Sturmian trees – a first investigation

Jean Berstel[1], Luc Boasson[2], Olivier Carton[2] and Isabelle Fagnot[1]

[1] Institut Gaspard–Monge, Université de Marne–la–Vallée and CNRS
[2] Liafa, Université Paris VII and CNRS

Journées Montoises 2006
Irisa, Rennes

Outline

- Definition and examples
- Rank and degree
- Slow automata
- Results
- A proof

Definition

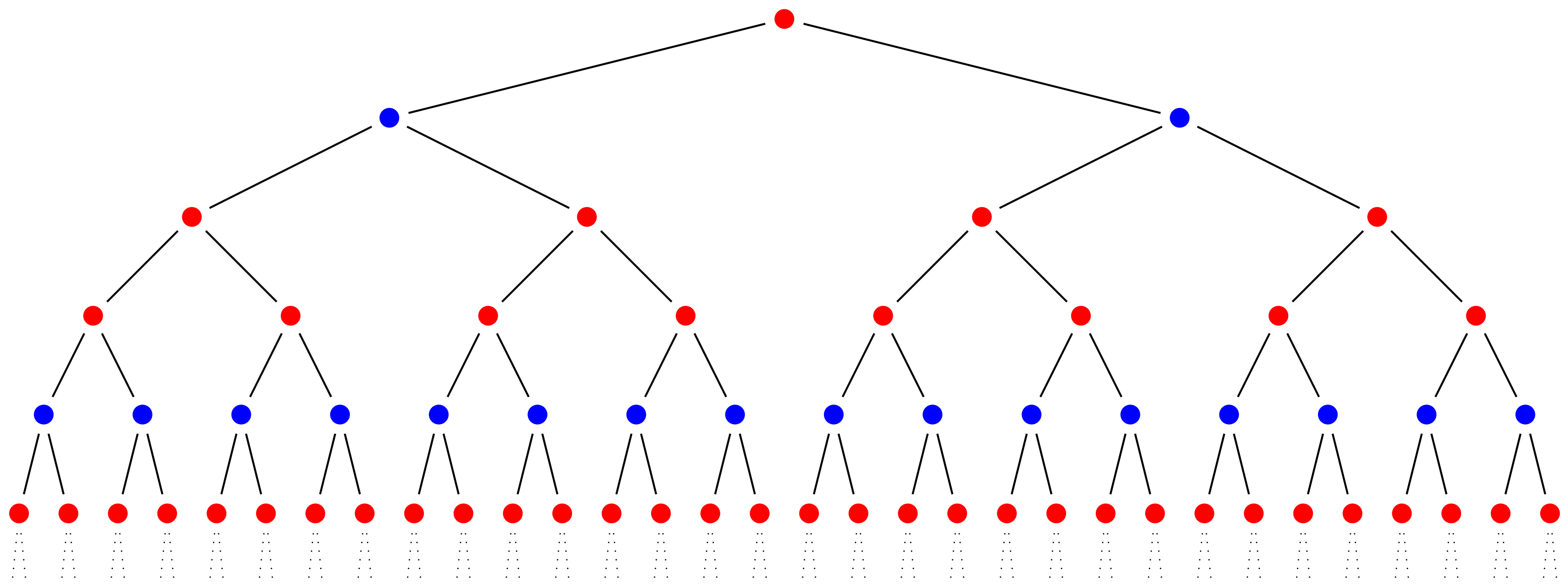
- An infinite word is Sturmian if it has $n+1$ distinct factors of length n for each $n \geq 0$.
- A (binary) labeled tree is Sturmian if it has $n+1$ distinct factor subtrees of height n for each $n \geq 0$.

Remarks

- The height of a subtree is the number of nodes on a path.
- Each node is a subtree of height 1
- The labeling alphabet has two elements.

