AN AUTOMATA-THEORETIC APPROACH TO THE STUDY OF THE INTERSECTION OF TWO SUBMONOIDS OF A FREE MONOID

LAURA GIAMBRUNO AND ANTONIO RESTIVO

ABSTRACT. We investigate the intersection of two finitely generated submonoids of the free monoid on a finite alphabet A. To this purpose, we consider automata that recognize such submonoids and we study the product automata recognizing their intersection. By using automata methods we obtain a new proof of a result of Karhumäki on the characterization of the intersection of two submonoids of rank two, in the case of prefix (or suffix) generators. In a more general setting, for an arbitrary number of generators, we prove that if H and K are two finitely generated submonoids of A^* generated by prefix sets such that the product automaton associated to $H \cap K$ has a given special property then $\widetilde{rk}(H \cap K) \leq \widetilde{rk}(H)\widetilde{rk}(K)$ where $\widetilde{rk}(L) = max(0, rk(L) - 1)$ for any submonoid L of A^*.

1. INTRODUCTION

The main purpose of this paper is to study the intersection of two submonoids of a free monoid by using an automata approach. We are, in particular, interested in submonoids of finite rank.

The study of the intersection of two submonoids of finite rank is not trivial at all. In fact, by a result of Latteux and Leguy ([8]), every language is regular if and only if the submonoid generated by it is obtained as omomorphic image of the intersection of two finitely generated monoids: *Let A be an alphabet and R a language of A^*. R is a regular language if and only if there exist two finite languages F_1, F_2 and a morphism g such that* $R^* = g({F_1}^* \cap {F_2}^*)$

Moreover the study of the intersection of two submonoids has been developed in many papers. It was first studied by Tilson in 1972 ([12]) who proved that the intersection of two free submonoids of A^*, the free monoid on a finite alphabet A, is free too.

Karhumäki in 1984 ([7]) deepened the study of the intersection of two submonoids generated by two elements by giving a characterization of such an intersection. In particular he proved that given two submonoids H and K of A^*, where A is a finite alphabet, if both H and K are of rank two, then $H \cap K$ is a submonoid generated either by at most two words or by a regular language of a special form. In particular if H and K are generated by prefix (or suffix) sets of two words, and $H \cap K$ is not finitely generated, then this intersection has the form $(\alpha\beta^*\gamma)^*$ where $\alpha, \beta, \gamma \in A^*$.

Recently Bruyére, Derencourt and Latteux ([2]) have studied the *meet* of two rational codes X and Y, defined as the base of the free monoid $X^* \cap Y^*$. They concentrate on the study of maximal rational codes such that their meet is yet a maximal rational code, showing with many examples the complex behavior of the meet. Finally they proved that any rational code is the meet of two rational maximal codes.

The starting motivation of this paper was the article of Karhumäki ([7]) that characterizes the intersection of two submonoids of the free monoid generated by two elements. The proof given by Karhumäki is long and provides no intuition on the real nature of the result. In this paper we prove, in particular, the result of Karhumäki in the case of two submonoids generated by prefix (or suffix) sets using a more intuitive approach based on automata.

When dealing with the intersection of two submonoids of finite rank it is natural to relate it to a more general problem in the theory of free groups known as the 'Hanna Neumann conjecture'. This conjecture deals with the problem of finding an upper bound of the rank of the intersection of two finitely generated subgroups.

In 1956 Hanna Neumann ([10]) proved that if H and K are two subgroups of finite rank of a free group F then $\widetilde{rk}(H \cap K) \leq 2\widetilde{rk}(H)\widetilde{rk}(K)$, where $\widetilde{rk}(T) = max(rk(T) - 1, 0)$ with $rk(T)$ the rank of a subgroup T of F. Then she made the following conjecture, known nowadays as the 'Hanna Neumann conjecture':

$$\widetilde{rk}(H \cap K) \leq \widetilde{rk}(H)\widetilde{rk}(K)$$

In 1991 Walter Neumann ([11]) formulated a stronger conjecture known as 'Strengthened Hanna Neumann conjecture' (in short SHN) and in 2002 Meakin and Weil ([9]) proved that SHN holds for the class of positively generated subgroups of the free group $F(A)$ on A, finite alphabet, that are generated by words on A^*. This last result suggested us to propose the problem of Hanna Neumann for finitely generated submonoids of a free monoid, in the case that their intersection is finitely generated.

Some of the basic tools in dealing with the Hanna Neumann conjecture for free groups makes use of the representation of subgroups of the free group by graphs (or automata). The same tools are still available when dealing with the intersection of two submonoids of the free monoid. For this purpose we refer to the correspondence (cf.[1]) between submonoids on the free monoid on a finite alphabet A and automata on A.

Through the study of the product of two automata associated to two finitely generated submonoids H and K, we were able to prove that if H and K are submonoids generated by prefix sets such that the product automaton associated to $H \cap K$ has a given special property then one has that $\widetilde{rk}(H \cap K) \le \widetilde{rk}(H)\widetilde{rk}(K)$.

And in the general case we have found a family of examples such that $rk(H \cap K) = 2^{log_2(rk(H)) log_2(rk(K))}$.

Moreover if the two submonoids H and K are generated by prefix sets of two elements, then their intersection $H \cap K$ either is of rank two, or it has the form $(\alpha\beta^*\gamma)^*$ where $\alpha, \beta, \gamma \in A^*$, that is the result of Karhumäki ([7]).

2. Automata and submonoids

We consider finite states automata on a finite alphabet A. For the notation we refer to [1]. Let A be a finite alphabet. Let us denote by ε the empty word of A^* and by A^+ the set of nonempty words on A. An *automaton* over A, $\mathcal{A} = (Q, I, T, \mathcal{F})$, consists of a finite set Q of *states*, of two subsets I and T of Q called sets of initial and final states, respectively, and of a set $\mathcal{F} \subseteq Q \times A \times Q$ whose elements are called *edges*. An edge $e = (x, a, y)$ is also denoted by $e : x \xrightarrow{a} y$. The letter a is called the label of the edge. We will say that the edge e goes out from x and comes in y.

Two edges $e : x \xrightarrow{a} y$ and $g : x' \xrightarrow{b} y'$ are *consecutive* if $y = x'$. A *path* in $\mathcal{A}$ is a finite sequence $p = p_1 p_2 \dots p_n$ of consecutive edges $p_i : x_i \xrightarrow{a_i} y_i$. We shall also write $p : x_1 \xrightarrow{w} y_n$ where $w = a_1 a_2 \dots a_n$ is the *label* of the path p. The path p is said to start at x_1 and end at y_n. We indicate with $i(p) = x_1$ the starting state and with $f(p) = y_n$ the ending state. The *length* of a path is the number of edges that compose it. For each state $x \in Q$ it is defined the *null path* starting and ending at x, denoted by $1_x : x \longrightarrow x$ having as label ε.

When two paths p and q are consecutive (i.e. $f(p) = i(q)$) then p and q can be concatenated and we call the resulting path pq. A *subpath* of a path p is a subsequence of consecutive edges. A subpath of a path p is a *prefix* of p if it starts at the same starting state of p. Given two paths p and q starting at the same state x, the *longest prefix path in common* between p and q is a path prefix of p and prefix of q that is the longest with this property. Analogously it can be defined, given two paths p and q ending at the same state x, the *longest suffix path in common* between p and q.

A path p is a *simple path* if all states in the path are distinct. Given two states u and v, if there exists a path from u to v then there exists also a simple path from u to v. A path p is a *cycle* if it is not the null path and if it starts and ends at the same state. We say that a path c is a cycle in x if it starts and ends at x. We moreover say that c is a cycle with basis x. A cycle c is a *simple cycle* if it has all the intermediate states distinct and different from the starting state.

We now introduce a new nomenclature for the cycles that do not visit the starting state in the intermediate states:

Definition 1. *Given a cycle c, we say that c is* simple in x *if it is a cycle in x such that no intermediate state is equal to x.*

Remark 2. We remark that a cycle that is simple in x is not in general a simple cycle.

Example 3. *In the automaton in figure 1 the cycle* $c : 1 \xrightarrow{a} 2 \xrightarrow{a} 2 \xrightarrow{b} 1$ *is a cycle simple in* 1 *but it is not a simple cycle. The cycle* $c' : 1 \xrightarrow{a} 2 \xrightarrow{b} 3 \xrightarrow{a} 1$ *is a simple cycle so, in particular, it is a cycle simple in* 1.

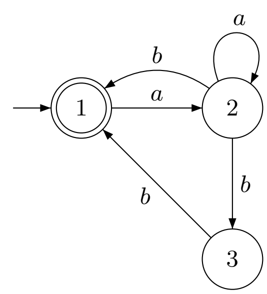

FIGURE 1. $\mathcal{A}$ automaton

The *language recognized by* $\mathcal{A}$, $L(\mathcal{A})$, is the set of words that are labels of paths from an initial state to a final one. A state x in $\mathcal{A}$ is *accessible* if there is a path starting at an initial state and ending at x. A state x in $\mathcal{A}$ is *coaccessible* if there is a path starting at x and ending in a final state. An automaton $\mathcal{A}$ is a *trim* automaton if all the states of the automaton are accessible and coaccessible. An automaton $\mathcal{A} = (Q, I, T, \mathcal{F})$ over A is *unambiguous* if, for each $x, y \in Q$, for each $w \in A^*$, there exists at most one path starting at x and ending at y with label w.

