One World Combinatorics on Words Seminar
June 25th, 2024

The Heinis spectrum

Julien Cassaigne

Institut de mathématiques de Marseille - CNRS, Marseille, France
julien.cassaigne@math.cnrs.fr

The Heinis spectrum

- Motivation
- Definition
- Examples
- Some properties
- Questions
- Tools

Infinite words, factors, complexity

$u \in A^{\mathbb{N}}$: an infinite word
(one may also consider a bi-infinite word $u \in A^{\mathbb{Z}}$, or a factorial language $L \subseteq A^*$, or a symbolic dynamical system $X \subseteq A^{\mathbb{Z}}$).

$w \in A^*$ is a factor of u if $w = u_k u_{k+1} u_{k+|w|-1}$ for some k.

$L(u)$: the set of factors of u, $L_n(u) = L(u) \cap A^n$.

$p_u(n) = \# L_n(u)$: the complexity function of u.

Questions on complexity

1. Given an infinite word u, compute its complexity function.

2. If a word has a given property (dynamical, combinatorial, etc.), what are the consequences on its complexity? and vice versa?

3. Given a function (or a class of functions), does there exist a word with such complexity? In this case, construct it explicitely.

4. Describe all words whose complexity function is in a given class.

Here we are interested in a question of type 3: among linear growths, which ones are possible for a complexity function?

Linear complexity

Many families of infinite words have a complexity function with linear growth $p(n) = O(n)$:

- automatic words
- primitive substitutive words
- Sturmian words: $p(n) = n + 1$
- Arnoux-Rauzy words: $p(n) = 2n + 1$
- codings of k-interval exchange transformations: $p(n) = (k-1)n+1$
- dendric words: $p(n) = (k - 1)n + 1$
- Rote words: $p(n) = 2n$ for $n \geq 1$
- paperfolding words: $p(n) = 4n$ for $n \geq 7$
- ...

Thue-Morse word

$u = abbabaabbaababbabaab\dots$ fixed point of $a \mapsto ab, b \mapsto ba$

Complexity (Brlek 1989)
$p(n+1) = 4n - 2.2^k$ if $2.2^k \leq n \leq 3.2^k$
$p(n+1) = 2n + 4.2^k$ if $3.2^k \leq n \leq 4.2^k$

$3n \leq p(n+1) \leq 10n/3$ for $n \geq 2$, sharp.

Typical for substitutive words: $p(n+1) - p(n)$ takes finitely many values, changing when n belongs to a certain sequence with exponential growth (the lengths of bispecial factors).

Heinis spectrum

Let $\alpha = \liminf_{n\to\infty} \frac{p(n)}{n}$ and $\beta = \limsup_{n\to\infty} \frac{p(n)}{n}$.

Theorem (Heinis 2001).
If $1 < \alpha < 2$, then $\beta - \alpha \geq \frac{(2-\alpha)(\alpha-1)}{\alpha}$.

In particular $1 < \alpha = \beta < 2$ is impossible. More generally, $\alpha = \beta \in \mathbb{R} \setminus \mathbb{N}$ is impossible (Cassaigne and Nicolas 2010).

The Heinis spectrum is the set of possible pairs (α, β):

$$\Omega = \{(\alpha, \beta) : u \in A^{\mathbb{N}}\} \subset [0, +\infty]^2$$

Question: what is the structure of Ω?

Examples

Ω contains:

