ON THE TRANSITION SEMIGROUPS OF CENTRALLY LABELED RAUZY GRAPHS

JORGE ALMEIDA AND ALFREDO COSTA

Abstract. Rauzy graphs of subshifts are endowed with an automaton structure. For Sturmian subshifts, it is shown that its transition semigroup is the syntactic semigroup of the language recognized by the automaton. A projective limit of the partial semigroups of nonzero regular elements of their transition semigroups is described. If the subshift is minimal, then this projective limit is isomorphic, as a partial semigroup, to the $\mathcal{J}$-class associated to it in the free pro-aperiodic semigroup.

1. Introduction

The *Rauzy graphs* of a subshift $\mathcal{X}$ are finite graphs that may be seen as approximations of the subshift [18, 8]. The vertices are windows (say of length $2n$), centered at the origin, to the biinfinite words in the subshift, and the edges correspond to the shift of origin in some biinfinite word. A natural labeling for Rauzy graphs is thus to label each edge by the letter from which the origin is moved. We call the result a *centrally labeled Rauzy graph*, which we view as a non-deterministic automaton in which all vertices are considered to be both initial and final. The language $F_n(\mathcal{X})$ it recognizes consists of all words all of whose factors of length $2n+1$ are blocks of the subshift, together with their factors.

Rauzy graphs of minimal subshifts may be very complicated. Indeed, by a result of Salimov [19], every sequence of finite strongly connected directed graphs with bounded in and out-degrees, has a subsequence in which each graph admits a uniform edge subdivision which is isomorphic to a Rauzy graph of some minimal subshift.

From the point of view of semigroup theory, a first question that comes up concerning centrally labeled Rauzy graphs is whether their transition semigroups $T_n(\mathcal{X})$ are of some significance. In particular, do they coincide with the syntactic semigroups of the languages $F_n(\mathcal{X})$? While the answer is negative in general, even for minimal subshifts, in one of the main results

2010 *Mathematics Subject Classification.* Primary 20M05, 20M35, 37B10; Secondary 20M07, 68R15.

Key words and phrases. Rauzy graph, minimal subshift, irreducible subshift, automaton, transition semigroup, free profinite semigroup, syntactic semigroup, aperiodic semigroup.

The authors acknowledge the support, respectively, of the Centro de Matemática da Universidade do Porto and of the Centro de Matemática da Universidade de Coimbra, financed by FCT through the programmes POCTI and POSI, with Portuguese and European Community structural funds, as well as the support of the FCT project PTDC/MAT/65481/2006, within the framework of programmes COMPETE and FEDER.

of this paper (Theorem 4.7), we show that the two semigroups coincide for Sturmian subshifts, a class which has deserved much attention [15, 13].

It turns out that there is a natural homomorphism from $T_m(\mathcal{X})$ onto $T_n(\mathcal{X})$ for $m \geq 2n$. In the case of an irreducible proper subshift, the semigroup $T_n(\mathcal{X})$ has a zero and all its nonzero regular elements are $\mathcal{J}$-equivalent, that is they generate the same ideal. We consider these $\mathcal{J}$-classes as partial semigroups. It is thus natural to consider the inverse system of partial semigroups of nonzero regular elements of $T_n(\mathcal{X})$, with the above connecting homomorphisms. Our second main result (Corollary 5.14) states that, for an arbitrary minimal subshift, its inverse limit is the $\mathcal{J}$-class $\mathcal{J}(\mathcal{X})$ of the free profinite aperiodic semigroup which is in natural correspondence with the subshift [3]. This result sheds some light over the distribution of idempotents in $\mathcal{J}(\mathcal{X})$. For a minimal subshift of sublinear complexity, we also show that in each $\mathcal{R}$-class and in each $\mathcal{L}$-class of $\mathcal{J}(\mathcal{X})$ there is only a bounded number of idempotents and for only countably many of them there may be more than one idempotent (Corollary 5.9). These idempotents turn out to be in natural bijection with the biinfinite words in $\mathcal{X}$.

The natural bijection between minimal subshifts over a finite alphabet and the regular $\mathcal{J}$-classes of relatively free profinite semigroups all of whose proper factors are nonregular holds for any pseudovariety V containing all finite local semilattices. For the absolutely free profinite semigroup, the maximal subgroups of $\mathcal{J}(\mathcal{X})$ have also been extensively investigated [3, 6]. The distribution of idempotents in $\mathcal{J}(\mathcal{X})$ is independent of the pseudovariety V. Thus, our present results are somewhat complementary to the study of the maximal subgroups.

2. Preliminaries

2.1. **Subshifts.** We assume familiarity with the basics of semigroup theory [17, 1]. Let A be a finite alphabet. All alphabets in this paper are finite. As in [1], if $w \in A^+$ and $|w| \geq n$, the prefix of length n of w is denoted by $\mathrm{i}_n(w)$; if $|w| < n$ then $\mathrm{i}_n(w) = w$. Dually, for suffixes one considers the map t_n.

Endow A with the discrete topology and $A^{\mathbb{Z}}$ with the product topology. The *shift* on $A^{\mathbb{Z}}$ is the homeomorphism $\sigma : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ sending $(x_i)_{i\in\mathbb{Z}}$ to $(x_{i+1})_{i\in\mathbb{Z}}$. A *subshift* of $A^{\mathbb{Z}}$ is a nonempty closed subspace $\mathcal{X}$ of $A^{\mathbb{Z}}$ such that $\sigma(\mathcal{X}) = \mathcal{X}$. Let $x \in A^{\mathbb{Z}}$. By a *factor* or *block* of $(x_i)_{i\in\mathbb{Z}}$ we mean a word $x_i x_{i+1} \cdots x_{i+n-1} x_{i+n}$ (briefly denoted by $x_{[i,i+n]}$), where $i \in \mathbb{Z}$ and $n \geq 1$. If $\mathcal{X}$ is a subset of $A^{\mathbb{Z}}$ then $L(\mathcal{X})$ denotes the set of factors of elements of $\mathcal{X}$, and $L_n(\mathcal{X})$ the set of elements of $L(\mathcal{X})$ with length n.

A subset K of a semigroup S is *factorial* if it is closed under taking factors of its factors; *prolongable* if, for every $s \in K$, there exist $u, v \in S$ such that $us, sv \in K$; *irreducible* if, for all $s, t \in K$, there exists $u \in S$ such that $sut \in K$. An irreducible set is prolongable. The map $\mathcal{X} \mapsto L(\mathcal{X})$ is a bijection from the set of subshifts of $A^{\mathbb{Z}}$ to the set of nonempty factorial prolongable languages of A^+. The subshift $\mathcal{X}$ is *irreducible* if $L(\mathcal{X})$ is irreducible. From hereon, $\mathcal{X}$ is a subshift of $A^{\mathbb{Z}}$.

The subshift $\mathcal{X}$ is *minimal* if it does not contain other subshifts. Minimal subshifts are irreducible. A *substitution* over an alphabet A is an endomorphism of A^+. It is said to be *primitive* if, for some $n \geq 1$, and for all $a \in A$, the words $\varphi^n(a)$ have the same letters and $\lim |\varphi^k(a)| = \infty$. For a primitive substitution, the set of factors of words in $\{\varphi^n(a) : n \geq 1\}$ is independent of $a \in A$; it is equal to $L(\mathcal{X}_\varphi)$ for a unique minimal subshift $\mathcal{X}_\varphi$. One has $\varphi(L(\mathcal{X}_\varphi)) \subseteq L(\mathcal{X}_\varphi)$. For the two-letter alphabet $A = \{a, b\}$, and for $u, v \in A^+$, denote by $[u, v]$ the unique endomorphism φ of A^+ such that $\varphi(a) = u$ and $\varphi(b) = v$.

The *Rauzy graph of order n of $\mathcal{X}$* is the directed graph whose set of edges is $L_{n+1}(\mathcal{X})$, whose set of vertices is $L_n(\mathcal{X})$, and such that an edge w has origin in $\mathrm{i}_n(w)$ and terminus in $\mathrm{t}_n(w)$. It is denoted by $\Sigma_n(\mathcal{X})$. Note that in $\Sigma_n(\mathcal{X})$ all vertices have positive in and out-degree. When $\mathcal{X} = A^{\mathbb{Z}}$ this graph is usually called the *De Bruijn graph* of order n.

All automata in this paper are finite, and every state is initial and final. Since the elements of $L_{2n+1}(\mathcal{X})$ have odd length, we can assign to each edge of $\Sigma_{2n}(\mathcal{X})$ its middle letter. This defines a nondeterministic automaton over the alphabet A, also denoted $\Sigma_{2n}(\mathcal{X})$, with transitions

$$a_1a_2\dots a_{2n} \xrightarrow{a_{n+1}} a_2\dots a_{2n}a_{2n+1}, \quad a_i \in A,$$

defined precisely when $a_1a_2\dots a_{2n}a_{2n+1}$ belongs to $L_{2n+1}(\mathcal{X})$. We call this automaton the *centrally labeled Rauzy graph of order $2n$ of $\mathcal{X}$*. See the example in Figure 1 of centrally labeled Rauzy graphs of the subshift defined given by the *Fibonacci substitution* $[ab, a]$ [15].

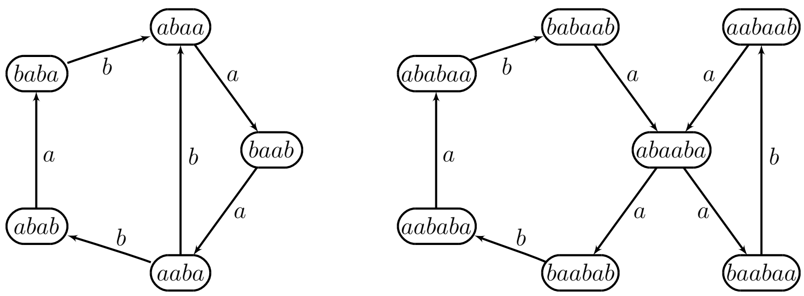

FIGURE 1. Centrally labeled Rauzy graphs of order 4 and 6 associated to the Fibonacci substitution $[ab, a]$.

Let us denote by $F_n(\mathcal{X})$ the language of the words that can be read in the automaton $\Sigma_{2n}(\mathcal{X})$. Observe that this language is also factorial and prolongable. Note also that $w \in F_n(\mathcal{X})$ if and only if $w \in L(\mathcal{X})$ or $|w| \geq 2n+1$ and all factors of w with length $2n+1$ belong to $L(\mathcal{X})$. In general, one has $L(\mathcal{X}) \subsetneq F_n(\mathcal{X})$. Since

$$F_1(\mathcal{X}) \supseteq F_2(\mathcal{X}) \supseteq F_3(\mathcal{X}) \supseteq \dots \quad \text{and} \quad \bigcap_{n\geq 1} F_n(\mathcal{X}) = L(\mathcal{X}) \tag{2.1}$$

the automatons $\Sigma_{2n}(\mathcal{X})$ may be considered as approximative devices for the study of $\mathcal{X}$.

The transition semigroup of $\Sigma_{2n}(\mathcal{X})$ will be denoted by $T_n(\mathcal{X})$, and the corresponding transition homomorphism by η_n. If $\mathcal{X} \neq A^{\mathbb{Z}}$ then the empty relation belongs to $T_n(\mathcal{X})$, and it is a zero of $T_n(\mathcal{X})$. Moreover, in that case we have $\eta_n^{-1}(0) = A^+ \setminus F_n(\mathcal{X})$. For not having to deal with the case $\mathcal{X} = A^{\mathbb{Z}}$ separately, we will always assume that $\mathcal{X} \subsetneq A^{\mathbb{Z}}$, if necessarily enlarging the alphabet. With this supposition, $T_n(\mathcal{X})$ does not depend on A.

The *right context* (respectively, *left context*) of a vertex p of an automaton, denoted by $C_R(p)$ (respectively, by $C_L(p)$) is the set of words that label paths beginning in p (respectively, ending in p).

For convenience, if u is a word of length $2n$, the letter of u at position i from the left is denoted u_{i-n-1}. For example, if u has length four, then $u = u_{-2}u_{-1}u_0u_1$. The word $u_k u_{k+1} \dots u_{k+l-1} u_{k+l}$ is denoted $u_{[k,k+l]}$.

The following remarks will be useful.

Remark 2.1. *The following properties of $\Sigma_{2n}(\mathcal{X})$ hold:*

(1) if $p \in \operatorname{Dom} \eta_n(u)$ then $\mathrm{i}_n(p)\, u \in F_n(\mathcal{X})$;
(2) if $q \in \operatorname{Im} \eta_n(u)$ then $u\, \mathrm{t}_n(q) \in F_n(\mathcal{X})$;
(3) if $(p, q) \in \eta_n(u)$ then $\mathrm{i}_n(p)\, u\, \mathrm{t}_n(q) \in F_n(\mathcal{X})$.