Example : uniform Sturmian tree

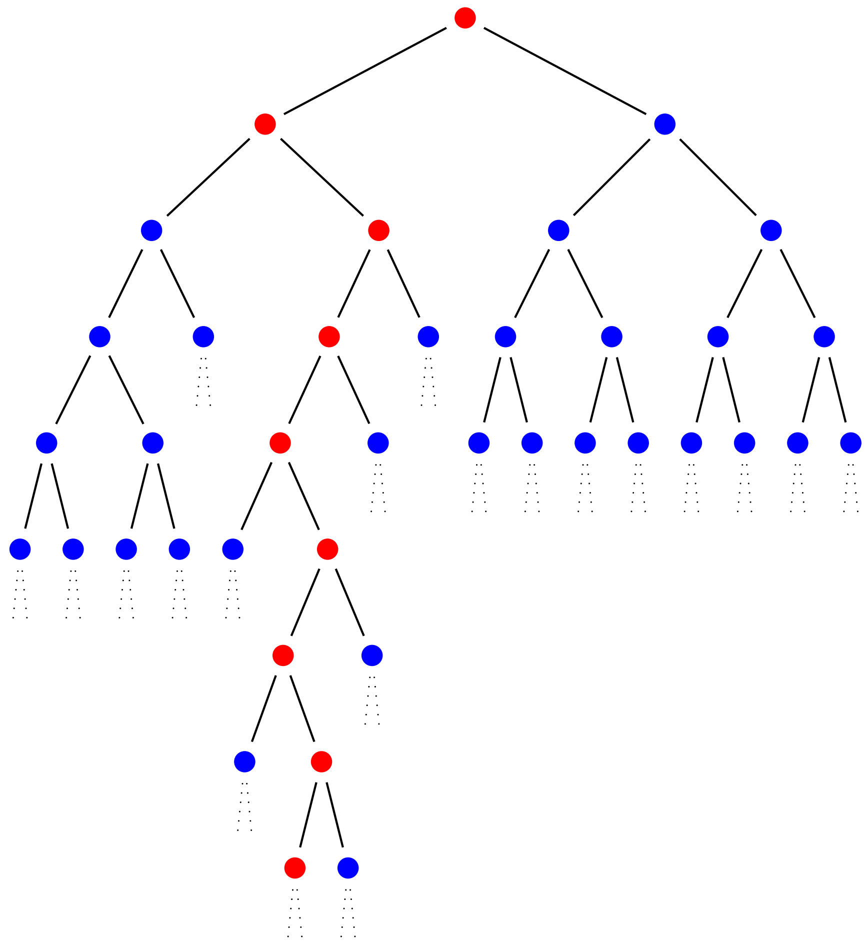
Take any Sturmian word (e.g. $abaaba\cdots$) and repeat it on each branch.



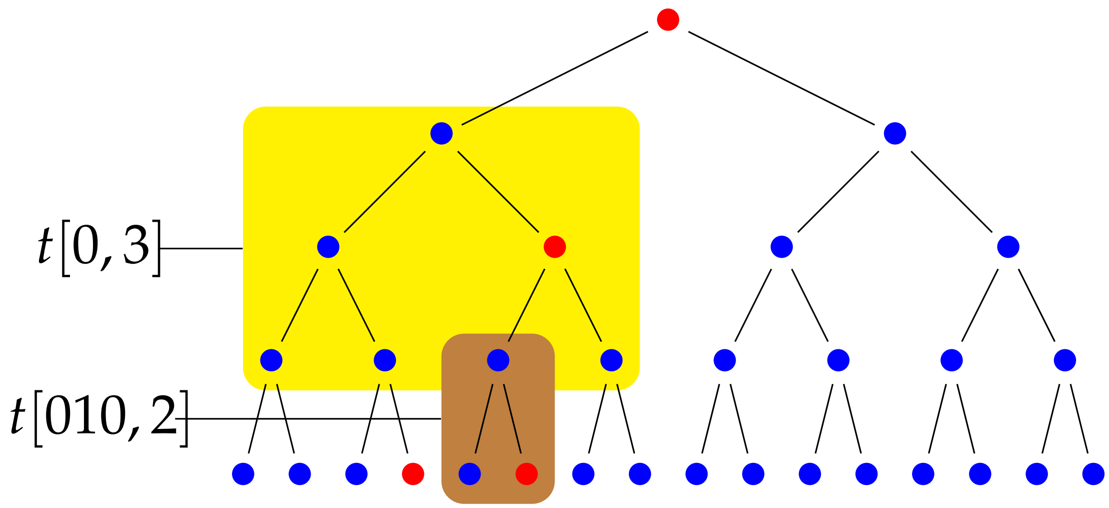
Node label a is represented by ●, and label b is represented by ●.

Example : indicator tree

Take any Sturmian word (e.g. $01001010\cdots$) and distinguish the *branch* labeled by this word.



Example : Dyck tree



- A tree is a mapping $t : \{0,1\}^* \to \{\bullet, \bullet\}$.
- Each node is a word over $\{0,1\}$.
- The language of the tree is the set of words labeled $\bullet$.
- $t[w,h]$ is the subtree rooted in w and of height h.

