Available online at www.sciencedirect.com
ScienceDirect
Discrete Applied Mathematics 155 (2007) 1014–1030
DISCRETE APPLIED MATHEMATICS
www.elsevier.com/locate/dam

On Sturmian graphs☆

Chiara Epifanio[a], Filippo Mignosi[b], Jeffrey Shallit[c], Ilaria Venturini[d]

[a]*Dipartimento di Matematica e Applicazioni, Università di Palermo, Italy*
[b]*Dipartimento di Informatica, Università dell'Aquila, Italy*
[c]*School of Computer Science, University of Waterloo, Ont., Canada*
[d]*TSI, ENST, Paris, France*

Received 14 March 2006; received in revised form 24 August 2006; accepted 8 November 2006
Available online 28 December 2006

Abstract

In this paper we define Sturmian graphs and we prove that all of them have a certain "counting" property. We show deep connections between this counting property and two conjectures, by Moser and by Zaremba, on the continued fraction expansion of real numbers. These graphs turn out to be the underlying graphs of compact directed acyclic word graphs of central Sturmian words. In order to prove this result, we give a characterization of the maximal repeats of central Sturmian words. We show also that, in analogy with the case of Sturmian words, these graphs converge to infinite ones.

Keywords: Sturmian words; CDAWGs; Continued fractions; Repeats

1. Introduction

Let Σ be a finite set of symbols, called an *alphabet*. A *word* or *string* w is a finite sequence $w = a_1a_2 \dots a_n$ of characters taken in the alphabet Σ, its length (i.e., the number of characters in the string) is defined to be n and it is denoted by $|w|$. We denote by Σ^* the set of words over Σ and by ε the empty word.

A word $u \in \Sigma^*$ is a *factor* (or *substring*) (resp., *prefix*, *suffix*) of a word w if there exist words $x, y \in \Sigma^*$ such that $w = xuy$ (resp., $w = uy$, $w = xu$). The factor (resp., prefix, suffix) is *proper* if $xy \neq \varepsilon$ (resp., $y \neq \varepsilon$, $x \neq \varepsilon$).

Sturmian words are aperiodic infinite words over a binary alphabet of minimal subword complexity, i.e., with exactly $n + 1$ factors of length n. They have been extensively studied for their properties and equivalent definitions. Moreover, the well-known Fibonacci word is Sturmian.

Among the different definitions, one is obtained by considering the intersections of a ray having an irrational slope $\alpha > 0$ with a square-lattice. The word obtained by coding each vertical intersection with an a, each horizontal intersection by a b and each corner with ab or ba is Sturmian. If the ray starts from the origin, the word obtained is called *characteristic*. Another way of constructing characteristic Sturmian words is by applying the *standard method*.

☆ Partially supported by MIUR National Project PRIN "Linguaggi Formali e Automi: teoria ed applicazioni".
E-mail addresses: epifanio@math.unipa.it (C. Epifanio), mignosi@di.univaq.it (F. Mignosi), shallit@graceland.math.uwaterloo.ca (J. Shallit), venturi@tsi.enst.fr (I. Venturini).

0166-218X/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2006.11.003

Define inductively the two sequences of words $\{A_n\}$ and $\{B_n\}$ by

$$\begin{cases} A_0 = a, \\ B_0 = b \end{cases}$$

and by the two *rules of Rauzy* [21]

$$R_1 : \begin{cases} A_{n+1} = A_n, \\ B_{n+1} = A_n B_n, \end{cases} \qquad R_2 : \begin{cases} A_{n+1} = B_n A_n, \\ B_{n+1} = B_n. \end{cases}$$

When each of the two rules is applied infinitely often, these two sequences converge to the same infinite word that is characteristic. Conversely, each characteristic word is obtained in this way.

Given a pair (A_n, B_n), we can associate with it its *directive sequence* (cf. [8]), that is, the sequence of integers $[a_0, a_1, \ldots, a_s]$ such that $\sum_{i=0}^{s} a_i = n$, representing the fact that the final sequences A_n and B_n are obtained by applying R_1 to A_0 and $B_0 a_0$ consecutive times, after that $R_2 a_1$ consecutive times, etc.

Words obtained by removing last two characters from A_n or B_n are called *central Sturmian words*.

Given a pair (A_n, B_n) having directive sequence $[a_0, a_1, \ldots, a_s]$, it is possible to recursively define $\max(|A_n|, |B_n|)$ as the $(s+1)$th element of the following sequence (l_j):

$$\begin{cases} l_0 = 1, \\ l_1 = a_0 + 1, \\ l_{j+1} = a_j \cdot l_j + l_{j-1}, \end{cases} \qquad j = 1 \ldots s.$$

For references on Sturmian words and their geometric representation see [13,17 Chapter 2].

If the directive sequence $[a_0, a_1, \ldots]$ is infinite, the infinite word to which A_n and B_n converge represents a ray having slope α, where α has $[a_0, a_1, \ldots]$ as its simple continued fraction expansion.

Let us recall some basic notation and results on continued fractions.

If α is a real number, we can expand α as a *simple continued fraction*

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}$$

which is usually abbreviated as $\alpha = [a_0, a_1, a_2, a_3, \ldots]$.

In this paper, we only discuss the case where a_0 is a non-negative integer and a_i is a positive integer for $i \geqslant 1$; the expansion may or may not terminate. For references to continued fractions, see [5,9 Chapter 10,19,22].

If α is irrational, this representation is infinite and unique. If α is rational, there are two possible finite representations. Indeed, it is well known that $[a_0, a_1, \ldots, a_{s-1}, a_s, 1] = [a_0, a_1, \ldots, a_{s-1}, a_s + 1]$.

The integers in the continued fraction expansion of a real number are called *partial quotients*.

Given the continued fraction expansion of α, it is possible to construct a sequence of rationals P_s/Q_s, called *convergents*, that converges to α, by the following rules:

$$\begin{cases} P_0 = a_0, & Q_0 = 1, \\ P_1 = a_1 \cdot a_0 + 1, & Q_1 = a_1, \\ P_{s+1} = a_{s+1} \cdot P_s + P_{s-1}, & Q_{s+1} = a_{s+1} \cdot Q_s + Q_{s-1}. \end{cases}$$

It is easy to see that $l_{j+1} = P_j + Q_j$.