Proof. For every vertex p of $\Sigma_{2n}(\mathcal{X})$ we have $\mathrm{i}_n(p) \in C_L(p)$. Therefore, if $p \in \operatorname{Dom} \eta_n(u)$ then there is path labeled $\mathrm{i}_n(p)\, u$ in $\Sigma_{2n}(\mathcal{X})$, which proves (1). The remaining properties can be checked similarly. □

Remark 2.2. *Let $u, v, w \in A^+$. If $uwv \in F_n(\mathcal{X})$ and u and v have length n, then there is a path labeled w from vertex $u\, \mathrm{i}_n(wv)$ to vertex $\mathrm{t}_n(uw)v$.*

Remark 2.3. *Let $p \in L_{2n}(\mathcal{X})$. Consider the automaton $\Sigma_{2n}(\mathcal{X})$. Let w be a word. Let $k = \min\{n, |w|\}$. Then $w \in C_R(p)$ implies $\mathrm{i}_k(w) = p_{[0,k-1]}$ and $w \in C_L(p)$ implies $\mathrm{t}_k(w) = p_{[-k,-1]}$.*

2.2. **Local automata.** Let ℓ and r be nonnegative integers. An automaton $\mathbb{A}$ is (ℓ, r)-*local* [9, Section 10.3] if for any paths $p \xrightarrow{u} q \xrightarrow{v} r$ and $p' \xrightarrow{u} q' \xrightarrow{v} r'$ with $|u| = \ell$ and $|v| = r$, one has $q = q'$. An automaton is *local* if it is (ℓ, r)-*local* for some $\ell, r \geq 0$. The well known De Bruijn automaton $\Sigma_{2n}(A^{\mathbb{Z}})$ is an important example of a (n, n)-local automaton. Since a subautomaton of a (ℓ, r)-local automaton is (ℓ, r)-local, the automaton $\Sigma_{2n}(\mathcal{X})$ is also (n, n)-local.

A word $u \in A^+$ is *synchronizing* for an automaton $\mathbb{A}$ over A, with transition homomorphism μ, if $\operatorname{Dom} \mu(u) \times \operatorname{Im} \mu(u) \subseteq \mu(u)$. We also say that $\mu(u)$ is a synchronizing element of the transition semigroup of $\mathbb{A}$. The following fact will be useful (cf. the proof of [9, Proposition 10.3.11]).

Lemma 2.4. *Suppose $\mathbb{A}$ is (ℓ, r)-local and let $u \in A^+$ be a word with length at least $\ell + r$. Then u is synchronizing.*

Moreover, one easily verifies that if u is synchronizing then $\mu(u)^2 = 0$ or $\mu(u)^2 = \mu(u)$. Therefore, the transition semigroup of a local automata is aperiodic.

Lemma 2.5. *Let $\mathbb{A}$ be a local (ℓ, r)-automaton with transition homomorphism μ. Suppose that $\mu(u)$ and $\mu(v)$ are not the empty relation. If $\mathrm{i}_{\ell+r}(u) = \mathrm{i}_{\ell+r}(v)$ and $\mathrm{t}_{\ell+r}(u) = \mathrm{t}_{\ell+r}(v)$ then $\mu(u) = \mu(v)$.*

Proof. Suppose that $|u|, |v| \geq \ell + r$ (otherwise $u = v$). Let $p \in \operatorname{Dom} \mu(u)$. Let $w \in A^*$ be such that $v = \mathrm{i}_{\ell+r}(v)w$. Since $\mu(v) \neq \emptyset$, there are states q, s, t

such that $(q,s) \in \mu(\mathrm{i}_{\ell+r}(v))$ and $(s,t) \in \mu(w)$. Since $\mathrm{i}_{\ell+r}(u) = \mathrm{i}_{\ell+r}(v)$ is a synchronizing word and $p \in \operatorname{Dom}\mu(\mathrm{i}_{\ell+r}(u))$, one has $(p,s) \in \mu(\mathrm{i}_{\ell+r}(v))$, thus $(p,t) \in \mu(v)$. Hence $\operatorname{Dom}\mu(u) \subseteq \operatorname{Dom}\mu(v)$. By symmetry, one concludes that $\operatorname{Dom}\mu(u) = \operatorname{Dom}\mu(v)$ and $\operatorname{Im}\mu(u) = \operatorname{Im}\mu(v)$ is obtained similarly. Hence $\mu(u) = \mu(v)$, since every word of length at least $\ell + r$ is synchronizing. □

If $\mathcal{X}$ is irreducible, then $\Sigma_{2n}(\mathcal{X})$ is strongly connected. It is in this context that the following result will be used. It is a special case of [9, Theorem 9.3.10], taking into account [9, Proposition 10.3.11].

Proposition 2.6. *Let $\mathbb{A}$ be a strongly connected (ℓ, r)-local automaton, with transition homomorphism μ. Then the set $\{\mu(u) : |u| \geq \ell + r\} \setminus \{0\}$ is a 0-minimum $\mathcal{J}$-class of the transition semigroup of $\mathbb{A}$, which is regular and contains all nonzero regular elements of $\mu(A^+)$.*

For an irreducible subshift $\mathcal{X}$, denote by $\mathcal{J}_n(\mathcal{X})$ the 0-minimum $\mathcal{J}$-class of $T_n(\mathcal{X})$.

3. Relationship between $T_m(\mathcal{X})$ and $T_n(\mathcal{X})$

To describe the relationship between $T_m(\mathcal{X})$ and $T_n(\mathcal{X})$ when $m \geq n$, we start with some preparatory remarks.

Remark 3.1. *For the case of centrally labeled Rauzy graphs, Lemma 2.5 states the following: given $u, v \in F_n(\mathcal{X})$, if $(\mathrm{i}_{2n}(u), \mathrm{t}_{2n}(u)) = (\mathrm{i}_{2n}(v), \mathrm{t}_{2n}(v))$ then $\eta_n(u) = \eta_n(v)$.*

The following consequence of Remark 2.3 will be useful.

Remark 3.2. *Let $u, v \in F_n(\mathcal{X})$. Suppose that $k \leq n$ and $|u|, |v| \geq k$. If $\eta_n(u) = \eta_n(v)$ then we have $(\mathrm{i}_k(u), \mathrm{t}_k(u)) = (\mathrm{i}_k(v), \mathrm{t}_k(v))$.*

Remark 3.2 cannot be extended to local automata: consider for example the local automaton with a unique state q and two loops at q labeled by two distinct letters. We proceed by giving examples of minimal subshifts defined by primitive substitutions showing how Remarks 3.1 and 3.2 cannot be improved.

Example 3.3. Consider the example of the Fibonacci substitution $[ab, a]$ (see Figure 1). Let $u = aba$ and $v = ababa$. Then $u, v \in L(\mathcal{X}_{[ab,a]})$.

For $n = 2$ we have $aba = \mathrm{i}_{2n-1}(u) = \mathrm{i}_{2n-1}(v) = \mathrm{t}_{2n-1}(u) = \mathrm{t}_{2n-1}(v)$, but $\eta_n(u) \neq \eta_n(v)$, since $(ba^2b, baba) \in \eta_n(u) \setminus \eta_n(v)$. Hence Remark 3.1 cannot be improved.

For $n = 3$ we have $\eta_n(v) = \{(u^2, u^2)\} = \eta_n(vu^2)$. Since $|v| > n$ and $\mathrm{t}_{n+1}(v) \neq \mathrm{t}_{n+1}(vu^2)$, this shows that Remark 3.2 cannot be extended to the case $k = n + 1$.

Example 3.4. Let $\varphi = [bababbaba, ba]$. The centrally labeled Rauzy graph $\Sigma_6(\mathcal{X}_\varphi)$ is represented in Figure 2. Then $\eta_3(bb) = \eta_3(bbabababb)$ but $\mathrm{i}_3(bb) \neq \mathrm{i}_3(bbabababb)$. Hence in Remark 3.2 the hypothesis $|u|, |v| \geq k$ cannot be dropped.

Let m and n be positive integers such that $m \geq n$. Let u be a word of length $2m$. Then $\pi_{m,n}(u)$ denotes the word $u_{[-n,n-1]}$ of length $2n$.

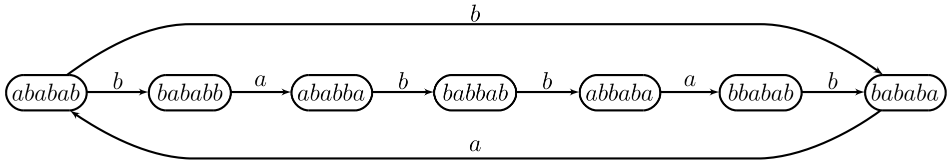

FIGURE 2. Rauzy graph associated to $\varphi = [bababbaba, ba]$.

Remark 3.5. *If* $p_0 \xrightarrow{a_1} p_1 \xrightarrow{a_2} \cdots \xrightarrow{a_{k-1}} p_{k-1} \xrightarrow{a_k} p_k$ *is a path in* $\Sigma_{2m}(\mathcal{X})$ *then* $\pi_{m,n}(p_0) \xrightarrow{a_1} \pi_{m,n}(p_1) \xrightarrow{a_2} \cdots \xrightarrow{a_{k-1}} \pi_{m,n}(p_{k-1}) \xrightarrow{a_k} \pi_{m,n}(p_k)$ *is a path in* $\Sigma_{2n}(\mathcal{X})$.

Proposition 3.6. *Let* m *and* n *be positive integers such that* $m \geq 2n$. *Let* $u, v \in F_m(\mathcal{X})$. *If* $\eta_m(u) = \eta_m(v)$ *then* $\eta_n(u) = \eta_n(v)$.

Proof. The case in which $|u| \geq 2n$ and $|v| \geq 2n$ is a direct consequence of Remarks 3.2 and 3.1.

Without loss of generality, we suppose that $|u| \leq |v|$ and $|u| < 2n$. Since $|u| < m$ and $\eta_m(u) = \eta_m(v) \neq \emptyset$, by Remark 2.3 we know that u is a prefix and a suffix of v. Hence, if $|u| = |v|$ then $u = v$. Assuming that $|u| < |v|$, there is a nonempty word z such that $v = uz$. Thus we have $\eta_m(u) = \eta_m(uz)$ and so $\eta_m(u) = \eta_m(u)\eta_m(z)^k$ holds for every positive integer k. Therefore, there is a word u' of length greater than $2|u|$ such that $\eta_m(u) = \eta_m(u')$. Again by Remark 2.3, we have $u' = uwu$ for some nonempty word w. Then $\eta_m(u) = \eta_m(u)\eta_m(wu)$, which implies $\eta_m(u) = \eta_m(u)\eta_m(wu)^k$ for all $k \geq 1$.

We claim that the equality

$$\eta_n(u(wu)^m) = \eta_n(u) \tag{3.1}$$

holds. This equality holds if and only if $\eta_n(u) \cup \eta_n(u(wu)^m) \subseteq \eta_n(u) \cap \eta_n(u(wu)^m)$. Therefore, since $\eta_n(u) \cup \eta_n(u(wu)^m) \subseteq \operatorname{Dom} \eta_n(u) \times \operatorname{Im} \eta_n(u)$, to establish (3.1) it suffices to prove that

$$\operatorname{Dom} \eta_n(u) \times \operatorname{Im} \eta_n(u) \subseteq \eta_n(u) \cap \eta_n(u(wu)^m). \tag{3.2}$$

We consider an arbitrary element (p, q) of $\operatorname{Dom} \eta_n(u) \times \operatorname{Im} \eta_n(u)$ and to prove that it belongs to $\eta_n(u) \cap \eta_n(u(wu)^m)$, we distinguish two cases.