Rational trees

A (complete) tree t is *rational* if it has finitely many suffixes (infinite subtrees). A node w is *rational* if it is the root of a rational subtree, it is *irrational* otherwise.

Theorem *A (complete) tree t is rational of and only if there is some integer h such that t has at most h distinct factors of height h.*

A. Carpi, A. De Luca, S. Varricchio, *Special factors and uniqueness conditions in rational trees*, Th. Comput. Systems 34:375-395, 2001.

The (minimal) automaton accepting the language of a tree is finite if and only if the tree is rational.

Rank and degree: rank

The *rank* of a tree t is the number of distinct rational nodes (subtrees) of t.

The rank is finite or infinite.

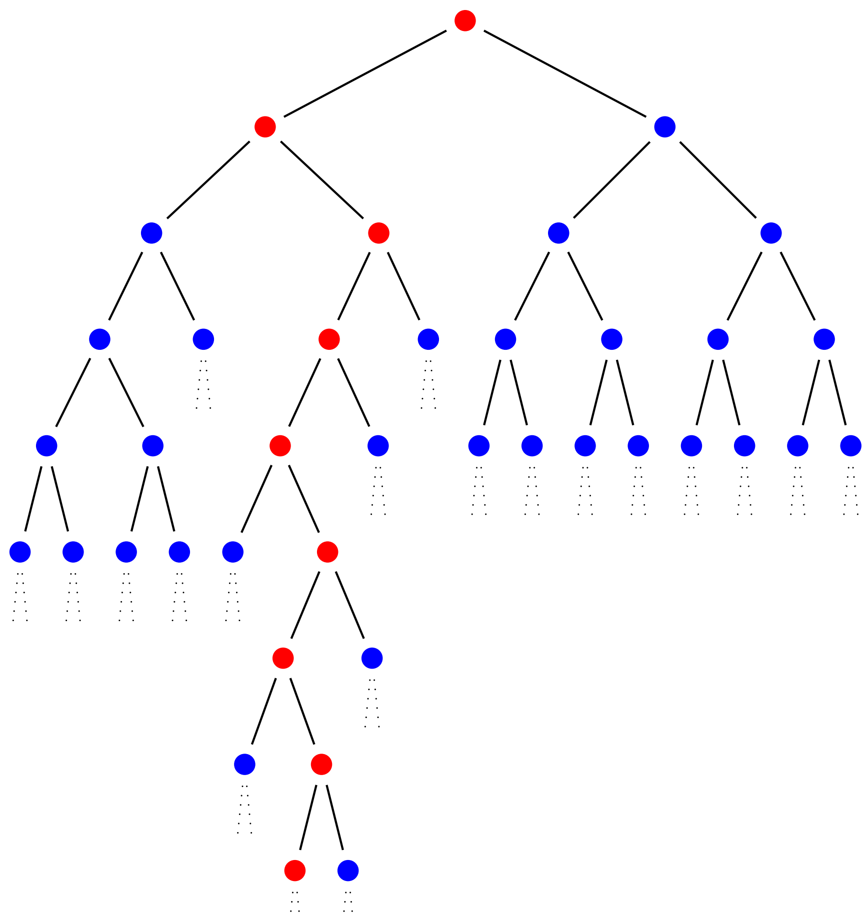
Example

- The rank of a uniform Sturmian tree is 0.
- The rank of an indicator Sturmian tree is 1.
- The rank of the Dyck tree is 1.

Rank of the indicator tree

Take any Sturmian word (e.g. $01001010\cdots$) and distinguish the *branch* labeled by this word.

The only rational tree is the tree rooted in the blue node •.



Rank and degree: degree

Recall that a node is irrational if it is the root of tree which is not rational.

- The *parent* of an irrational node is irrational.
- One at least of the *children* of an irrational node is irrational.
- An infinite *path* is irrational if all its nodes are irrational.