The *directed acyclic word graph* of a word w, *DAWG*(w), is the smallest finite state automaton that recognizes all the suffixes of the word. DAWGs have linear size and can be built in linear time with respect to the size of the word. They are involved in several combinatorial algorithms on strings and have many applications, such as full-text indexing. If the last letter in w is a letter \$ that does not appear elsewhere in w, DAWG(w) coincides, apart from the set of final states, with the factor automaton of w, i.e., with the minimal deterministic automaton that recognizes the factors of w. In fact, while in the factor automaton every state is final, in the DAWG the only final state is the last one in every topological order. Blumer et al. (cf. [1–3]) first introduced the *compact directed acyclic word graph* of a word w, *CDAWG*(w), a space efficient variant of *DAWG*(w), obtained by compacting it. Arcs in the obtained structure are labeled by representations of the factors of the word. More precisely, each arc is labeled by the initial position and the length of the factor represented by the arc. For references on CDAWGs, see also [6,7,11,12].

In this paper we define a new data structure, the Sturmian graph of a directive sequence $[a_0, \ldots, a_s]$, $G([a_0, \ldots, a_s])$, and we show how it coincides with the CDAWG of the word w obtained by the longest word in the pair (A_n, B_n) with directive sequence $[a_0, \ldots, a_s]$ replacing last two letters with a \$ symbol, where the label of each arc is replaced by the length of the factor it represents. More exactly, we show that $G([a_0, \ldots, a_s])$ coincides with the CDAWG of the word obtained in such a way, where arcs are labeled only by the lengths of the factors they represent. Moreover, we prove that, analogously to Sturmian central words, the Sturmian graph $G([a_0, \ldots, a_s, 1])$ of directive sequence $[a_0, \ldots, a_s, 1]$ coincides with the one $G([a_0, \ldots, a_s + 1])$ of directive sequence $[a_0, \ldots, a_s + 1]$ and that $G([0, a_1, \ldots, a_s]) = G([a_1, \ldots, a_s])$. Finally, we prove that Sturmian graphs have a certain counting property.

The paper is organized as follows. In the next section we introduce our new data structure and prove some results on it. In Section 3, we show how Sturmian graphs turn out to be the underlying graphs of compact directed acyclic word graphs, CDAWGs, of central Sturmian words. Finally, in Section 4 we show how Sturmian graphs converge, in analogy with the case of Sturmian words, to infinite ones.

2. Special (or finite) Sturmian graphs

A *weighted DAG* is a directed acyclic graph, where each arc is weighted by a real number. Arcs are represented by triples (p, c, q), that means that there exists an arc from state p to state q of weight c.

For any rational $P/Q = [a_0, \ldots, a_s]$ with $\sum_{i=0}^{s} a_i \geqslant 2$ we inductively define a graph $G(P/Q) = G([a_0, \ldots, a_s])$ that we call the *Sturmian graph of* $P/Q = [a_0, \ldots, a_s]$. This graph is a weighted DAG where weights are positive integers.

If $a_0 = 0$ we set $G([a_0, a_1, \ldots, a_s]) = G([a_1, \ldots, a_s])$. Therefore, in what follows we suppose that $a_0 \geqslant 1$.

The first Sturmian graph—the base case—is the graph $G([1, 1]) = G([2])$. It consists of only two states and two arcs, both going from state 1 to the final state F and having weights, respectively, 1 and 2. It can be seen in Fig. 1.

To give the inductive step, let us recall the definition of the sequence (l_j):

$$\begin{cases} l_0 = 1, \\ l_1 = a_0 + 1, \\ l_{j+1} = a_j \cdot l_j + l_{j-1}. \end{cases}$$

Given the Sturmian graph of $[a_0, \ldots, a_s]$, $s \geqslant 0$, $\sum_{i=0}^{s} a_i \geqslant 2$, $G([a_0, \ldots, a_s])$, we define the Sturmian graph $G([a_0, \ldots, a_s, 1])$ in the following way: each arc of maximal weight in $G([a_0, \ldots, a_s])$ (all of them end at the final state) is split into one arc of that weight minus 1 from the same outgoing state to a new state (the same for each arc) and two arcs from this new state towards the final one, one labeled 1 and the other labeled $l_s + 1$.

Moreover, if $a_s = 1$, then for each state of out-degree 2, except the new one, one must add a new outgoing arc labeled $l_s + 1$ towards the final state, with the exception of the new state that has already one such arc.

As $[a_0, a_1, \ldots, a_s, 1] = [a_0, a_1, \ldots, a_s + 1]$, the previously defined inductive step let us construct every Sturmian graph $G([a_0, \ldots, a_k])$, $k \geqslant 0$.

Let us give some examples. Fig. 2 shows graphs $G([3, 1])$ and $G([1, 1, 1, 1])$. The first one is obtained starting from $G([3])$, inductively built from the base case $G([2])$, being $G([3]) = G([2, 1])$. The second one is derived starting from $G([1, 1, 1])$ that, in turn, comes from the base case $G([1, 1])$.

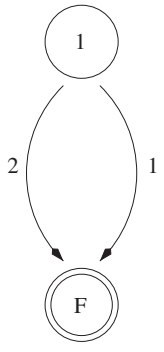

Fig. 1. Base case.

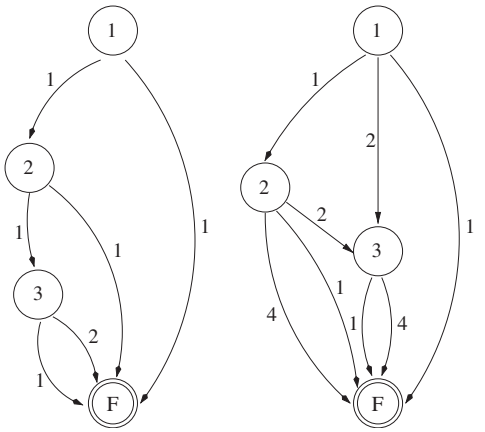

Fig. 2. Graphs $G([3, 1])$ and $G([1, 1, 1, 1])$.