We first suppose that $n < |u| < 2n$. Then there are words u_1, u_2 and u_3 such that $u = u_1u_2u_3$, $|u_1| = |u_3| = |u| - n$ and $|u_2| = 2n - |u|$. By hypothesis, there are in $\Sigma_{2n}(\mathcal{X})$ paths of the following forms:

$$p \xrightarrow{u_1} p' \xrightarrow{u_2u_3} \bullet \qquad \text{and} \qquad \bullet \xrightarrow{u_1u_2} q' \xrightarrow{u_3} q. \tag{3.3}$$

And since $|u_1u_2| = |u_2u_3| = n$, by Remark 2.3 we have $p' = xu$ and $q' = uy$ for some $x, y \in A^+$. As $xu \in L(\mathcal{X})$ and $uy \in L(\mathcal{X})$, there are in $\Sigma_{2m}(\mathcal{X})$ paths

$$\bullet \xrightarrow{x} \bar{p} \xrightarrow{u} \bullet \qquad \text{and} \qquad \bullet \xrightarrow{u} \bar{q} \xrightarrow{y} \bullet.$$

Since $\eta_m(u) = \eta_m(u)\eta_m(wu)^k$ for all $k \geq 1$, the relation $\eta_m(u)$ is synchronizing by Lemma 2.4. Therefore, there is a path in $\Sigma_{2m}(\mathcal{X})$ as follows:

$$\bullet \xrightarrow{x} \bar{p} \xrightarrow{u_1} p'' \xrightarrow{u_2} q'' \xrightarrow{u_3} \bar{q} \xrightarrow{y} \bullet. \tag{3.4}$$

Then, since $\eta_m(u(wu)^m) = \eta_m(u)$, there is also in $\Sigma_{2m}(\mathcal{X})$ a path of the following form:

$$\bullet \xrightarrow{x} \bar{p} \xrightarrow{u_1} p''' \xrightarrow{u_2u_3(wu)^{m-1}wu_1u_2} q''' \xrightarrow{u_3} \bar{q} \xrightarrow{y} \bullet. \tag{3.5}$$

From (3.4) and (3.5) and Remark 3.5, we conclude the existence of the following paths in $\Sigma_{2n}(\mathcal{X})$:

$$\bullet \xrightarrow{x} \pi_{m,n}(\bar{p}) \xrightarrow{u_1} \pi_{m,n}(p'') \xrightarrow{u_2} \pi_{m,n}(q'') \xrightarrow{u_3} \pi_{m,n}(\bar{q}) \xrightarrow{y} \bullet \tag{3.6}$$

(3.7)

$$\bullet \xrightarrow{x} \pi_{m,n}(\bar{p}) \xrightarrow{u_1} \pi_{m,n}(p''') \xrightarrow{u_2u_3(wu)^{m-1}wu_1u_2} \pi_{m,n}(q''') \xrightarrow{u_3} \pi_{m,n}(\bar{q}) \xrightarrow{y} \bullet$$

Since $p' = xu$ and $|u_2u_3| = n$, we have $|xu_1| = n$. Similarly, $|u_3y| = n$. Therefore, from (3.6), (3.7) and Remark 2.3, one concludes that $\pi_{m,n}(p''') = \pi_{m,n}(p'') = p'$ and $\pi_{m,n}(q''') = \pi_{m,n}(q'') = q'$. Then, again by (3.6) and (3.7), we have $(p', q') \in \eta_n(u_2) \cap \eta_n(u_2u_3(wu)^{m-1}wu_1u_2)$. Since $(p, p') \in \eta_n(u_1)$ and $(q', q) \in \eta_n(u_3)$ (by (3.3)), we conclude that $(p, q) \in \eta_n(u) \cap \eta_n(u(wu)^m)$.

It remains to consider the case $|u| \leq n$. Let $p' \in \pi^{-1}_{m,n}(p)$ and $q' \in \pi^{-1}_{m,n}(q)$. Since $|u| \leq n$ and $p \in \operatorname{Dom} \eta_n(u)$, we know that u is a prefix of the suffix of length n of p. Hence, u is a prefix of the suffix of length m of p'. Since in p' begin paths of arbitrarily large length bigger or equal to n, and their labels all start by u, we conclude that $p' \in \operatorname{Dom} \eta_m(u)$. Similarly, one can show that $q' \in \operatorname{Im} \eta_m(u)$. Because $\eta_m(u)$ is synchronizing, we have $(p', q') \in \eta_m(u)$, thus $(p, q) \in \eta_n(u)$. And since $\eta_m(u) = \eta_m(u(wu)^m)$, we also have $(p, q) \in \eta_n(u(wu)^m)$. This concludes the proof of (3.2), and therefore the proof of the claim that (3.1) holds.

The above argument also shows that there is a nonempty word w' such that $\eta_m(v) = \eta_m(v)\eta_m(w'v)^k$ for all $k \geq 1$ and

$$\eta_n(v) = \eta_n(v(w'v)^m). \tag{3.8}$$

In particular, we have

$$\eta_m(u(wu)^m) = \eta_m(u) = \eta_m(v) = \eta_m(v(w'v)^m).$$

Then, since $|u(wu)^m| > 2n$ and $|v(w'v)^m| > 2n$, it follows from the observation made in the first paragraph of this proof that $\eta_n(u(wu)^m) = \eta_n(v(w'v)^m)$. Hence $\eta_n(u) = \eta_n(v)$, by (3.1) and (3.8). □

Example 3.7. We give an example in which $\eta_3(u) = \eta_3(v)$ and $\eta_2(u) \neq \eta_2(v)$, thus showing the need of condition $m \geq 2n$ in Proposition 3.6. Let A be the two-letter alphabet $\{a, b\}$. Let $u = b^2a^2b^2$ and $v = b^2ab^4u$. Consider the periodic subshift $\mathcal{P}$ whose set of blocks are the factors of elements of $(a^2b^7ab^7ab^2a^2bab^7)^+$. Then $\eta_3(u) = \{(b^5a, ab^5)\} = \eta_3(v)$ but $(bab^2, b^4) \in \eta_2(u) \setminus \eta_2(v)$. See Figure 3 for the corresponding Rauzy graphs.

Proposition 3.8. *Let $\mathcal{X}$ be an irreducible subshift. Let m and n be positive integers such that $m \geq 2n$. If $\eta_m(v) \in \mathcal{J}_m(\mathcal{X})$ then $\eta_n(v) \in \mathcal{J}_n(\mathcal{X})$.*

Proof. If $\eta_m(v) \in \mathcal{J}_m(\mathcal{X})$ then $\eta_m(v) = \eta_m(u)$ for some word u such that $|u| \geq 2n$, because $\mathcal{J}_m(\mathcal{X})$ is regular. In view of Proposition 3.6, it follows that $\eta_n(v) = \eta_n(u)$ so that, by Proposition 2.6, we have $\eta_n(v) \in \mathcal{J}_n(\mathcal{X})$. □

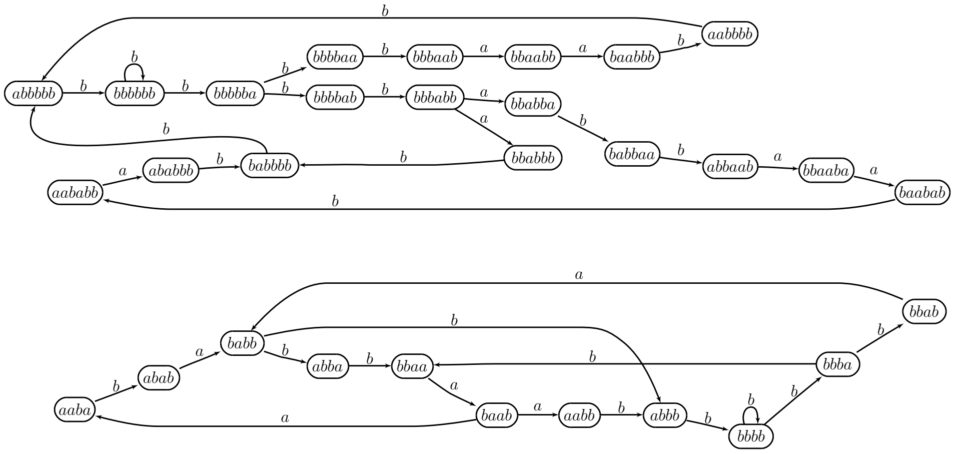

FIGURE 3. Rauzy graphs of $\mathcal{P}$.

4. Relationship with the syntactic semigroup

If u is a word on the two-letter alphabet $A = \{a, b\}$ such that every word of length three is a factor of u, and φ is the primitive substitution $[u, a]$, then $L_3(\mathcal{X}_\varphi) = A^3$. Therefore, the syntactic semigroup of $F_1(\mathcal{X}_\varphi)$ is trivial, but $\eta_1(a) \neq \eta_1(b)$. So, in general $T_n(\mathcal{X})$ does not coincide with the syntactic semigroup of $F_n(\mathcal{X})$, even for minimal subshifts. In this section we prove they are equal when $\mathcal{X}$ is a *Sturmian subshift* [15], a widely studied type of subshift whose definition we next recall (for more informations, see [15, Chapter 2] or [13, Chapter 6]).

An element u of $L(\mathcal{X})$ is *right special* if there are at least two distinct letters a and b such that $ua, ub \in L(\mathcal{X})$. *Left special* words are defined dually. A subshift of $\{a, b\}^{\mathbb{Z}}$ is *Sturmian* if for every n there is a unique right special word in $L_n(\mathcal{X})$. This means that all but one of the vertices of the Rauzy graph $\Sigma_n(\mathcal{X})$ have out-degree one, the exceptional one having out-degree two (Figure 4). Equivalently, all vertices of $\Sigma_n(\mathcal{X})$ have in-degree one, except one, with in-degree two, that is, for every n there is a unique left special word in $L_n(\mathcal{X})$. Sturmian subshifts are minimal subshifts.

A primitive substitution $[u, v]$ over $\{a, b\}$ defines a Sturmian subshift precisely when $\{u, v\}$ generates the free group on $\{a, b\}$ [15, Theorem 2.3.23]. Therefore, the Fibonacci substitution and the substitution $[bababbaba, ba]$ (cf. Figure 2) are Sturmian, while the *Prouhet-Thue-Morse substitution* $[ab, ba]$ (see [15]) is not.

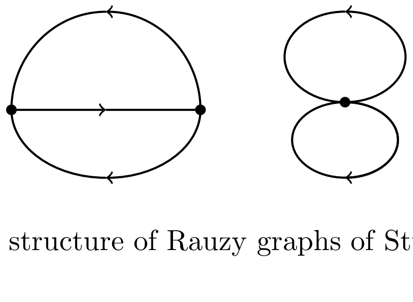

FIGURE 4. Cycle structure of Rauzy graphs of Sturmian subshifts.

For a subshift $\mathcal{X}$, we denote by $\equiv_n$ the syntactic congruence of $F_n(\mathcal{X})$. We proceed to deduce a series of technical lemmas relating $\equiv_n$ with η_n, culminating in the proof that η_n is the syntactic homomorphism in the case of a Sturmian subshift (Theorem 4.7).

Lemma 4.1. *Let $\mathcal{X}$ be a subshift. Let $u, v \in F_n(\mathcal{X})$ be words of length greater than or equal to n such that $u \equiv_n v$. Then $\eta_n(u) = \eta_n(v)$ if and only if $\mathrm{i}_n(u) = \mathrm{i}_n(v)$ and $\mathrm{t}_n(u) = \mathrm{t}_n(v)$.*

Proof. The "only if" part is in Remark 3.2. Conversely, suppose $\mathrm{i}_n(u) = \mathrm{i}_n(v)$ and $\mathrm{t}_n(u) = \mathrm{t}_n(v)$. Let $(p, q) \in \eta_n(u)$. Then $\mathrm{i}_n(p)\, u\, \mathrm{t}_n(q) \in F_n(\mathcal{X})$ by Remark 2.1. Since $u \equiv_n v$, the word $w = \mathrm{i}_n(p)\, v\, \mathrm{t}_n(q)$ also belongs to $F_n(\mathcal{X})$. By Remark 2.3 and our hypothesis, $\mathrm{t}_n(p) = \mathrm{i}_n(u) = \mathrm{i}_n(v)$ and $\mathrm{i}_n(q) = \mathrm{t}_n(u) = \mathrm{t}_n(v)$. Therefore $\mathrm{i}_{2n}(w) = p$ and $\mathrm{t}_{2n}(w) = q$. Hence v labels a path in $\Sigma_{2n}(\mathcal{X})$ from p to q by Remark 2.2, that is, we have $(p, q) \in \eta_n(v)$. By symmetry, $\eta_n(u) = \eta_n(v)$. □

According to the next lemma, in the case of irreducible subshifts, to prove that η_n and the syntactic homomorphism are equal, it suffices to show that they take the same value for sufficiently large words.

Lemma 4.2. *Let $\mathcal{X}$ be an irreducible subshift of $A^{\mathbb{Z}}$ and let n and N be positive integers. Suppose that, for all $u, v \in F_n(\mathcal{X})$ such that $|u| > N$ and $|v| > N$, we have $\eta_n(u) = \eta_n(v)$ whenever $u \equiv_n v$. Then, for all $u, v \in A^+$, we have $\eta_n(u) = \eta_n(v)$ if and only if $u \equiv_n v$.*

Proof. Since η_n recognizes $F_n(\mathcal{X})$, one has $u \equiv_n v$ whenever $\eta_n(u) = \eta_n(v)$.

