A First Investigation of Sturmian Trees

Jean Berstel[2], Luc Boasson[1]

Olivier Carton[1], Isabelle Fagnot[2]

[1]LIAFA, CNRS
Université Paris 7

[2]IGM, CNRS
Université de Marne-la-Vallée

Atelier de Combinatoire, Montréal
12 mars 2007

Outline

Sturmian words

Proposition (Hedlund & Morse)

An infinite word x *ultimately periodic iff there is an integer* n *such that* x *has at most* n *distinct factors of length* n*.*

Definition

An infinite word is Sturmian if it has $n + 1$ distinct factors of length n for each $n \geq 0$.

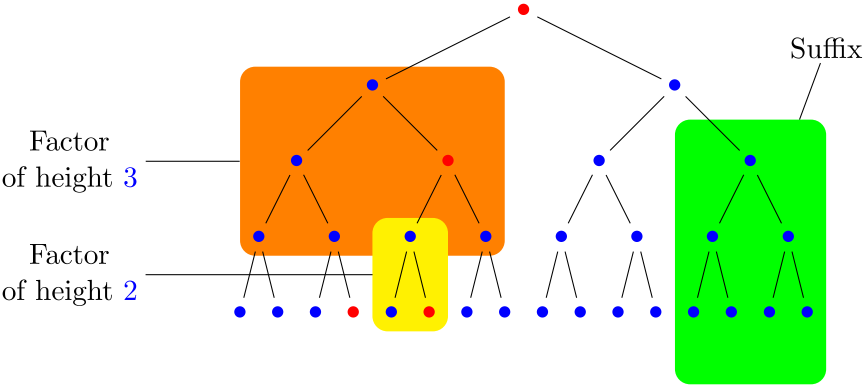
Remark

Sturmian words are non ultimately periodic words with the smallest complexity.

Factors in a tree

Definition

A factor of height h of a tree t is a subtree of height h that occurs in t.
A suffix of tree t is an infinite subtree of t.



Rational tree

Definition

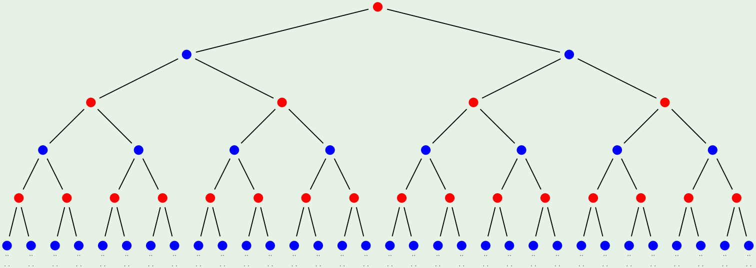
A tree is rational if it has a finite number of distinct suffixes.

Proposition (Carpi De Luca, Varricchio)

A complete tree t is rational if there is some integer h such that t has at most h distinct factors of height h.

Example (Two distinct suffixes)

Red nodes at even levels, blue nodes at odd levels.



Sturmian tree

Definition

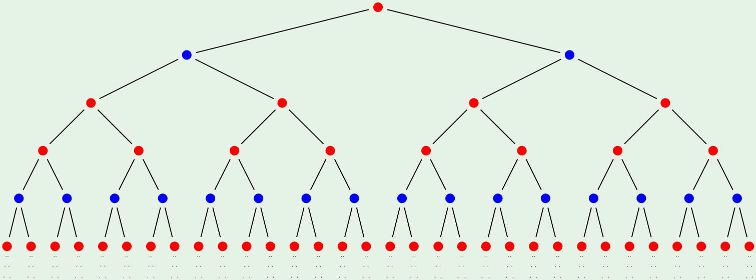
A tree is Sturmian if it has $h+1$ distinct factors of height h for each h.

Remark

Sturmian trees are irrational trees with the smallest complexity.

Example (Easy one: uniform tree)

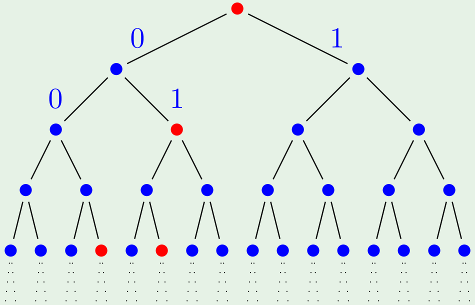
An Sturmian word $x = abaaba\ldots$ is repeated on each branch.