An automaton $\mathcal{A}$ over A is a *deterministic automaton* if $card(I) = 1$ and if, for each state x and for each $a \in A$, there is at most one edge starting at x with label a. If $\mathcal{A} = (Q, i, T, \mathcal{F})$ is a deterministic automaton it can be defined the function $\delta : Q \times A \longrightarrow Q$ such that $\delta(x, a) = y$ if $x \xrightarrow{a} y \in \mathcal{F}$ and $\delta(x, a) = \emptyset$ otherwise. This function δ is extended to words in A^* by setting, for all $x \in Q$, $\delta(x, \varepsilon) = x$ and, for $w \in A^*$ and $a \in A$, $\delta(x, wa) = \delta(\delta(x, w), a)$. This function is called the *transition function.* With this notation we have that $L(\mathcal{A}) = \{w \in A^* \mid \delta(i, w) \in F\}$.

Let $\mathcal{A} = (Q, I, T, \mathcal{F})$ be an automaton. We say that a state $x \in Q$ is a *branch point going out* (in short *bpo*) if the number of edges going out from x is at least two. We say that $x \in Q$ is a *branch point going in* (in short *bpi*) if the number of edges coming in x is at least two.

From now on we will consider automata with non empty languages.

We say that $\mathcal{A}$ is a *monoidal automaton* if it is a trim automaton with a unique final state equal to a unique initial one. Such a special state is denoted by 1. Let us note that in a monoidal automaton, for each state x, there exists a simple path from 1 to x and a simple path from x to 1.

It is easy to prove the following (cf.[1]):

Proposition 4. *Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be a monoidal automaton. The automaton $\mathcal{A}$ recognizes the submonoid generated by the set of labels of the cycles that are simple in* 1.

In general given a submonoid H of A^* there exists a unique minimal set of generators of H(cf.[1]). We define the rank of H as the cardinality of the minimal set of generators. It is denoted by $rk(H)$. We say that a submonoid H of A^* is cyclic if $rk(H) = 1$. We also define the reduced rank of a submonoid H as $\widetilde{rk}(H) = max(rk(H) - 1, 0)$. A submonoid $H \subseteq A^*$ is free if there exists a set of generators X such that every element in H can be factorized in a unique way in words of X. If H is free the minimal set of generators of H is called the *basis* of H.

Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be a monoidal automaton. Let us denote by $C_{\mathcal{A}}$ the set of cycles that are simple in 1 and by $Y_{\mathcal{A}}$ the set of their labels. In general if $\mathcal{A}$ is monoidal then $Y_{\mathcal{A}}$ is not the minimal set of generators (see an example in figure 2). If we suppose that $\mathcal{A}$ is unambiguous then $Y_{\mathcal{A}}$ is the minimal set of generators and moreover the submonoid generated by $\mathcal{A}$ is free.

Proposition 5. *Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be an unambiguous monoidal automaton. The automaton $\mathcal{A}$ recognizes a free submonoid with basis $Y_{\mathcal{A}}$.*

Proof. Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be an unambiguous monoidal automaton. By prop.4 the set of labels of cycles that are simple in 1, $Y_{\mathcal{A}}$, is a set of generators for $L(\mathcal{A})$. Moreover, by the unambiguity of $\mathcal{A}$, every element in $L(\mathcal{A})$ can be factorized in a unique way in words of $Y_{\mathcal{A}}$, otherwise there would be a word in $L(\mathcal{A})$ label of two different cycles in 1. □

So to a monoidal automaton $\mathcal{A}$ is associated the submonoid $H = L(\mathcal{A})$ of A^*. Conversely to each submonoid X^* of A^* generated by a finite set X it is associated $\mathcal{F}_X$ the *flower automaton of* X (cf.[3],[1]). It is built in the following way. First we build $\mathcal{S}_X$ the *solar automaton* recognizing X in this way: we build one automaton for each word $x \in X$ with $|x| + 1$ states and merge all the initial states. Note that this automaton is a tree with root the initial state 1. Then we merge all the final states with the initial state 1. Doing this we obtain the flower automaton of X. Such an automaton is a monoidal automaton recognizing X^* such that all the cycles visit the unique initial-final state 1, all the cycles that are simple in 1 intersect themselves only in 1 and have as labels the words of X. See an example in figure 3.

Let us define now a class of monoidal automata recognizing finitely generated submonoids: the class of semi-flower automata.

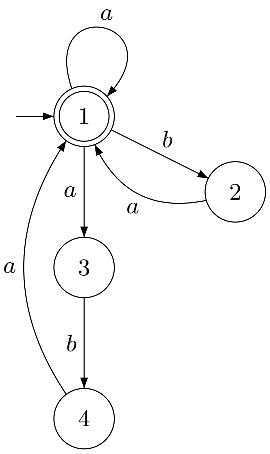

FIGURE 2. $\mathcal{A}$ monoidal automaton with $Y_{\mathcal{A}} = \{a, ba, aba\}$ and $L(\mathcal{A}) = \{a, ba\}^*$

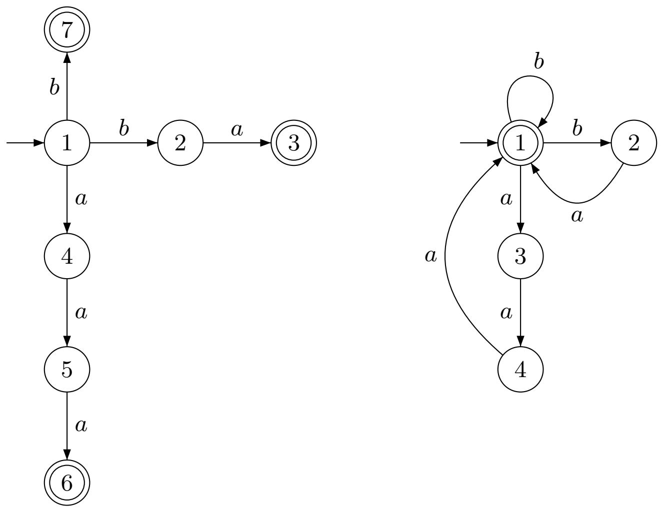

FIGURE 3. $\mathcal{S}_X$, $\mathcal{F}_X$) for $X = \{b, ba, aaa\}$

Definition 6. *Let $\mathcal{A}$ be an automaton. $\mathcal{A}$ is a semi-flower automaton if it is a monoidal automaton such that all the cycles visit the unique initial-final state.*

In a semi-flower automaton the cycles that are simple in 1 intersect themselves not necessarily only in 1. In particular given a finite set $X \subseteq A^*$ the flower automaton associated to X is a semi-flower automaton in which all the cycles that are simple in 1 intersect themselves only in 1. It is interesting to observe that in a semi-flower automaton every cycle that is simple in 1 is in particular a simple cycle. It is easy to prove the following proposition:

Proposition 7. *Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be a semi-flower automaton, then $\mathcal{A}$ recognizes a finitely generated submonoid.*

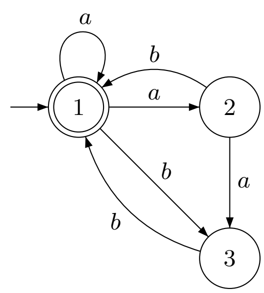

FIGURE 4. $\mathcal{A}$ semi-flower automaton, $L(\mathcal{A}) = \{a, ab, aab, bb\}^*$

An example of semi-flower automaton is shown in figure 4. The reverse of Proposition 7 is not true in general as it is shown in the example of figure 5.

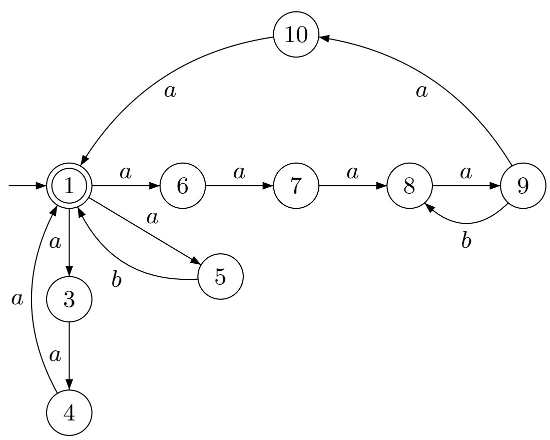

FIGURE 5. $\mathcal{A}$ not semi-flower automaton, $L(\mathcal{A}) = \{ab, aaa\}^*$

However, with the supplementary hypothesis of unambiguity we get also the reverse as stated in the following proposition:

Proposition 8. *Let $\mathcal{A}$ be an unambiguous monoidal automaton such that $L(\mathcal{A}) = H$. The submonoid H is finitely generated if, and only if $\mathcal{A}$ is a semi-flower automaton.*

Proof. Let $\mathcal{A}$ be an unambiguous monoidal automaton that recognizes H finitely generated submonoid. By prop.5, H is free with basis $Y_{\mathcal{A}}$, the set of labels of the cycles that are simple in 1. By hypothesis,

$$|Y_{\mathcal{A}}| = rk(H) < \infty$$

If there exists a cycle c not visiting 1 then, since $\mathcal{A}$ is monoidal, there exists an infinite number of different cycles that are simple in 1, that is a contradiction.
The other implication is proved in prop.7. □

3. Semiflower-automata with at most one bpi

We have defined a branch point going out, in short bpo, (resp. branch point going in, in short bpi) of an automaton as a state x of the automaton where the number of edges going out from x is at least two (resp. the number of edges coming in x is at least two).

Let $\mathcal{A}$ be a monoidal automaton. Let $BPI(\mathcal{A})$ be the set of vertices of $\mathcal{A}$ that are bpi and let $BPO(\mathcal{A})$ be the set of vertices of $\mathcal{A}$ that are bpo.Let us consider now, for each $i \geq 2$, the bpo's of an automaton in which there are i different edges going out from them.
Let $\mathcal{A} = (Q, I, F, \mathcal{F})$ be an automaton over A. For each state $x \in Q$, let m_x be the number of edges going out from x. It is easy to see that $|\mathcal{F}| - |Q| = \sum_{x \in Q}(m_x - 1)$. For each $i \geq 2$, let us consider the set

$$BPO_i(\mathcal{A}) = \{x \in Q \mid m_x = i\}$$

For simplicity of notation, we will moreover consider the sets

$$BPO_0(\mathcal{A}) = \{x \in Q \mid m_x = 0\}, \quad BPO_1(\mathcal{A}) = \{x \in Q \mid m_x = 1\}$$

When no confusion arises we will write BPO_i in place of $BPO_i(\mathcal{A})$.