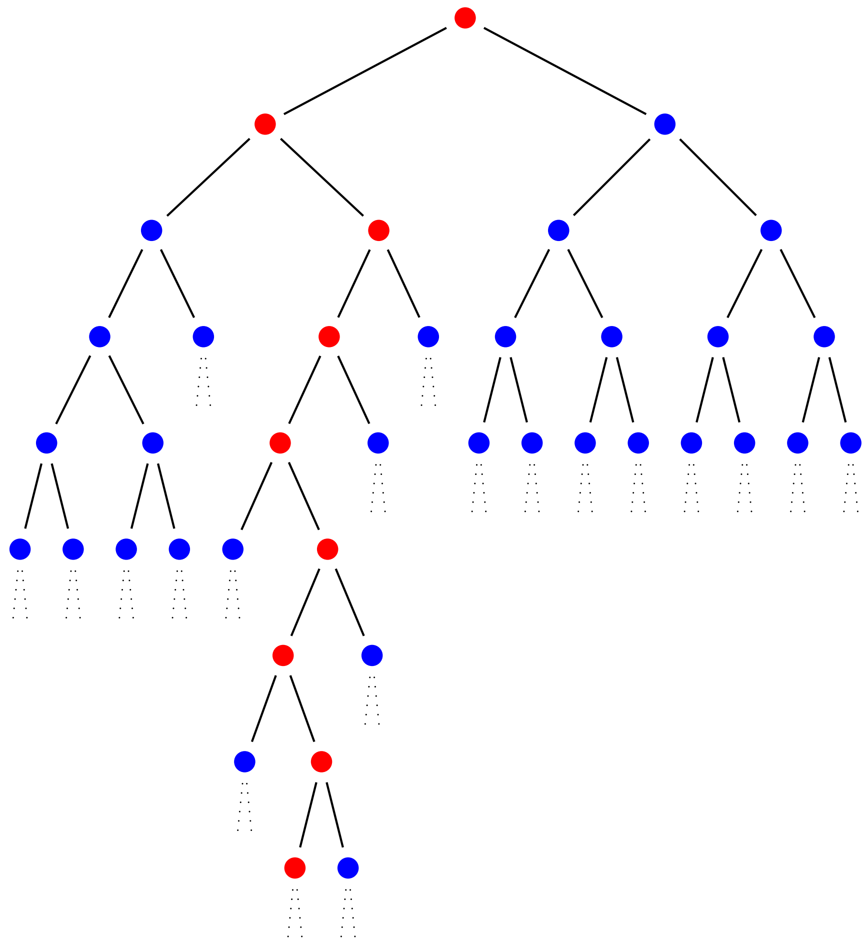
The *degree* of a tree is the number of its distinct irrational paths.

Example

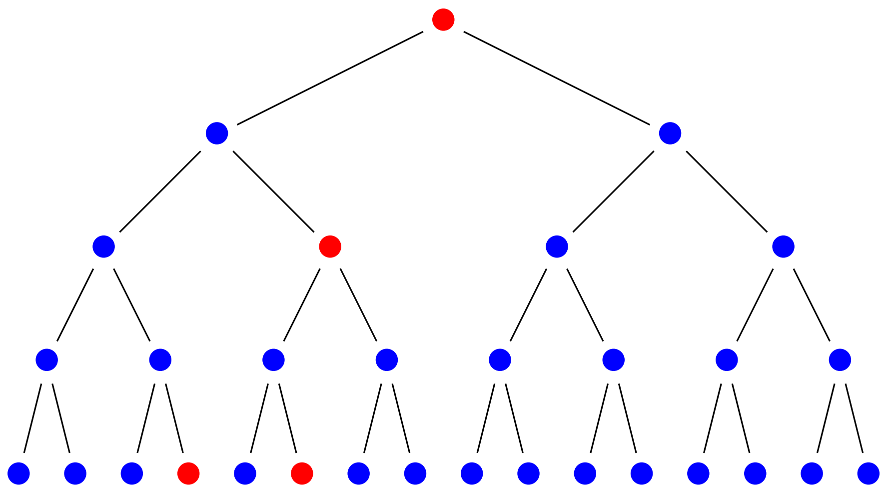
- The degree of a uniform Sturmian tree is ∞.
- The degree of an indicator Sturmian tree is 1.
- The degree of the Dyck tree is ∞.

Degree of the indicator tree

Take any Sturmian word (e.g. $01001010\cdots$) and distinguish the *branch* labelled by this word. This red path is the only irrational path.



Degree of the Dyck tree



Every (prefix of a) Dyck word extends to an infinite irrational path by concatenating some infinite product of distinct Dyck words.

Rank and degree: results

degree	rank	
	finite	infinite
1	are characterized	example in paper
≥ 2, finite	proved to be empty	example in paper
infinite	example of Dyck tree	example not given here

Main result in red.

Slow automata

Given a minimal deterministic complete automaton $\mathcal{A}$ (finite or infinite) over the alphabet $D = \{0, 1\}$ with final states F, the *Moore equivalence* of order h is defined by

$$q \sim_h q' \iff (\forall w \in D^{<h} : q \cdot w \in F \Leftrightarrow q' \cdot w \in F)$$

If the language recognized by $\mathcal{A}$ is not regular, then each equivalence $\sim_h$ is a strict refinement of the preceeding.