Example (Unexpected one: Dyck tree)

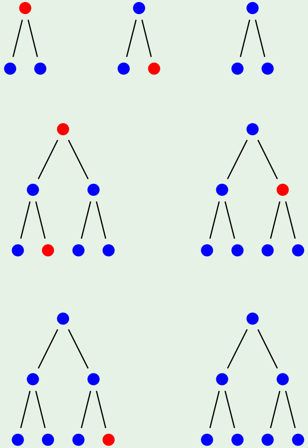
A node is • if it is a Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



$D_2^* = \{\varepsilon, 01, 0101, 0011, \dots\}$

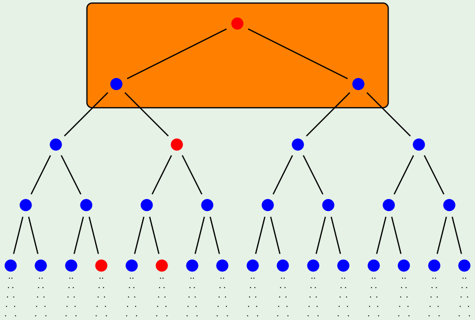
Its factors



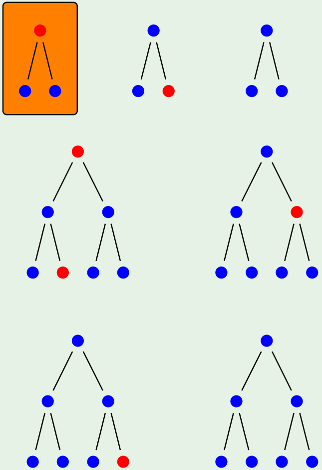
Example (Unexpected one: Dyck tree)

A node is • if it is a Dyck word over the alphabet $\{0,1\}$.

The Dyck tree



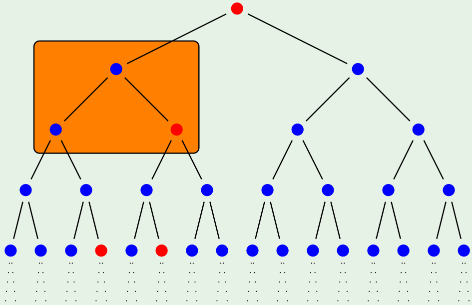
Its factors



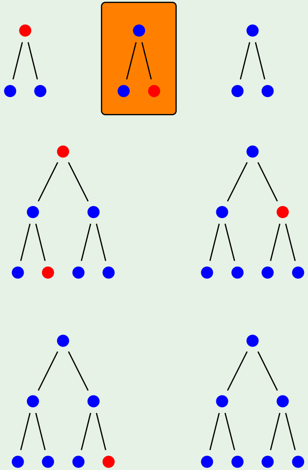
Example (Unexpected one: Dyck tree)

A node is • if it is a Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



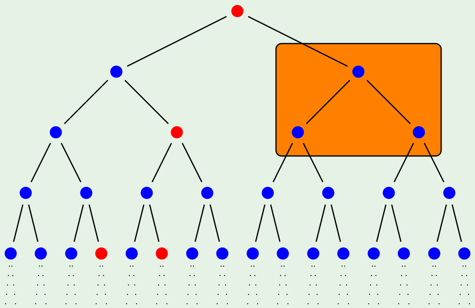
Its factors



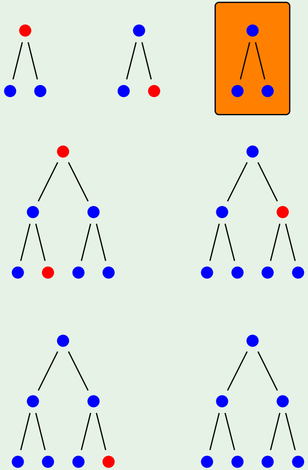
Example (Unexpected one: Dyck tree)