Conversely, let u and v be elements of A^+ such that $u \equiv_n v$. Since $\eta_n(x) = \emptyset$ if and only if $x \in A^+ \setminus F_n(\mathcal{X})$, we are reduced to the case where $u, v \in F_n(\mathcal{X})$. Moreover, we may as well assume that $|u| \leq |v|$.

Since u belongs to the prolongable language $F_n(\mathcal{X})$, there is $x \in A^+$ such that $ux \in F_n(\mathcal{X})$ and $|x| > N$. Note that $ux \equiv_n vx$, thus $\eta_n(ux) = \eta_n(vx)$, by hypothesis. Hence $\mathrm{i}_n(ux) = \mathrm{i}_n(vx)$, by Remark 3.2. If $|u| \geq n$, then $\mathrm{i}_n(u) = \mathrm{i}_n(v)$ and, dually, $\mathrm{t}_n(u) = \mathrm{t}_n(v)$; whence $\eta_n(u) = \eta_n(v)$, by Lemma 4.1.

It remains to consider the case $|u| < n$. Since $\mathrm{i}_n(ux) = \mathrm{i}_n(vx)$ and $|u| \leq \min\{n-1, |v|\}$, we know that u is a prefix of v. If $u = v$ then we are done, hence we may assume that $v = uz$ for some $z \in A^+$. Then $u \equiv_n uz \equiv_n uz^k$, for all $k \geq 1$, whence there is a word w with length greater than $2n$ such that $u \equiv_n v \equiv_n w$. Hence, we may assume as well that $|v| > 2n$.

Let $(p, q) \in \eta_n(u)$. Recall that we have $\eta_n(ux) = \eta_n(vx)$. In particular, $p \in \operatorname{Dom} \eta_n(v)$. Dually, $q \in \operatorname{Im} \eta_n(v)$. Since $|v| \geq 2n$, the word v is synchronizing for $\Sigma_{2n}(\mathcal{X})$ by Lemma 2.4, thus $(p, q) \in \eta_n(v)$. Therefore $\eta_n(u) \subseteq \eta_n(v)$.

Conversely, let $(p, q) \in \eta_n(v)$. Then $\mathrm{i}_n(p)\, v\, \mathrm{t}_n(q) \in F_n(\mathcal{X})$ by Remark 2.1. Since $u \equiv_n v$, the word $u' = \mathrm{i}_n(p)\, u\, \mathrm{t}_n(q)$ is in $F_n(\mathcal{X})$. The words $p' = \mathrm{i}_{2n}(u')$ and $q' = \mathrm{t}_{2n}(u')$ are such that $(p', q') \in \eta_n(u)$ by Remark 2.2. As we already proved that $\eta_n(u) \subseteq \eta_n(v)$, it follows that $(p', q') \in \eta_n(v)$. From Remark 2.3 we deduce that $\mathrm{t}_n(p') = \mathrm{i}_n(v) = \mathrm{t}_n(p)$ and since $\mathrm{i}_n(p)$ is a prefix of p', we conclude that $p = p'$. Similarly $q = q'$. Hence $\eta_n(u) = \eta_n(v)$. □

The following is a simple combinatorial property of Sturmian subshifts.

Lemma 4.3. *Let $\mathcal{X}$ be a Sturmian subshift. For each positive integer n there is a word s_n in $L_n(\mathcal{X})$ such $|L_n(\mathcal{X})s_n \cap L_{2n}(\mathcal{X})| = 1$.*

Proof. Should the lemma fail then there would be some positive integer n such that $|L_{2n}(\mathcal{X})| \geq 2|L_n(\mathcal{X})|$. However, this inequality is impossible because the number of factors of length m of a Sturmian subshift is $m + 1$ [15]. □

Let $\mathcal{X}$ be a Sturmian subshift. We shall denote by z_n the unique word in $L_n(\mathcal{X})s_n \cap L_{2n}(\mathcal{X})$, where s_n is a word as in Lemma 4.3. The unique left special and right special factors in $L_{2n}(\mathcal{X})$ will be respectively denoted by l_n and r_n.

Given two states p and q of a strongly connected graph G, we denote by $d(p, q)$ the minimal length of a (possibly empty) path from p to q. Note that, since we are considering oriented paths, we may have $d(p, q) \neq d(q, p)$.

Lemma 4.4. *Let $\mathcal{X}$ be a Sturmian subshift. Let q_1 and q_2 be vertices of $\Sigma_{2n}(\mathcal{X})$. If $C_R(q_1) \subseteq C_R(q_2)$ then $d(q_2, z_n) \leq d(q_1, z_n)$. Moreover, if $C_R(q_1) = C_R(q_2)$ then $d(q_1, l_n) = d(q_2, l_n)$.*

Proof. Suppose that $C_R(q_1) \subseteq C_R(q_2)$. Let ρ_i be a path in $\Sigma_{2n}(\mathcal{X})$ from q_i to z_n of minimum length. Let u be the label of ρ_1. By Remark 2.1, since $u\,\mathrm{t}_n(z_n) \in C_R(q_1)$ and $C_R(q_1) \subseteq C_R(q_2)$, we know that the word $v = \mathrm{i}_n(q_2)u\mathrm{t}_n(z_n)$ belongs to $F_n(\mathcal{X})$. Moreover, by Remark 2.3,

$$\mathrm{t}_n(q_1) = \mathrm{i}_n(u\,\mathrm{t}_n(z_n)) = \mathrm{t}_n(q_2), \tag{4.1}$$

whence $\mathrm{i}_{2n}(v) = q_2$. On the other hand, by the definition of z_n and the choice of v, we have $\mathrm{t}_{2n}(v) = z_n$. Hence, there is a path labeled u from q_2 to z_n by Remark 2.2. Therefore, $d(q_2, z_n) \leq |u| = d(q_1, z_n)$.

If $C_R(q_1) = C_R(q_2)$, then $d(q_1, z_n) = d(q_2, z_n)$, by the first part of the lemma. Hence, if q_1 and q_2 are in the same simple cycle (since $\mathcal{X}$ is Sturmian, each of its Rauzy graphs has exactly two simple cycles), then $q_1 = q_2$. By the same reason, if q_1 and q_2 are in distinct simple cycles, then l_n is a common vertex of ρ_1 and ρ_2, and $d(q_1, l_n) = d(q_2, l_n)$. □

The next result provides a sufficient condition for two vertices in $\Sigma_{2n}(\mathcal{X})$ to have the same right context.

Lemma 4.5. *Let $\mathcal{X}$ be a subshift. Consider its labeled Rauzy graph of order $2n$. Let $u, v \in F_n(\mathcal{X})$ be words of length greater than or equal to n such that $u \equiv_n v$. If $u \in C_L(p)$, $v \in C_L(q)$ and $\mathrm{t}_n(p) = \mathrm{t}_n(q)$, then $C_R(p) = C_R(q)$.*

Proof. Let $w \in C_R(p)$ be such that $|w| > n$. Then $w = \mathrm{t}_n(p)w'$, for some w', by Remark 2.3. Since $u \in C_L(p)$, we have $u\,\mathrm{t}_n(p)w' \in F_n(\mathcal{X})$. Hence $v\,\mathrm{t}_n(p)w' \in F_n(\mathcal{X})$, because $u \equiv_n v$. Since $v \in C_L(q)$ and $|v| \geq n$, we have $\mathrm{t}_n(v) = \mathrm{i}_n(q)$, by Remark 2.3. Moreover, since $\mathrm{t}_n(p) = \mathrm{t}_n(q)$, from $v\,\mathrm{t}_n(p)w' \in F_n(\mathcal{X})$ we deduce by Remark 2.2 that $qw' \in F_n(\mathcal{X})$ and so $w \in C_R(q)$. Since every element of $C_R(p)$ is the prefix of an element of $C_R(p)$ with length greater than n, and every prefix of an element of $C_R(q)$ also belongs to $C_R(q)$, this proves $C_R(p) \subseteq C_R(q)$. By symmetry, we get $C_R(p) = C_R(q)$. □

The following lemma isolates a rather technical tool used later in the proof of Theorem 4.7.

Lemma 4.6. *Let $\mathcal{X}$ be a Sturmian subshift. Consider its labeled Rauzy graph of order $2n$. Let $v_1, v_2 \in F_n(\mathcal{X})$ be words of length greater than or equal to n such that $v_1 \equiv_n v_2$ and $\mathrm{t}_n(v_1) \neq \mathrm{t}_n(v_2)$. For every $q_1 \in L_{2n}(\mathcal{X})$ such that $v_1 \in C_L(q_1)$, there is $q_2 \in L_{2n}(\mathcal{X})$ such that q_1 and q_2 are in distinct simple paths from r_n to l_n, $v_2 \in C_L(q_2)$, $d(q_1, l_n) = d(q_2, l_n) > 0$ and $\mathrm{t}_n(q_1) = \mathrm{t}_n(q_2)$.*

Proof. As $v_1 \in C_L(q_1)$, we have $v_1\mathrm{t}_n(q_1) \in F_n(\mathcal{X})$. Hence, from $v_1 \equiv_n v_2$ we get $v_2\mathrm{t}_n(q_1) \in F_n(\mathcal{X})$. Therefore, for $q_2 = \mathrm{t}_n(v_2)\mathrm{t}_n(q_1)$, we have $q_2 \in L_{2n}(\mathcal{X})$ and $v_2 \in C_L(q_2)$. By Lemma 4.5, we know that $C_R(q_1) = C_R(q_2)$. Therefore, $d(q_1, l_n) = d(q_2, l_n)$ by Lemma 4.4. Moreover, since $\mathrm{i}_n(q_1) = \mathrm{t}_n(v_1)$, $\mathrm{i}_n(q_2) = \mathrm{t}_n(v_2)$ and $\mathrm{t}_n(v_1) \neq \mathrm{t}_n(v_2)$, we have $q_1 \neq q_2$, thus $d(q_1, l_n)$ and $d(q_2, l_n)$ are positive.

The hypothesis of $\mathcal{X}$ being Sturmian guarantees that there are precisely two simple paths ρ_1 and ρ_2 from r_n to l_n. On the other hand, there is a unique (possibly empty) simple path ρ_0 from l_n to r_n.

Since $d(q_1, l_n) = d(q_2, l_n)$ and $q_1 \neq q_2$, one cannot have q_1 and q_2 both in ρ_1 or both on ρ_2. For the same reason, if $l_n \neq r_n$ then the vertices q_1 and q_2 are not both in ρ_0.

Hence, to prove that q_1 and q_2 are in distinct simple paths from r_n to l_n, it suffices to prove that is not possible that one of them is in $\rho_0 \setminus \{l_n, r_n\}$ and the other one is not. Without loss of generality, we assume that $q_2 \in \rho_0 \setminus \{l_n, r_n\}$ and $q_1 \in \rho_1 \setminus \rho_0$.

Since $d(q_2, l_n) = d(q_1, l_n)$, the shortest path from q_2 to l_n includes ρ_2, thus $|\rho_2| < |\rho_1|$. The sum $|\rho_0| + |\rho_1| + |\rho_2|$ is equal to $|L_{2n+1}(\mathcal{X})|$, the number of edges of $\Sigma_{2n}(\mathcal{X})$. Since $\mathcal{X}$ is Sturmian, we have $|L_{2n+1}(\mathcal{X})| = 2n + 2$. Let π be the simple cycle starting in q_2 which passes trough q_1. Then $|\pi| = |\rho_0| + |\rho_1|$. If $|\pi| \leq n + 1$ then $|\rho_2| < |\rho_1| < n + 1$, and so $|\rho_0| + |\rho_1| + |\rho_2| \leq (n + 1) + (n - 1) = 2n$, which is absurd. Therefore, $|\pi| \geq n + 2$. Let p be the last vertex in the path π' of length n which is a prefix path of π. In figure 5 one represents the four possible positions I, II, III and IV for p with respect to the positions of q_1 and q_2 and to the paths ρ_0 and ρ_1.

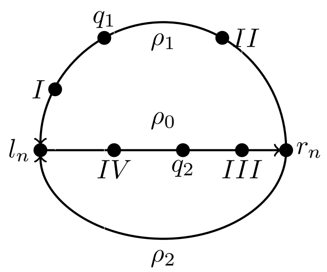

FIGURE 5. Positions I to IV for vertex p in the proof of Lemma 4.6.

The label of π' is $\mathrm{i}_n(p)$. Since $\mathrm{t}_n(p) \in C_R(p)$, we clearly have $p \in C_R(q_2)$. Whence, there is a path labeled p starting in q_1, because $C_R(q_1) = C_R(q_2)$.

In particular, there is a path ζ from q_1 to p labeled $\mathrm{i}_n(p)$ (cf. Remark 2.2). In π there is a unique path of length n ending in p (namely π'), and so ζ is not contained in π. This, together with the fact that $q_1 \in \rho_1 \setminus \rho_0$, implies that ζ contains $\rho_0 \cup \rho_2$.