$(0, 0)$: periodic words

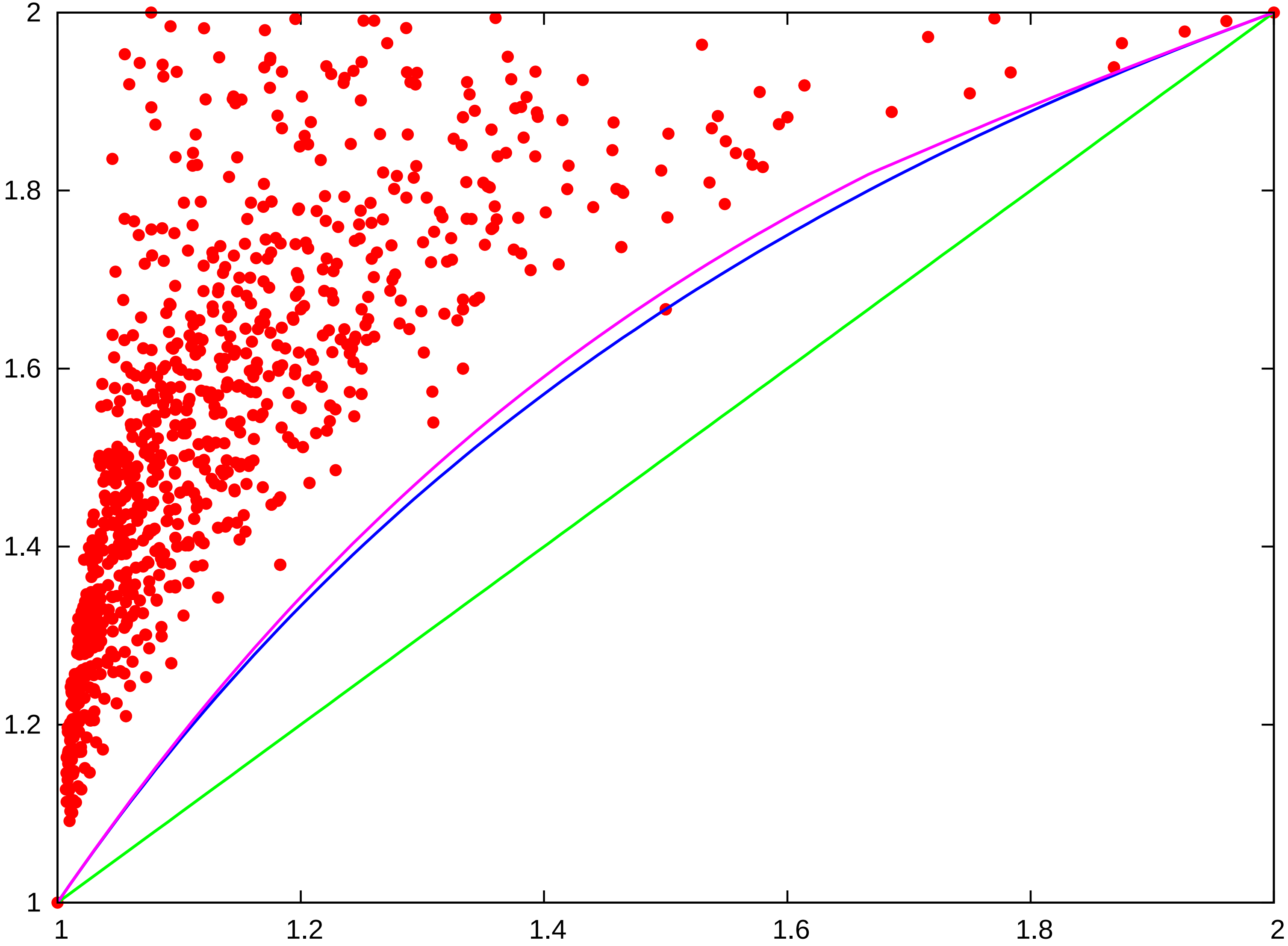
$(1, 1)$: Sturmian words

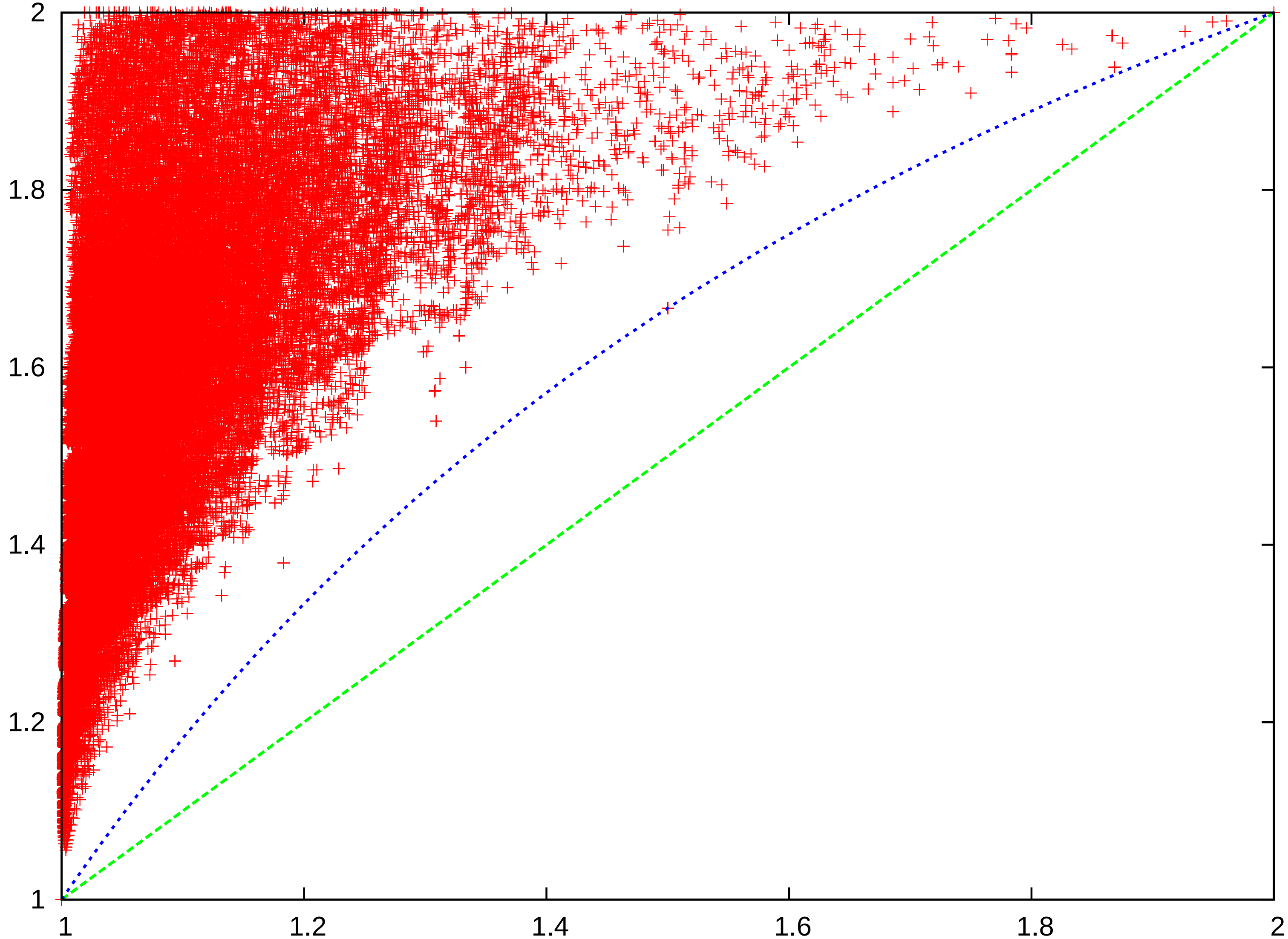
(k, k): codings of interval exchange transformations

$(3, 10/3)$: Thue-Morse

(∞, ∞): Champernowne

And also $(1, \infty)$, $(3/2, 5/3)$, $(\frac{1+\sqrt{2}}{2}, \frac{3+\sqrt{2}}{3})$, ...