A node is • if it is a Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



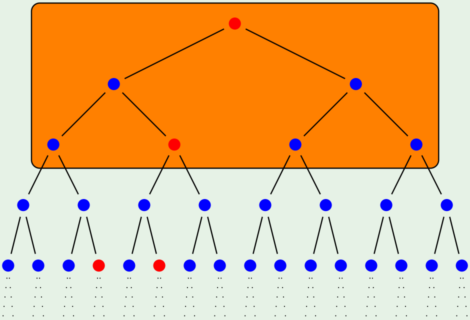
Its factors



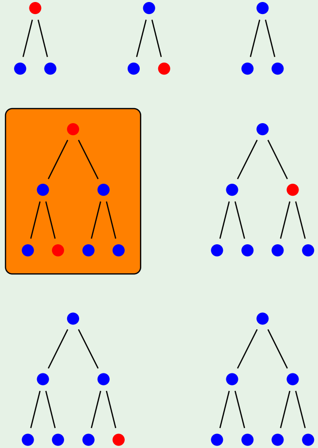
Example (Unexpected one: Dyck tree)

A node is • if it is a Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



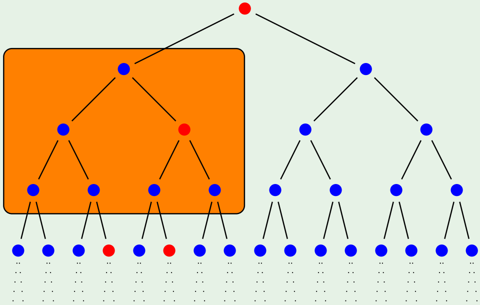
Its factors



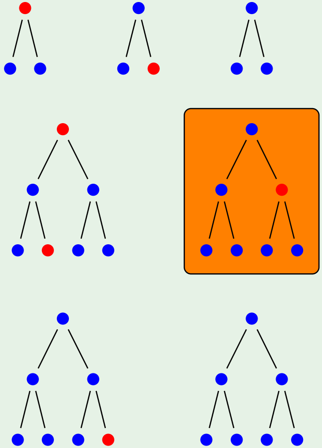
Example (Unexpected one: Dyck tree)

A node is ● if it is a Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



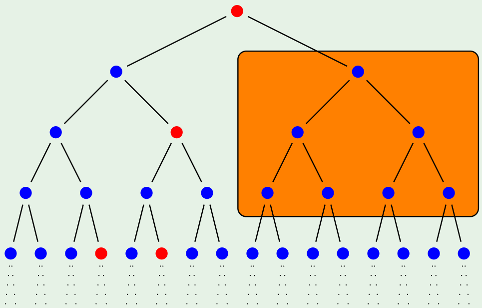
Its factors



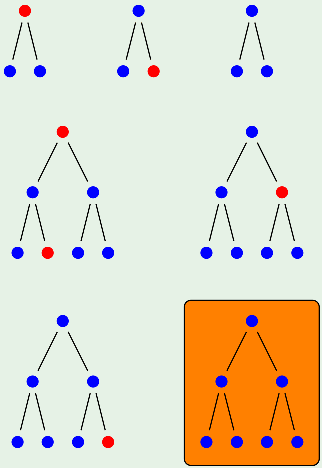
Example (Unexpected one: Dyck tree)

A node is • if it is a Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



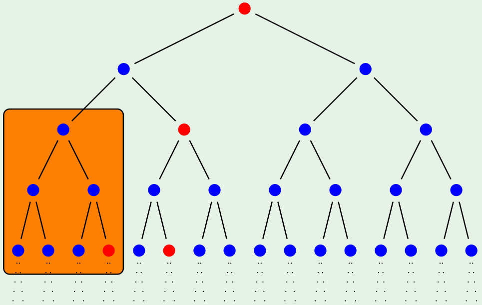
Its factors



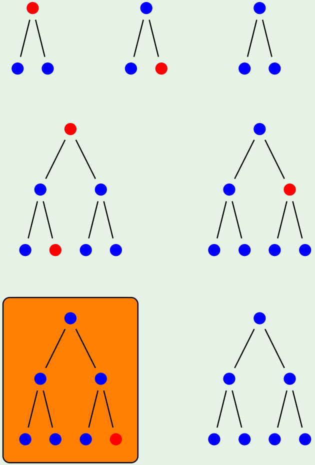
Example (Unexpected one: Dyck tree)

