Workshop on Words and Complexity
Villeurbanne, February 19-23, 2018

Words with complexity $p(n) = n + o(n)$

Julien Cassaigne

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

Joint work with Ali Aberkane (Marseille, 2003)
and Mitali Thatte (IISER Pune, 2018)

Words with complexity $p(n) = n + o(n)$

- Dynamics of Rauzy graphs for Sturmian words
- Recurrent words with $p(n) = n + o(n)$
- Non-recurrent words with $p(n) = n + o(n)$

Infinite words and their factors

$u \in A^{\mathbb{N}}$: an infinite word
(one may also consider bi-infinite words $u \in A^{\mathbb{Z}}$, or subshifts).

$w \in A^*$ 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.

Our goal

$p_u(n) = O(1)$ if and only if u is eventually periodic.

$p_u(n) = n + 1$ if and only if u is Sturmian.

$p_u(n) = n + O(1)$ if and only if u is quasi-Sturmian:
$u = wh(v)$ with w finite word and h injective morphism.

General problem: find other complexity classes of infinite words that can be explicitely described.

Here we study the class $p(n) = n + o(n)$, i.e.,

$$\lim \frac{p(n)}{n} = 1 \quad .$$

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.

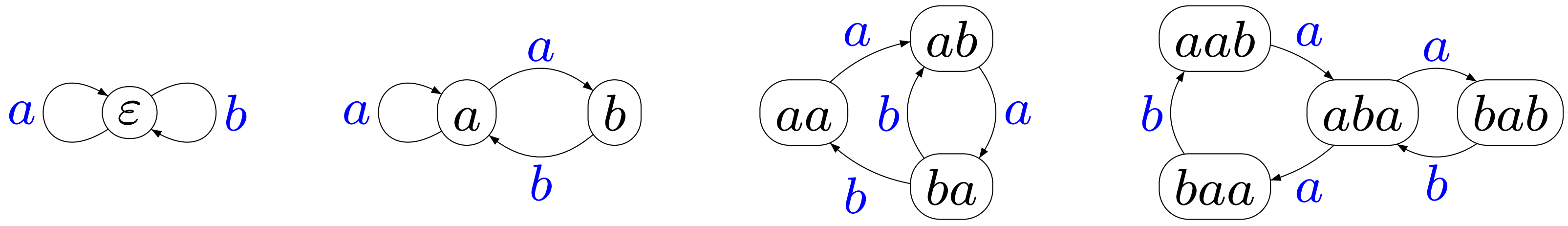
Example: Fibonacci word

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

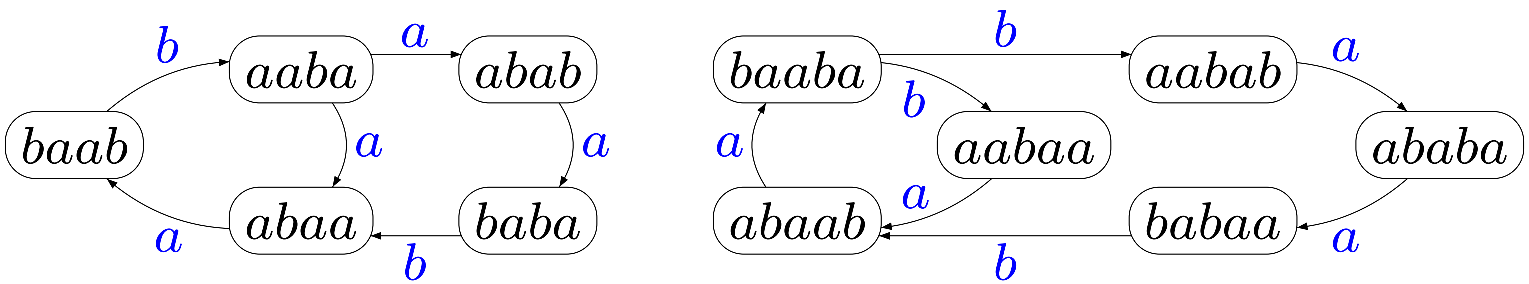
It is a Sturmian word: $p(n) = n + 1$ for all n.