An automaton $\mathcal{A}$ is *slow* if each $\sim_h$ has at most $h + 1$ distinct equivalence classes.

Proposition *Let t be a complete tree and let $\mathcal{A}$ be an automaton over D accepting the language of t, with initial state i. For any words $w, w' \in D^*$ and any positive integer h, one has*

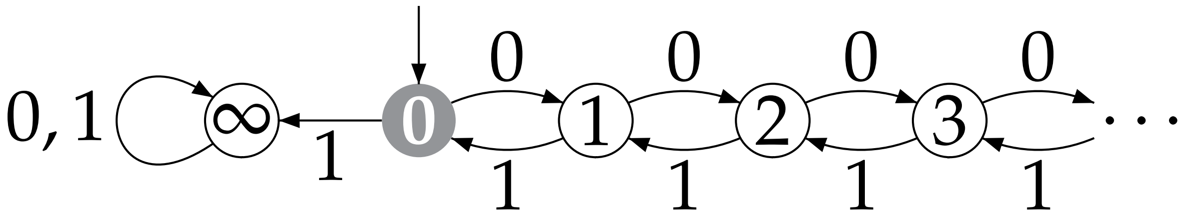
$$i \cdot w \sim_h i \cdot w' \iff t[w, h] = t[w', h] .$$

Corollary *Let t be a complete tree and let $\mathcal{A}$ be an automaton over D accepting the language of t. The tree t is Sturmian iff each equivalence relation $\sim_h$ has $h + 1$ classes.*

Corollary *A complete tree t is Sturmian iff the minimal automaton of its language is infinite and slow.*

A first example of slow automata

Automaton of the Dyck language. State 0 is both the initial and the unique terminal state.



Moore equivalences:

$0 \mid 123 \cdots \infty$

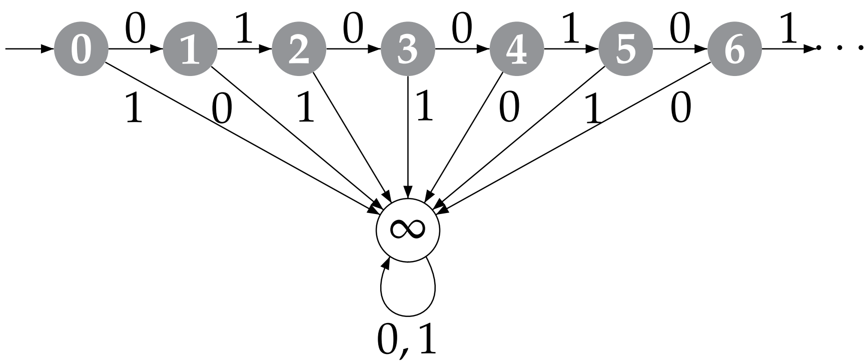
$0 \mid 1 \mid 23 \cdots \infty$

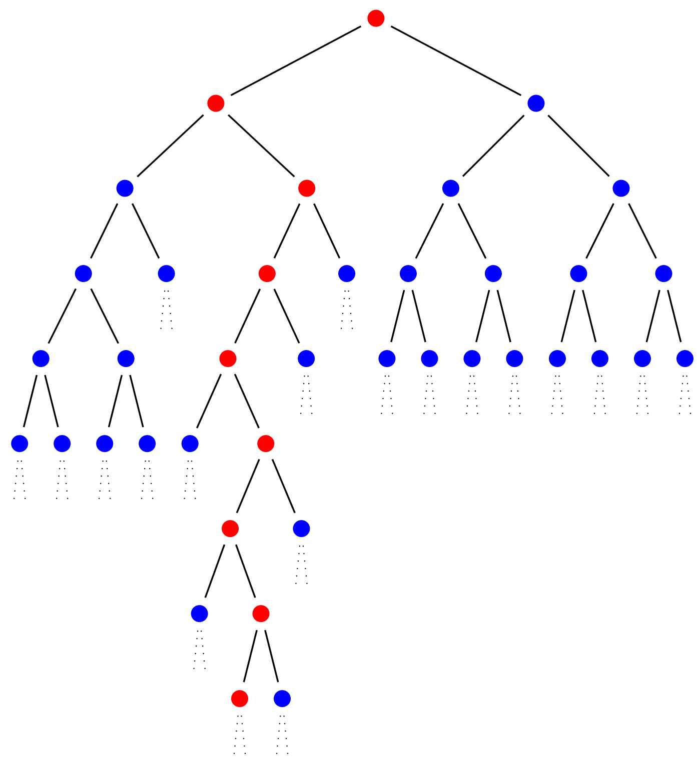
$0 \mid 1 \mid 2 \mid 3 \cdots \infty$

etc.