2
1.8
1.6
1.4
1.2
1
1
1.2
1.4
1.6
1.8
2



2
1.8
1.6
1.4
1.2
1
1
1.2
1.4
1.6
1.8
2

Isolated points in Ω

By the theorem of Morse and Hedlund, $(0,0)$ is the only point for which $\alpha < 1$, hence it is isolated.

Theorem (Turki 2016). If

$$\beta < \min\left(\frac{5\alpha^2 - 3\alpha}{2\alpha^2 - \alpha + 1}, \frac{4\alpha}{2+\alpha}\right)$$

then $(\alpha, \beta) = (\frac{3}{2}, \frac{5}{3})$.

Corollary. $(\frac{3}{2}, \frac{5}{3})$ is an isolated point in Ω.

It is attained by the fixed point of $a \mapsto ab$, $b \mapsto aa$ (period-doubling word).

Accumulation point in Ω

Theorem (Aberkane 2001). For $\ell \in \mathbb{N}$, let u be the fixed point of $\sigma: a \mapsto ab, b \mapsto (ab)^\ell aa$. Then $\alpha - 1 \sim 1/\ell^2$ and $\beta - 1 \sim 1/\ell$.

Corollary. $(1,1)$ is an accumulation point in Ω.

Questions

- Describe families of points in Ω (Kaitlyn Loyd, to appear).
- Find other isolated points in Ω (Firas Ben Ramdhane, in progress).
- Does Ω have non-empty interior?
- Is $\{1\} \times [1, \infty] \subset \Omega$?
- What does one obtain when restricting to words with a given property? For instance (Boshernitzan 1984), for words generating non-uniquely ergodic systems:
  $\Omega_{\text{nonUE}} \subseteq \Omega \cap [2, \infty] \times [3, \infty]$.
- Is Ω compact?
- Are points corresponding to purely substitutive words dense in Ω?

Rauzy graphs

(Rauzy 1983)
For each $n \in \mathbb{N}$, the Rauzy graph G_n is the directed graph with

- vertices: $L_n(u)$,
- edges: $L_{n+1}(u)$,
- $x \xrightarrow{z} y$ if x is a prefix of z and y is a suffix of z.

Edges may be labelled in several ways.
Here we choose the first letter of z.

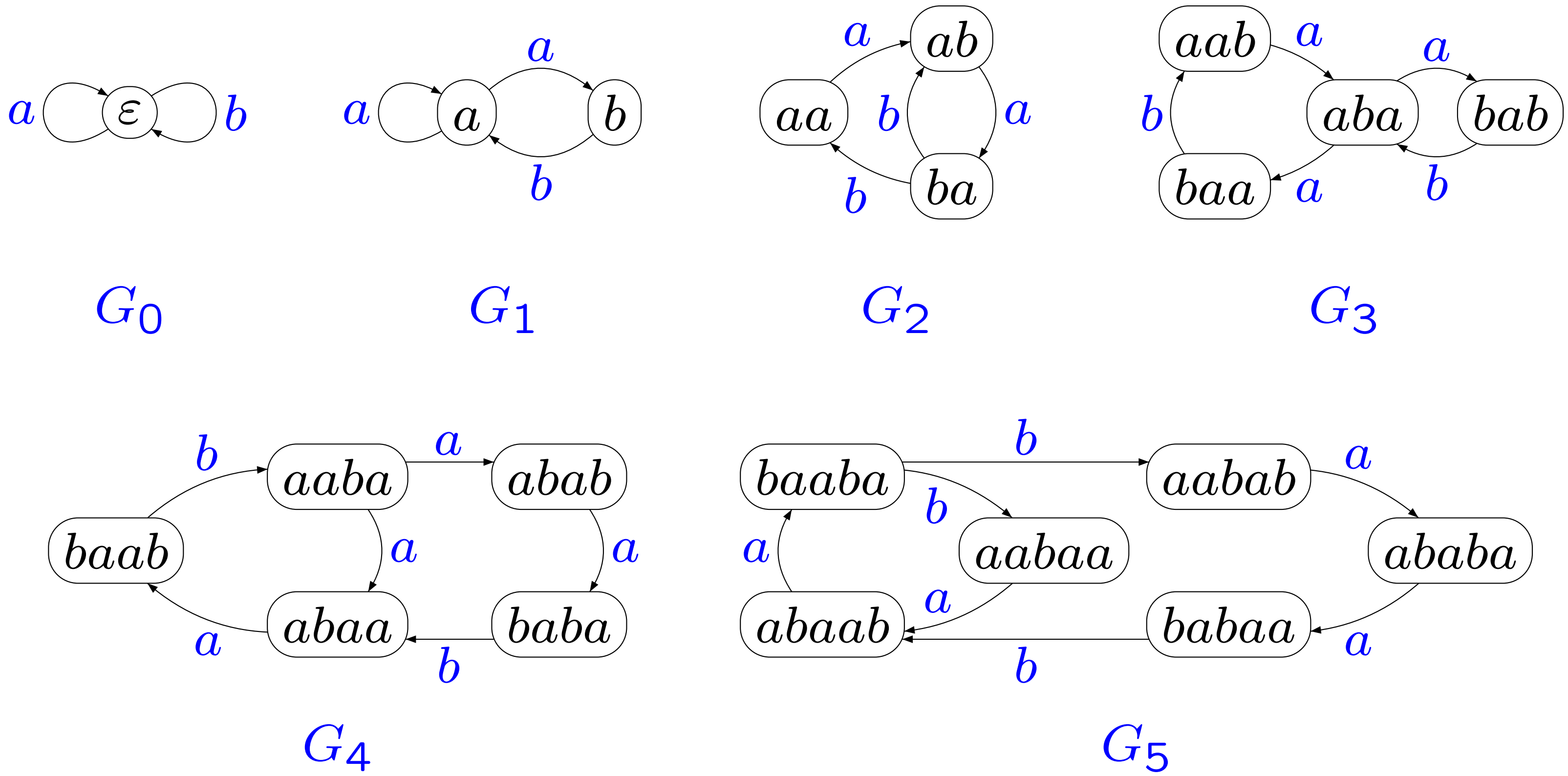
Example: Fibonacci word

Let $u = abaababaabaababaababaabaababaabaab\ldots$ be the Fibonacci word.
It is the fixed point of the substitution $a \mapsto ab$, $b \mapsto a$.