Proposition 1. *The Sturmian graph* $G([a_0, \ldots, a_s, 1])$, $s \geqslant 0$, $n = \sum_{i=0}^{s} a_i + 1 \geqslant 2$, *contains exactly n states, among them* a_s *of out-degree* 2, *and* $3(n-1) - a_s$ *arcs.*

Proof. Let us proceed by induction on $n = \sum_{i=0}^{s} a_i + 1 \geqslant 2$. The claim is easily verified in the base case $G([1, 1])$.

Let us consider the inductive step. The Sturmian graph $G([a_0, \ldots, a_s, 1])$ is built starting from the Sturmian graph $G([a_0, \ldots, a_s])$. Let us distinguish two cases:

$a_s > 1$: In this case $G([a_0, \ldots, a_s, 1])$ has exactly one more state than $G([a_0, \ldots, a_s])$, that moreover has out-degree 2. Therefore it has one more state, two more arcs and one more state of out-degree 2 of $G([a_0, \ldots, a_s]) = G([a_0, \ldots, a_s - 1, 1])$. But, by the inductive hypothesis, $G([a_0, \ldots, a_s - 1, 1])$ has $n-1$ states, $a_s - 1$ states of out-degree 2 and $3(n-1-1) - (a_s - 1)$ arcs and the claim is easily verified.

$a_s = 1$: In this case $G([a_0, \ldots, a_s]) = G([a_0, \ldots, a_{s-1}, 1])$. $G([a_0, \ldots, a_s, 1])$ has one more state than $G([a_0, \ldots, a_s])$, that is the unique of out-degree two, because to each of the a_{s-1} states of $G([a_0, \ldots, a_{s-1}, 1])$ having out-degree 2 we added one new outgoing arc. The number of added arcs is $2 + a_{s-1}$. Therefore the number of states of $G([a_0, \ldots, a_s, 1])$ is n, $a_s = 1$ state of out-degree 2 and $3(n-1-1) - a_{s-1} + 2 + a_{s-1}$ arcs, and the claim is proved. □

Let us give some definitions that will be useful in what follows. The first definition holds both for finite and infinite DAGs.

Definition 2. A DAG having a unique smallest state with respect to the order induced by the arcs is called *semi-normalized*. If it has also a unique greatest state it is called *normalized*. The smallest state is called the *initial state* and the greatest is called the *final state*.

Note that any normalized DAG is also semi-normalized. Note also that any DAG can always be semi-normalized by adding at most one new state and can be normalized by adding at most two new states.

Definition 3. A normalized weighted DAG G has the (h, k)-counting property, or, in short, it counts from h to k if any path from the initial state to the final one has weight in the range $h \ldots k$ and for any i, $h \leqslant i \leqslant k$ there exists just one unique path from the initial state to the final one having weight i. A semi-normalized weighted DAG G' has the (h, k)-counting property, or, in short, it counts from h to k if any non-empty path from the initial state has weight w in the range $h \ldots k$ and for any i, $h \leqslant i \leqslant k$ there exists just one unique path that starts from the initial state and has weight i.

Remark 4. Note that a normalized graph is also semi-normalized and that it can have the counting property as semi-normalized but not as normalized. Indeed, if G' is semi-normalized and it counts from 1 to n, then we can build a

normalized DAG G that counts from 1 to $n+1$ in the following way: add a final state F to G', and, for any state $q \in G'$ add also an arc (q, F) labeled by 1. If G' has out-degree at most l then G has out-degree at most $l+1$.

Suppose, conversely, that G is normalized with final state F with positive integer weights, that it counts from 1 to $n+1$ and that from any state q there is an arc (q, F). Then we can build a DAG G' semi-normalized that counts from 1 to n in the following way: for any arc (q, F) decrease its label by 1, and, if this label is now 0, erase the arc. If G has out-degree at most $l+1$ then G' has out-degree at most l.

It is easy to prove that Sturmian graphs are normalized weighted DAG with positive integer weights and out-degree at most 3.

Indeed, Sturmian graphs turn out to have also the $(1, n)$-counting property for some n, as pointed out by the following theorem, whose proof is given in the next section. The proof is based on the fact that Sturmian graphs are CDAWGs of central Sturmian words and any CDAWG has the counting property, i.e., it is a direct consequence of Proposition 18 and Theorem 19 together with its remark.

Theorem 5. *The Sturmian graph* $G(P/Q = [a_0, \ldots, a_s])$, *with* $\gcd(P, Q) = 1$, *can count from* 1 *up to* $P + Q - 1$.

The reader can check in Fig. 2 that $G(\frac{5}{3} = [1, 1, 1, 1])$ can count from 1 up to 7.

Remark 6. Notice that in Sturmian graphs having final state F, from any state q there is an arc (q, F). Therefore, we can apply the procedure described in Remark 4 and obtain a semi-normalized DAG $G'(P/Q)$ with positive integer weights, of out-degree at most 2, that can count from 0 up to $P + Q - 2$. By extension, these graphs are also called Sturmian graphs.

We are now interested in the "inverse problem".

Problem 7. *Given a positive integer* m, *find a normalized DAG with positive integer weights*, *where each state has out-degree at most* 3, *having minimal number of states and that can count from* 1 *up to* m.

The same problem can be analogously stated for semi-normalized DAGs with out-degree at most 2.

If we do not impose a bound on the out-degree, above problem has the trivial solution given by a graph having just the initial and the final states and m arcs labeled from 1 to m going from the initial to the final state.

If we ask to any state to have out-degree 2 instead of 3, then next proposition shows that above problem has an easy solution. Therefore, the hypothesis on the out-degree 3 makes sense.

Proposition 8. *For any integer* $m \geqslant 2$ *there are at most two* (*up to isomorphism*) *normalized DAGs with positive integer weights*, *where each state has out-degree at most* 2, *that can count from* 1 *up to* m. *They have*, *respectively*, m *and* $m+1$ *states*.