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

$$u = abaababaabaababaababaabaababaabaababaababaabaababaababa\ldots$$



G_0 G_1 G_2 G_3



G_4 G_5

Rauzy graphs and special factors

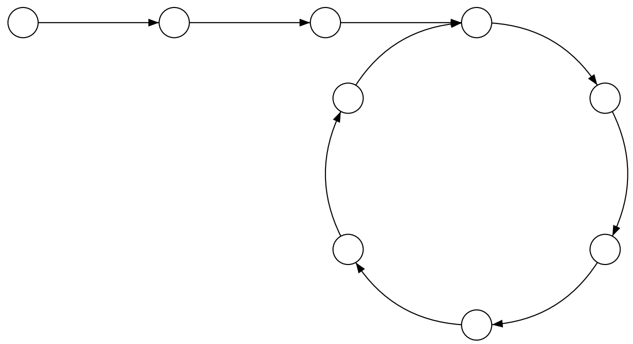
A factor $w \in L(u)$ is right special (for u) if there exist distinct letters a and b such that $wa \in L(u)$ and $wb \in L(u)$.

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).

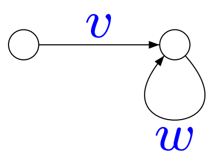
Shape of a Rauzy graph

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

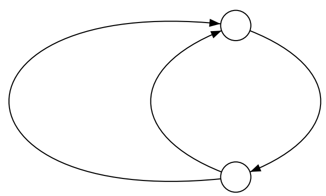
A Sturmian word is a word such that $p(n) = n + 1$ for all n (the smallest possible complexity for a non-periodic word).

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

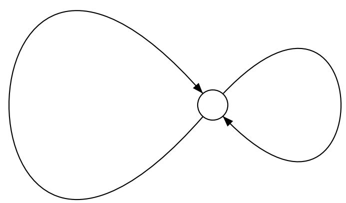
$s(n) = (n + 2) - (n + 1) = 1$: there is one left special factor l and one right special factor r of length n. Therefore only two shapes are possible for G_n:

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

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

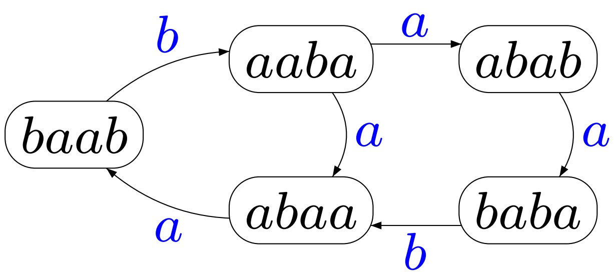
G_{n+1} is always a subgraph of $D(G_n)$. Often $G_{n+1} = D(G_n)$, in particular when u is recurrent and there is no bispecial factor (a factor that is both left special and right special).

Evolution without bispecial factor

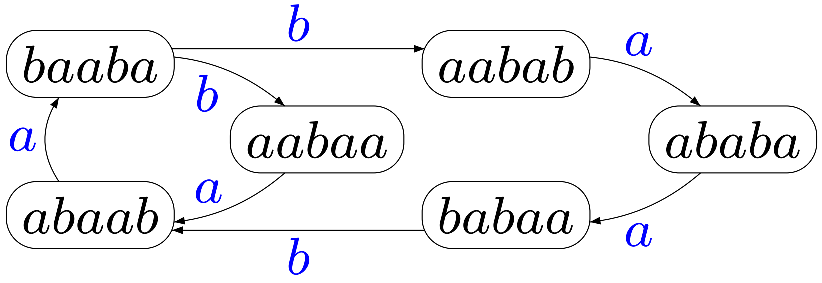
When there is no bispecial factor, $G_{n+1} = D(G_n)$ can be deduced from G_n without any additional information.

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):



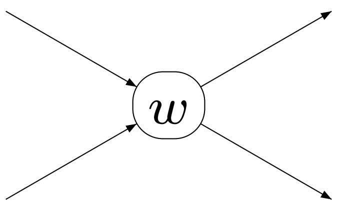
G_4



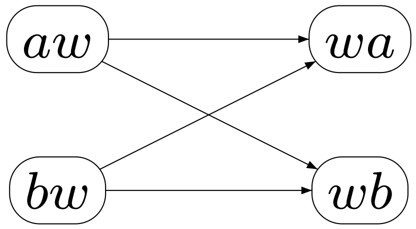
G_5

Bispecial factor burst

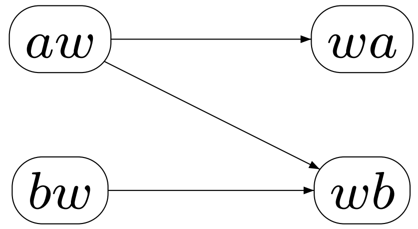
A bispecial factor is a factor that is both left special and right special. For simplicity assume a binary alphabet $A = \{a, b\}$.