It is a Sturmian word: $p(n) = n + 1$ for all n.

So G_n has $n + 1$ vertices and $n + 2$ edges.

$u = abaababaabaababaababaabaababaabaababaababaabaababaababa\ldots$



Rauzy graphs as automata

G_n can be viewed as a nondeterministic finite automaton, where all states are initial and final. Then:

$$L(u) \subseteq L(G_n)$$

$$L(u) \cap A^{\leq n+1} = L(G_n) \cap A^{\leq n+1}$$

$$L(G_{n+1}) \subseteq L(G_n)$$

$$L(u) = \bigcap_{n \in \mathbb{N}} L(G_n)$$

Rauzy graphs and special factors

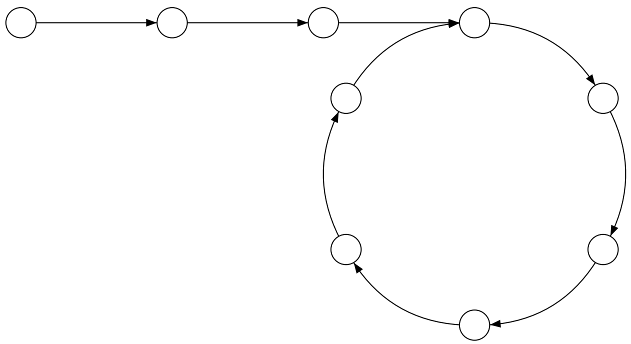
A factor $w \in L(u)$ is right special (for u) if there exist distinct letters a and b such that $wa \in L(u)$ and $wb \in L(u)$.

In G_n:
right special factor = vertex with more than one outgoing edge
left special factor = vertex with more than one incoming edge.

On a binary alphabet:
the number of right special factors is $s(n) = p(n+1) - p(n)$;
the number of left special factors is $s(n)$ or $s(n)+1$ (in the case where one vertex has no incoming edge).

Rauzy graphs for eventually periodic words

If u is eventually periodic, for n large enough G_n looks like this:



The length of the cycle is the period of u; the length of the tail its preperiod (if u is purely periodic there is no tail).

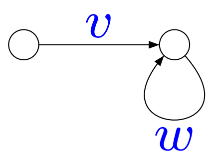
Shape of a Rauzy graph

The shape of a Rauzy graph is the graph obtained by removing all vertices with indegree and outdegree 1. Branches

$$x_0 \xrightarrow{a_1} x_1 \xrightarrow{a_2} x_2 \cdots x_{k-1} \xrightarrow{a_k} x_k$$

are replaced with a single edge $x_0 \xrightarrow{a_1 a_2 \ldots a_k} x_k$ labelled with a word.

If u is eventually (but not purely) periodic, for n large the shape of G_n is:



where $u = vw^\omega$.

Rauzy graphs for Sturmian words

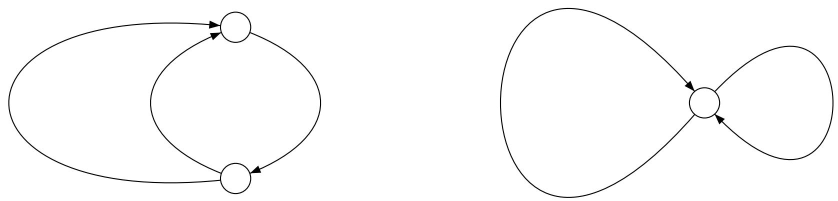
A Sturmian word is a word such that $p(n) = n + 1$ for all n (the smallest possible complexity for a non-periodic word).

Such a word is always recurrent: every factor occurs infinitely often. As a consequence, its Rauzy graphs are strongly connected.

$s(n) = (n + 2) - (n + 1) = 1$: there is one left special factor l and one right special factor r of length n. Therefore only two shapes are possible for G_n:

Rauzy graphs for Sturmian words

$s(n) = (n+2) - (n+1) = 1$: there is one left special factor l and one right special factor r of length n. Therefore only two shapes are possible for G_n:



Case 1: $l \neq r$ Case 2: $l = r$

Evolution from G_n to G_{n+1}

If $G = (V, E)$ is a directed graph, then its line graph is the graph $D(G) = (V', E')$ with $V' = E$ and

$$E' = \{(e_1, e_2) : \text{head}(e_1) = \text{tail}(e_2)\} \ .$$

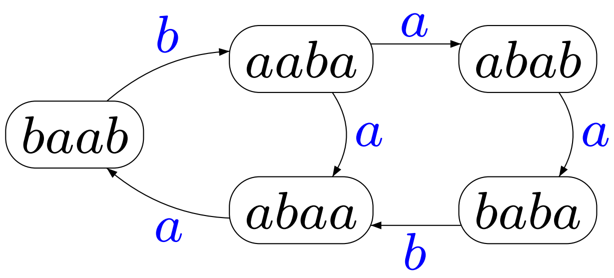
G_{n+1} is always a subgraph of $D(G_n)$. Often $G_{n+1} = D(G_n)$, in particular when u is recurrent (we assume this usually) and there is no bispecial factor (a factor that is both left special and right special).