Proof. Let G be a DAG satisfying the hypotheses of the proposition, for a fixed $m \geqslant 2$. For any integer i, $1 \leqslant i \leqslant m$, let $\tilde{G}_i$ be the subgraph of G of all the states and arcs included in any path from the initial state to the final state and weight smaller than or equal to i. In order to simplify the notation in this proof we consider isomorphic graphs to be equal. We claim that if $i < m$, $\tilde{G}_i$ is isomorphic to the graph in Fig. 3 that has $i+1$ states, including the final one.

The proof of the claim is by induction on i.

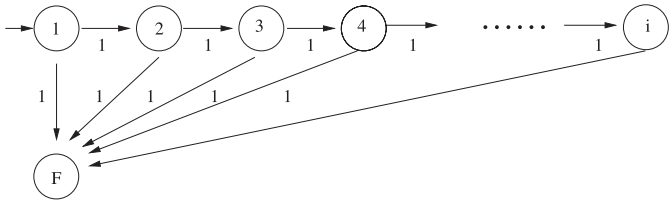

Fig. 3. Graph $\tilde{G}_i$.

If $i = 1$, $\tilde{G}_1$ must be the graph having two states, the initial one and the final one, and one arc of weight 1 and in this case the statement of the claim is true.

Now suppose that the statement of the claim is true for $i-1$ and let us prove it for i. Since in $\tilde{G}_{i-1}$ any state, except state $i-1$, has out-degree 2 and since there is a path in G from state 1 to F of weight i, this path has to go through an arc leaving state $i-1$. There are only two possible cases. Either this arc reaches the final state and has weight 2, or it goes to another state, that we call state $i-1$. In the first case, all states, but the final one, have two outgoing arcs, contradicting the fact that there is a path in G of weight $i+1 \leqslant m$, since by hypothesis $i < m$. Therefore, the only possible case is the second one. Since weights are positive, the arc from $i-2$ to $i-1$ has weight 1 and $i-1$ has an outgoing arc of same weight to F and the claim is proved.

To complete the proof of the proposition, simply notice that $G = \tilde{G}_m$ can be obtained by $\tilde{G}_{m-1}$ by using the same argument of the inductive step of previous claim, where, now, both cases are possible and give rise to exactly two non-isomorphic graphs. □

We do not know whether Problem 7 can be settled in time polynomial in $\log m$ (recall that the number of bits needed to describe m is $O(\log m)$). We do not even know whether the minimal number of states is $O(\log m)$, and, concerning this fact, we make the following conjecture.

Conjecture 9. Given a number m, the minimal number of states of a normalized DAG with positive integer weights, where each state has out-degree at most 3, that can count from 1 up to m, is $O(\log m)$.

For some special classes of numbers above conjecture is a consequence of Theorem 5 and Proposition 1. For instance, if f_s is the sth Fibonacci number, then $G(f_{s+2}/f_{s+1})$ has $s+1$ states, because $(f_{s+2}/f_{s+1}) = [a_0, a_1, \ldots, a_s]$ with, for any i, $0 \leqslant i \leqslant s$, $a_i = 1$. Since it is well known that $f_s = O(\varphi^s)$, where φ is the golden ratio, Conjecture 9 holds.

By using Theorem 5 and Proposition 1, with the same ideas used to prove above conjecture for $m = f_s - 1$ we can prove the following proposition.

Proposition 10. *If there exists an integer K such that for every integer $m \geqslant 1$ there exist integers $1 \leqslant p < q$ with $\gcd(p, q) = 1$ and $p + q = m$ such that every partial quotient in the continued fraction expansion of p/q is $\leqslant K$ then Conjecture 9 is true.*

We conjecture further that the hypothesis of previous proposition always holds.

Conjecture 11. There exists an integer K such that for every integer $m \geqslant 1$ there exist integers $1 \leqslant p < q$ with $\gcd(p, q) = 1$ and $p + q = m$ such that every partial quotient in the continued fraction expansion of p/q is $\leqslant K$.

We do not know if this conjecture is true, but it turns out to be equivalent to the following celebrated conjecture of Zaremba:

Conjecture 12 (*Zaremba [23]*). There exists an integer K such that for every integer $m \geqslant 1$ there exists an integer i, $1 \leqslant i \leqslant m$, $\gcd(i, m) = 1$, such that every partial quotient in the continued fraction expansion of i/m is $\leqslant K$.

In [4] it is reported that Zaremba's conjecture has been verified with constant $K = 5$ up to 3 200 000 by Knuth.

Proposition 13. *Conjecture 11 and Zaremba's conjecture are logically equivalent. The same K can be used in both cases.*

Proof. Suppose Conjecture 11 holds. Given m, let $\alpha := p/q = [0, a_1, a_2, \ldots, a_s]$ have partial quotients bounded by K, and $m = p + q$. Now consider $1/(\alpha + 1) = q/(p + q) = [0, 1, a_1, a_2, \ldots, a_s]$. Letting $i = q$ in Zaremba's conjecture, we have found a fraction with denominator m where the partial quotients are bounded by K.

On the other hand, suppose Zaremba's conjecture holds. Given m, let i be such that $\beta := i/m = [0, a_1, a_2, \ldots, a_s]$, where the partial quotients are bounded by K. $i \geqslant m/2$, then $a_1 = 1$, so consider $(m - i)/i = [0, a_2, \ldots, a_s]$. Now take

$p = m - i$, $q = i$. (In the case where $i = 1$, $m = 2$, take the expansion $[0, 1, 1]$.) If $i < m/2$, then $a_1 > 1$, so consider $i/(m - i) = [0, a_1 - 1, a_2, \ldots, a_s]$. Now take $p = i$, $q = m - i$. In both cases we have found the desired numbers satisfying our conjecture. □

Alternatively we can consider the sum of the partial quotients. Moser made the following conjecture that is weaker than Zaremba's one, in the sense that if Zaremba's conjecture is true then also next conjecture is true.

Conjecture 14 (*Moser*). There exists a constant c such that for all integers $m \geqslant 2$ there exists an integer i, $0 \leqslant i \leqslant m$, $\gcd(i, m) = 1$, such that $\sum_j a_j \leqslant c \log m$, where $i/m = [a_0, a_1, a_2, \ldots, a_s]$.