If $p \notin \rho_0$ (that is, if p occupies a position like I or II in Figure 5) then, as $\rho_0 \cup \rho_2$ is contained in ζ, there is a path ζ' from q_2 to p, contained in $\pi \cap \zeta$, with length less than n. Therefore, we have two distinct paths in π from q_2 to p with length less than that of π, namely π' and ζ'. But this is impossible, because π is a simple cycle.

Therefore $p \in \rho_0$, that is, p occupies a position like III or IV in Figure 5. If p occupies a position of type III, then π' is strictly contained in ζ, contradicting that π' and ζ have length n. Therefore p occupies a position of type IV, whence ρ_1 is contained in π', and so $|\rho_1| \leq n$. On the other hand we have $|\rho_0| + |\rho_2| \leq n$, since $\rho_0 \cup \rho_2$ is contained in ζ. But put together, these inequalities contradict $|\rho_0| + |\rho_1| + |\rho_2| = 2n + 2$.

Therefore, q_1 and q_2 are in distinct simple paths from r_n to l_n, otherwise we reach a contradiction. □

We are finally ready to prove the main result of this section.

Theorem 4.7. *If $\mathcal{X}$ is a Sturmian subshift then $T_n(\mathcal{X})$ is isomorphic to the syntactic semigroup of $F_n(\mathcal{X})$.*

Proof. Let v_1 and v_2 be elements of $F_n(\mathcal{X})$ such that $v_1 \equiv_n v_2$ and $|v_1|, |v_2| > n$. We want to prove that $\mathrm{i}_n(v_1) = \mathrm{i}_n(v_2)$ and $\mathrm{t}_n(v_1) = \mathrm{t}_n(v_2)$. Indeed, if we prove this, then $\eta_n(v_1) = \eta_n(v_2)$ by Lemma 4.1; since v_1 and v_2 were chosen arbitrarily among words in $F_n(\mathcal{X})$ with length greater than n, the theorem then follows from Lemma 4.2. The reader may wish to refer to Figure 6, where various vertices constructed in the remainder of the proof are represented.

Suppose that $\mathrm{t}_n(v_1) \neq \mathrm{t}_n(v_2)$. Let $q_1 \in L_{2n}(\mathcal{X})$ be such $v_1 \in C_L(q_1)$ and choose q_2 as in Lemma 4.6. For each $i \in \{1, 2\}$, let ρ_i be the simple path from r_n to l_n to which q_i belongs. Note that $\rho_1 \neq \rho_2$ by Lemma 4.6. Without loss of generality, we may assume $|\rho_1| \leq |\rho_2|$. We further assume that $d(r_n, q_1) + d(r_n, q_2)$ is minimum for all such pairs (q_1, q_2). Let q_3 be the unique element of ρ_2 such that $d(r_n, q_3) = d(r_n, q_1)$.

Since $d(r_n, q_3) + d(q_2, l_n)$ is equal to $d(r_n, q_1) + d(q_1, l_n)$, which is the length of ρ_1, and since q_3 and q_2 lie in ρ_2, we must have

$$d(r_n, q_3) \leq d(r_n, q_2). \tag{4.2}$$

and so

$$d(q_2, l_n) \leq d(q_3, l_n). \tag{4.3}$$

We claim that $v_1 \in C_L(q_3)$. We know ρ_1 and ρ_2 are the only two simple paths from r_n to l_n, because $\mathcal{X}$ is Sturmian. For the same reason, the number of edges of $\Sigma_{2n}(\mathcal{X})$ is $2n+2$, and so $|\rho_1| \leq n+1$. And since $q_1 \neq l_n$, we thus have $d(r_n, q_1) \leq n$. There are unique paths π and π' with length $d(r_n, q_1)$ from r_n to q_1 and q_3 respectively. Note that the labels of π and π' are equal to the prefix of length $d(r_n, q_1)$ of $\mathrm{t}_n(r_n)$, by Remark 2.3. Let ρ be a path ending in q_1 labeled v_1. Since $|v_1| > d(r_n, q_1)$, we have $\rho = \rho'\pi$

for some path ρ'. Then $\rho'\pi'$ is also a path labeled by v_1, and it ends in q_3, which proves the claim.

By Lemma 4.6, we know there is a vertex q_4 in ρ_1 such that $v_2 \in C_L(q_4)$, $d(q_3, l_n) = d(q_4, l_n)$, and $\mathrm{t}_n(q_3) = \mathrm{t}_n(q_4)$. Then

$$\begin{aligned} d(r_n, q_4) + d(q_4, l_n) = d(r_n, l_n) &= d(r_n, q_1) + d(q_1, l_n) \\ &= d(r_n, q_1) + d(q_2, l_n) \\ &\leq d(r_n, q_1) + d(q_3, l_n) \quad \text{(by (4.3))} \\ &= d(r_n, q_1) + d(q_4, l_n). \end{aligned} \tag{4.4}$$

Comparing the first and last member of (4.4), we conclude that

$$d(r_n, q_4) \leq d(r_n, q_1). \tag{4.5}$$

Then, by (4.2) and (4.5), we have $d(r_n, q_3)+d(r_n, q_4) \leq d(r_n, q_1)+d(r_n, q_2)$.

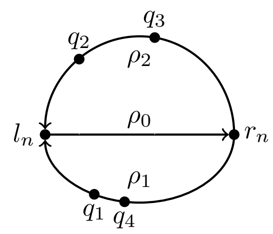

FIGURE 6. Diagram for the proof of Theorem 4.7

By minimality of $d(r_n, q_1) + d(r_n, q_2)$, it follows from (4.2) and (4.5) that $d(r_n, q_4) = d(r_n, q_1)$, thus $q_1 = q_4$. Since $v_1 \in C_L(q_1)$ and $v_2 \in C_L(q_4)$, that implies $\mathrm{t}_n(v_1) = \mathrm{t}_n(v_2)$ by Remark 2.3, which contradicts the assumption that $\mathrm{t}_n(v_1) \neq \mathrm{t}_n(v_2)$. Hence, the equality $\mathrm{t}_n(v_1) = \mathrm{t}_n(v_2)$ holds. Dually, $\mathrm{i}_n(v_1) = \mathrm{i}_n(v_2)$. □

The *Arnoux-Rauzy subshifts* [8] generalize Sturmian subshifts, since their Rauzy graphs have just one vertex with out-degree greater than one, and also just one vertex with in-degree greater than one. The proof of Lemma 4.4 stands for Arnoux-Rauzy subshifts, and most arguments used in the proofs of Lemma 4.6 and Theorem 4.7 also hold for them. However, occasionally the tighter conditions satisfied by Sturmian subshifts were used. No counter-example was found among the cases calculated by us to the validity of the extension of Theorem 4.7 to Arnoux-Rauzy subshifts, or to the Prouhet-Thue-Morse subshift, a problem we leave open.

5. Relationship with the structure of relatively free profinite semigroups

5.1. Relatively free profinite semigroups and subshifts.

A *pseudovariety of semigroups* is a class of finite semigroups closed under taking subsemigroups, homomorphic images and finitary direct products. Let V be a pseudovariety of semigroups. A topological semigroup is a *pro-*V *semigroup* if it is a projective limit of semigroups of V, finite semigroups being endowed with the discrete topology. The pseudovariety of all finite semigroups is denoted S. Pro-S semigroups are called *profinite*. See [4] for details about pro-V semigroups not included in this paper.

The category of pro-V semigroups has a free object $\overline{\Omega}_A\mathsf{V}$, the *free pro-*$\mathsf{V}$ *semigroup*. From hereon we will suppose that V contains the pseudovariety consisting of all finite semigroups whose local submonoids are semilattices. This hypothesis is essential for all results from hereon. It is also sufficient to guarantee that A^+ embeds into $\overline{\Omega}_A\mathsf{V}$ as a discrete dense subspace which we identify with its image. The elements of $\overline{\Omega}_A\mathsf{V}$ are called *pseudowords*.

Lemma 5.1 ([7, Lemma 8.2])**.** *Let $u,v \in \overline{\Omega}_A\mathsf{V}$. If x is a finite factor of uv then either x is a factor of u, or of v, or $x = sp$ for some suffix s of u and some prefix p of v. In particular, if x is a finite factor of uwv and $w \in \overline{\Omega}_A\mathsf{V} \setminus A^+$, then x is a factor of uw or of wv.*

There is a close interplay between the structure of $\overline{\Omega}_A\mathsf{V}$ and subshifts [4, 3, 10, 5, 11]. We summarize some of its aspects. Let $\mathcal{X}$ be a subshift of $A^{\mathbb{Z}}$. The *mirage* of $\mathcal{X}$ is the set of pseudowords of $\overline{\Omega}_A\mathsf{V}$ whose finite factors belong to $L(\mathcal{X})$. One has $\bigcap_{n\geq 1}\overline{F_n(\mathcal{X})} = \mathcal{M}(\mathcal{X})$ (compare with (2.1)). In particular, the inclusion $\overline{L(\mathcal{X})} \subseteq \mathcal{M}(\mathcal{X})$ always holds, but in general the reverse inclusion fails [10]. The irreducible factorial closed subsets of a compact semigroup contain a unique $\mathcal{J}$-minimal $\mathcal{J}$-class [11]. In case $\mathcal{X}$ is irreducible, the set of elements $\mathcal{J}$-equivalent to $\overline{L(\mathcal{X})}$ and the set $\mathcal{M}(\mathcal{X})$ are factorial and irreducible. We denote by $\mathcal{J}(\mathcal{X})$ and $\mathcal{JM}(\mathcal{X})$ their minimal $\mathcal{J}$-classes, respectively. If $\mathcal{X}$ is minimal then $\mathcal{JM}(\mathcal{X}) = \mathcal{J}(\mathcal{X}) = \mathcal{M}(\mathcal{X})\setminus A^+$ and $\mathcal{JM}(\mathcal{X})$ is a maximal regular $\mathcal{J}$-class; moreover, every maximal regular $\mathcal{J}$-class is of this form [3]. From hereon, all subshifts are irreducible. Sometimes we shall want to compare different pseudovarieties, and so to emphasize that $\mathcal{JM}(\mathcal{X})$ is a subset of $\overline{\Omega}_A\mathsf{V}$ we will denote it by $\mathcal{JM}_{\mathsf{V}}(\mathcal{X})$.

It will be convenient to consider the topological spaces $A^{\mathbb{Z}^-}$ and $A^{\mathbb{N}}$, endowed with the product topology defined by the discrete space A, where $\mathbb{Z}^-$ and $\mathbb{N}$ denote the sets of negative and nonnegative integers, respectively. For $A^{\mathbb{N}}$, we will also require the noninvertible shift action $\sigma_+ : A^{\mathbb{N}} \to A^{\mathbb{N}}$, which maps $(x_i)_{i\in\mathbb{N}}$ to $(x_{i+1})_{i\in\mathbb{N}}$. The function $\nu : A^{\mathbb{Z}^-} \times A^{\mathbb{N}} \to A^{\mathbb{Z}}$ sending (y,z) to the bi-infinite word $y.z$ is a homeomorphism; we say that y is the *negative ray* of $y.z$, and that z is the *nonnegative ray*.

If $w \in \overline{\Omega}_A\mathsf{V} \setminus A^+$, then there is a unique $u \in A^n$ such that $w = uv$ for some $v \in \overline{\Omega}_A\mathsf{V}$. Extending the notation for finite words, denote u by $\mathrm{i}_n(w)$. We can also analogously extend t_n to $\overline{\Omega}_A\mathsf{V}$. The maps i_n and t_n are continuous functions from $\overline{\Omega}_A\mathsf{V}$ to the discrete space A^+ [1, Section 5.2]. If $w \in \overline{\Omega}_A\mathsf{V} \setminus A^+$, then we denote by $\overrightarrow{w}$ the unique element x of $A^{\mathbb{N}}$ such that $\mathrm{i}_n(w) = x_{[0,n-1]}$ for all $n \geq 1$, and by $\overleftarrow{w}$ the unique element y of $A^{\mathbb{Z}^-}$ such that $\mathrm{t}_n(w) = x_{[-n,-1]}$ for all $n \geq 1$. The maps $w \mapsto \overrightarrow{w}$ and $w \mapsto \overleftarrow{w}$ are continuous.

