikzcd}
(\varepsilon, \mathcal{E}_{4,4}(\varepsilon)) \arrow[r, "(\varepsilon, aba)"] & (aba, \mathcal{E}_{4,4}(aba)) \arrow[r, "(\varepsilon, aba)"] & G \arrow[loop right, "(\varepsilon, aba)"] \\
(a, \mathcal{E}_{4,4}(a)) \arrow[r, "(\varepsilon, aba)"'] & (abaaba, \mathcal{E}_{4,4}(abaaba)) \arrow[r, "(\varepsilon, aba)"'] & H \arrow[loop right, "(\varepsilon, aba)"]
\end{tikzcd}$$

Figure 4: Graph of graphs of the Fibonacci set S.

Since any extension graphs of a bispecial word in S_σ is obtained by projection of its uniform generalized uniform graph, we have the following corollary.

Corollary 28. *Given a primitive aperiodic morphism $\sigma : A^* \to A^*$, one can decide whether S_σ is acyclic and whether it is connected. Therefore we can decide whether it is a tree set.*

We can acutally decide any property of S_σ that depends only on the shape of the extension graphs of words in S_σ. Given a particular extension graph, we can also decide whether there exists some word in S_σ for which it is the extension graph and we can exactly describe all words in S_σ for which this is the case.

Example 29. Let σ_F and S be as in Example 1. The only two possible extension graphs of a bispecial word in S are showed in Figure 5: the graph on the left is obtained by projection of the graphs $E_{4,4}(a), E_{4,4}(abaaba)$ and H, while the graph of the right is obtained as projection of the graphs $E_{4,4}(\varepsilon), E_{4,4}(aba)$ and G (see Example 27).

Figure 5: The two possible extension graphs of a word of S.

By checking this two graphs we can see that the Fibonacci set S is a planar tree set with respect to the orders $a <_1 b$ and $b <_2 a$.

References

[BDFD$^+$15a] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. Acyclic, connected and tree sets. *Monatshefte für Mathematik*, 176(4):521–550, 2015.

[BDFD$^+$15b] V. Berthé, C. De Felice, F. Dolce, J. Leroy, D. Perrin, C. Reutenauer, and G. Rindone. Bifix codes and interval exchanges. *Journal of Pure and Applied Algebra*, 219(7):2781 – 2798, 2015.

[Cas97] J. Cassaigne. Complexity and special factors. (complexité et facteurs spéciaux.). *Bulletin of the Belgian Mathematical Society - Simon Stevin*, 4(1):67–88, 1997.

[DL] F. Durand and J. Leroy. The constant of recognizability is computable for primitive morphisms. *Preprint*.

[DP16] F. Dolce and D. Perrin. Neutral and tree sets of arbitrary characteristic. *Theoretical Computer Science*, pages –, 2016.

[HJ90] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[Klo12] K. Klouda. Bispecial factors in circular non-pushy d0l languages. *Theor. Comput. Sci.*, 445:63–74, 2012.

[KM16] K. Klouda and K. Medková. Synchronizing delay for binary uniform morphisms. *Theoret. Comput. Sci.*, 615:12–22, 2016.

[Mos92] B. Mossé. Puissances de mots et reconnaissabilité des points fixes d'une substitution. *Theoretical Computer Science*, 99:327–334, 1992.

[Mos96] B. Mossé. Reconnaissabilité des substitutions et complexité des suites automatiques. 124:329–346, 1996.