As above, Moser's conjecture is logically equivalent to a similar conjecture about the sum of p and q.
Indeed, in analogy to Proposition 10 we have the following proposition:

Proposition 15. *If Moser's conjecture is true then Conjecture* 9 *is also true.*

Larcher [15, Corollary 2] proved that Moser's conjecture holds if $\log m$ is replaced by $(\log m)(\log \log m)^2$. Hence we get

Proposition 16. *There exists a constant c such that for all integers $m \geqslant 2$ there exist integers p, q with* $\gcd(p, q) = 1$ *and $p + q = m$ such that $p/q = [a_0, a_1, \ldots, a_s]$ and $\sum_i a_i < c(\log m)(\log \log m)^2$.*

This result implies a weak form of our conjecture.

Corollary 17. *Given a number m, there exists a constant c such that the minimal number of states of a normalized DAG with positive integer weights, where each state has out-degree at most* 3 *and that can count from* 1 *up to m, is smaller than $c(\log m)(\log \log m)^2$.*

3. Indexing, DAWGS and Sturmian graphs

The *directed acyclic word graph* of a word w, $DAWG(w)$, is the smallest finite state automaton that recognizes all the suffixes of the word. If the empty suffix is allowed then the initial state is also final. DAWGs have several applications, such as indexing. Blumer et al. (cf. [1–3]) introduced the *compact directed acyclic word graph* of a word w, $CDAWG(w)$, that is obtained by compacting $DAWG(w)$, i.e., by deleting all states of out-degree 1 and their corresponding edges, joining all consecutive arcs in a path including such states in an unique arc. Thus arcs are labeled by representations of the factors of the word. More precisely, each arc is labeled by the initial position and the length of the factor represented by the arc. For a reference on CDAWGs, see also [6,7,10–12]. We just recall that the underlining DAG of the CDAWG of w, that is $CDAWG(w)$ without labels, is a semi-normalized one. If the last character of w is a symbol never encountered before in w, then the underlining DAG of the CDAWG of w is a normalized one, i.e., it has also a unique final state.

In this section we show how the Sturmian DAG $G([a_0, \ldots, a_s])$ defined in last section coincides with the CDAWG of the word w obtained by the longest word in the pair (A_n, B_n) of directive sequence $[a_0, \ldots, a_s]$ replacing last two letters with a \$ symbol, where the label of each arc is replaced by the length of the factor it represents.

CDAWGs can be used in indexing. Indeed a CDAWG of a word w can give the list of all occurrences of a factor u of w in time proportional to the size of this list. Indeed, by reading the factor u in the CDAWG we reach a position t in it. This position either can be a state or can correspond to a proper prefix of the word representing the label of an arc. Each final occurrence of the required factor u is the length of w plus one minus the word-length of any path from position t to any final state. The reason of this relies on the fact that all possible paths represent all non-empty suffixes of the word w that have u as a prefix.

Since the empty word ε is a prefix of any non-empty suffix of w, the list of its final occurrences in w (with the exception of occurrence $|w| + 1$ that is not considered here as a valid suffix) is the set $\{1, 2, ..., |w|\}$. Moreover, there is an unique suffix, and hence a unique path, from the initial state to the final state having any fixed j, $1 \leqslant |w|$. Therefore, we have proved the following proposition.

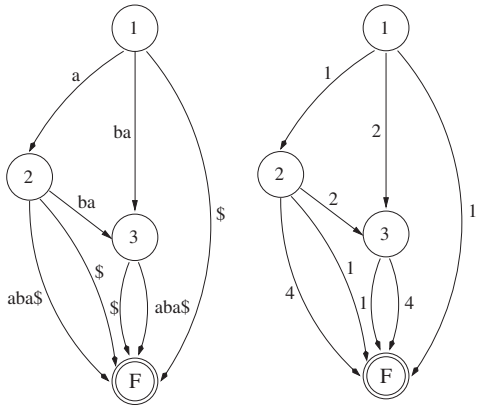

Fig. 4. The CDAWG of word *abaaba*\$ and the one obtained by coding each arc with the length of the factor it represents.

Proposition 18. *Suppose that the last character of w is a symbol never encountered before in w. If we label each arc of CDAWG(w) just with the length of the factor it represents, the obtained weighted DAG can count from* 1 *up to* $|w|$.

Let us go into the details of studying DAWGS of Sturmian words. Consider the s-uple $[a_0, \ldots, a_s]$ and apply the Rauzy rule $R_1 a_0$ times to the pair $(A_0, B_0) = (a, b)$. We obtain a pair (A_{a_0}, B_{a_0}). Let us now apply the rule $R_2 a_1$ times to (A_{a_0}, B_{a_0}), in such a way we obtain $(A_{a_0+a_1}, B_{a_0+a_1})$. Let us continue by alternating the two rules and at the end we obtain the pair $(A_{a_0+\cdots+a_s}, B_{a_0+\cdots+a_s})$ with the directive sequence $[a_0, \ldots, a_s]$. Pick the longest of these two words and replace the last two letters with a \$ symbol. The word obtained is the one whose CDAWG we are interested in. Recall that these words without the final dollar sign are the central Sturmian words and are deeply studied in the literature (cf. [17, Chapter 3]). Concerning directive sequences, we notice that the central Sturmian word associated with $[0, a_1, \ldots, a_s]$ is equal to the one having directive sequence $[a_1, \ldots, a_s]$, up to an exchange of a's and b's.

Let us give an example. Consider the directive sequence [1, 1, 1, 1], the word we obtain is *abaaba*\$. In fact

$$(a, b) \to_{R_1} (a, ab) \to_{R_2} (aba, ab) \to_{R_1} (aba, abaab) \to_{R_2} (abaababa, abaab).$$

Fig. 4 shows *CDAWG*(*abaaba*\$) and, next to it, the DAG obtained by it labeling each arc only with the length of the factor it represents. In order to give a better idea of which factor each arc represents we have labeled each arc not with the initial position and the length of the factor, but by the factor itself. This kind of representation is also used in Fig. 6. Remember that it is not the right representation, because this last representation requires, in the worst case, quadratic space, while the right one requires only linear space.