A node is ● if it is a Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



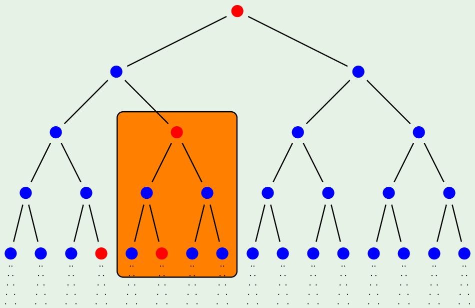
Its factors



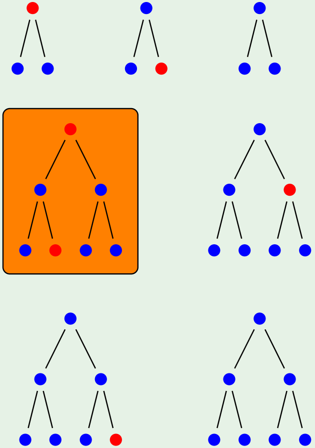
Example (Unexpected one: Dyck tree)

A node is ● if it is a Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



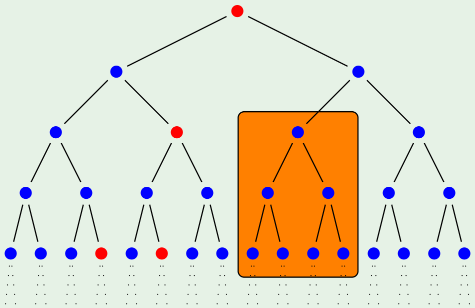
Its factors



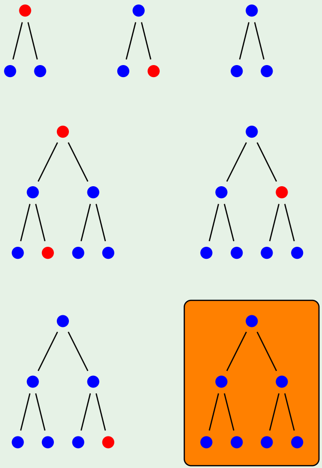
Example (Unexpected one: Dyck tree)

A node is $\bullet$ if it is a Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



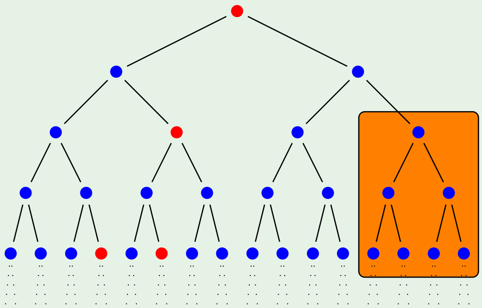
Its factors



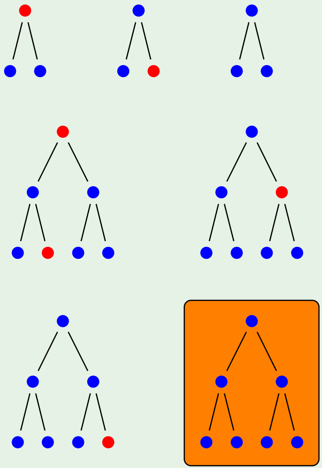
Example (Unexpected one: Dyck tree)

A node is ● if it is a Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



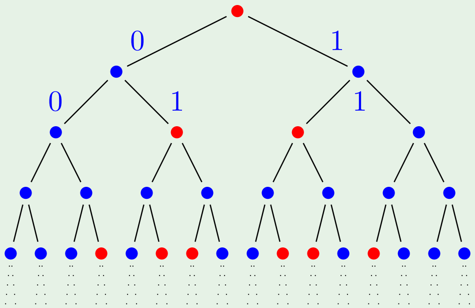
Its factors



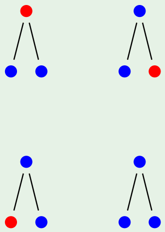
Example (Two-sided Dyck tree)

A node is • if it is a tow-sided Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



Its four factors of height 2

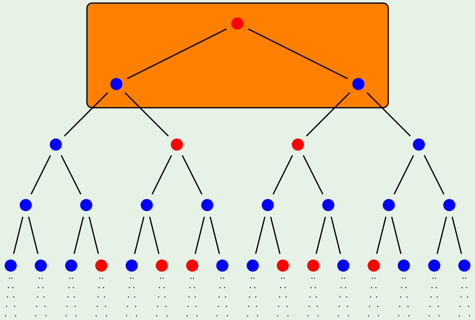


$$D_2^* = \{\varepsilon, 01, 10, 0011, 0101, 0110, 1001, 1010, 1100, \dots\}$$

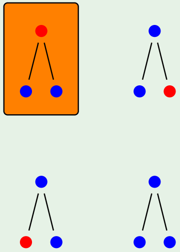
Example (Two-sided Dyck tree)