Example 9. *In figure* **??** *we have $\mathcal{A}$, a monoidal automaton on $A = \{a, b, c\}$ with $BPI(\mathcal{A}) = \{1\}$ and $BPO(\mathcal{A}) = \{1, 2\}$. One has that $BPO_1(\mathcal{A}) = \{3\}$, $BPO_2(\mathcal{A}) = \{1\}$ and $BPO_3(\mathcal{A}) = \{2\}$.*

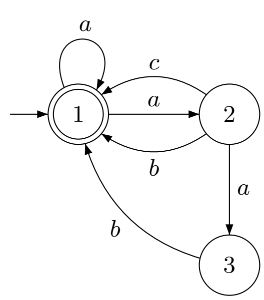

Figure 6. $\mathcal{A}$ monoidal automaton on $A = \{a, b, c\}$.

Let now

$$m_{\mathcal{A}} = max_{x \in Q}\{m_x\}$$

be the maximum number of edges going out from states of $\mathcal{A}$.

Since $Q = \cup_{i \geq 0} BPO_i$ and, for each $i \neq j$, $BPO_i \cap BPO_j = \emptyset$, then

$$|\mathcal{F}| - |Q| = \sum_{i=0,\dots,m_{\mathcal{A}}} \left(\sum_{x \in BPO_i} (m_x - 1) \right)$$

For each $i = 0, \dots m_{\mathcal{A}}$, if $x \in BPO_i$ then $m_x - 1 = i - 1$ and we get

$$|\mathcal{F}| - |Q| = \sum_{i=0,\dots,m_{\mathcal{A}}} |BPO_i|(i-1)$$

If $\mathcal{A}$ is a monoidal automaton then $BPO_0 = \emptyset$ as the following proposition implies:

Proposition 10. *Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be a monoidal automaton. One has $|\mathcal{F}| - |Q| = \sum_{i=2,\dots,m_{\mathcal{A}}} |BPO_i|(i-1)$.*

Proof. Since, by hypothesis, $L(\mathcal{A}) \neq \{\varepsilon\}$ and $\mathcal{A}$ is trim with a unique final state equal to the initial one, then $1 \leq m_x \leq m_{\mathcal{A}}$. □

If, moreover, $\mathcal{A}$ is a deterministic monoidal automaton on the alphabet A of cardinality n it trivially follows that:

Corollary 11. *Let A be an alphabet of cardinality n. Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be a deterministic monoidal automaton. One has*

$$|\mathcal{F}| - |Q| = \sum_{i=2,\dots,n} |BPO_i|(i-1)$$

In the example in figure **??** we have that $|\mathcal{F}| - |Q| = 6 - 3 = 3$ and

$$\sum_{i=2,\dots,n} |BPO_i|(i-1) = |BPO_2| + 2|BPO_3| = 3$$

so the equality of corollary 17 holds.

We have the following proposition that links the existence of bpi with the existence of bpo.

Proposition 12. *Let $\mathcal{A}$ be a monoidal automaton. $BPI(\mathcal{A}) = \emptyset$ if and only if $BPO(\mathcal{A}) = \emptyset$.*

Proof. Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be a monoidal automaton with $BPO(\mathcal{A}) \neq \emptyset$. So there exists $x \in Q$ bpo. Let e_1 and e_2 the two edges starting at x and let us consider p a simple path from $f(e_1)$ to 1 and q a simple path from $f(e_2)$ to 1. Let us consider r the longest suffix path in common between $e_1 p$ and $e_2 q$. The initial vertex of r is a bpi and so $BPI(\mathcal{A}) \neq \emptyset$.

The converse is proved in an analogue way. □

We have also the following characterization of submonoids such that every monoidal automaton recognizing it has not bpi.

Theorem 13. *Let $\mathcal{A}$ be a monoidal automaton. $BPI(\mathcal{A}) = \emptyset$ if and only if $L(\mathcal{A})$ is cyclic.*

Proof. Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be a monoidal automaton such that $BPI(\mathcal{A}) = \emptyset$ then, by prop.9, $BPO(\mathcal{A}) = \emptyset$. If $|Q| = 1$ then trivially $L(\mathcal{A})$ is cyclic. Let $|Q| > 1$ and let $x \in Q$, $x \neq 1$. Let p be a simple path from 1 to x and q a simple path from x to 1 then the cycle pq is, in particular, simple in 1. It is the unique cycle that is simple in 1. In fact if by contradiction there is another cycle that is simple in 1, let it be c, then considering r the longest suffix path in common between pq and c we get that the initial state of r is a bpo that is a contradiction. So if u is the label of pq then $L(\mathcal{A}) = \{u\}^*$.

Conversely let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be a monoidal automaton such that $L(\mathcal{A})$ is cyclic. Then $BPI(\mathcal{A}) = \emptyset$ otherwise there exists $x \in BPO(\mathcal{A})$ and so there exist two different cycles that are simple in 1, that is a contradiction. □

The semi-flower automata with a unique bpi have also interesting properties as we will see. The automaton in figure 4 is a semi-flower automaton such that $BPI(\mathcal{A}) = \{1, 3\}$. The automaton in figure 6 is a semi-flower automaton such that $BPI(\mathcal{A}) = \{1\}$.

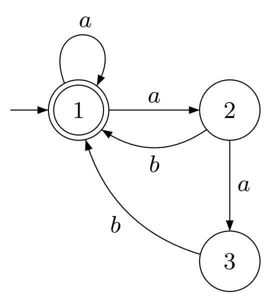

FIGURE 7. $\mathcal{A}$ semi-flower automaton, $|BPI(\mathcal{A})| = 1$

It is interesting to note that if $\mathcal{A}$ is a monoidal automaton with initial-final state 1 such that $BPI(\mathcal{A}) = \{1\}$ then $\mathcal{A}$ is a semi-flower automaton as explained in the following proposition:

Proposition 14. *Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be a monoidal automaton such that $BPI(\mathcal{A}) = \{1\}$. It follows that $\mathcal{A}$ is a semi-flower automaton.*

Proof. Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be a monoidal automaton having $BPI(\mathcal{A}) = \{1\}$. Let c be a cycle in $\mathcal{A}$. If c is a cycle in 1 then obviously c visits 1. So let c be a cycle in $x \neq 1$. Let us p be a simple path from 1 to x and let r be the longest suffix path in common between p and c. If p and c are not suffix each other then $i(r)$ is a bpi different from 1 that is a contradiction. Since p is a simple path then c cannot be a proper suffix of p so it must be p a proper suffix of c and c visits 1. □

In general if $\mathcal{A}$ is a monoidal automaton such that $|BPI(\mathcal{A})| = 1$ then $\mathcal{A}$ is not necessarily a semi-flower automaton as we can see in the example in figure 7.

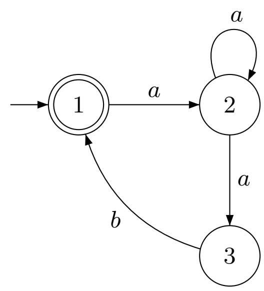

FIGURE 8. $\mathcal{A}$ monoidal automaton with a unique bpi that is not a semi-flower automaton

Let us see now, given a semi-flower automaton with exactly one bpi, how to link the rank of the submonoid generated by it with the characteristics of the automaton. Let us give before some definitions and propositions.

Let us consider now a graph $\Gamma = (V, E)$. We recall that an undirected graph Γ is a *tree* if it is connected and acyclic. Given $x, z \in V$ we say that x is reachable from z if there exists a path in Γ from z to x. There exist different algorithms for visiting a graph. The most known is the *breadth-first search* (BFS in shorts). Given a graph $\Gamma = (V, E)$ and a distinguished *source* vertex s, BFS systematically explores the edges of Γ to "discover" every vertex that is reachable from s. It also produces a tree T called *breadth-first tree* with root s that contains all such reachable vertices. We are in particular interested in the properties of this tree. For the procedure of breadth-first search and breadth-first tree we recall [4]. We recall here the properties of this tree.

Proposition 15. *Let $\Gamma = (V, E)$ be a directed graph. Let $s \in V$. The breadth-first tree $T = (V', E')$ of Γ in s, obtained applying the BFS to Γ in s, is such that:*

1) T is a tree

2) V' consists of all vertices reachable from s

3) For all $v \in V'$ there is a unique simple path from s to v in T, that is also the shortest path from s to v in Γ.

In figure 8 we see an example of a semi-flower automaton with pointed out the edges of the breadth-first tree in 1.

If $\Gamma = (V, E, i, f)$ is a directed multigraph then there exists also a tree subgraph of Γ satisfying the properties in prop. 12.

In fact we can consider $\Gamma_1 = (V_1, E_1)$ the subgraph of Γ obtained from Γ in this way. For each $u, v \in V$, let $E(u, v) = \{e \in E \mid i(e) = u \ , f(e) = v\}$. For each $u, v \in V$, if $E(u, v) \neq \emptyset$ then let us consider $x \in E(u, v)$ and let us call it $x_{uv} = x$. Let $V_1 = V$ and $E_1 = \{x_{uv} \ \forall u, v \in V\}$.