As we can see, the DAG obtained coincides with the Sturmian graph $G([1, 1, 1, 1])$, i.e., with the Sturmian graph of the same sequence from which we have obtained word *abaaba*\$.

What is surprising in CDAWGs of Sturmian words is that they have a relatively "small" number of nodes, compared to the length of the word itself, as shown by Proposition 1 and next theorem.

Theorem 19. *Let w_n be the word obtained by replacing in the longest word of the pair (A_n, B_n) of directive sequence $[a_0, \ldots, a_s, 1]$ the last two letters with a \$ symbol and CDAWG(w_n) be its CDAWG. Now code each arc with the length of the factor it represents. The obtained DAG always coincides with the Sturmian graph $G([a_0, \ldots, a_s, 1])$.*

The proof of this theorem involves some definitions and preliminary results. Let us begin with the definitions of maximal pair and maximal repeat that is essential for our proof.

Definition 20. A *maximal pair* in a word w is a pair of identical factors u_1, u_2 in w such that the character to the immediate left (or right) of u_1 is different from the character to the immediate left (right) of u_2. That is, extending u_1 and u_2 in either direction would destroy the equality of the two strings. A maximal pair is represented by the triple (i_1, i_2, l), where i_1 and i_2, $i_1 < i_2$, give the starting positions of the two factors and l gives their length.

Definition 21. A maximal repeat in w is a factor u of w that appears in a maximal pair in w.

Proposition 22. *If w is a palindrome, then all its palindromic prefixes are maximal repeats of w.*

Proof. Let $w = uv$ be a palindromic word and u be a palindromic prefix of w. Then $w = uv = \widetilde{uv} = \tilde{v}\tilde{u} = \tilde{v}u$ and u is also a suffix. Therefore $(1, |w| - |u| + 1, |u|)$ is a maximal pair and u is a maximal repeat. □

Before going on, we recall that given a word $w = a_1 a_2 \cdots a_n$, an integer $p \geqslant 1$ is a *period* of w if $a_i = a_{i+p}$, for $i = 1, \ldots, n - p$. The smallest period of w is called *the* period of w. Let us define the following set:

$$Per = \{w : \exists p, q \text{periods of } w \text{ such that } \gcd(p, q) = 1 \text{ and } |w| \geqslant p + q - 2\}.$$

The set of the words of *Per* having length $p + q - 2$ coincides with the set of central Sturmian words, that are the words of the form A_n or B_n without the last two letters for some n and some sequence $[a_0, a_1 \ldots]$ (cf. [17, Theorem 2.2.11]). Any word in *Per* is a palindrome (cf. [17, Corollary 2.2.9]) and every palindromic prefix of a word in *Per* is again in *Per* (cf. [17, Corollary 2.2.10]). The following lemma belongs to the folklore of Sturmian words:

Lemma 23. *The central Sturmian word with directive sequence $[a_0, \ldots, a_s, 1]$ coincides with the word with directive sequence $[a_0, \ldots, a_s + 1]$.*

Theorem 24. *Let w_n be the word obtained by replacing in the longest word of the pair (A_n, B_n) of directive sequence $[a_0, \ldots, a_s, 1]$ the last two letters with a \$ symbol. All the maximal repeats of w_n are the prefixes of w_n of length strictly smaller than $|w_n| - 1$ that belong to Per.*

Proof. First of all we notice that \$ is negligible in the study of the maximal repeats of w_n, i.e., the set of maximal repeats of w_n coincides with one of the words obtained by removing the last character in w_n. Therefore, in what follows in this proof (and in this proof only) we consider that w_n no longer has the symbol \$ as the last character.

We recall that a word u is *right* (*resp.*, *left*) *special* in a word $\mathbf{x}$ over a binary alphabet $\{a, b\}$ if both ua and ub (resp., au and bu) are factors of $\mathbf{x}$. The word u is *bispecial* if it is both left special and right special.

First of all, we claim that w_n is bispecial in any infinite characteristic Sturmian word $\mathbf{x}$ that has w_n as a prefix. By [18, Corollary 2.11] it follows that the set of factors of any infinite Sturmian word $\mathbf{x}$ is closed by reversal and a factor u of $\mathbf{x}$ is right special if and only if it is of the form $u = \tilde{p}$, where p is a prefix of $\mathbf{x}$ (cf. also [8, Proposition 9]) when $\mathbf{x}$ is characteristic. As w_n is a palindrome and it is a prefix of $\mathbf{x}$ then it is right special, i.e., $w_n a$ and $w_n b$ are factors of $\mathbf{x}$. Moreover, since the set of factors of $\mathbf{x}$ is closed by reversal and w_n is a palindrome, also aw_n and bw_n are factors of $\mathbf{x}$. Therefore, the claim is proved.

Now we claim that if u is a maximal repeat and it is a factor of w_n then it is bispecial in any infinite Sturmian word $\mathbf{x}$ that has w_n as a prefix. Indeed, if none of the two occurrences of u is a prefix or a suffix of w_n then u is trivially bispecial in any infinite characteristic Sturmian word $\mathbf{x}$ that has w_n as a prefix. If one of the two occurrences is a prefix (resp., suffix) of w_n then, since w_n is left (resp., right) special in $\mathbf{x}$ then any of its prefixes (resp., suffixes) is left (resp., right) special. The fact that u is also right (resp., left) special comes from the fact that u is a maximal repeat and the two occurrences of it are followed (resp., preceded) by a different character, unless the other occurrence is a suffix (resp., prefix) of w_n. In all cases u is bispecial and also this second claim is proved.