Another example of slow automata

Automaton accepting the prefixes of $01001010\cdots$. All states are final excepted ∞.





Equivalence classes:

$\infty \mid 012\cdots$

$\infty \mid 0235\cdots \mid 146\cdots$

$\infty \mid 035\cdots \mid 2\cdots \mid 146\cdots$

Lazy paths

A *lazy path* in a N state finite minimal automaton is a path

$$\pi : q_0 \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \cdots q_{h-1} \xrightarrow{a_{h-1}} q_h$$

of length h, where

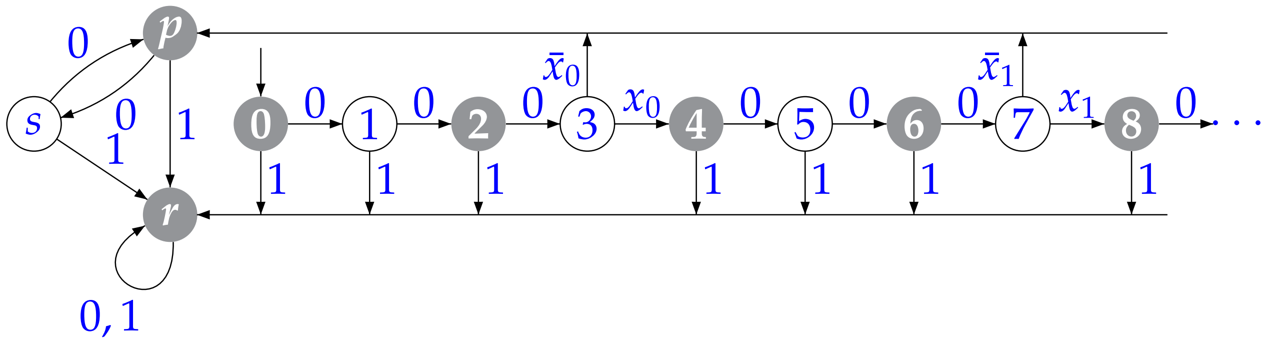
- q_0 and q_h are the two states which are separated in the *last step* in Moore's algorithm
- $q_{h-1} \cdot \bar{a}_{h-1} = q_0$ or q_h.

The first of these conditions means that $q_0 \sim_{N-2} q_h$ and $q_0 \not\sim_{N-1} q_h$.

The second property means that state $q_{h-1} \cdot \bar{a}_{h-1}$ cannot be separated from state $q_{h-1} \cdot a_{h-1}$ before the very last step of the Moore algorithm.

A typical example

A slow automaton $\hat{\mathcal{A}}$ for the Fibonacci word $x_0x_1\cdots = 01001010\cdots$. The final states are $p, r, 0, 2, 4, \cdots$.

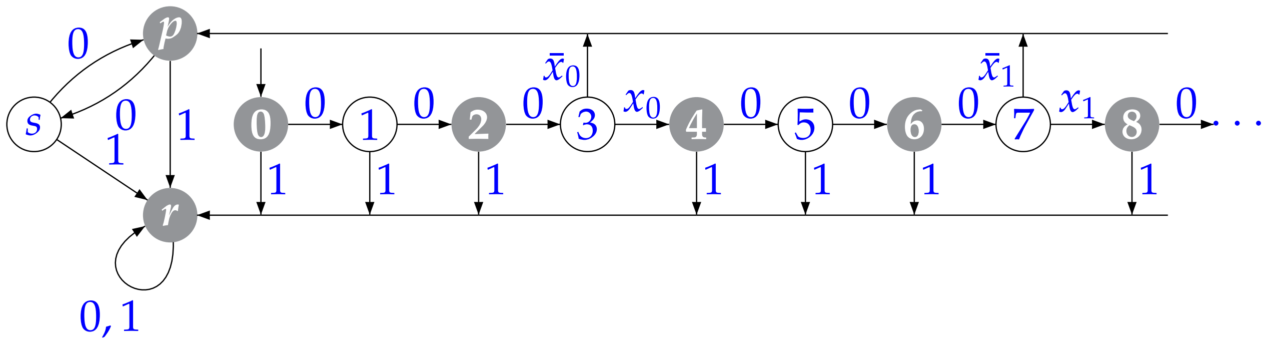


Recall that for a lazy path $\pi : q_0 \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \cdots q_{h-1} \xrightarrow{a_{h-1}} q_h$: q_0 and q_h are separated in the *last step* and $q_{h-1} \cdot \bar{a}_{h-1} = q_0$ or q_h.