G_n
(w is bispecial)



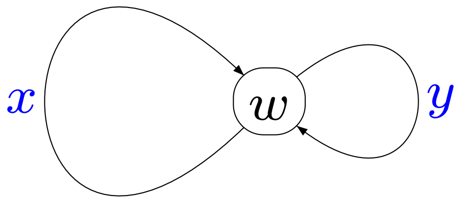
$D(G_n)$
(w yields 4 edges)



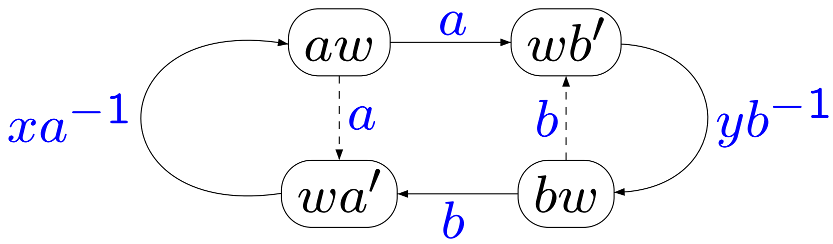
G_{n+1}
(edges may be deleted)

Evolution for Sturmian words

Assume that there is a bispecial factor of length n.

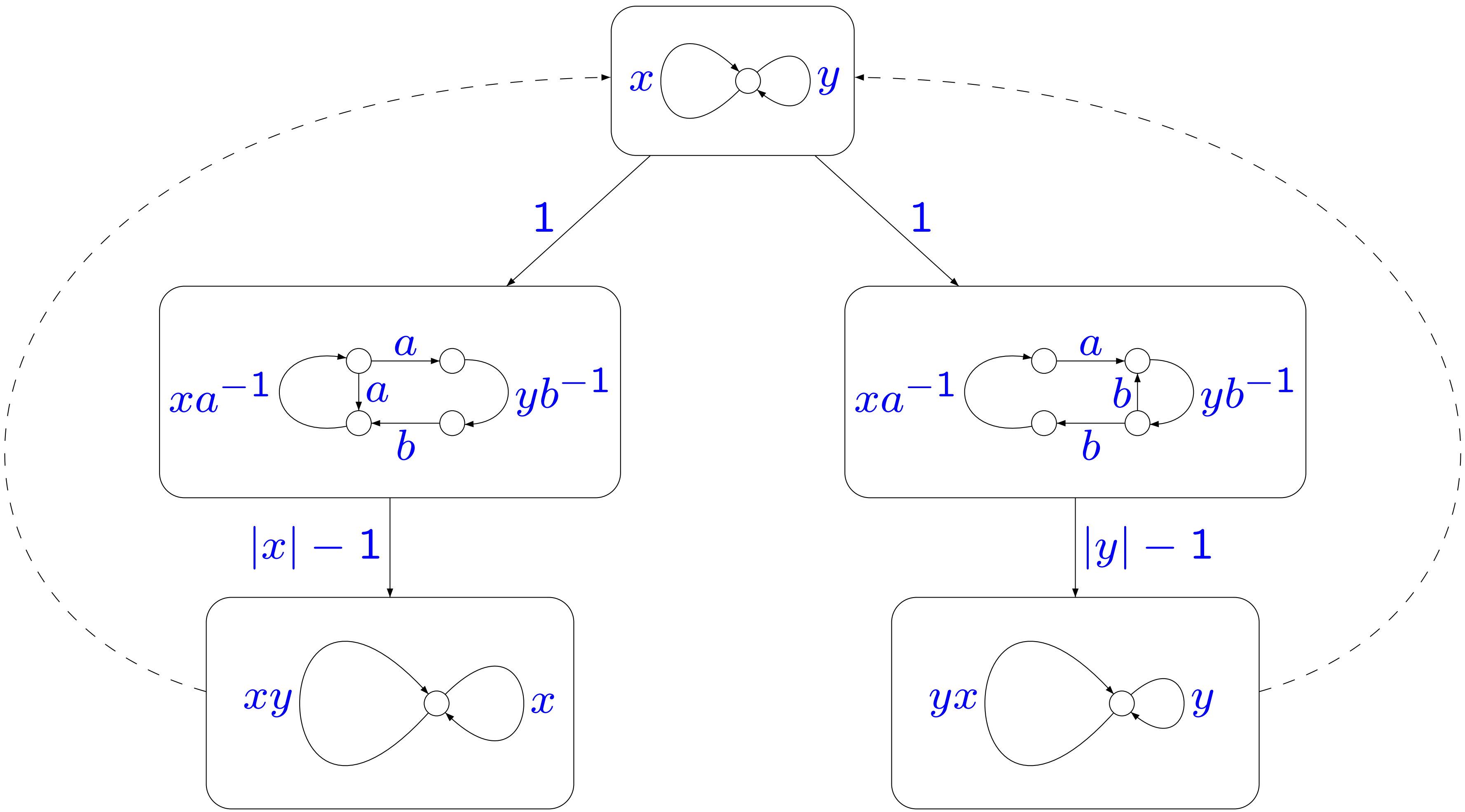


G_n



$D(G_n)$

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
a
a
xa−1
yb−1
b
a
b
xa−1
yb−1
b
|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 and G_{n+1}.

An s-adic interpretation

Let $\varphi(a) = ab$, $\varphi(b) = a$, $\psi(a) = ba$, $\psi(b) = b$. Then there is a sequence of morphisms $(\sigma_i) \in \{\varphi, \psi\}^{\mathbb{N}}$ 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.