By [8, Proposition 9], u belongs to *Per*. Since u is right special, it is the reversal of a prefix of $\mathbf{x}$. Since u is in *Per*, it is a palindromic word and, so, it is a prefix of $\mathbf{x}$ and therefore it is a prefix of w_n, and the proof is complete. □

Proposition 25. *If w_n is the word obtained by replacing in the longest word of the pair (A_n, B_n) of directive sequence $[a_0, \ldots, a_s, 1]$ the last two letters with a \$ symbol, then CDAWG(w_n) is isomorphic to the labeled graph $\Gamma(w_n)$ whose states are w_n, its prefixes of length is strictly smaller than $|w_n| - 1$ belonging to Per and the empty word ε, that represents the initial state. There is an arc in $\Gamma(w_n)$ from u to u' labeled $v \neq \varepsilon$ if and only if v is such that uv is a suffix of u' and there is no state u''* with $|u''| < |u'|$ *such that uv'' is a suffix of u'', with $v'' \neq \varepsilon$ prefix of v.*

Proof. Theorem 1 of [20] says that a non-empty word u is a maximal repeat in a word w if and only if it is the longest string reaching each internal state of *CDAWG*(w) (non-initial, not representing the whole word). By Theorem 24

we know that all the maximal repeats of w_n are its prefixes of length strictly smaller than $|w_n| - 1$ belonging to *Per*. Therefore, we can conclude that the longest strings reaching each internal state of $CDAWG(w_n)$ are the prefixes of w_n of length strictly smaller than$|w_n| - 1$ belonging to *Per*. The final state is the one corresponding to the whole word w_n and the initial one is the one corresponding to ε. For any state q of $CDAWG(w_n)$, we denote by u_q the longest string label of a path from the initial state to q. Therefore, the mapping that associates each state q to the state u_q is a bijection from the set of states of $CDAWG(w_n)$ to the one of $\Gamma(w_n)$. It remains to prove that this bijection preserves the transition function, i.e., that the two labeled graphs are isomorphic. Let (q, q') be an arc labeled v in $CDAWG(w_n)$, we want to prove that there exists an arc $(u_q, u_{q'})$ in $\Gamma(w_n)$. By hypothesis, we know that $u_q v$ is the label of a path from the initial state to q' in $CDAWG(w_n)$. Since $u_{q'}$ is the longest string label of such a path, we obtain that $|u_{q'}| \geqslant |u_q v|$. Moreover, since by definition of $CDAWG(w_n)$, there exists a word y such that both $u_{q'} y$ and $u_q vy$ are suffixes of w_n, then we obtain that one between $u_{q'}$ and $u_q v$ is a suffix of the other one. By above inequality, this implies that the second one is a suffix of the first one. Obviously, there is no state $u_{q''}$ with $|u_{q''}| < |u_{q'}|$such that $u_q v''$ is a suffix of $u_{q''}$, with $v'' \neq \varepsilon$ prefix of v, because if not so, there would be two outgoing arcs from q with labels beginning with the same letter. This is impossible in CDAWGs. We want to prove, now, that for any arc labeled v from u_q to $u_{q'}$ in $\Gamma(w_n)$, there exists an arc (q, q') in $CDAWG(w_n)$ labeled v. Since by hypothesis both u_q and $u_{q'}$ are prefixes of w_n and $|u_q| < |u_{q'}|$, then u_q is a prefix of $u_{q'}$. Therefore, there exists a path labeled v from q to q' in $CDAWG(w_n)$. We want to prove that this path has length 1, i.e. that it consists merely of an arc. Let us suppose that this path has length greater than 1. This implies that there exists a prefix v'' of v, $v'' \neq \varepsilon$, label of an arc (q, q''), $q'' \neq q'$. Since $u_{q''}$ is the longest string label of a path from the initial state to q'', then $|u_q v''| \leqslant |u_{q''}|$.By construction of $\Gamma(w_n)$ this implies that $u_q v''$ is a prefix and, consequently, a suffix of $u_{q''}$. This contradicts the hypothesis that there is no u'' with $|u''| < |u'|$ such that uv'' is a suffix of u'', with $v'' \neq \varepsilon$ prefix of v. Therefore there exists an arc (q, q') labeled v.

Let w_n be the word obtained by replacing in the longest word of the pair (A_n, B_n) of directive sequence $[a_0, \ldots, a_s, 1]$ the last two letters with a \$ symbol and $CDAWG(w_n)$ be its CDAWG. In what follows we give an inductive characterization of $CDAWG(w_n)$, whose correctness is proved in Proposition 26.

Given the directive sequence [1, 1], the pair (A_{1+1}, B_{1+1}) obtained by applying each of the Rauzy rules R_1 and R_2 once to the pair (a, b), obtaining $(A_{1+1}, B_{1+1}) = (aba, ab)$. Let us pick the longest between the two words and replace last two letters with a \$ symbol. We obtain word $a\$$. The CDAWG of $a\$$ is represented in Fig. 5. It is the same of the CDAWG corresponding to the directive sequence [2] and represents the base case.

Before introducing the inductive step, let us give some new notation. Let w be a word of length $|w| \geqslant 2$, we denote $w^=$ the word obtained by w deleting its last two letters and by $w\$$, the concatenation of w and the \$ symbol. Moreover, given two words $u = \alpha_1 \ldots \alpha_i$ (resp., $u = \alpha_i \ldots \alpha_j$) and $v = \alpha_1 \ldots \alpha_j$ the first one being prefix (resp., suffix) of the second one, we denote by $u^{-1}v$ (resp., vu^{-1}) the factor $\alpha_{i+1} \ldots \alpha_j$ (resp., $\alpha_1 \ldots \alpha_{i-1}$) of v. Finally, given a pair (A_n, B_n), we denote by M_n the longest between the two words A_n and B_n.

Given the CDAWG corresponding to the directive sequence $[a_0, \ldots, a_s]$, $s \geqslant 0$, $\sum_{i=0}^{s} a_i \geqslant 2$, we define the CDAWG corresponding to the directive sequence $[a_0, \ldots, a_s, 1]$ in the following way, depending on the value of s.

(1) If s is even, i.e., we have just applied the rule R_1, then each arc whose label corresponds to the factor $S_n\$ = (M_{n-1}^=)^{-1} B_n^=\$$ is split in an arc labeled S_n from the same outgoing state towards a new state and two arcs from

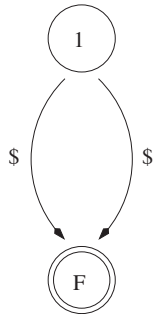

Fig. 5. $CDAWG(a\$)$.