Lemma 5.2. *If $u,v \in \mathcal{JM}(\mathcal{X})$, then $uv \in \mathcal{JM}(\mathcal{X})$ if and only if $\overleftarrow{u}.\overrightarrow{v} \in \mathcal{X}$.*

Proof. Note that, since $u,v \in \mathcal{JM}(\mathcal{X})$, we have $uv \in \mathcal{JM}(\mathcal{X})$ if and only if $uv \in \mathcal{M}(\mathcal{X})$, by the definition of $\mathcal{JM}(\mathcal{X})$. By Lemma 5.1, every finite factor of uv is either a finite factor of u, of v, or of $\overleftarrow{u}.\overrightarrow{v}$. Thus, since $u,v \in \mathcal{M}(\mathcal{X})$, we have $uv \in \mathcal{JM}(\mathcal{X})$ if and only if all finite factors of $\overleftarrow{u}.\overrightarrow{v}$ belong to $L(\mathcal{X})$, that is, if and only if $\overleftarrow{u}.\overrightarrow{v} \in \mathcal{X}$. □

Lemma 5.3. *Let $w \in \mathcal{JM}(\mathcal{X})$. Then w is a group element if and only if $\overleftarrow{w}.\overrightarrow{w} \in \mathcal{X}$. Moreover, if $x \in \mathcal{X}$ then there is a maximal subgroup H of $\mathcal{JM}(\mathcal{X})$ such that $\overleftarrow{w}.\overrightarrow{w} = x$ for every $w \in H$.*

Proof. The first part follows from Lemma 5.2 and the fact that w is a group element if and only if $w^2 \in \mathcal{JM}(\mathcal{X})$. Let $x \in \mathcal{X}$. Let u and v be accumulation points of the sequences $(x_{[-n,-1]})_n$ and $(x_{[0,n-1]})_n$. Then u and v belong to $\overline{L(\mathcal{X})}$, and $\overleftarrow{u}.\overrightarrow{v} = x$. Since $\mathcal{M}(\mathcal{X})$ is irreducible, there is a pseudoword z such that $vzu \in \mathcal{JM}(\mathcal{X})$. And since $\overleftarrow{u}.\overrightarrow{v} \in \mathcal{X}$, the idempotent $e = (vzu)^\omega$ belongs to $\mathcal{JM}(\mathcal{X})$ by Lemma 5.2. Note that $\overleftarrow{e}.\overrightarrow{e} = x$. Since $\mathcal{H}$-equivalent elements have the same prefixes and suffixes, it follows that $\overleftarrow{w}.\overrightarrow{w} = x$ for every element w in the maximal subgroup containing e. □

For $\mathcal{X} \subseteq A^{\mathbb{Z}}$, denote $\overleftarrow{\mathcal{X}}$ and $\overrightarrow{\mathcal{X}}$ respectively the first and second components of $\nu^{-1}(\mathcal{X})$. One consequence of Lemma 5.3 is that the mappings $u \in \mathcal{JM}(\mathcal{X}) \mapsto \overrightarrow{u} \in \overrightarrow{\mathcal{X}}$ and $u \in \mathcal{JM}(\mathcal{X}) \mapsto \overleftarrow{u} \in \overleftarrow{\mathcal{X}}$ are surjective.

5.2. **The case of minimal subshifts.** Maximal regular elements of $\overline{\Omega}_A\mathsf{V}$ have the remarkable property that finite prefixes and finite suffixes determine their $\mathcal{R}$-classes and $\mathcal{L}$-classes, respectively. More precisely, we have the following result.

Lemma 5.4 ([2] and [5, Lemma 6.6])**.** *Let $\mathcal{X}$ be a minimal subshift. If $u, v \in \mathcal{JM}(\mathcal{X})$, then $u \mathrel{\mathcal{R}} v$ if and only if $\overrightarrow{u} = \overrightarrow{v}$, and dually, $u \mathrel{\mathcal{L}} v$ if and only if $\overleftarrow{u} = \overleftarrow{v}$. Hence $u \mathrel{\mathcal{H}} v$ if and only if $\overleftarrow{u}.\overrightarrow{u} = \overleftarrow{v}.\overrightarrow{v}$.*

Therefore, if $\mathcal{X}$ is minimal, then nonnegative rays of elements of $\mathcal{X}$ parameterize $\mathcal{R}$-classes, negative rays parameterize $\mathcal{L}$-classes, and to understand how groups are distributed in $\mathcal{JM}(\mathcal{X})$ is the same as to understand for which pairs $(y, z) \in \overleftarrow{\mathcal{X}} \times \overrightarrow{\mathcal{X}}$ we have $y.z \in \mathcal{X}$.

Let $z \in \overrightarrow{\mathcal{X}}$. The *degree of* z is the number of letters a such that $az \in \overrightarrow{\mathcal{X}}$. If z has degree greater than one, then we say z is a *left special* element of $\overrightarrow{\mathcal{X}}$.

Every finite prefix of a left special element of $\overrightarrow{\mathcal{X}}$ is a left special factor of $\mathcal{X}$. For a finite set S of elements of $\mathcal{X}$ there is a positive integer N such that, for all $m \geq N$, the prefixes of elements of S with length m are pairwise distinct. From these facts we get the following remark.

Remark 5.5. *Let $\mathcal{X}$ be a minimal subshift and k a positive integer. If there are infinitely many positive integers n such that $\mathcal{X}$ has at most k left special factors of length n, then there are at most k left special elements of $\overrightarrow{\mathcal{X}}$.*

Remark 5.5 relates with the following theorem, where one gets information about $\mathcal{JM}(\mathcal{X})$ from the number of left special nonnegative rays.

Theorem 5.6. *Let $\mathcal{X}$ be a minimal subshift for which there are at most k left special nonnegative rays of $\mathcal{X}$ with degree less than or equal to d. Then each $\mathcal{R}$-class of $\mathcal{JM}(\mathcal{X})$ has at most d^k idempotents, and if $\mathcal{X}$ is not periodic then the set of $\mathcal{R}$-classes with more than one idempotent is countable.*

In the proof of Theorem 5.6 we use the following auxiliary lemma, which is probably folklore. We include its proof for the sake of completeness.

Lemma 5.7. *In the conditions of Theorem 5.6, for each* $z \in \overrightarrow{\mathcal{X}}$ *there are at most* d^k *elements of* $\mathcal{X}$ *having* z *as a nonnegative ray.*

Proof. If $\mathcal{X}$ is periodic, or equivalently, if $d = 1$, then the lemma clearly holds. Henceforth we assume that $\mathcal{X}$ is not periodic. Consider the graph G defined as follows: the set of vertices is $\bigcup_{n\in\mathbb{Z}} \sigma_+^{-n}(z)$, for each vertex x there is a unique edge from x to $\sigma_+(x)$, and there are no other edges. The graph G is acyclic, otherwise $\mathcal{X}$ would contain a periodic element, contradicting the assumption that $\mathcal{X}$ is minimal non-periodic. Every vertex has out-degree one, and in-degree at most d. Moreover, in a path there are at most k distinct vertices with in-degree greater than one, otherwise the path would contain a cycle since by hypothesis $\overrightarrow{\mathcal{X}}$ has at most k left special elements. The number of elements of $\mathcal{X}$ having z as nonnegative ray is precisely the cardinal of the set P of right infinite paths in G ending in z, that is, infinite paths of the form $\cdots \to \bullet \to \bullet \to \bullet \to z$.

Let T be the tree defined as follows: the set of vertices are the vertices of elements of P with in-degree greater than one, and an edge from x to y is a (nonempty) path in G from x to y. Since a path in T has at most k elements and every vertex has at most in-degree d, the tree T has at most d^k vertices with in-degree zero. This concludes the proof because the number of these vertices is precisely the cardinal of P. $\square$

Proof of Theorem 5.6. Let R be an $\mathcal{R}$-class of $\mathcal{J}\mathcal{M}(\mathcal{X})$. Let z be the corresponding element of $\overrightarrow{\mathcal{X}}$ given by the parameterization of $\mathcal{R}$-classes resulting from Lemma 5.4. By Lemmas 5.3 and 5.4, there are as many idempotents in R as elements of $\mathcal{X}$ having z as nonnegative ray. Therefore, by Lemma 5.7, it only remains to prove that if $\mathcal{X}$ is not periodic then there is only a countable set of $\mathcal{R}$-classes with at least two idempotents. By Lemma 5.3 and 5.4, this amounts to prove that if Z is the set of elements z of $\overrightarrow{\mathcal{X}}$ such that z is the nonnegative ray of at least two elements of $\mathcal{X}$, then Z is countable. Note that $\sigma^+(Z) \subseteq Z$, thus Z is infinite, since $\mathcal{X}$ is not periodic. Let $z \in Z$. Then, denoting by F the (by hypothesis, finite) set of left special elements of $\overrightarrow{\mathcal{X}}$, there are $z' \in F$ and $m \geq 0$ such that $\sigma_+^m(z') = z$. Therefore, Z is contained in the set $\bigcup_{n\geq 0} \sigma_+^n(F)$, which is at most countable since F is finite. $\square$

Example 5.8. If $\mathcal{X}$ is a Sturmian subshift of $\{a, b\}^{\mathbb{Z}}$, then (see [13, Subsection 6.1.3]) there are $x, x' \in \mathcal{X}$ such that

- $x = \cdots x_{-4}x_{-3}x_{-2}a.bx_1x_2x_3\cdots$;
- $x' = \cdots x_{-4}x_{-3}x_{-2}b.ax_1x_2x_3\cdots$;
- if $w, w' \in \mathcal{X}$ have a common (negative or nonnegative) ray, then $w = w'$ or $\{w, w'\} = \{\sigma^n(x), \sigma^n(x')\}$ for some $n \in \mathbb{Z}$.

The conditions of Theorem 5.6 and its dual are satisfied, with $k = 1$ and $d = 2$, whence each $\mathcal{R}$-class and each $\mathcal{L}$-class of $\mathcal{J}\mathcal{M}(\mathcal{X})$ has at most two idempotents, and only a countable number of them have two idempotents.

The $\mathcal{R}$-classes (respectively, $\mathcal{L}$-classes) with two idempotents are those parameterized by the nonnegative rays of elements of the form $\sigma^n(x)$ with $n \geq 1$ (respectively, of the form $\sigma^{-n}(x)$ with $n \geq 1$). See Figure 7.

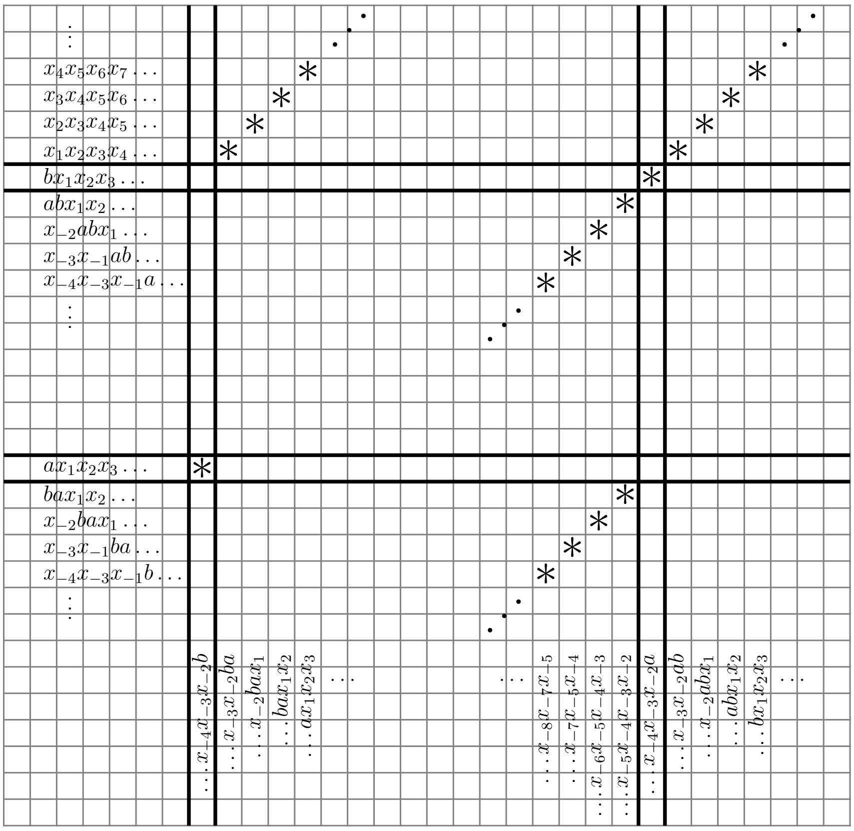

FIGURE 7. Picture representing a portion of $\mathcal{JM}(\mathcal{X})$ when $\mathcal{X}$ is Sturmian. Rows represent $\mathcal{R}$-classes, columns represent $\mathcal{L}$-classes, and stars indicate $\mathcal{H}$-classes with idempotents. In each row/column is indicated the nonnegative/negative ray parameterizing it.