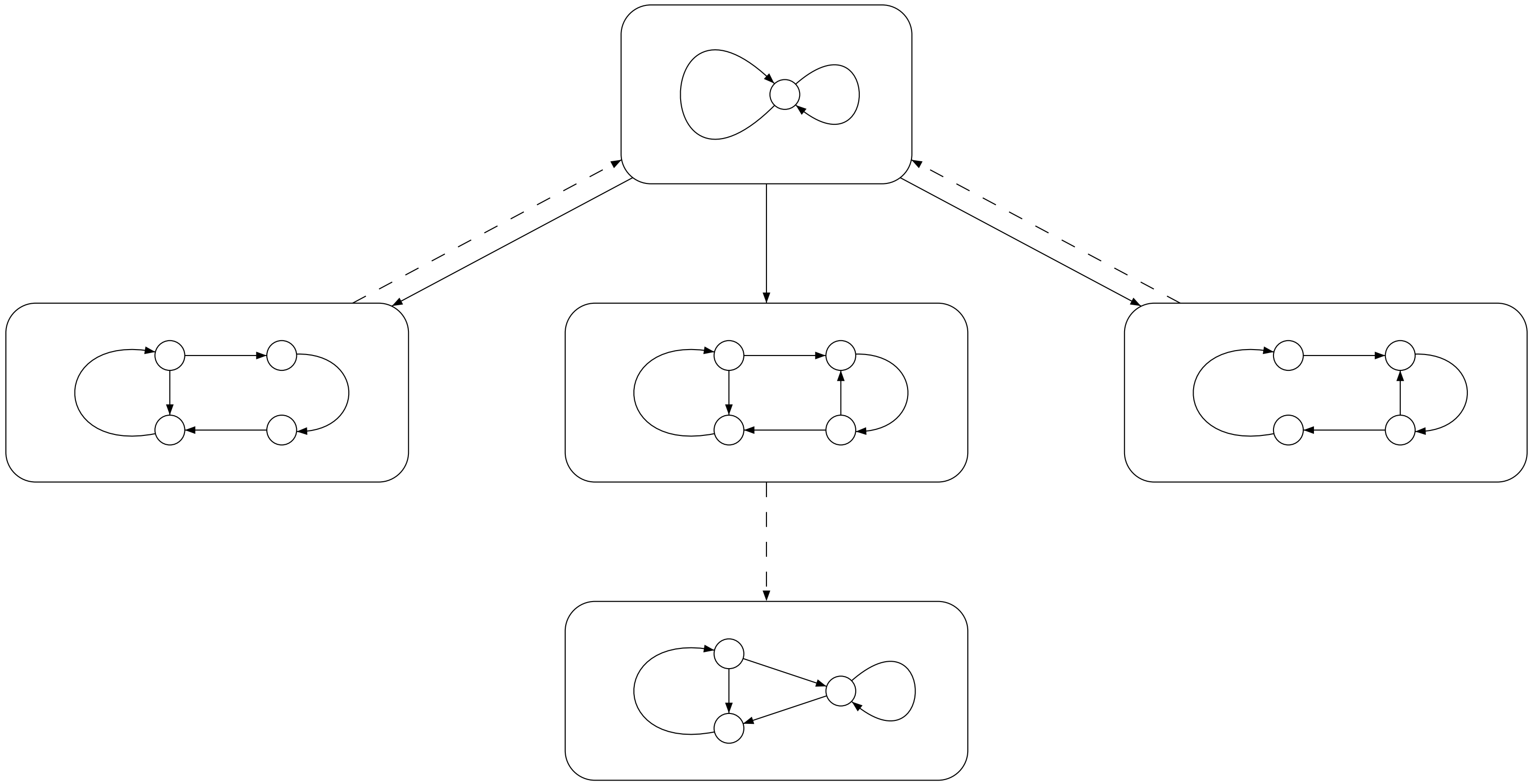
Recurrent words with $p(n) = n + o(n)$