Evolution without bispecial factor

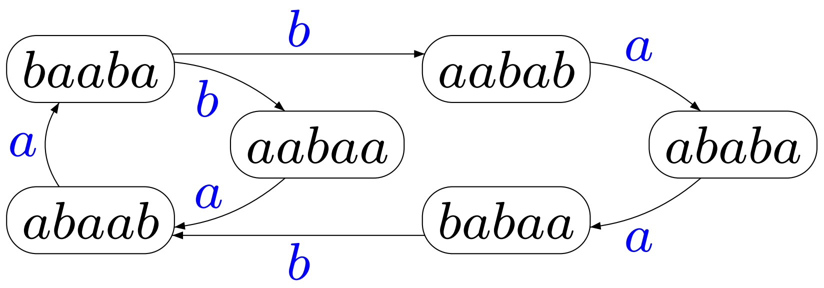
When there is no bispecial factor, $G_{n+1} = D(G_n)$ can be deduced from G_n without any additional information.

G_n and G_{n+1} have the same shape. The lengths of branches may increase or decrease by 1. At least one branch shrinks, so eventually a bispecial factor will occur in a later graph.

Example (Fibonacci):



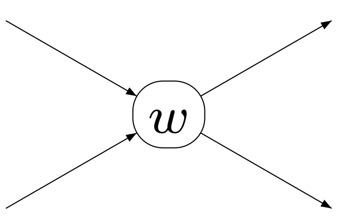
G_4



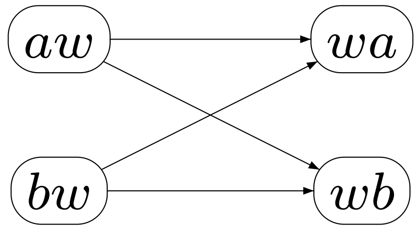
G_5

Bispecial factor burst

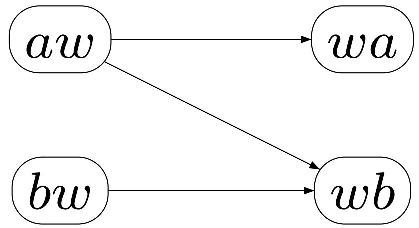
A bispecial factor is a factor that is both left special and right special.
For simplicity assume a binary alphabet $A = \{a, b\}$.



G_n
(w is bispecial)



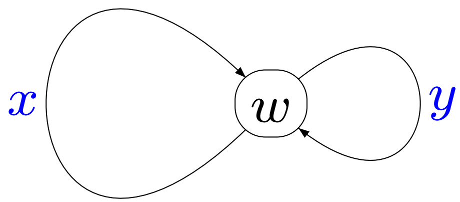
$D(G_n)$
(w yields 4 edges)



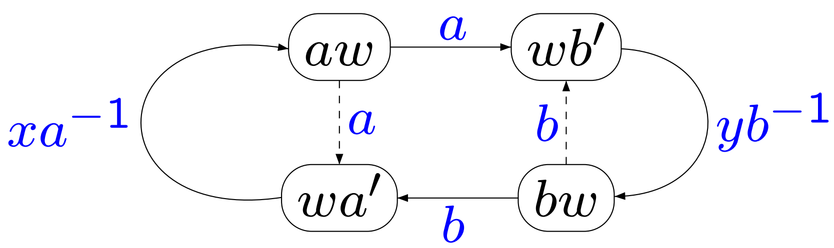
G_{n+1}
(edges may be deleted)

Evolution for Sturmian words

Assume that there is a bispecial factor of length n.