A node is ● if it is a tow-sided Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



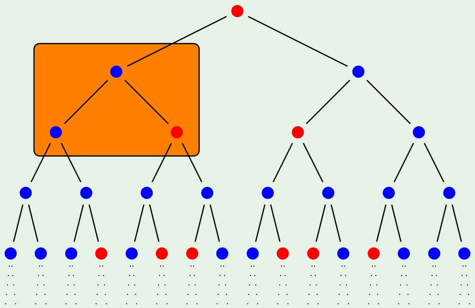
Its four factors of height 2



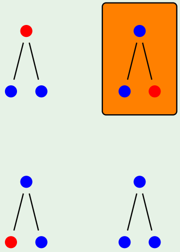
Example (Two-sided Dyck tree)

A node is • if it is a tow-sided Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



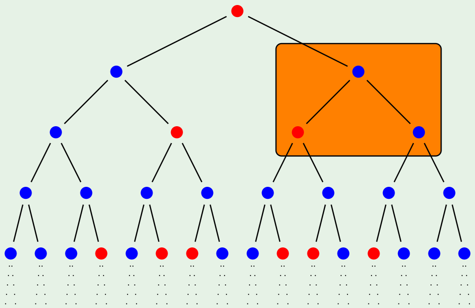
Its four factors of height 2



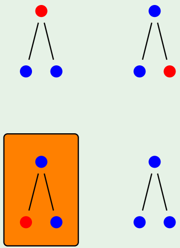
Example (Two-sided Dyck tree)

A node is ● if it is a tow-sided Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



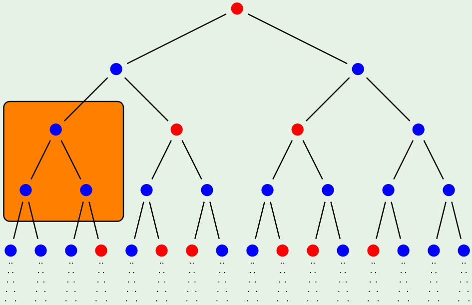
Its four factors of height 2



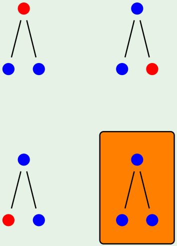
Example (Two-sided Dyck tree)

A node is ● if it is a tow-sided Dyck word over the alphabet $\{0, 1\}$.

The Dyck tree



Its four factors of height 2



Moore equivalence

Let $\mathcal{A}$ a (infinite) minimal deterministic automaton over $D = \{0, 1\}$ with states Q and final states F automaton.

Definition

The Moore equivalence $\sim_h$ of order h is

$$q \sim_1 q' \iff (q \in F \Leftrightarrow q' \in F)$$
$$q \sim_{h+1} q' \iff (q \sim_h q') \text{ and } (\forall a \in D \;\; q \cdot a \sim_h q' \cdot a)$$

Remarks

- $q \not\sim_h q'$ if there is a word w of length h such that $q \cdot w \in F$ and $q' \cdot w \notin F$ (or $q \cdot w \notin F$ and $q' \cdot w \in F$).
- Any two distinct states q, q' of a minimal automaton will be eventually *separated*: $q \not\sim_h q'$ for some h, but need not to be *isolated*, that is singleton equivalence classes.

Slow automata

Definition

An infinite automaton is slow iff the Moore equivalence $\sim_h$ of order h has $h+1$ classes for each h.

Remark

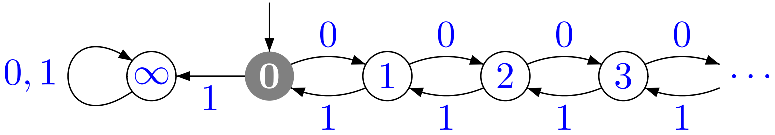
In a slow automaton, exactly one equivalence class of $\sim_h$ is split into two classes of $\sim_{h+1}$.

Proposition

A tree t is Sturmian iff the minimal automaton of its language is slow.