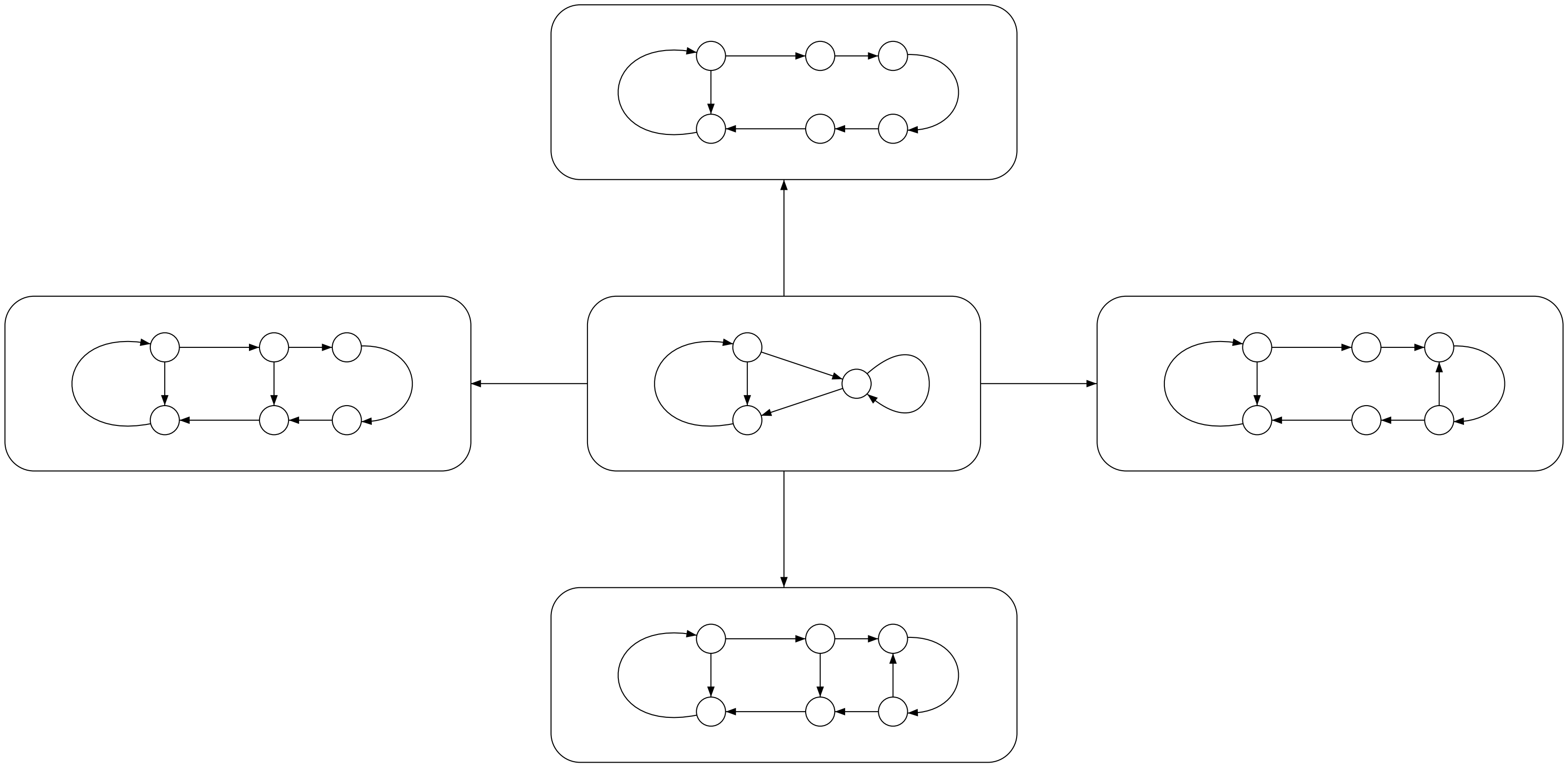
Assume that u is recurrent and $p(n) = n + o(n)$.