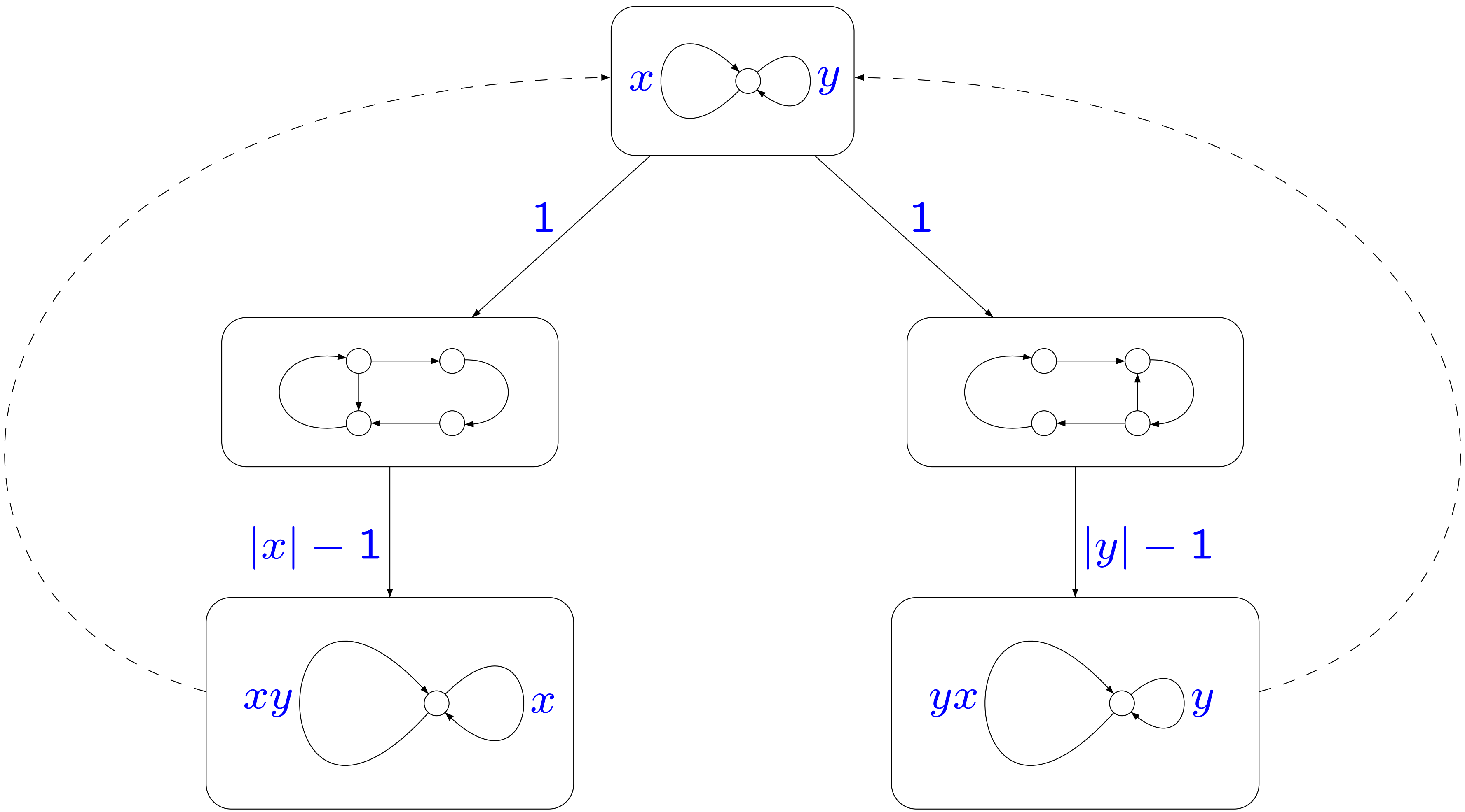


G_n



$D(G_n)$

To obtain G_{n+1}, one of the dashed vertical edges has to be removed from $D(G_n)$ (exactly one to get $p(n+2) = n+3$ edges; and the horizontal edges are needed for strong connectedness). So two evolutions are possible.



x
y
1
1
$|x| - 1$
$|y| - 1$
xy
x
yx
y

Recurrence formulas

Let n_i be the length of the i-th bispecial factor ($n_0 = 0$).

Let x_i, y_i be the labels of the loops of G_{n_i}, with $|x_i| \geq |y_i|$, $x_0 = a$, $y_0 = b$. Then

$$\begin{cases} n_{i+1} = n_i + |x_i| \\ x_{i+1} = x_i y_i \\ y_{i+1} = x_i \end{cases} \qquad \text{or} \qquad \begin{cases} n_{i+1} = n_i + |y_i| \\ x_{i+1} = y_i x_i \\ y_{i+1} = y_i \end{cases}$$

depending on the type of evolution between G_{n_i} and $G_{n_{i+1}}$.

An s-adic interpretation

Let $\varphi(a) = ab$, $\varphi(b) = a$, $\psi(a) = ba$, $\psi(b) = b$. Then there is a sequence of substitutions $(\sigma_i) \in \{\varphi, \psi\}^{\mathbb{N}}$ such that $x_i = \tau_i(a)$, $y_i = \tau_i(b)$, with $\tau_i = \sigma_0 \circ \sigma_1 \circ \cdots \circ \sigma_{i-1}$.

The infinite word

$$\hat{u} = \lim_{i \to \infty} \tau_i(a)$$

is such that $L(\hat{u}) = L(u)$ (actually $\hat{u}$ is standard Sturmian).

(σ_i) is an s-adic representation of $\hat{u}$.

(σ_i) has a strong connection with the continued fraction expansion of the slope of u.