Applying the BFS procedure to the graph Γ_1 we obtain a breadth-first tree $T = (V', E')$ such that:

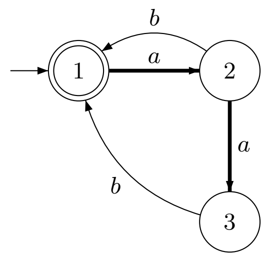

FIGURE 9. $\mathcal{A}$ semi-flower automaton. T BFS tree in 1

1)T is a tree subgraph of Γ_1 and so of Γ

2) For all $v \in V$ there is a unique simple path from s to v in T, that is also a shortest path in Γ_1 and so in Γ.

All this results are resumed in the following proposition:

Proposition 16. *Let $\Gamma = (V, E)$ be a directed multigraph and let $s \in V$. There exists a tree T, subgraph of Γ, such that for all $v \in V$ reachable from s there is a unique simple path from s to v in T, that is also the shortest path from s to v in Γ.*

Let us give now a theorem that link the rank of the submonoid recognized by a semi-flower automaton with a unique bpi to the multigraph associated to it.

Theorem 17. *If $\mathcal{A}$ is a semi-flower automaton with V as set of vertices and E as set of edges such that $|BPI(\mathcal{A})| = 1$ then $rk(L(\mathcal{A})) \leq |E| - |V| + 1$.*

Proof. Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be a semi-flower automaton with a unique bpi. Let $BPI(\mathcal{A}) = \{x\}$. By prop.4 , a set of generators for $L(\mathcal{A})$ is $Y_{\mathcal{A}}$, the finite set of labels of the cycles that are simple in 1. So, we have to count the number of cycles that are simple in 1. Let us prove that there is a bijection between $C_{\mathcal{A}}$, the set of cycles that are simple in 1, and $E(x)$, the set of edges ending at x. Then we will have $rk(L(\mathcal{A})) \leq |Y_{\mathcal{A}}| \leq |C_{\mathcal{A}}| = |E(x)|$.

1.*If $c \in C_{\mathcal{A}}$ then c visits x.*

If $x = 1$ it is done. Let so $x \neq 1$ and let q be a simple path in $\mathcal{A}$ from x to 1. If we consider the longest suffix path in common between q and c we get that if c does not visit x then there exists a bpi different from x (the proof is analogous of that one of prop.11).

2.*There is a unique simple path in $\mathcal{A}$ from x to 1.*

If $x = 1$ then the unique simple path from 1 to 1 is the null path. Let so $x \neq 1$. If, by contradiction, there exists q' simple path from x to 1, $q' \neq q$, then q and q' cannot be suffix each other since they are simple paths with the same initial vertex. So if we consider t the longest suffix path in common between q and q' we have that $i(r)$ is a bpi different from x.

3. *For each edge e ending at x there exists a unique simple path*

$$p_e : 1 \longrightarrow i(e)$$

If by contradiction there exists another simple path q_e from 1 to $i(e)$ then we get as before another bpi considering the longest suffix path in common between p_e and q_e.
For each $e \in E(x)$, let p_e be the unique simple path from 1 to $i(e)$.

Let us define now a map from $C_{\mathcal{A}}$ to $E(x)$. Let

$$\varphi : C_{\mathcal{A}} \longrightarrow E(x)$$

be such that

$$for\ each\ c \in C_{\mathcal{A}},\ \ \varphi(c) = e_c$$

where $e_c \in \mathcal{F}$ is the unique edge in c ending at x.

-φ is well defined since c is a simple cycle so it visits x only once.

-φ is injective. Let $c_1, c_2 \in C_{\mathcal{A}}$ $c_1 \neq c_2$ such that $\varphi(c_1) = \varphi(c_2) = e_c$. Then $c_1 = c_2 = p_{e_c} e_c q$.

-φ is surjective. For each $e \in E(x)$, let p be the simple path from 1 to $i(e)$ then $peq \in C_{\mathcal{A}}$ and $\varphi(peq) = e$.

Since φ is a bijection then:

$$|C_{\mathcal{A}}| = |E(x)|$$

Let us consider now $\mathcal{A}$ as a directed multigraph and denote with V the set of vertices and with E the set of edges of $\mathcal{A}$. Since every $v \in V$ is reachable from 1 then by prop.13 there exists a tree $T = (V(T), E(T))$, undirected subgraph of $\mathcal{A}$, such that $V(T) = V$ and for all $v \in V$ there is a unique simple path from 1 to v in T, that is also a path in $\mathcal{A}$. Let us consider now two cases: when the unique bpi in $\mathcal{A}$ is $x = 1$ and when $x \neq 1$. Let us note that, for each e ending at x, p_e is the unique simple path from 1 to v in T.

(1)Let $x = 1$.
Let us prove that

$$E \setminus E(T) = E(1)$$

-$E \setminus E(T) \subseteq E(1)$. In fact, let $e \in E \setminus E(T)$ then $f(e) = 1$ otherwise $p_e e$ is a not null simple (since $\mathcal{A}$ is semi-flower) path in T from 1 to $f(e)$. Since $f(e)$ is not a bpi then e is contained in p_e and then in T, contradiction!

-$E(1) \subseteq E \setminus E(T)$. In fact, if, by contradiction, there exists $e \in E(T)$ with $f(e) = 1$ then $p_e e$ is a not null cycle in T, contradiction!

(2) Let $x \neq 1$.

- *There is a unique edge in T ending at x*

Let e_x be an the edge in T ending at x and let by contradiction f be another edge in T ending at x, $f \neq e_x$. The paths $p_f f$ and $p_{e_x} e_x$ are simple since $\mathcal{A}$ is semi-flower. So, they are different simple paths from 1 to x in T, that is a contradiction by prop.13.

We will show now that all edges of $\mathcal{A}$ missing in T are those ones ending at x plus e_q, the end edge of q, minus e_x.

$$E \setminus E(T) = (E(x) \cup \{e_q\}) \setminus \{e_x\}$$

-$\{e_q\} \subseteq E \setminus E(T)$. Otherwise the path $p_{e_q}e_q$ is a not null cycle in T.

-$(E(x) \setminus \{e_x\}) \subseteq E \setminus E(T)$. Let $f \in E(x)$, $f \neq e_x$. As explained before e_x is the only edge in T arriving in x and so $f \notin E(T)$.

- $E \setminus E(T) \subseteq (E(x) \cup \{e_q\}) \setminus \{e_x\}$. Let $e \in E \setminus E(T)$. Let us suppose that $f(e) \neq x$ and $e \neq e_q$. By hypothesis $f(e)$ is not a bpi and $f(e) \neq 1$. Since $f(e) \neq 1$ the path $p_e e$ is a not null path and since $f(e)$ is not a bpi then e is contained in $p_e e$ and then in T, contradiction!

So we have that

$$|E \setminus E(T)| = |(E(x) \cup \{e_q\}) \setminus \{e_x\}| = |E(x)| = |C_{\mathcal{A}}|$$

and by the properties of trees we have

$$|E(T)| = |V(T)| - 1 = |V| - 1$$

and so by prop. 7

$$|C_{\mathcal{A}}| = |E \setminus E(T)| = |E| - |V| + 1$$

So

$$rk(H) \leq |C_{\mathcal{A}}| = |E| - |V| + 1$$

□

Let us note that the statement of theorem 14 is not true without the hypothesis of $|BPI(\mathcal{A})| = 1$ as we can see in figure 9.

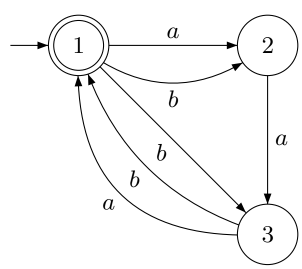

FIGURE 10. $\mathcal{A}$ semi-flower automaton, BPI($\mathcal{A}$) $= \{1, 2, 3\}$ and $L(\mathcal{A}) = \{aab, aaa, bab, baa, bb, ba\}^* > |E| - |V| + 1 = 4$

It follows from the proof of theorem 14 that in the unambiguous case the following holds:

Proposition 18. *If $\mathcal{A}$ is an unambiguous semi-flower automaton with V as set of vertices and E as set of edges such that $|BPI(\mathcal{A})| = 1$ then $rk(L(\mathcal{A})) = |E| - |V| + 1$.*

We remark that a similar result holds for free groups (cf.[9]): if $\mathcal{A}$ is an inverse automaton with V as set of vertices and E as set of edges recognizing a subgroup H then $rk(H) = |E| - |V| + 1$.

Let us consider now deterministic semi-flower automata. The submonoids recognized by these automata are generated by finite prefix sets.

Proposition 19. *Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be a deterministic semi-flower automaton. $\mathcal{A}$ recognizes a free submonoid generated by a finite prefix set.*

Proof. Let $\mathcal{A} = (Q, 1, 1, \mathcal{F})$ be a deterministic semi-flower automaton. By prop.5 and prop.7 , $L(\mathcal{A})$ is a free finitely generated submonoid with base the set of labels of the simple cycles in 1. Let us prove that this is a prefix set. By contradiction, let u, v be labels of c_u and c_v, simple cycles in 1, such that $v = uw$ with w not empty word. $\mathcal{A}$ is deterministic so $c_v = c_u c_w$ with c_w the cycle with label w. This is a contradiction because c_v is simple. □

In general given a submonoid X^* of A^* generated by a finite prefix set X we can easily construct an automaton recognizing X^*: the *literal automaton of* X^* (cf.[1]). It is the automaton

$$\mathcal{A}_X = (Q, \varepsilon, \varepsilon, \delta)$$

where $Q = X(A^+)^{-1}$ is the set of the proper prefixes of X and where

$$\delta(u, a) = \begin{cases} ua & \text{if } ua \in X(A^+)^{-1} \\ \varepsilon & \text{if } ua \in X \\ \emptyset & \text{otherwise} \end{cases}$$

It is immediate that $L(\mathcal{A}_X) = X^*$.