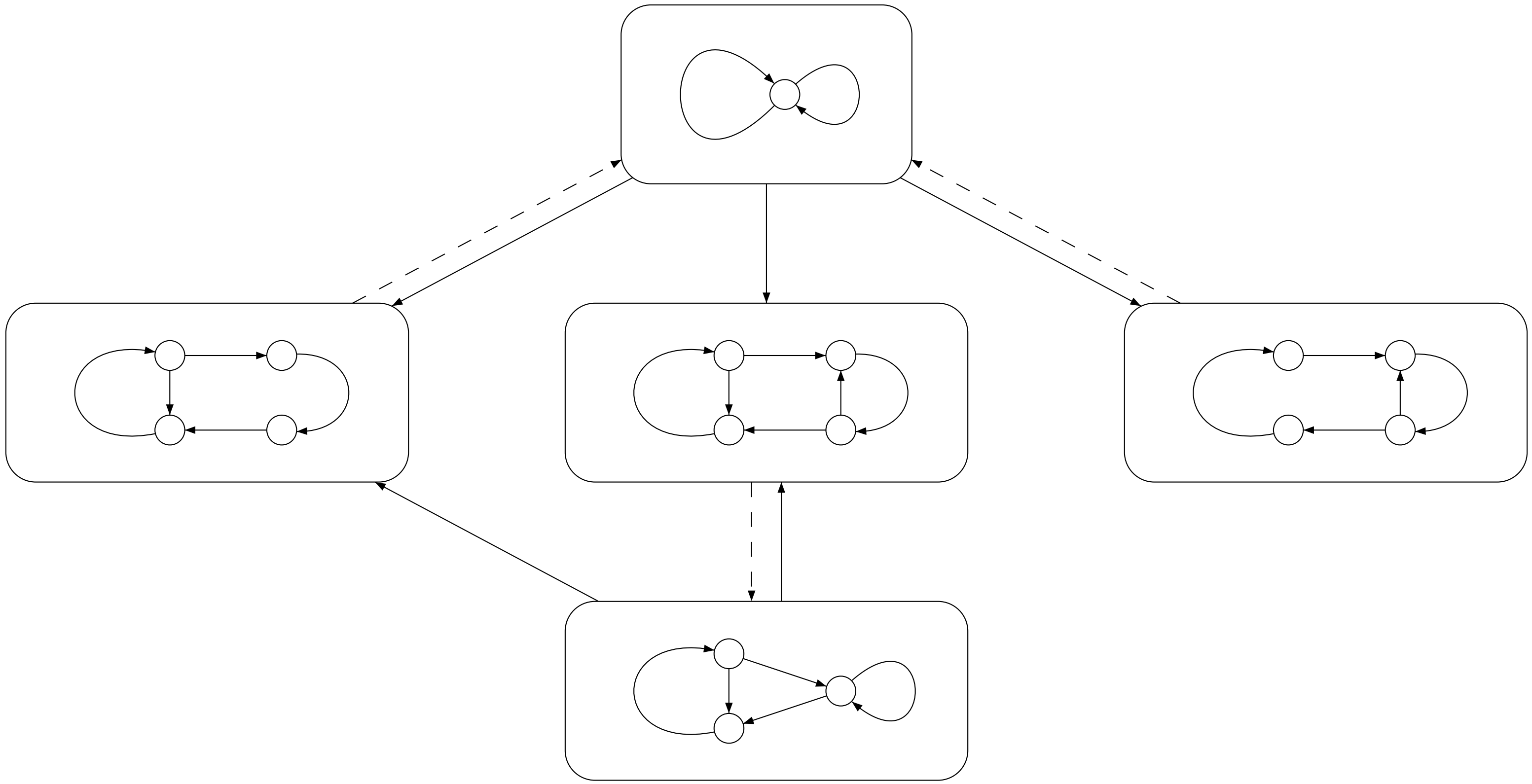
Idea 1: $s(n) = p(n+1) - p(n) = 1$ for infinitely many n: infinitely often G_n has a Sturmian shape.