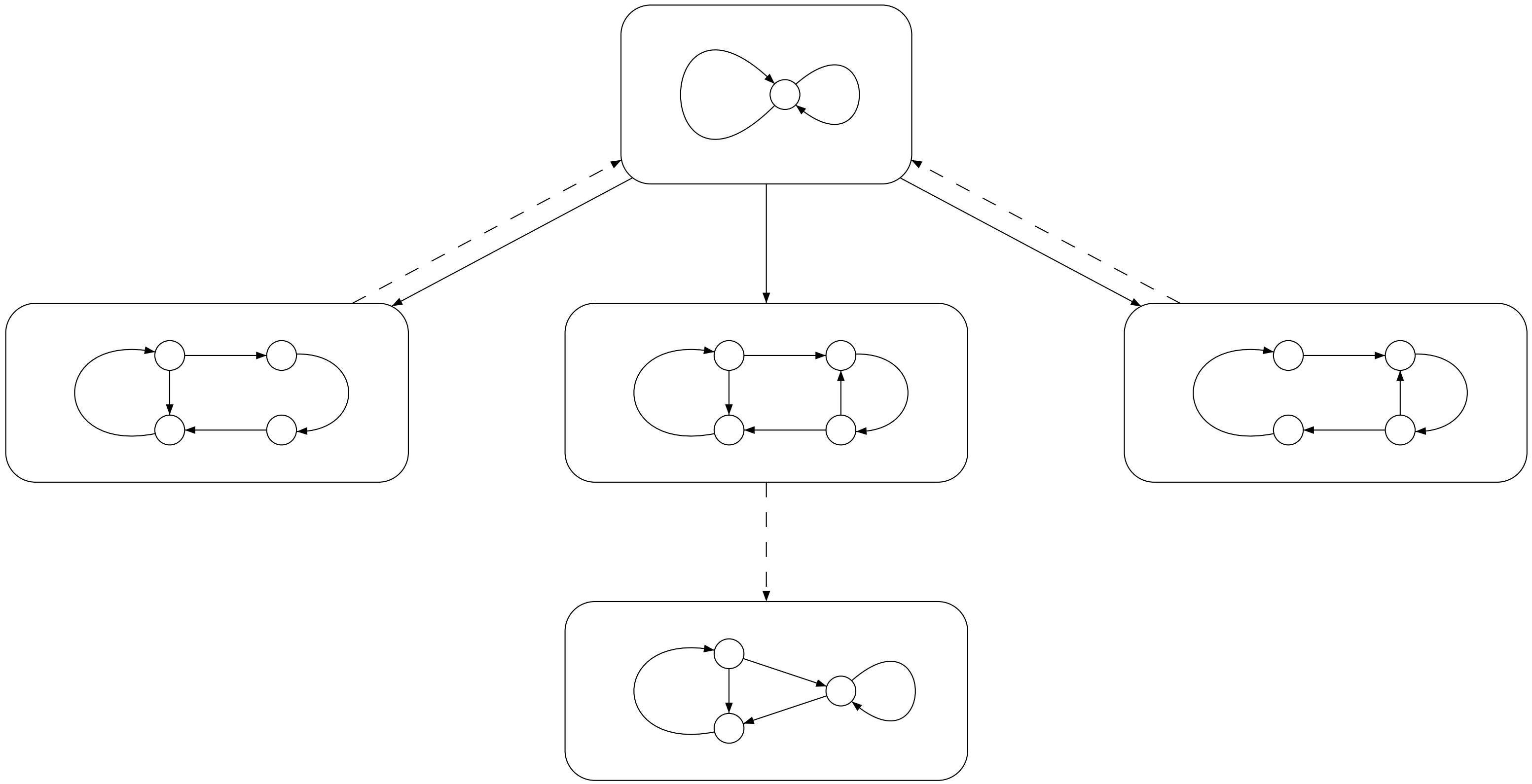
Very low complexity

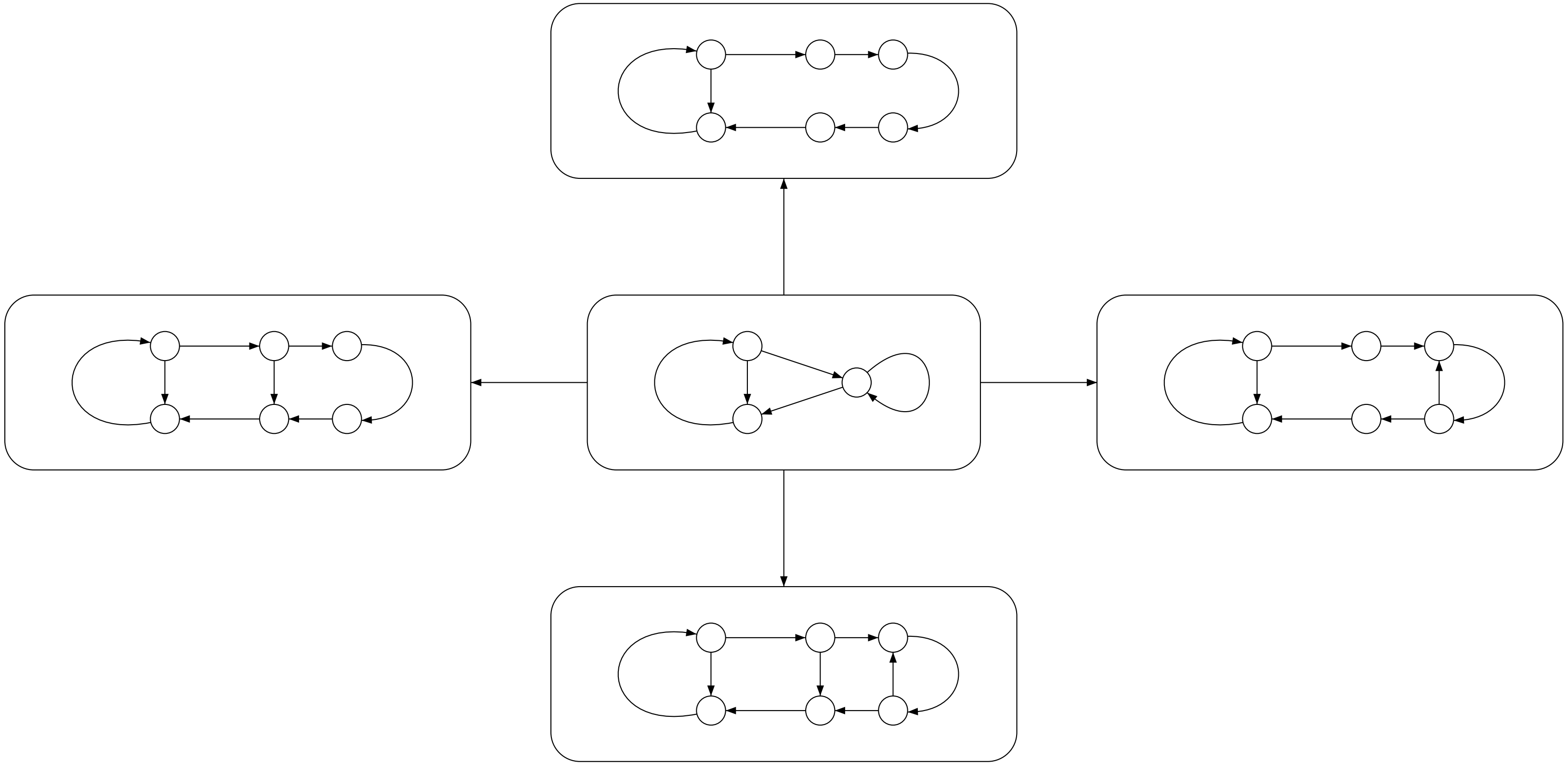
Let's describe graphs for a slightly larger class.