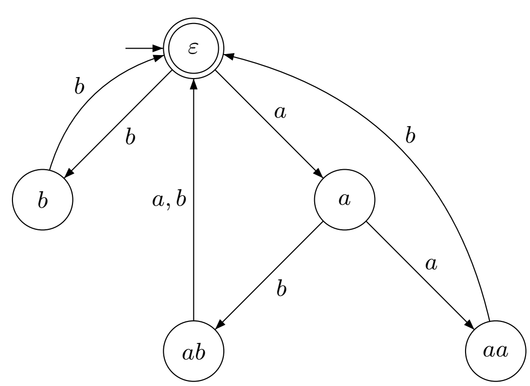

FIGURE 11. $\mathcal{A}_X$ with $X = \{bb, aba, abb, aaa\}$

Proposition 20. *Let X be a finite prefix set then $\mathcal{A}_X$ is a deterministic semi-flower automaton with at most the state ε as bpi.*

Proof. Let $\mathcal{A}_X = (Q, \varepsilon, \varepsilon, \delta)$. By construction $\mathcal{A}_X$ is a deterministic monoidal automaton. Since $L(\mathcal{A}_X) = X^*$ and X is finite then, by prop.8, $\mathcal{A}_X$ is a semi-flower automaton. If it has not bpi then it is done. Let us suppose that there exists at least one bpi.

Let us suppose, by contradiction, that there exists $u \in Q$ bpi for $\mathcal{A}_X$ and $u \neq \varepsilon$. So in $\mathcal{A}_X$ there exist two different edges ending at u, $e_1 : v_1 \xrightarrow{a} u$ and $e_2 : v_2 \xrightarrow{b} u$. Since $u \neq \varepsilon$ it is $u = v_1 a = v_2 b$. Since a, b are letters in A it follows that $a = b$ then $v_1 = v_2$ and so $e_1 = e_2$, contradiction since $e_1 \neq e_2$. □

So we have that if $\mathcal{A}$ is a deterministic semi-flower automaton with at most one bpi then $L(\mathcal{A})$ is a submonoid generated by a finite prefix set and conversely given a finite prefix set X there exists a deterministic semi-flower automaton with at most the initial vertex as bpi recognizing X^*. So the class of deterministic semi-flower automata with at most the initial vertex as bpi recognizes the class of submonoids generated by finite prefix sets.

In the example in figure 10 $\mathcal{A}_X$ is a deterministic semi-flower automaton with a unique bpi (the vertex ε), with three bpo (the vertices ε, a and ab) and $rk(X^*) = 4 = |BPO_2| + 1$.

Let us note that in all figures every edge with label a, b has to be understood as two edges with labels a and b, respectively.

4. Intersection of two submonoids

In this section we investigate the intersection of two finitely generated submonoids of A^* by studying the product of two deterministic semi-flower automata recognizing the two given submonoids.

For the definitions and the properties the of product automaton we recall [5]. It is well known that the product of two automata recognizing submonoids recognizes the intersection of the two submonoids.

Some properties of two given automata $\mathcal{A}_1$ and $\mathcal{A}_2$ are saved in the product $\mathcal{A}_1 \times \mathcal{A}_2$ as shown in the following proposition:

Proposition 21. *Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two automata.*

(1) *If $\mathcal{A}_1$ and $\mathcal{A}_2$ are monoidal automata then $\mathcal{A}_1 \times \mathcal{A}_2$ is a monoidal automaton.*
(2) *If $\mathcal{A}_1$ and $\mathcal{A}_2$ are unambiguous automata then $\mathcal{A}_1 \times \mathcal{A}_2$ is a unambiguous automaton.*
(3) *If $\mathcal{A}_1$ and $\mathcal{A}_2$ are deterministic automata then $\mathcal{A}_1 \times \mathcal{A}_2$ is a deterministic automaton.*

On the other hand the product of two trim automata is not necessarily a trim automaton. Moreover the product of two semi-flower automata is not necessarily a semi-flower automaton (see example in figure 11). This is in agreement with the fact that the intersection of two finitely generated submonoids is not necessarily finitely generated. Further, if the product of two semi-flower automata with a unique bpi is a semi-flower automaton, then it is not necessarily a semi-flower automaton with a unique bpi (see example in figure 12).

Let now $\mathcal{A}_1$ and $\mathcal{A}_2$ be two deterministic automata on A

$$\mathcal{A}_1 = (Q_1, q_1, F_1, \delta_1),\ \mathcal{A}_2 = (Q_2, q_2, F_2, \delta_2)$$

The product automaton is defined as

$$\mathcal{A}_1 \times \mathcal{A}_2 = (Q_1 \times Q_2, (q_1, q_2), F_1 \times F_2, \delta)$$

with for each $(x, y) \in Q_1 \times Q_2$, for each $a \in A$

$$\delta((x, y), a) = (\delta_1(x, a), \delta_2(y, a))$$

Let us give now a lemma that links the bpo in the product of two deterministic automata with the bpo of the respective automata:

Lemma 22. *Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two deterministic automata on A, alphabet of cardinality n. One has $BPO_t(\mathcal{A}_1 \times \mathcal{A}_2) \subseteq \cup_{t \leq r,s \leq n}(BPO_r(\mathcal{A}_1), BPO_s(\mathcal{A}_2))$.*

Proof. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two deterministic automata on $A = \{\alpha_1, \ldots, \alpha_n\}$, $\mathcal{A}_1 = (Q_1, q_1, F_1, \delta_1)$ and $\mathcal{A}_2 = (Q_2, q_2, F_2, \delta_2)$.

If $(x, y) \in BPO_t(\mathcal{A}_1 \times \mathcal{A}_2)$ then there exist t different edges going out from (x, y). So, for each $i = 1, \ldots, t$, there exists $(x_i, y_i) \in Q_1 \times Q_2$, $\alpha_i \in A$ such that

$$\delta((x, y), \alpha_i) = (x_i, y_i)$$

And, since $\mathcal{A}_1 \times \mathcal{A}_2$ is deterministic, for $i \neq j$ $\alpha_i \neq \alpha_j$.

By definition, for each $i = 1, \ldots, t$

$$\delta((x, y), \alpha_i) = (\delta(x, \alpha_i), \delta(y, \alpha_i)) = (x_i, y_i)$$

Then for each $i = 1, \ldots, t$,

$$\delta(x, \alpha_i) = x_i, \quad \delta(y, \alpha_i) = y_i$$

Since $\alpha_i \neq \alpha_j$, for each i, j with $i \neq j$, then there are at least t different edges going out from x and so $x \in BPO_r(\mathcal{A}_1)$ with $r \geq t$. Analogously $y \in BPO_s(\mathcal{A}_2)$ with $s \geq t$ and so it follows the thesis. □

Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be deterministic automata on A alphabet of cardinality n. Let $a_i := |BPO_i(\mathcal{A}_1)|$ and $b_i := |BPO_i(\mathcal{A}_2)|$ for each $i = 1, \ldots, n$. In terms of cardinality of the sets of bpo's the lemma 22 became:

Corollary 23. $|BPO_t(\mathcal{A}_1 \times \mathcal{A}_2)| \leq \sum_{t \leq r,s \leq n} a_r b_s$.

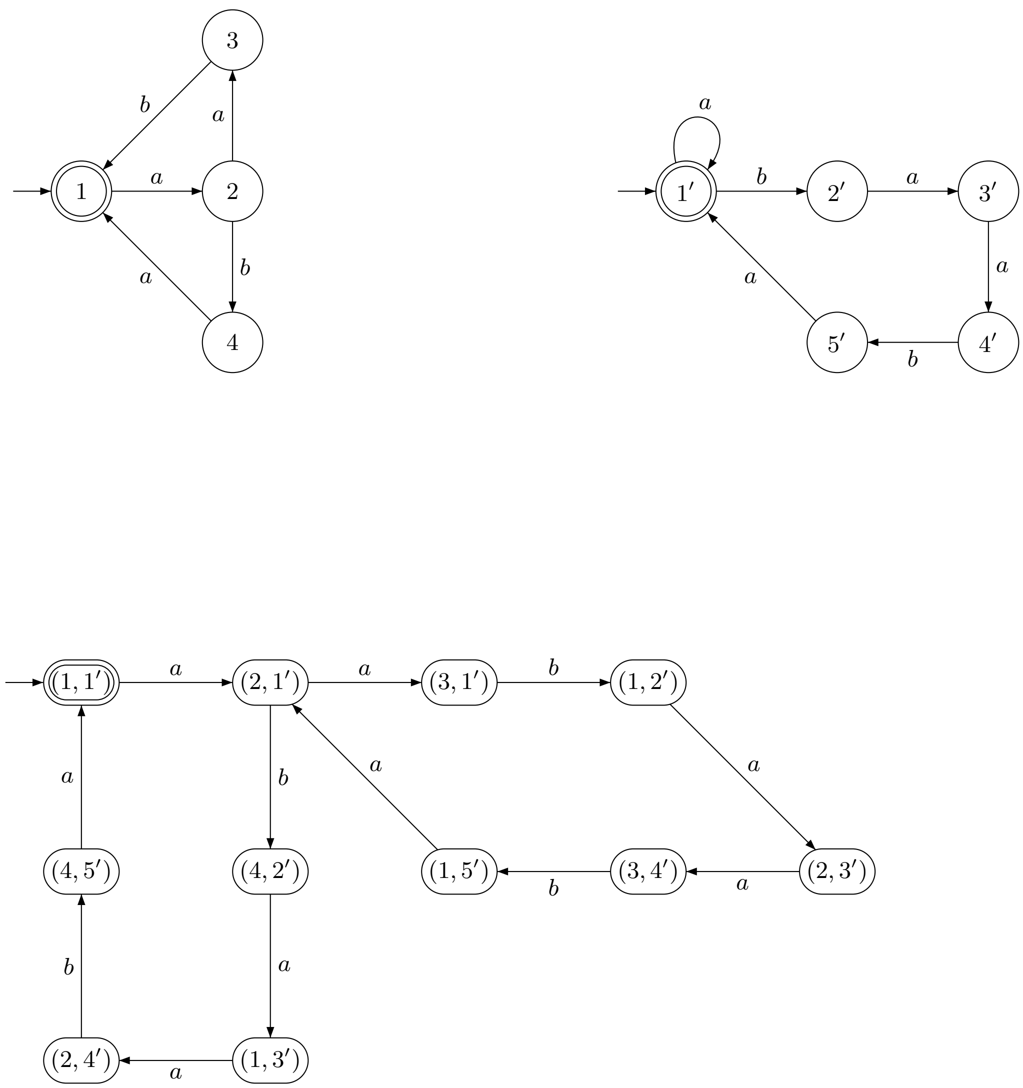

FIGURE 12. $\mathcal{A}_H$, $\mathcal{A}_K$ and $\mathcal{A}_H \times \mathcal{A}_K$ with $H = \{aab, aba\}^*$ and $K = \{a, baaba\}^*$