In this example:

- The subautomaton $\mathcal{A}$ with states $\{p, q, r\}$ is slow, and p, r are separated in the last step.
- The path $\pi : p \xrightarrow{0} s \xrightarrow{0} p \xrightarrow{0} s \xrightarrow{1} r$ is lazy: indeed $s \xrightarrow{0} p$. Here $h = 4$.

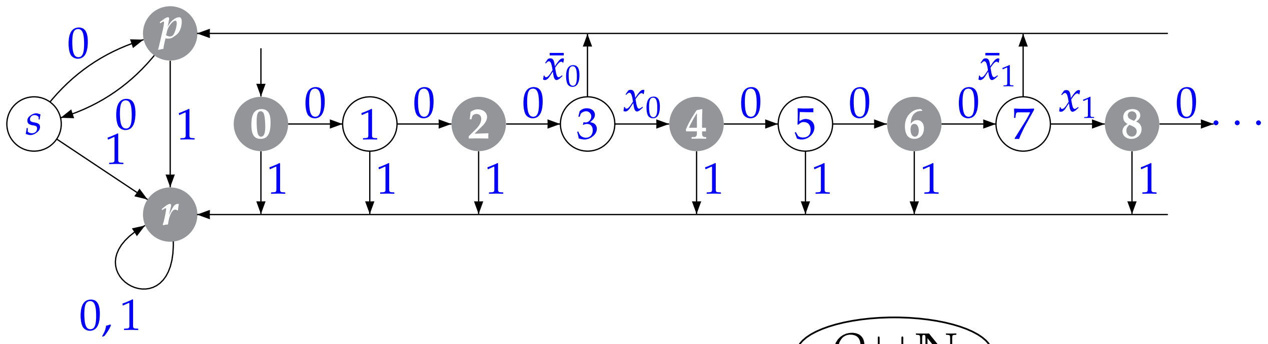
The example continued



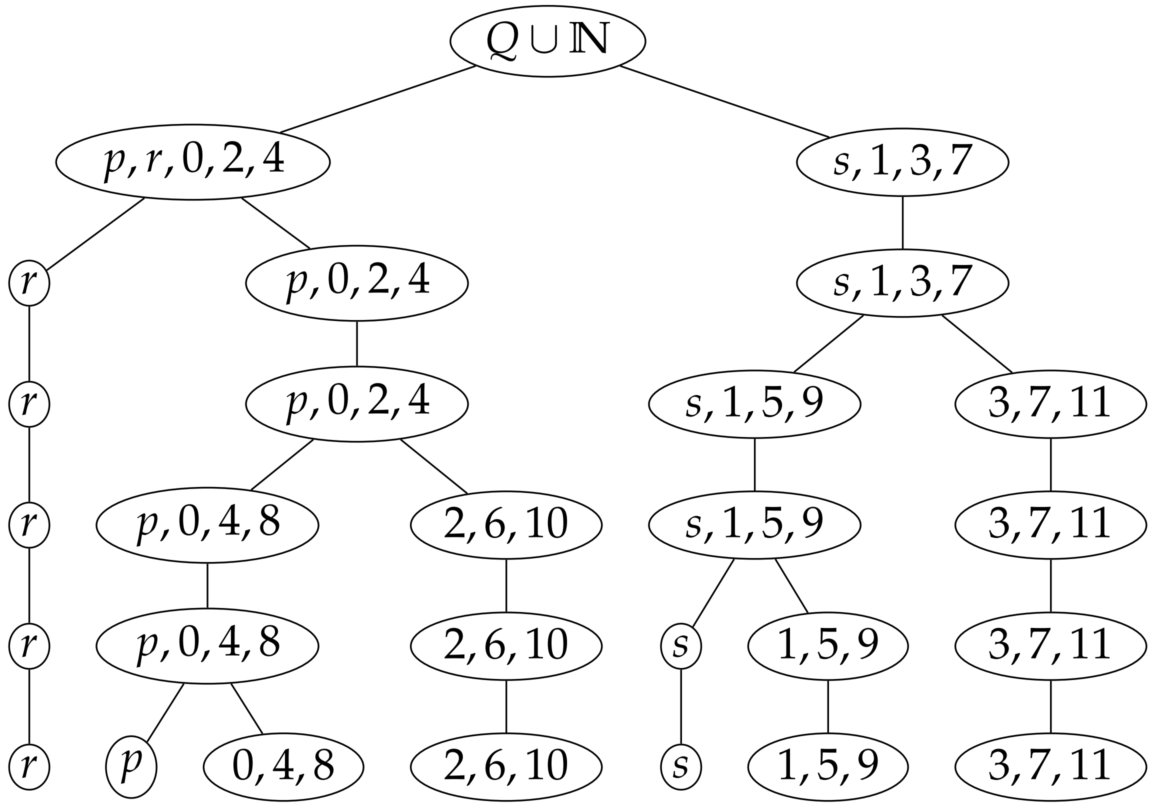
- The patterns $(0, 0, 0, 1)$ and $(0, 0, 0, 0)$ of the lazy path $\pi : p \xrightarrow{0} s \xrightarrow{0} p \xrightarrow{0} s \xrightarrow{1} r$ is repeated to build the infinite path.
- The choice of the final 1 or 0 is driven by the Fibonacci word $x_0 x_1 \cdots = 01 \cdots$.