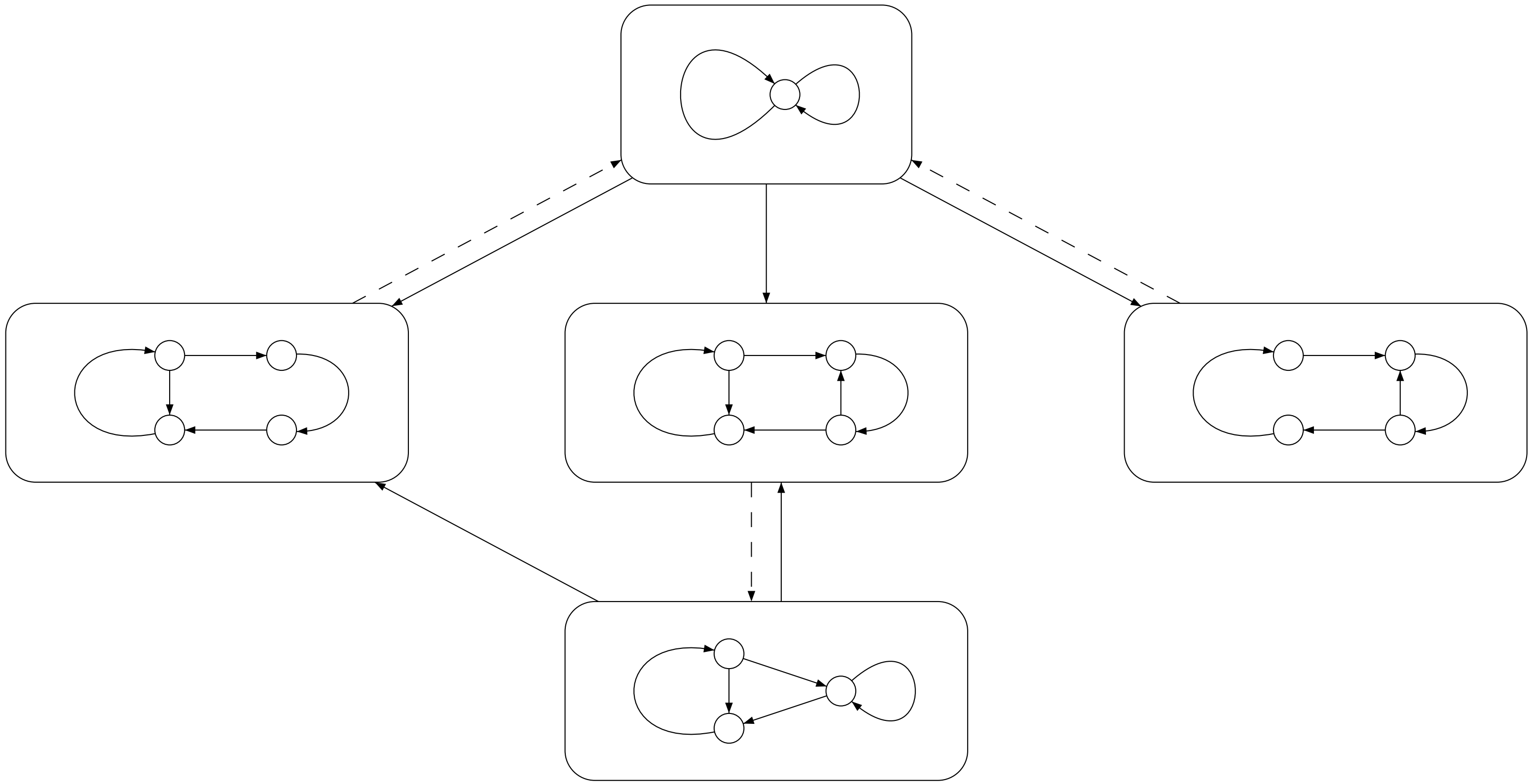
Idea: if $p(n+1) - p(n) = 1$ for infinitely many n, then infinitely often G_n will have a Sturmian shape. This is for instance the case if $\alpha = \liminf p(n)/n < 2$.

If $p(n) \leq 4n/3+1$ (Aberkane 2001), or more generally if $\beta(2+\alpha) < 4\alpha$ (Turki 2016), then the possible evolutions between 8-shaped graphs are completely described, they correspond to substitutions φ_m ($m \geq 1$) and ψ, where $\varphi_m(a) = ab^m$, $\varphi_m(b) = a$, $\psi(a) = ba$, $\psi(b) = b$.

Note that $\psi^q \circ \varphi_m$: $a \mapsto b^q a b^m, b \mapsto b^q a$.
Up to word conjugacy, this substitution is the same as $\tau_{m+q+1,q+1}$ (Creutz and Pavlov 2023).







Recurrent words with $p(n) = n + o(n)$

Recall that $\varphi_m : a \mapsto ab^m, b \mapsto a$ ($m \geq 1$) and $\psi : a \mapsto ba, b \mapsto b$.

Theorem (Aberkane 2003).
Let u be recurrent. Then $p_u(n) = n + o(n)$ if and only if u has the same factors as $\hat{u} = \lim_{i \to \infty} \tau_0 \circ \sigma_1 \circ \cdots \circ \sigma_i(a)$ where τ_0 is any non-periodic substitution, and $\sigma_j = \psi^{q_j} \circ \varphi_{m_j}$, with $m_j \geq 1$, $q_j \geq 0$, and

$$\lim_{\substack{j \to \infty \\ m_j \neq 1}} \frac{q_j}{m_j - 1} = +\infty \quad .$$