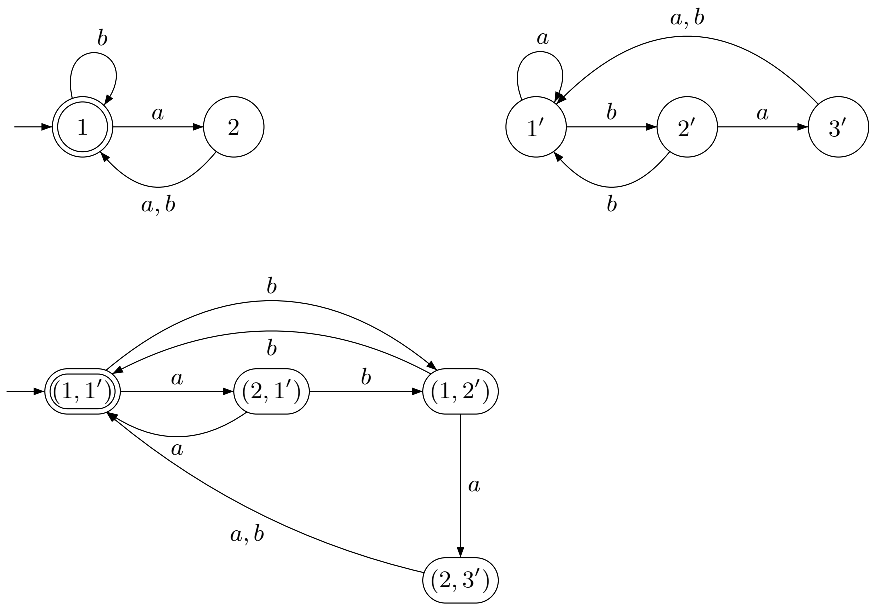

FIGURE 13. $\mathcal{A}_H$, $\mathcal{A}_K$ and $\mathcal{A}_H \times \mathcal{A}_K$ with $H = \{b, aa, ab\}^*$ and $K = \{a, bb, baa, bab\}^*$

Let $\mathcal{A}$ be an automaton. We call $\mathcal{A}^T$ the automaton obtained from $\mathcal{A}$ taking only the accessible and co-accessible states.

Let us give now a lemma that will allow us to prove the inequality of Hanna Neumann for submonoids recognized by deterministic semi-flower automata with a unique bpi. Let $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$ be two subsets of the natural numbers $\mathbb{N}$.

Let

$$P_{AB} = \sum_{t=2,\ldots,n} (t-1)\Big(\sum_{t \le r \le n} a_r \sum_{t \le s \le n} b_s\Big)$$

and

$$Q_{AB} = \Big(\sum_{i=2,\ldots,n} (i-1)a_i\Big)\Big(\sum_{j=2,\ldots,n} (j-1)b_j\Big)$$

Lemma 24. *Let $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$ be subsets of $\mathbb{N}$. One has $P_{AB} \le Q_{AB}$. Moreover if there exists $k, l > 2$ such that $a_k \neq 0$ and $b_l \neq 0$ then $P_{AB} < Q_{AB}$.*

Proof. Developing P_{AB} and Q_{AB} we get

$$P_{AB} = \sum_{2 \le k, l \le n} \alpha_{kl} a_k b_l$$

and

$$Q_{AB} = \sum_{2 \le k,l \le n} \beta_{kl} a_k b_l$$

Our thesis is that $\alpha_{kl} \le \beta_{kl}$, for each k, l.

Let us write $P_{AB} = P_2 + 2P_3 + \ldots + (n-1)P_n$ where $P_t = P_t' P_t''$ with $P_t' = \sum_{t \le r \le n} a_r$ and $P_t'' = \sum_{t \le s \le n} b_s$.

Let $k, l \in \{2, \ldots, n\}$. Then a_k appears in P_i' if and only if $i \le k$ and b_l appears in P_i'' if and only if $i \le l$. Then $a_k b_l$ appears in $P_i' P_i''$ if and only if $i = 1, \ldots, \overline{k}$ where $\overline{k} = min(k, l)$. So $a_k b_l$ appears in P_{AB} with coefficient

$$\alpha_{kl} = 1 + 2 + \ldots + (\overline{k} - 1)$$

Let $Q = Q'Q''$ with $Q' = \sum_{i=2,\ldots,n} (i-1)a_i$ and $Q'' = \sum_{i=2,\ldots,n} (i-1)b_i$. Let $k, l \in \{2, \ldots, n\}$. Then a_k appears in Q' with coefficient $k-1$ and b_l appears in Q'' with coefficient $l-1$ so

$$\beta_{kl} = (k-1)(l-1)$$

It is $1+2+\ldots+(\overline{k}-1) < (\overline{k}-1)(\overline{k}-1)$ for $\overline{k} > 2$ and $1+2+\ldots+(\overline{k}-1) = (\overline{k}-1)(\overline{k}-1)$ for $\overline{k} = 2$.

So $\alpha_{kl} \le \beta_{kl}$, for each $k, l \ge 2$. Moreover $\alpha_{kl} < \beta_{kl}$ for $\overline{k} > 2$. So we get that $P_{AB} \le Q_{AB}$.

Let us suppose that there exists $k > 2$ such that $a_k \ne 0$ and there exists $l > 2$ such that $b_l \ne 0$. Since $\alpha_{kl} < \beta_{kl}$ and $a_k b_l \ne 0$ we get $P_{AB} < Q_{AB}$. □

Let now H and K be submonoids finitely generated by prefix sets. Let $\mathcal{A}_H$ and $\mathcal{A}_K$ be deterministic semi-flower automata with a unique bpi recognizing H and K, respectively. Since $\mathcal{A}_H$ and $\mathcal{A}_K$ are deterministic monoidal automata then $\mathcal{A}_H \times \mathcal{A}_K$ is still a deterministic monoidal automaton.

Remark 25. If $(\mathcal{A}_H \times \mathcal{A}_K)^T$ is a semi-flower automaton that has not bpi then $\widetilde{rk}(H \cap K) \le \widetilde{rk}(H)\widetilde{rk}(K)$. In fact, if $H \cap K = \emptyset$ then $\widetilde{rk}(H \cap K) = 0$ otherwise by prop.10, $H \cap K$ is cyclic and so $\widetilde{rk}(H \cap K) = 0$.

If $(\mathcal{A}_H \times \mathcal{A}_K)^T$ is a semi-flower automaton with a unique bpi we get the Hanna Neumann inequality as it is stated in the following theorem:

Theorem 26. *If $(\mathcal{A}_H \times \mathcal{A}_K)^T$ is a semi-flower automaton with a unique bpi then $\widetilde{rk}(H \cap K) \le \widetilde{rk}(H)\widetilde{rk}(K)$. Moreover the strict inequality holds if there exist $i, j > 2$ such that $BPO_i(\mathcal{A}_H) \ne \emptyset$ and $BPO_j(\mathcal{A}_K) \ne \emptyset$.*

Proof. Let $(\mathcal{A}_H \times \mathcal{A}_K)^T$ be a deterministic semi-flower automaton with a unique bpi. Since $BPO_t((\mathcal{A}_H \times \mathcal{A}_K)^T) \subseteq BPO_t(\mathcal{A}_H \times \mathcal{A}_K)$, for each $t = 1, \ldots, n$, applying cor.23 and theorem 18 we get:

$$\widetilde{rk}(H \cap K) \le \sum_{t=2\ldots n} (t-1)(\sum_{t \le r \le n} a_r \sum_{t \le s \le n} b_s)$$

On the other hand by theorem 18 it is

$$\widetilde{rk}(H)\widetilde{rk}(K) = (\sum_{i=2,\dots n}(i-1)a_i)(\sum_{j=2,\dots n}(j-1)b_j)$$

Applying lemma 24 to the sets $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_n\}$ we get that $\widetilde{rk}(H \cap K) \leq \widetilde{rk}(H)\widetilde{rk}(K)$.

Let us suppose that there exist $i, j > 2$ such that $BPO_i(\mathcal{A}_H) \neq \emptyset$ and $BPO_j(\mathcal{A}_K) \neq \emptyset$. Then $|BPO_i(\mathcal{A}_H)| = a_i \neq 0$, $i > 2$ and $|BPO_j(\mathcal{A}_H)| = b_j \neq 0$, $j > 2$. And by lemma 24 we get that $\widetilde{rk}(H \cap K) < \widetilde{rk}(H)\widetilde{rk}(K)$. □

If H and K are submonoids finitely generated by prefix sets such that $(\mathcal{A}_H \times \mathcal{A}_K)^T$ is a deterministic semi-flower automaton with more than one bpi then it is not more true that $\widetilde{rk}(H \cap K) \leq \widetilde{rk}(H)\widetilde{rk}(K)$.