Idea 2: intervals on which $s(n) \geq 2$ are short (length $o(n)$): graph shapes with several right special factors cannot last long.

More precisely: Let n_0 be such that $p(n) \leq 1.1n$ when $n \geq n_0$. Then in any interval $[n, 1.2n]$ with $n \geq n_0$, there is n' such that $s(n') = 1$.







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

Theorem 1 (Aberkane 2003).
If $p_u(n) = n + o(n)$ and u is recurrent, then for n large enough the possible evolutions between 8-shaped graphs follow the previous figure, they correspond to morphisms φ_m $(m \geq 1)$ and ψ, where $\varphi_m(a) = ab^m$, $\varphi_m(b) = a$, $\psi(a) = ba$, $\psi(b) = b$.

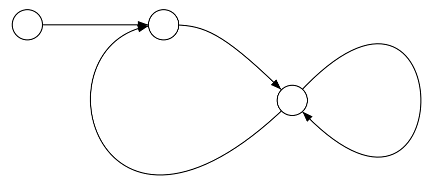
Theorem 2 (Aberkane 2003).
Let u be recurrent. Then $p_u(n) = n + o(n)$ if and only if u has the same factors as $\hat{u} = \lim_{i \to \infty} \tau_0 \circ \sigma_1 \circ \cdots \circ \sigma_i(a)$ where $\sigma_j = \psi^{q_j} \circ \varphi_{m_j}$, with $m_j \geq 1$, $q_j \geq 0$, and

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

Non-recurrent words with $p(n) = n + o(n)$

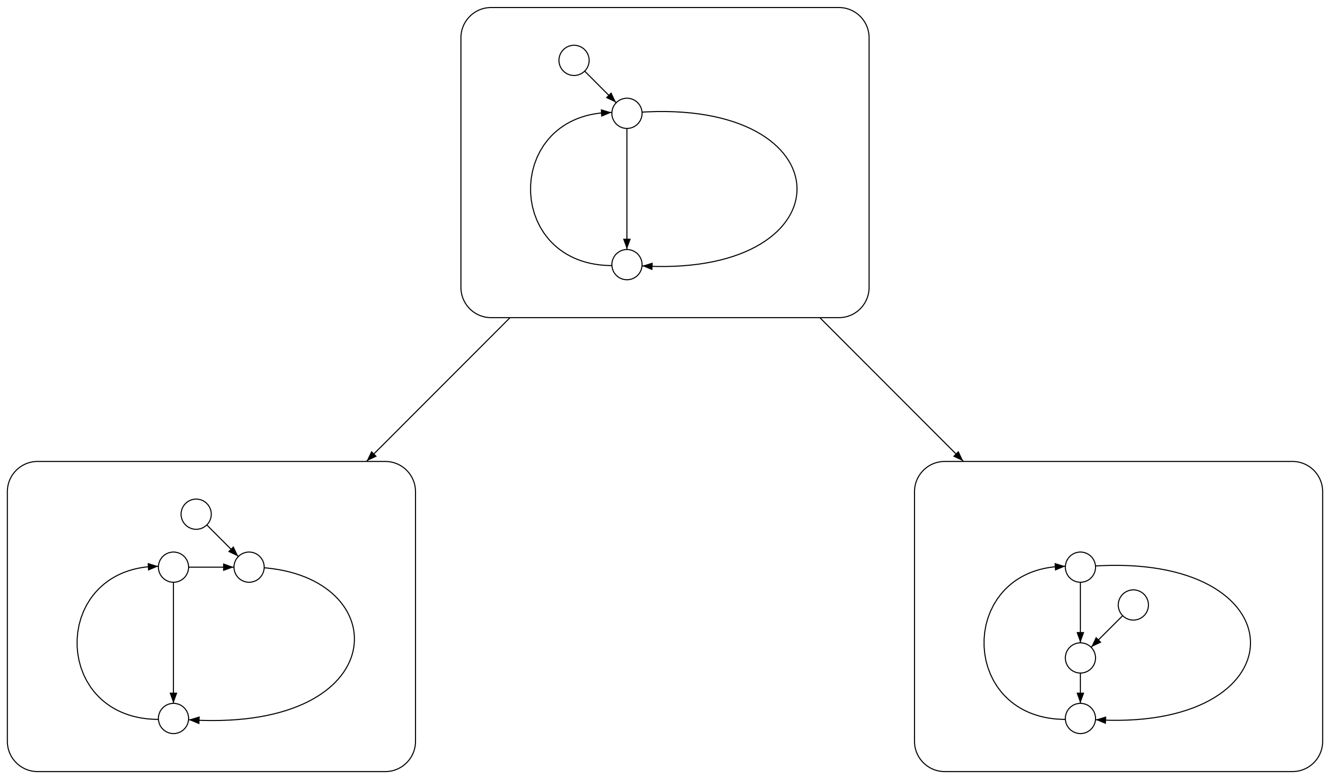
Work in progress...