A first slow automaton

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

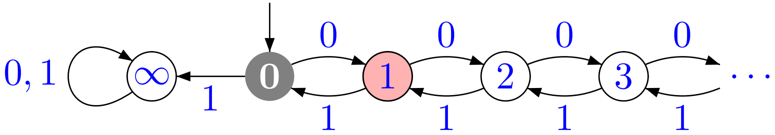


Moore equivalences:

$0 \mid 12345 \cdots \infty$

A first slow automaton

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



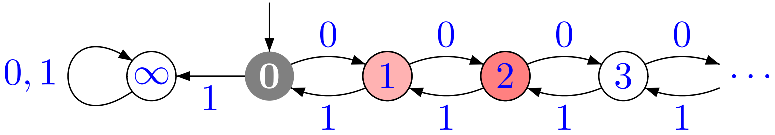
Moore equivalences:

$0 \mid 12345 \cdots \infty$

$0 \mid 1 \mid 2345 \cdots \infty$

A first slow automaton

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



Moore equivalences:

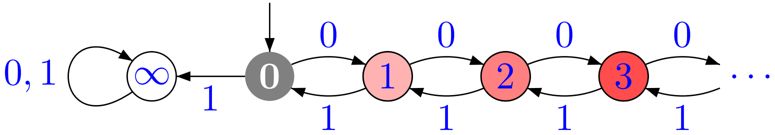
$0 \mid 12345 \cdots \infty$

$0 \mid 1 \mid 2345 \cdots \infty$

$0 \mid 1 \mid 2 \mid 345 \cdots \infty$

A first slow automaton

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



Moore equivalences:

$0 \mid 12345 \cdots \infty$

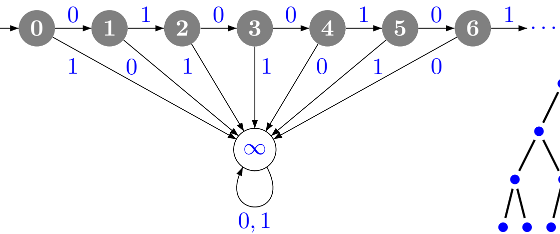
$0 \mid 1 \mid 2345 \cdots \infty$

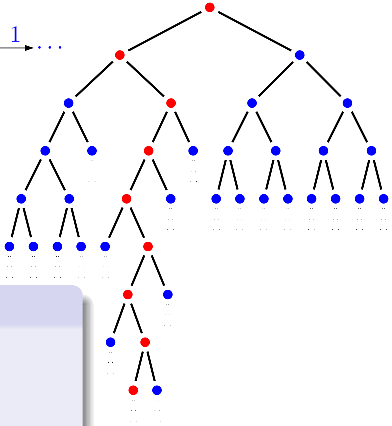
$0 \mid 1 \mid 2 \mid 345 \cdots \infty$

$0 \mid 1 \mid 2 \mid 3 \mid 45 \cdots \infty$

Another slow automaton

Automaton accepting the prefixes of $01001010\cdots$.



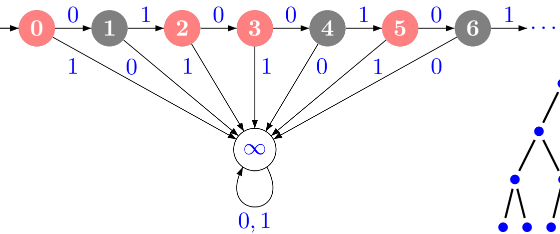


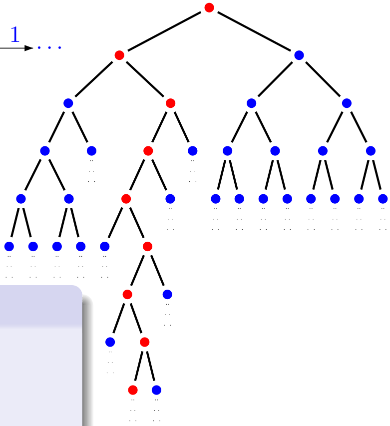
Equivalence classes:

$\infty \mid 012\cdots$

Another slow automaton

Automaton accepting the prefixes of $01001010\cdots$.