There is a family of examples such that $rk(H \cap K) = 2^{log_2(rk(H))log_2(rk(K))}$:

Example 27. *Let p and q be two positive coprime integers. Let A be a binary alphabet and let $H = A^p$ and $K = A^q$, the set of words of length p and q respectively.*
One has $rk(H) = 2^p$ and so $p = log_2(rk(H))$. It is $H \cap K = A^{pq}$ and $rk(H \cap K) = 2^{pq} = 2^{log_2(rk(H))log_2(rk(K))}$.

See examples in figures 13 and 14

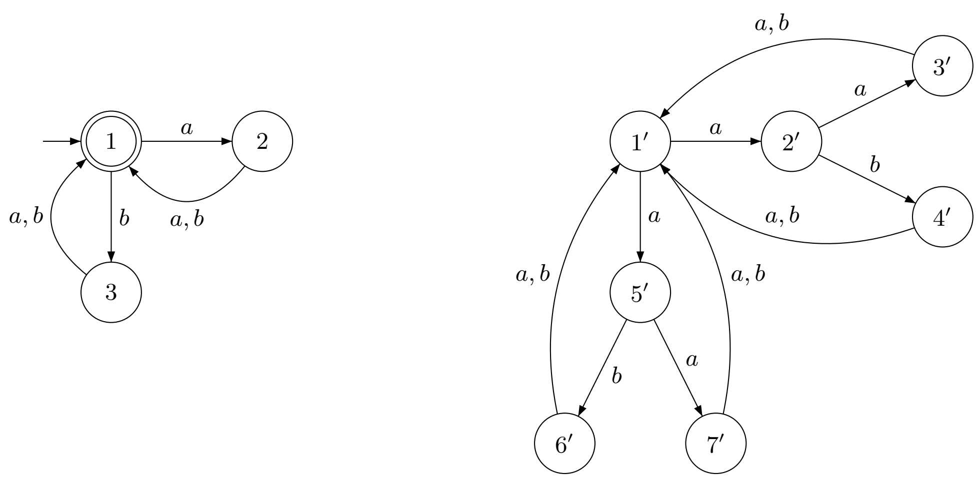

FIGURE 14. $\mathcal{A}_{A^2}$, $\mathcal{A}_{A^3}$ with A^2 the set of words in $A = \{a, b\}$ of length 2 and A^3 the set of words in $A = \{a, b\}$ of length 3

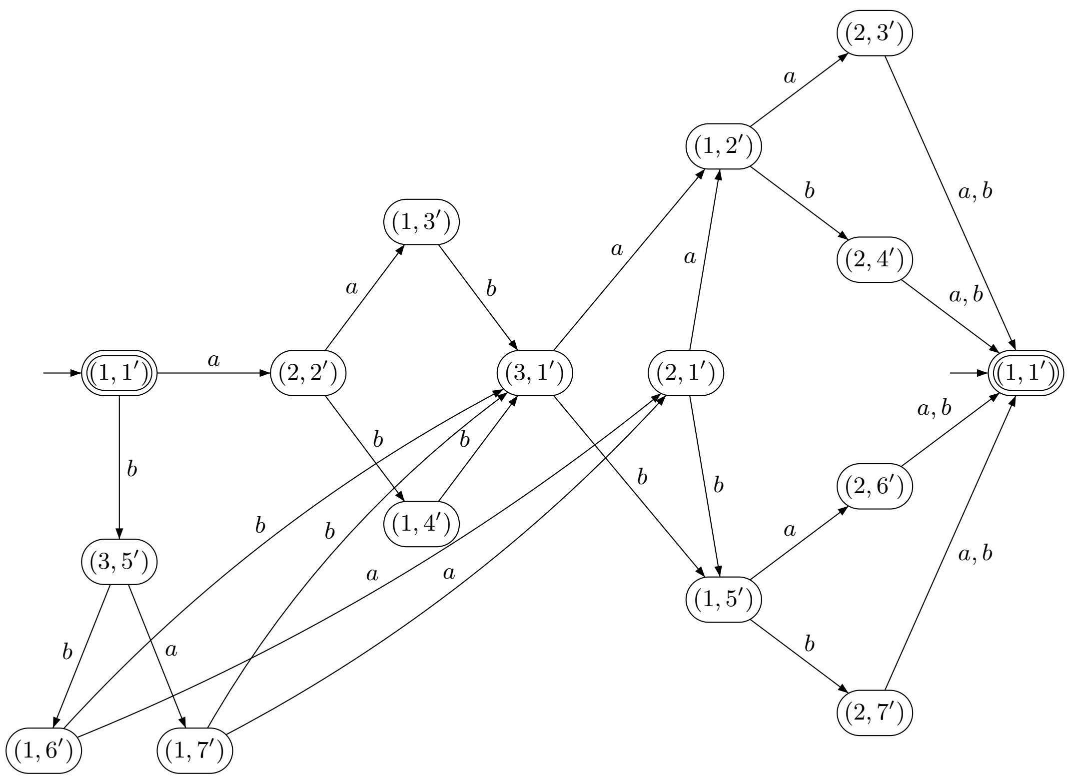

FIGURE 15. $\mathcal{A}_{A^2} \times \mathcal{A}_{A^3}$

5. Prefix case with two generators

If H and K are submonoids generated by prefix sets of two elements of A^* we get the result of Karhumäki (cf.[7]). In particular we prove that there exists a product automaton with at most one bpi, so in the case of finite rank of the intersection we obtain that the rank of the intersection is less than or equal 2.

Let H be a submonoid of A^* generated by a finite prefix set of two elements X. Let $\mathcal{A}_X$ be the literal automaton of X^*, let us call it $\mathcal{A}_H$ for simplicity of notation. In the beginning we prove that $\mathcal{A}_H$ has a unique bpo with two edges starting at it and all the other vertices have just one edge ending at them.

Proposition 28. *Let H be a submonoid generated by a prefix set of two elements. Let $\mathcal{A}_H = (Q, 1, 1, \delta)$ be the literal automaton of H.*

For each $x \in Q$, it is $1 \leq m_x \leq 2$. Moreover there is at most one vertex $y \in Q$ such that $m_y = 2$.

Proof. Let H be a submonoid finitely generated by a prefix set of two elements $X = \{u, v\}$. Let $\mathcal{A}_H = (Q, 1, 1, \delta)$ be the literal automaton of H. Since $\mathcal{A}_H$ is monoidal then $1 \leq m_x$. Let $Q = X(A^+)^{-1} = \{u_1, \ldots, u_n\}$.

Let 's observe that, for each $u_i \in Q$, there exists an edge starting at u_i with label a if and only if $u_i a$ is a prefix of u or a prefix of v.

Let $\overline{u}$ be the longest prefix in common between u and v, $u = \overline{u}u_1$ and $v = \overline{u}v_1$ with $u_1 \in aA^*$, $v_1 \in bA^*$, $a, b \in A$ and $a \neq b$. The vertex $\overline{u}$ is a bpo for $\mathcal{A}_H$ since $\overline{u}a$ is a prefix of u and $\overline{u}b$ is a prefix of v and $a \neq b$. So $m_{\overline{u}} = 2$.

Let now $w \in Q$, $w \neq \overline{u}$. If $|w| < |\overline{u}|$ then w is a proper prefix of $\overline{u}$ and so $\overline{u} = wu'$ with $u' \in cA^*$, $c \in A$. So $e : w \xrightarrow{c} wc$ (or $e : w \xrightarrow{c} 1$ if $wc = \varepsilon$) is the unique edge starting from w and we get $m_w = 1$. If $|w| > |\overline{u}|$ then either w is a prefix of u or a prefix of v, it cannot be a prefix of both since $\overline{u}$ is the longest prefix in common between u and v. So there is a unique edge starting from w and $m_w = 1$. □

For the product automaton of two literal automata, associated to submonoids generated by prefix sets of two elements, it holds the following:

Lemma 29. *Let H and K be submonoids generated by prefix sets of two elements such that $H \cap K \neq \emptyset$. Let $(\mathcal{A}_H \times \mathcal{A}_K)^T = (Q, 1, 1, \delta)$.*

For each $w \in Q$, it is $1 \leq m_w \leq 2$. Moreover there is at most one vertex $z \in Q$ such that $m_z = 2$.

Proof. Since $(\mathcal{A}_H \times \mathcal{A}_K)^T$ is monoidal and $H \cap K \neq \emptyset$ then, for each $x \in Q$, it is $m_x \geq 1$. Let $\mathcal{A} = (\mathcal{A}_H \times \mathcal{A}_K)^T$. Let $a_i := |BPO_i(\mathcal{A}_H)|$ and $b_i := |BPO_i(\mathcal{A}_K)|$ for each $i = 1, \ldots, n$.
By cor.23 we have that for each $t \geq 0$

$$|BPO_t(\mathcal{A}^T)| \leq \sum_{t \leq r, s \leq n} a_r b_s$$

By prop. 28 we get $a_2 = 1$, $b_2 = 1$ and for each $s > 2$ $a_s = 0$ and $b_s = 0$.