A subshift $\mathcal{X}$ has *sublinear complexity* if the mapping $n \mapsto |L_n(\mathcal{X})|$ (known as the *complexity function*) is bounded by a monomial of degree one. For more information about the complexity function of a subshift, see the survey [12] and the more recent book [13]. Arnoux-Rauzy subshifts (and therefore Sturmian subshifts) have sublinear complexity. All minimal subshifts defined by primitive substitutions have sublinear complexity [16]. Suppose that the minimal subshift $\mathcal{X}$ has sublinear complexity. Then it has a finite number of left and right special rays [13, Proposition 5.1.12]. Applying Theorem 5.6, we thus obtain the following result.

Corollary 5.9. *Let $\mathcal{X}$ be a minimal subshift of sublinear complexity. Then the $\mathcal{R}$-classes and $\mathcal{L}$-classes of $\mathcal{JM}(\mathcal{X})$ have a bounded number of idempotents, and either none of them (in case $\mathcal{X}$ is periodic) or only a countably* *infinite number of them (in case $\mathcal{X}$ is not periodic) has more than one idempotent.* □

5.3. **The $\mathcal{J}$-classes $\mathcal{J}_n(\mathcal{X})$ as finite approximations of $\mathcal{JM}(\mathcal{X})$.** Let S be a set endowed with a partial binary operation $*$. Adding to S a new element, denoted by 0, we obtain the set S^0. Extend the operation $*$ to an operation $\circledast$ in S^0, by letting $s \circledast t = s * t$ if $s * t$ is defined and $s \circledast t = 0$ otherwise. We say that $(S, *)$ is a *partial semigroup* if $(S^0, \circledast)$ is a semigroup. A *homomorphism* from a partial semigroup S to a partial semigroup T is a function $\varphi : S \to T$ such that, for all $s_1, s_2 \in S$, if $s_1 * s_2$ is defined then $\varphi(s_1) * \varphi(s_2)$ is defined and $\varphi(s_1 * s_2) = \varphi(s_1) * \varphi(s_2)$. This defines the category of partial semigroups.

In the sequel $\mathcal{X}$ is always an irreducible subshift of $A^{\mathbb{Z}}$. Let m and n be positive integers such that $m \geq 2n$. By Proposition 3.6, the following map is a well-defined homomorphism of partial semigroups:

$$\begin{aligned} \psi_{m,n} : T_m(\mathcal{X}) \setminus \{0\} &\to T_n(\mathcal{X}) \setminus \{0\} \\ \eta_m(u) &\mapsto \eta_n(u), \quad u \in F_m(\mathcal{X}). \end{aligned}$$

Lemma 5.10. *If $m \geq 2n$ then $\psi_{m,n}(\mathcal{J}_m(\mathcal{X})) = \mathcal{J}_n(\mathcal{X})$.*

Proof. Let s be an element of $\mathcal{J}_m(\mathcal{X})$. There is some idempotent $e \in \mathcal{J}_m(\mathcal{X})$ such that $s = es$, and therefore $\psi_{m,n}(s) = \psi_{m,n}(e)\psi_{m,n}(s)$. Since $\psi_{m,n}(e)$ is an idempotent and the (n,n)-local Rauzy automaton $\Sigma_{2n}(\mathcal{X})$ satisfies the hypothesis of Proposition 2.6, we conclude that $\psi_{m,n}(s) \in \mathcal{J}_n(\mathcal{X})$.

On the other hand, let $t \in \mathcal{J}_n(\mathcal{X})$. By Proposition 2.6, there is $u \in L(\mathcal{X})$ such that $|u| \geq 2m$ and $t = \eta_n(u)$. Then $\eta_m(u) \in \mathcal{J}_m(\mathcal{X})$ by Proposition 2.6, and $\psi_{m,n}(\eta_m(u)) = t$. □

We endow $\mathbb{Z}^+$ with the following partial order: $n \preceq m$ if and only if $2n \leq m$. Consider the following inverse system:

$$\mathcal{D}(\mathcal{X}) = \{\psi_{m,n} : \mathcal{J}_m(\mathcal{X}) \to \mathcal{J}_n(\mathcal{X}) \,|\, n, m \in \mathbb{Z}^+, n \preceq m\}.$$

Let $u \in \mathcal{JM}(\mathcal{X})$. Let $W_n(u)$ be the set of elements w of A^* such that $\mathrm{i}_{2n}(u) \cdot w \cdot \mathrm{t}_{2n}(u)$ belongs to $L(\mathcal{X})$. Note that $W_n(u)$ is nonempty because $\mathcal{X}$ is irreducible. For each $n \geq 1$, choose an element $\theta_n(u) \in W_n(u)$. Note that $\eta_n\big(\mathrm{i}_{2n}(u) \cdot \theta_n(u) \cdot \mathrm{t}_{2n}(u)\big) \in \mathcal{J}_n(\mathcal{X})$, by Proposition 2.6. If $m \geq 2n$ then applying Remark 3.1, we obtain:

$$\begin{aligned} \psi_{m,n}\Big[\eta_m\big(\mathrm{i}_{2m}(u) \cdot \theta_m(u) \cdot \mathrm{t}_{2m}(u)\big)\Big] &= \eta_n\big(\mathrm{i}_{2m}(u) \cdot \theta_m(u) \cdot \mathrm{t}_{2m}(u)\big) \\ &= \eta_n\big(\mathrm{i}_{2n}(u) \cdot \theta_n(u) \cdot \mathrm{t}_{2n}(u)\big). \end{aligned}$$

Therefore, the following mapping is a well-defined function:

$$\begin{aligned} \psi : \mathcal{JM}(\mathcal{X}) &\to \varprojlim \mathcal{D}(\mathcal{X}) \\ u &\mapsto \Big(\eta_n\big(\mathrm{i}_{2n}(u) \cdot \theta_n(u) \cdot \mathrm{t}_{2n}(u)\big)\Big)_n. \end{aligned}$$

The component function $\psi_n : \mathcal{JM}(\mathcal{X}) \to \mathcal{J}_n(\mathcal{X})$ is continuous, since i_{2n} and t_{2n} are continuous and $\psi_n(u)$ is independent of the choice of $\theta_n(u)$ among elements of $W_n(u)$, by Remark 3.1. Hence ψ is continuous.

The continuity of ψ suggests to consider the following notion: a *compact partial semigroup* is a partial semigroup $(S,*)$ endowed with a compact topology such that the set

$$D(S) = \{(s,t) \in S \times S \,:\, s * t \text{ is defined}\}$$

is closed in $S \times S$, and if $(s_i,t_i)_{i\in I}$ is a net converging in $D(S)$ to (s,t), then $(s_i * t_i)_{i\in I}$ converges to $s*t$. For example, every $\mathcal{J}$-class of a compact semigroup is a compact partial semigroup under the induced operation. One may be tempted to include in the definition of compact partial semigroup that $D(S)$ must be open, because then S^0 would be a compact semigroup, with 0 as an isolated point. However, under such definition $\mathcal{JM}(\mathcal{X})$ would not in general be a compact partial semigroup as shown in the following proposition.

Proposition 5.11. *Let τ be the Prouhet-Thue-Morse substitution $[ab,ba]$. The set $D(\mathcal{JM}(\mathcal{X}_\tau))$ is not open.*

Proof. The words $\tau^n(a)$ and $\tau^n(a^2)$ belong to $L(\mathcal{X}_\tau)$ for every $n \geq 1$. Since $\overline{L(\mathcal{X}_\tau)}$ is prolongable, there are infinite pseudowords u_n, v_n such that $\alpha_n = u_n\tau^{n!}(a)$ and $\beta_n = \tau^{n!}(a^2)v_n$ belong to $\overline{L(\mathcal{X}_\tau)}$. The word $\tau^{n!}(a)^3$ is a factor of $\alpha_n\beta_n$. It is well known that $L(\mathcal{X}_\tau)$ does not contain cubes [15, Proposition 3.1.1], whence $\alpha_n\beta_n \notin \mathcal{M}(\mathcal{X}_\tau)$.

Let (u,v) be a accumulation point in $\overline{\Omega}_A\mathsf{V} \times \overline{\Omega}_A\mathsf{V}$ of the sequence (u_n,v_n). The endomorphism $\tau : A^+ \to A^+$ has a unique extension to a continuous endomorphism $\overline{\Omega}_A\mathsf{V} \to \overline{\Omega}_A\mathsf{V}$. The monoid of endomorphisms of $\overline{\Omega}_A\mathsf{V}$, endowed with the pointwise convergence topology, is profinite [14]. For an element s of a profinite semigroup S, the sequence $(s^{n!})_n$ converges to an idempotent, denoted s^ω. The pseudowords $\alpha = u\tau^\omega(a)$ and $\beta = \tau^\omega(a^2)v$ are accumulation points of the sequences α_n and β_n, respectively. Note that $\alpha\beta = u\tau^\omega(a)^3v$. Since $\tau^\omega(a)$ is an infinite pseudoword and the pseudowords $u\tau^\omega(a)$, $\tau^\omega(a)^2$ and $\tau^\omega(a)v$ belong to $\mathcal{M}(\mathcal{X}_\tau)$, from Lemma 5.1 we deduce that $\alpha\beta \in \mathcal{M}(\mathcal{X}_\tau)$.

Because $\mathcal{X}_\tau$ is minimal, we have $\mathcal{M}(\mathcal{X}_\tau) \setminus A^+ = \mathcal{JM}(\mathcal{X}_\tau)$. Hence (α_n,β_n) has a subsequence of elements of $\mathcal{JM}(\mathcal{X}_\tau) \times \mathcal{JM}(\mathcal{X}_\tau) \setminus D(\mathcal{JM}(\mathcal{X}_\tau))$ converging to an element of $D(\mathcal{JM}(\mathcal{X}_\tau))$. Therefore $D(\mathcal{JM}(\mathcal{X}_\tau))$ is not open, even as a subset of $\mathcal{JM}(\mathcal{X}_\tau)$. □

Denote by $E(X)$ the set of idempotents of a subset X of a semigroup S. The inverse image of an idempotent by an onto homomorphism of finite semigroups contains at least one idempotent. This fails for the onto homomorphism $\psi_{m,n} : \mathcal{J}_m(\mathcal{X}) \to \mathcal{J}_n(\mathcal{X})$ of partial finite semigroups. As an example, for the Sturmian substitution $\varphi = [aabaa, aabaaa]$, one can easily check that $|E(\mathcal{J}_4(\mathcal{X}_\varphi))| = 11 < 15 = |E(\mathcal{J}_2(\mathcal{X}_\varphi))|$ by direct computation. The following lemma characterizes the idempotents whose inverse image by $\psi_{m,n}$ contains some idempotent.

Lemma 5.12. *Let s be an idempotent element of $\mathcal{J}_n(\mathcal{X})$. Let $m \geq 2n$. Then $s = \psi_n(e)$ for some idempotent e of $\mathcal{JM}(\mathcal{X})$ if and only if $s = \psi_{m,n}(t)$ for some idempotent t of $\mathcal{J}_m(\mathcal{X})$.*

Proof. If e is an idempotent of $\mathcal{JM}(\mathcal{X})$ then $\psi_m(e)$ is an idempotent of $\mathcal{J}_m(\mathcal{X})$ and $\psi_{m,n}(\psi_m(e)) = \psi_n(e)$. Conversely, suppose that $s = \psi_{m,n}(t)$ for some idempotent t of $\mathcal{J}_m(\mathcal{X})$. By Proposition 2.6, there is an element u of $L(\mathcal{X})$ such that $\eta_m(u) = t$ and $|u| \geq 2m$. Since t is idempotent, u^2 belongs to $F_m(\mathcal{X})$, thus $\mathrm{t}_m(u)\,\mathrm{i}_m(u) \in L(\mathcal{X})$. Therefore, by Lemma 5.3, there is an idempotent e in $\mathcal{JM}(\mathcal{X})$ such that $\mathrm{t}_m(e) = \mathrm{t}_m(u)$ and $\mathrm{i}_m(e) = \mathrm{i}_m(u)$. Then, $\psi_n(e) = \eta_n\big(\mathrm{i}_{2n}(e)\cdot\theta_n(e)\cdot\mathrm{t}_{2n}(e)\big) = \eta_n(u)$ by Remark 3.1. On the other hand we have the following sequence of equalities

$$\eta_n(u) = \psi_{m,n}(\eta_m(u)) = \psi_{m,n}(t) = s,$$

thus $s = \psi_n(e)$. □

Theorem 5.13. *The mapping $\psi : \mathcal{JM}(\mathcal{X}) \to \varprojlim \mathcal{D}(\mathcal{X})$ is an onto homomorphism of partial semigroups. The kernel of ψ is the kernel of the map $w \in \mathcal{JM}(\mathcal{X}) \mapsto \overleftarrow{w}.\overrightarrow{w}$. Moreover, ψ maps $E(\mathcal{JM}(\mathcal{X}))$ onto $E\big(\varprojlim \mathcal{D}(\mathcal{X})\big)$.*

Proof. Let u and v be elements of $\mathcal{JM}(\mathcal{X})$ such that $uv \in \mathcal{JM}(\mathcal{X})$. Then, for every $n \geq 1$, we have $\mathrm{t}_{2n}(u)\,\mathrm{i}_{2n}(v) \in L(\mathcal{X})$, and therefore there exist $x_n, y_n \in A^+$ such that

$$\mathrm{i}_{2n}(u)\,x_n\mathrm{t}_{2n}(u)\,\mathrm{i}_{2n}(v)\,y_n\mathrm{t}_{2n}(v) \in L(\mathcal{X}).$$

By Remark 3.1 we have

$$\begin{aligned}\psi_n(uv) &= \eta_n\big[\mathrm{i}_{2n}(u)\,x_n\mathrm{t}_{2n}(u)\,\mathrm{i}_{2n}(v)\,y_n\mathrm{t}_{2n}(v)\big]\\ &= \eta_n\big[\mathrm{i}_{2n}(u)\,x_n\mathrm{t}_{2n}(u)\big]\cdot\eta_n\big[\mathrm{i}_{2n}(v)\,y_n\mathrm{t}_{2n}(v)\big]\\ &= \psi_n(u)\psi_n(v).\end{aligned}$$

Hence $\psi(uv) = \psi(u)\psi(v)$, and so ψ is a homomorphism of partial semigroups.