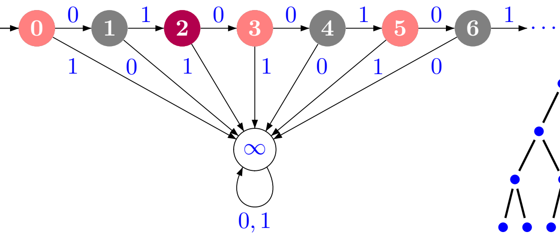
Equivalence classes:

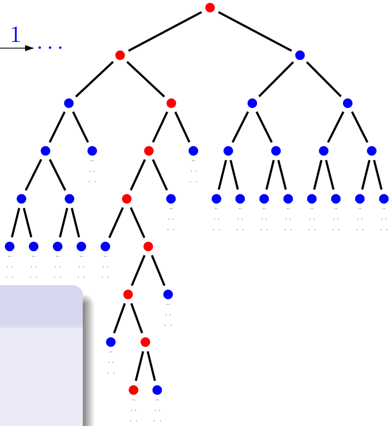
$\infty \mid 012\cdots$

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

Another slow automaton

Automaton accepting the prefixes of 01001010 · · · .





Equivalence classes:

∞ | 012 · · ·

∞ | 0235 · · · | 146 · · ·

∞ | 035 · · · | 2 · · · | 146 · · ·

Rank and degree

Definition

- A node is called irrational if the infinite subtree rooted in this node is not rational.
- The rank is the number of distinct rational subtrees.
- The degree is the number of branches of irrational nodes.

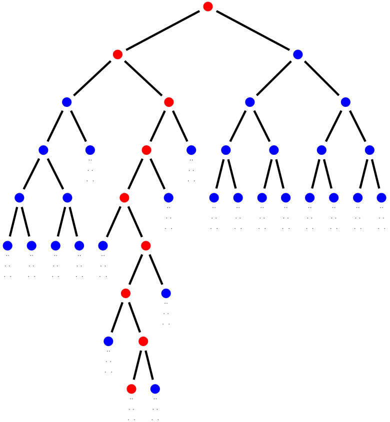
Examples

- The uniform tree has rank 0 and degree ∞.
- The Dyck tree has rank 1 and degree ∞.
- The indicator tree of a Sturmian word has rank 1 and degree 1.

Rank and degree of the indicator tree

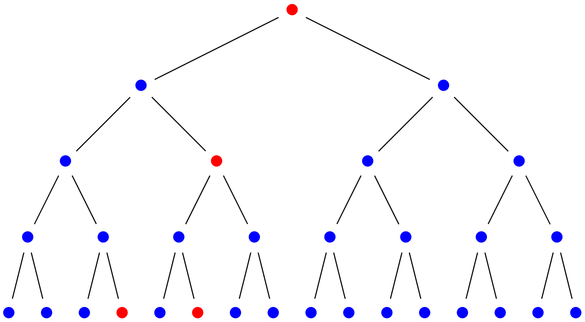
Take any Sturmian word (e.g. $01001010\cdots$) and distinguish the branch labeled by this word.

The only rational tree is the tree rooted in the blue node.
The only irrational path is composed of the red nodes.



Rank and degree of the Dyck tree

The only rational subtree is composed of blue nodes only.



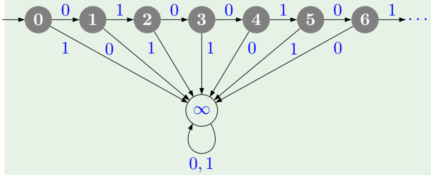
The degree is infinite because every (prefix of a) Dyck word extends to an infinite irrational path by concatenating some infinite product of distinct Dyck words.

Results

degree	rank	
	finite	infinite
1	characterized	example later
≥ 2, finite	proved to be empty	example in full paper
infinite	example of Dyck tree	example in full paper

Characterization: a generalization of the indicator tree

Example (Indicator tree)

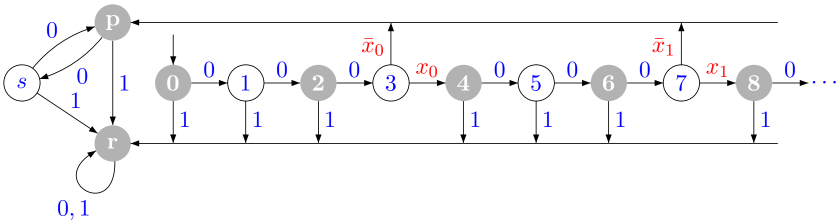


General situation

- More than one rational subtrees
- The infinite path is interleaved with a fixed finite path.

A typical example

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



- 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: and p, r are separated in the last step (+ some technical condition).

A characterization

Definition

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

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 1 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.

The big example

Example (Rank ∞ and degree 1)