For each $t > 2$ we obtain $|BPO_t(\mathcal{A}^T)| \leq 0$ and so, for each $x \in Q$, $m_x \leq 2$. For $t = 2$ we get $|BPO_2(\mathcal{A}^T)| \leq a_2 b_2 = 1$ and so there is only one $x \in Q$ such that $m_x = 2$. □

We can now prove the result of Karhumäki (cf.[7]):

Theorem 30. *Let H and K be submonoids generated by prefix sets of two elements.*

1. *If $H \cap K$ is finitely generated then $rk(H \cap K)$ is at most two.*
2. *If $H \cap K$ is not finitely generated then there exist $\alpha, \beta, \gamma \in A^*$ such that $H \cap K = (\alpha(\beta)^*\gamma)^*$.*

Proof. Let H and K be submonoids generated by prefix sets of two elements. We denote $\mathcal{A} := (\mathcal{A}_H \times \mathcal{A}_K)^T$.
(1)*Let $rk(H \cap K) < \infty$*

Since $\mathcal{A}$ is a deterministic monoidal automaton recognizing $H \cap K$, that is finitely generated, then, by prop.8, $\mathcal{A}$ is a semi-flower automaton.
Let us prove that $\mathcal{A}$ has at most one bpi. If $\mathcal{A}$ has exactly one bpi then, by theorem 26, it will follow $\widetilde{rk}(H \cap K) \leq \widetilde{rk}(H)\widetilde{rk}(K)$.

-If there are not bpi then either $H \cap K = \emptyset$ or, by theorem 10, $H \cap K$ is a cyclic submonoid. In all cases $rk(H \cap K)$ is at most two.

- Let us suppose now that there is at least one bpi, we'll prove that it is the only one.

By contradiction, let x and x' be two different bpi of $\mathcal{A}$. Let e_1 and e_2 be the edges ending at x. Let us consider the simple paths p and q from 1 to $i(e_1)$ and from 1 to $i(e_2)$, respectively. Let p_1 be the longest prefix path in common between pe_1 and qe_2, $pe_1 = p_1p_2$ and $qe_2 = p_1q_2$. The paths pe_1 and qe_2 are simple paths since $\mathcal{A}$ is semi-flower and since they are with the same end state then $f(p_1)$ is a bpo. Let us call it y.

Let us consider now the bpi x'. Let e_3 and e_4 be the edges ending at x'. Analogously as before there exist r_1, r_2, s_2 simple paths such that $r = r_1r_2$ and $s = r_1s_2$ are paths starting at 1 and ending at x', $r_2 = {r'}_2e_3$, $s_2 = {s'}_2e_4$ and $y' = f(r_1)$ is a bpo.

If $y \neq y'$ there is a contradiction since, by lemma 29, there is at most a bpo. Therefore $y = y'$ and let f_1 and f_2 be the edges starting at y. The paths p_2 and q_2 have to contain as beginning edges f_1 and f_2 or f_2 and f_1. Analogously for r_2 and s_2. Let us consider the paths between p_2, q_2, r_2 and s_2 starting with the edge f_1, and suppose that they are p_2 and r_2. If we consider t and t' the longest prefix paths in common between p_2, r_2 and q_2, s_2, respectively, then if either t or t' is not a proper prefix of p_2, r_2 and q_2, s_2, respectively, then the final vertex of t or t' is a bpo different from y, which is a contradiction, by lemma 29.

Therefore, let us assume that $r_2 = p_2r'$. It follows that $s_2 = q_2s'$, otherwise q_2 would visit x' and so q_2r' would visit x' two times, which is a contradiction since q_2 and r' do not visit 1.

Therefore $r_2 = p_2r'$ and $s_2 = q_2s'$. So r' and s' are two different simple paths starting at x and ending at the same vertex x'. If t'' is the longest prefix path in common between them, then $f(t'')$ is a bpo different from y, contradiction!

Analogously it is proved assuming $p_2 = r_2r'$.

(2)Let $rk(H \cap K) = \infty$

By prop.7, $\mathcal{A}$ is not a semi-flower automaton and so there is a cycle c not visiting 1. Let c be a cycle in x. We may assume that c is simple otherwise we take as c the first simple cycle in which c is decomposed. We denote by v the label of c.

Let p be a simple path from 1 to x, u its label. Let q be a simple path from x to 1, w its label. So the cycle

$$pcq : 1 \xrightarrow{u} x \xrightarrow{v} x \xrightarrow{w} 1$$

is simple in 1. Moreover, for each $n > 0$, the cycle $pc^n q$ is simple in 1.

Let p_1 be the longest prefix path in common between c and q, $c = p_1 c_1$ and $q = p_1 q_1$. The paths c_1 and q_1 are not null paths so $i(c_1) = i(q_1) = x_1$ is a bpo. Let v_1 be the label of p_1, v_2 the label of c_1 and w_1 the label of q_1.

We get

$$pp_1 : 1 \xrightarrow{uv_1} x_1, \quad c_1 p_1 : x_1 \xrightarrow{v_2 v_1} x_1, \quad q_1 : x_1 \xrightarrow{w_1} 1$$

with x_1 bpo.

Let $\alpha = pp_1$, $\beta = c_1 p_1$ and $\gamma = q_1$. If α is not a simple path then let us take as α the simple path from 1 to $f(\alpha)$. Analogously for γ and if β is not a simple cycle then we'll take as β the simple cycle from $i(\beta)$ to $f(\beta)$. So we have that

$$\alpha\beta\gamma : 1 \xrightarrow{uv_1} x_1 \xrightarrow{v_2 u_1} x_1 \xrightarrow{w_1} 1$$

We have found that ${C_{\mathcal{A}}}' = \{\alpha(\beta)^n\gamma \mid n \geq 0\} \subseteq C_{\mathcal{A}}$ the set of cycles that are simple in 1.

We'll prove now that these are the only cycles in $\mathcal{A}$ that are simple in 1. Let us suppose, by contradiction, that there exists a cycle d that is simple in 1 and such that $d \notin {C_{\mathcal{A}}}'$.

For each $c_i = \alpha(\beta)^i\gamma \in {C_{\mathcal{A}}}'$, let p_i be the longest prefix path in common between d and c_i. Let $P = \{p_i \mid i \geq 0\}$. Since $|p_i| \leq |d|$, for each i, we may consider $p_j \in P$ such that $|p_j| = max_{i \geq 0}\{|p_i|\}$. Then $d = p_j d'$ and $c_j = p_j c'$ and $f(p_j)$, the end vertex of p_j, is a bpo.

If $f(p_j) \neq f(\alpha)$ then there are two different bpo's in $\mathcal{A}$ and this is a contradiction to lemma 29.

If $f(p_j) = f(\alpha)$ then there exists $k \geq 0$ such that $p_j = \alpha\beta^k$ since p_j is a prefix of c_j and α, β and γ are simple. Let e_γ be the initial edge of γ and e_β be the initial edge of β. Since, by lemma 29, every vertex has at most two edges going out from it then there are only e_γ and e_β going out from $f(p_j)$. It follows that $d = \alpha\beta^k e_\gamma x$ and $c_j = \alpha\beta^k e_\beta y$ or $d = \alpha\beta^k e_\beta x$ and $c_j = \alpha\beta^k e_\gamma y$ for some $x, y \in A^*$. In the first case c_k and d have a prefix in common longer than p_j contradiction! In the second case c_{k+1} and d have a prefix in common longer than p_j contradiction! □

It follows trivially the same result for H and K submonoids finitely generated by suffix sets of two elements.

6. CONCLUSIONS

The results reported in the present paper show that the automata's tools we have introduced are useful for studying the intersection of two submonoids of a free monoid. We think that such methods could be extended to more general cases than the ones considered in this paper, in order to discover upper bounds on the rank of the intersection, when it is finite, depending on the ranks of the submonoids.

Two research's directions appear of particular interest. A first direction is to relate properties of the product of non deterministic automata to study the intersection of two submonoids generated by finite non prefix sets of words. Another possible research's direction is to consider, even in the prefix case, the semi-flower automata with a fixed number of bpi's greater than one.

As a more general problem, one could study the intersection of a finite number of submonoids of rank two trying to discover in the case of finite rank of the intersection a result analogous of that one of Karhumäki ([7]).

References

[1] J. Berstel and D. Perrin. *Theory of Codes.* Academic Press, 1985.

[2] V. Bruyére, D. Derencourt, M. Latteux The meet operation in the lattice of codes *Theoretical Computer Science*, 191:117–129, 1998.

[3] J. Clement, J. Duval, G.Guaiana, D. Perrin, G. Rindone Paarsing with a finite dictionary *Theoretical Computer Science*, 340:432–442, 2005.

[4] H.Cormen, E. Leiserson, L. Rivest *Introduction to Algorithms.* The MIT press, 1990.

[5] J.E. Hopcroft, J.D.Ullman *Introduction to Automata Theory, Languages and Computation.* Addison-Weisley Publishing Company, 1979.

[6] A.G. Howson. On the intersection of finitely generated free groups. *J. London Math. Soc:*, 29:428–434, 1954.

[7] J. Karhumäki A note on intersection of free submonoids of a free monoid. *Semigroup forum*, 29:183–205, 1984.

[8] M. Latteux and J. Leguy On the composition of morphism and inverse morphisms *Lecture Notes in Computer Science*, 154:420-432, 1983.

[9] J.Meakin and P.Weil. Sugroups of free groups: a contribution to the hanna neumann conjecture. *Geometriae Dedicata*, 94:33–43, 2002.

[10] H. Neumann. On intersections of finitely generated subgroups of free groups. *Publ. Math.Debrecen*, 4:186–189, 1956.

[11] W.D. Neumann. On intersections of finitely generated subgroups of free groups. *Lect. Notes in Math.*, 1456:161–170, 1990.

[12] B. Tilson. The intersection of free submonoids of a free monoid is free. *Semigroup forum*, 4:345–350, 1972.

Dipartimento di Matematica e Applicazioni, Università di Palermo, Via Archirafi 34, 90123 Palermo, Italy

E-mail address: lgiambr@math.unipa.it

Dipartimento di Matematica e Applicazioni, Università di Palermo, Via Archirafi 34, 90123 Palermo, Italy

E-mail address: restivo@math.unipa.it