Idea 1 still works, except that a tail has to be added.

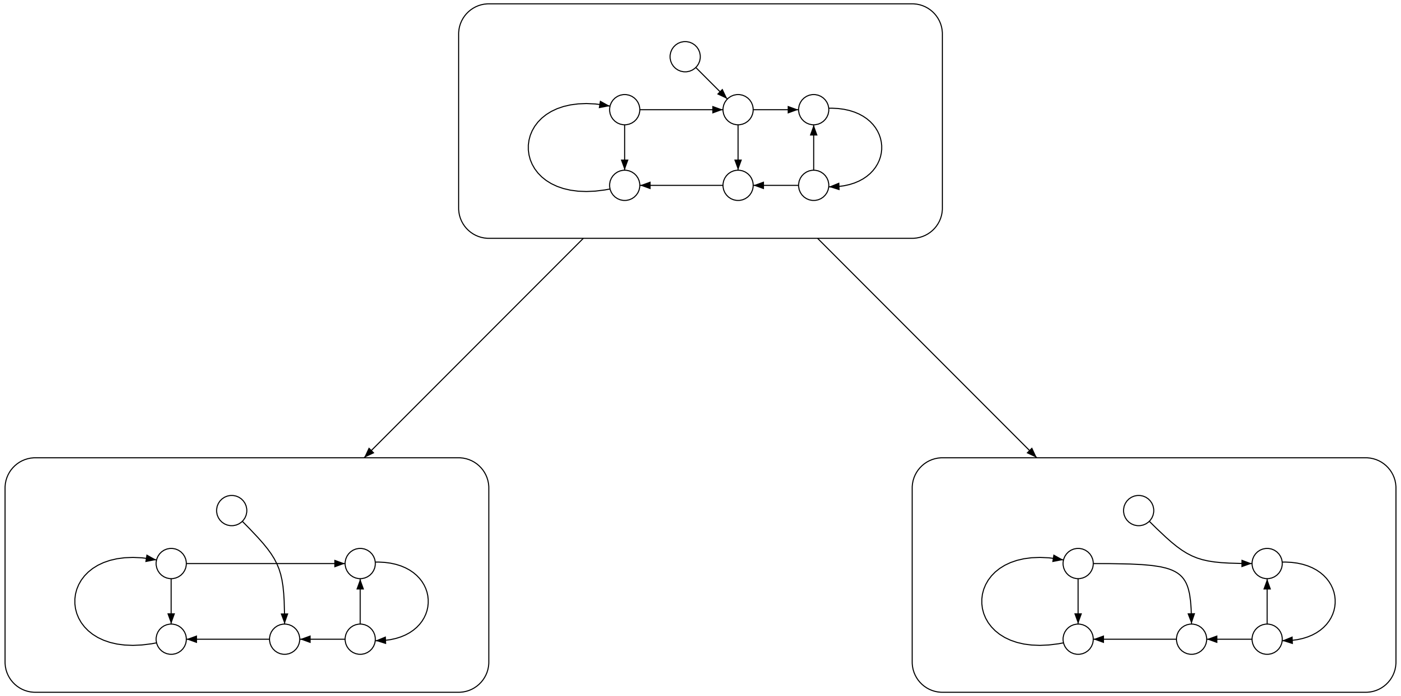


Idea 2 still works, but the tail may interact with right special factors.

Evolutions with the tail unchanged



Evolutions changing the tail



Partial results

Theorem 3.
If $p(n) = n + o(n)$, then for n large enough, $s(n) \leq 3$.

A finite list of possible graph shapes can be established, as well as possible transitions between them.

To be done: encode the evolutions with morphisms (on an alphabet with 4 letters to include the tail and where it is attached), to generalize Theorem 1.

Open: find a necessary and sufficient condition as in Theorem 2.