Let $u, v \in \mathcal{JM}(\mathcal{X})$ be such that $\psi(u) = \psi(v)$. For every positive integer n, the equality $\psi_n(u) = \psi_n(v)$ implies $\mathrm{i}_n(u) = \mathrm{i}_n(v)$ and $\mathrm{t}_n(u) = \mathrm{t}_n(v)$, by Remark 3.2. Hence $\overleftarrow{u}.\overrightarrow{u} = \overleftarrow{v}.\overrightarrow{v}$. Conversely, if $\overleftarrow{u}.\overrightarrow{u} = \overleftarrow{v}.\overrightarrow{v}$ then $\psi(u) = \psi(v)$ by Remark 3.1.

We claim that the mapping $\psi_n : \mathcal{JM}(\mathcal{X}) \to \mathcal{J}_n(\mathcal{X})$ is onto. To prove it, let $v \in F_n(\mathcal{X})$ be such that $|v| \geq 2n$. Then $\mathrm{i}_{2n}(v)$ and $\mathrm{t}_{2n}(v)$ are elements of $L(\mathcal{X})$. Since $\mathcal{M}(\mathcal{X})$ is irreducible, it follows that for all $w \in \mathcal{JM}(\mathcal{X})$ there are $z, t \in \overline{\Omega}_A\mathsf{V}$ such that the pseudoword $u = \mathrm{i}_{2n}(v)\,zwt\,\mathrm{t}_{2n}(v)$ belongs to $\mathcal{M}(\mathcal{X})$. Then $u \in \mathcal{JM}(\mathcal{X})$ by the definition of $\mathcal{JM}(\mathcal{X})$, and $\psi_n(u) = \eta_n(v)$ by Remark 3.1, which establishes the claim. If follows from well-known properties of projective limits of compact spaces that the mapping ψ is onto (the result we use is formulated in [5, Proposition 2.1]).

Finally, let $s = (s_n)_n$ be an idempotent element of $\varprojlim \mathcal{D}(\mathcal{X})$. By Lemma 5.12, for each n there is an idempotent $e_n \in \mathcal{JM}(\mathcal{X})$ such that $\psi_n(e_n) = s_n$. Since $s_n = \psi_{m,n}(\psi_m(e_m)) = \psi_n(e_m)$ it follows that

$$e_m \in \bigcap_{n \preceq m} (\psi_n^{-1}(s_n) \cap E(\mathcal{JM}(\mathcal{X}))).$$

Therefore, the family $(\psi_n^{-1}(s_n) \cap E(\mathcal{JM}(\mathcal{X})))_{n\geq 1}$ of closed sets has the finite intersection property, thus it has nonempty intersection, since $\mathcal{JM}(\mathcal{X})$ is compact. An element of the intersection of the family is an idempotent e such that $\psi(e) = s$. □

If $\mathcal{X}$ is minimal, then the kernel of the mapping $w \in \mathcal{JM}(\mathcal{X}) \mapsto \overleftarrow{w}.\overrightarrow{w}$ is the restriction of Green's relation $\mathcal{H}$ to $\mathcal{JM}(\mathcal{X})$, by Lemma 5.4. Therefore, from Theorem 5.13 we deduce the following result.

Corollary 5.14. *If $\mathcal{X}$ is minimal and V is a pseudovariety of aperiodic semigroups then $\psi : \mathcal{JM}_{\mathsf{V}}(\mathcal{X}) \to \varprojlim \mathcal{D}(\mathcal{X})$ is a continuous isomorphism of compact partial semigroups.* □

Hence, when $\mathcal{X}$ is minimal and $\mathsf{V} \subseteq \mathsf{A}$ (where A is the pseudovariety of finite aperiodic semigroups) then $\mathcal{JM}_{\mathsf{V}}(\mathcal{X})$ is isomorphic to $\mathcal{JM}_{\mathsf{A}}(\mathcal{X})$. This can be deduced more directly from Lemma 5.4: if U and W are pseudovarieties such that $\mathsf{U} \subseteq \mathsf{W}$ then we may consider a canonical continuous onto homomorphism $p_{\mathsf{W},\mathsf{U}} : \overline{\Omega}_A\mathsf{W} \to \overline{\Omega}_A\mathsf{U}$, and it is easy to prove that $p_{\mathsf{W},\mathsf{U}}(\mathcal{JM}_{\mathsf{W}}(\mathcal{X})) = \mathcal{JM}_{\mathsf{U}}(\mathcal{X})$; since $\overrightarrow{\mathcal{X}}$ and $\overleftarrow{\mathcal{X}}$ respectively parameterize $\mathcal{R}$-classes and $\mathcal{L}$-classes of both $\mathcal{JM}_{\mathsf{W}}(\mathcal{X})$ and $\mathcal{JM}_{\mathsf{U}}(\mathcal{X})$, the kernel of the restriction $p_{\mathsf{W},\mathsf{U}}|\mathcal{JM}_{\mathsf{W}}$ is contained in $\mathcal{H}$.

6. Final remarks

Searching for how the maximal subgroups of $\mathcal{JM}(\mathcal{X})$ are spread, with the help of the approximation of $\mathcal{JM}(\mathcal{X})$ via the projective limit of the partial semigroups $\mathcal{J}_n(\mathcal{X})$, one is lead to investigate how the idempotents from $\mathcal{J}_{2n}(\mathcal{X})$ are related with the idempotents of $\mathcal{J}_n(\mathcal{X})$, for an infinity of values of n. Hence, it seems to be worthy to investigate how the semigroups $T_n(\mathcal{X})$ evolve with n. For the subshift $\mathcal{X}_{[ab,a]}$ associated with the Fibonacci substitution, Figure 8 displays $\mathcal{J}_8(\mathcal{X}_{[ab,a]})$ and $\mathcal{J}_4(\mathcal{X}_{[ab,a]})$, with the idempotents of $\mathcal{J}_8(\mathcal{X}_{[ab,a]})$ grouped in the kernel classes of the restriction $\psi_{8,4} : E(\mathcal{J}_8(\mathcal{X}_{[ab,a]})) \to E(\mathcal{J}_4(\mathcal{X}_{[ab,a]}))$. Each number in the egg-box

											15	15
									13	13	14	
									13	13		
								11	12			
						9	9	10				
						9	9					
				7	7	8						
				7	7							
			5	6								
	3	3	4									
	3	3										
1	2											
1												

							15
						13	14
					11	12	
				9	10	*	
			7	8			
		5	6				
	3	4	*				
1	2						

FIGURE 8. Egg-box diagrams of $\mathcal{J}_8(\mathcal{X}_{[ab,a]})$ and $\mathcal{J}_4(\mathcal{X}_{[ab,a]})$.

diagram of $\mathcal{J}_8(\mathcal{X}_{[ab,a]})$ represents the idempotent in the kernel class of $\psi_{8,4} : E(\mathcal{J}_8(\mathcal{X}_{[ab,a]})) \to E(\mathcal{J}_4(\mathcal{X}_{[ab,a]}))$ determined by the idempotent of $\mathcal{J}_4(\mathcal{X}_{[ab,a]})$ represented by the same number in the egg-box diagram of $\mathcal{J}_4(\mathcal{X}_{[ab,a]})$. The stars represent idempotents in $\mathcal{J}_4(\mathcal{X}_{[ab,a]})$ which are not images of idempotents of $\mathcal{J}_8(\mathcal{X}_{[ab,a]})$.

References

1. J. Almeida, *Finite semigroups and universal algebra*, World Scientific, Singapore, 1995, English translation.
2. ______, *Symbolic dynamics in free profinite semigroups*, RIMS Kokyuroku, no. 1366, April 2004, pp. 1–12.
3. ______, *Profinite groups associated with weakly primitive substitutions*, Fundamentalnaya i Prikladnaya Matematika (Fundamental and Applied Mathematics) **11** (2005), no. 3, 13–48, In Russian. English version in J. Math. Sciences **144**, No. 2 (2007) 3881–3903.
4. ______, *Profinite semigroups and applications*, Structural Theory of Automata, Semigroups, and Universal Algebra (New York) (Valery B. Kudryavtsev and Ivo G. Rosenberg, eds.), NATO Science Series II: Mathematics, Physics and Chemistry, vol. 207, Springer, 2005, Proceedings of the NATO Advanced Study Institute on Structural Theory of Automata, Semigroups and Universal Algebra, Montréal, Québec, Canada, 7-18 July 2003, pp. 1–45.
5. J. Almeida and A. Costa, *Infinite-vertex free profinite semigroupoids and symbolic dynamics*, J. Pure Appl. Algebra **213** (2009), 605–631.
6. ______, *Presentations of Schützenberger groups of minimal subshifts*, Tech. Report CMUP 2010-1, Univ. Porto, 2010.
7. J. Almeida and M. V. Volkov, *Subword complexity of profinite words and subgroups of free profinite semigroups*, Int. J. Algebra Comput. **16** (2006), 221–258.
8. J. Berstel, *Recent results on extensions of Sturmian words*, Int. J. Algebra Comput. **12** (2002), 371–385.
9. J. Berstel, D. Perrin, and Ch. Reutenauer, *Codes and automata*, Encyclopedia of Mathematics and its Applications, vol. 129, Cambridge University Press, Cambridge, 2010.
10. A. Costa, *Conjugacy invariants of subshifts: an approach from profinite semigroup theory*, Int. J. Algebra Comput. **16** (2006), no. 4, 629–655.
11. A. Costa and B. Steinberg, *Profinite groups associated to sofic shifts are free.*, Proc. London Math. Soc. **102** (2011), 341–369.
12. S. Ferenczi, *Complexity of sequences and dynamical systems*, Discrete Math. **206** (1999), 145–154.
13. N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, Lecture Notes in Mathematics, vol. 1794, Springer-Verlag, Berlin, 2002.
14. R. P. Hunter, *Some remarks on subgroups defined by the Bohr compactification*, Semigroup Forum **26** (1983), 125–137.
15. M. Lothaire, *Algebraic combinatorics on words*, Cambridge University Press, Cambridge, UK, 2002.
16. J.-J. Pansiot, *Complexité des facteurs des mots infinis engendrés par morphismes itérés*, Automata, Languages and Programming (Antwerp, 1984) (Berlin), Lect. Notes Comput. Sci., no. 172, Springer, 1984, pp. 380–389.
17. J.-E. Pin, *Varieties of formal languages*, Plenum, London, 1986, English translation.
18. G. Rauzy, *Suites à termes dans un alphabet fini*, Seminar on number theory, 1982–1983 (Talence, 1982/1983), Univ. Bordeaux I, Talence, 1983, pp. Exp. No. 25, 16.
19. P. V. Salimov, *On Rauzy graph sequences of infinite words*, J. Appl. Indust. Math. (2010), 127–135, Originally published in Russian in Diskretnyi Analiz i Issledovanie Operatsii, 2008, Vol. 15, No. 5, pp. 6175.

CMUP, Departamento de Matemática, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal.
E-mail address: jalmeida@fc.up.pt

CMUC, Department of Mathematics, University of Coimbra, 3001-454 Coimbra, Portugal.
E-mail address: amgc@mat.uc.pt