Definition The automaton $\hat{A} = \hat{A}(\pi, x)$ is the *extension* of the slow automaton $\mathcal{A}$ by the lazy path π and the infinite word x.

The example continued (2)



Moore equivalences:



A characterization

Proposition *Let $\hat{\mathcal{A}} = \mathcal{A}(\pi, x)$ be the extension of the finite slow automaton $\mathcal{A}$ by a lazy path π and an infinite word x. If the word x is Sturmian, then $\hat{\mathcal{A}}$ defines a tree t which is Sturmian, of degree 1, and of finite rank.*

The converse is the main result:

Theorem *Let t be a Sturmian tree of degree one having finite rank, and let $\hat{\mathcal{A}}$ be the minimal automaton of the language of t. Then $\hat{\mathcal{A}}$ is the extension of a slow finite automaton $\mathcal{A}$ by a lazy path π and a Sturmian word x, i.e. $\hat{\mathcal{A}} = \mathcal{A}(\pi, x)$.*

A constraint

Proposition *The degree of a Sturmian tree with finite rank is either one or infinite.*

There exist Sturmian trees of finite degree greater than one and they must have infinite rank.

A class of a Moore equivalence $\sim_n$ is *irrational* if it is composed only of irrational states, and it is *rational* otherwise.

Proposition *The degree of a Sturmian tree t with finite rank is 1 or infinite.*

Assume t has finite degree $d > 1$.

A node w of t is a *fork* if both $w0$ and $w1$ are irrational nodes. t has exactly $d-1$ fork nodes.

A state of the minimal automaton of t is a *fork state* if it is the state of a fork node. The automaton has at most $d-1$ fork states.

Claim: For large enough n, a class of $\sim_n$ containing a fork state is a singleton.

Let H be such that each fork state is a singleton class of $\sim_H$.

The Nerode equivalence and $\sim_H$ coincide for these states: two fork nodes in the tree t define the same state in the automaton if and only if they are the roots of the same subtree of height H.

There are infinitely many occurrences of any subtree of height H in a Sturmian tree.

There are infinitely many nodes in t that correspond to the same fork state, so there are infinitely many fork nodes in t, contradiction.

Proof of the claim

Claim: For large enough n, a class of $\sim_n$ containing a fork state is a singleton.

Lemma *Let t be a Sturmian tree with finite rank. Either there is an integer n such that all rational classes of $\sim_n$ are singletons, or there is an integer n such that all irrational obtained by splitting a class of $\sim'_n$ for $n' \geq n$ are singleton classes.*

Either there is an integer n such that all rational states are singletons for $\sim_n$. Then a class of $\sim_n$ containing a fork state contains only fork states since indeed a state that is not a fork state maps to a rational state by at least one letter, whereas a fork state does not.
So any class containing a fork state is finite, and will be split eventually into singleton classes.

In the other case, irrational states will be in singleton classes for large enough n. Again, since there are only finitely many fork states, each of these will be constructed at some step in the Moore algorithm.

Final remarks

- Slow finite automata may have intrinsic properties.
- Sturmian trees may have rational nodes, and this makes their investigation difficult.
- Rauzy graphs exist for Sturmian trees.
- The Moore equivalences for a Sturmian word are in bijection with the Rauzy graphs.