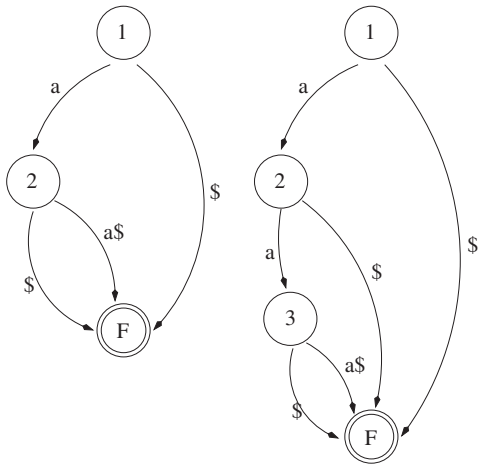

Fig. 6. CDAWGs corresponding to directive sequences [3] and [3, 1].

this new state towards the final one, labeled \$ and $T_n = (B_n^=)^{-1} A_{n+1}^= \$$. Moreover, if $a_s = 1$, then for each state of out-degree 2, except the new one, there is a new outgoing arc labeled T_n towards the final state.

(2) If s is odd, i.e., we have just applied the rule R_2, then each arc whose label corresponds to the factor $S'_n\$ = (M_{n-1}^=)^{-1} A_n^= \$$ is split in an arc labeled S'_n from the same outgoing state towards a new state and two arcs from this new state towards the final one, labeled \$ and $T'_n = (A_n^=)^{-1} B_{n+1}^= \$$. Moreover, if $a_s = 1$, then for each state of out-degree 2, except the new one, there is a new outgoing arc labeled T'_n towards the final state.

In such a way by Lemma 23 we are able to construct a CDAWG corresponding to a generic word w_n having as directive sequence $[a_0, \ldots, a_k]$, $k \geqslant 0$. And the inductive characterization of $CDAWG(w_n)$ is now complete. □

The correctness of previous characterization is proved in next proposition, but first let us give some examples.

Fig. 6 shows CDAWGs corresponding to directive sequences [3] and [3, 1], the first one obtained directly from the base case, and the second one obtained starting from the first one.

Proposition 26. *The previous inductive characterization of $CDAWG(w_n)$ is correct.*

Proof. The proof is by induction on n.

The base case is for $n \leqslant a_0 + 1$. In this case the statement is trivially true because w_n is the *nth* power of a single letter.

Now let us consider the inductive step on n. We suppose the statement is true for w_n and we want to prove it for w_{n+1}. We suppose, without loss of generality by Lemma 23, that w_n corresponds to the directive sequence $[a_0, \ldots, a_s]$ and w_{n+1} to the directive sequence $[a_0, \ldots, a_s, 1]$. Since the statement is true in the base case, we can suppose $s \geqslant 1$.

By Proposition 25 $CDAWG(w_n)$ is isomorphic to a labeled graph whose states are the empty word ε (which represents the initial state) the whole w_n (which represents the final state), and all its prefixes that are central Sturmian words of length smaller than $|w_n| - 1$. By using the same technique of [17, Corollary 2.2.10], all these last words are those obtained by removing the \$ symbol from each w_i, $2 \leqslant i < n$. Let us set, in order to simplify the arguments, that in the following: $w_1 = \$$, so that $w_1 \$^{-1} = \varepsilon$. If we apply one of the two Rauzy rules once, we would obtain $CDAWG(w_{n+1})$ that is isomorphic to a labeled graph whose states are the empty word ε (which represents the initial state), the whole w_{n+1} (which represents the final state), and all its prefixes that are central Sturmian words of length smaller than $|w_{n+1}| - 1$, i.e., those obtained by removing the \$ symbol from each w_i, $2 \leqslant i < n + 1$. These states are exactly the same as those belonging to the graph to which $CDAWG(w_n)$ is isomorphic, except for the final state, that in this case represents w_{n+1} and not w_n, and an additional state that is the one corresponding to w_n without the \$ symbol. Therefore, the inductive characterization is correct with respect to the set of states of the graph.

Now let us examine the arcs. Proposition 25 says also that there is an arc from u to u' labeled $v \neq \varepsilon$ if and only if v is such that uv is a suffix of u' and there is no state u'' with $|u''| < |u'|$ such that uv'' is a suffix of u'', with $v'' \neq \varepsilon$

prefix of v. First of all this implies that each state in $CDAWG(w_{n+1})$, except the final one, has an outgoing arc labeled \$ towards the final one. Second, any arc of $CDAWG(w_n)$ that does not end in the final state of $CDAWG(w_n)$ is an arc in $CDAWG(w_{n+1})$.

Now let us consider the arcs going to the final state of $CDAWG(w_n)$.

If there was an arc $(w_i\$^{-1}, w_n)$ to the final state of $CDAWG(w_n)$ labeled by $v \neq \varepsilon$, then there is an arc $(w_i\$^{-1}, w_n\$^{-1})$ in $CDAWG(w_{n+1})$ labeled by $v\$^{-1}$. Moreover, from the state $w_n\$^{-1}$ of $CDAWG(w_{n+1})$ there is an arc to the corresponding final state labeled by $(w_n\$^{-1})^{-1}w_{n+1}$, and this latter word coincides with the word T_n or T'_n depending on the case of the inductive characterization.

To complete the proof we have to see if there are more arcs than those already described.

Since w_n is a suffix of w_{n+1}, we have that if $|w_i\$^{-1}v| \leqslant |w_n|$ then $w_i v\$^{-1}$ is a suffix of $w_n\$^{-1}$. Therefore, by definition, there are no arcs $(w_i\$^{-1}, w_{n+1})$ in $CDAWG(w_{n+1})$ labeled by v.

Hence, we consider the remaining case $|w_i\$^{-1}v| > |w_n|$. We have to distinguish two different subcases: $a_s > 1$ or $a_s = 1$.

If $a_s > 1$, we have to prove that even in the case $|w_i\$^{-1}v| > |w_n|$ there are no arcs $(w_i\$^{-1}, w_{n+1})$ in $CDAWG(w_{n+1})$ labeled by v. If $w_i\$^{-1}v$, $1 \leqslant i < n$, is a suffix of w_{n+1} then there is a prefix v'' of v such that $w_i\$^{-1}v''$ is suffix of $w_n\$^{-1}$.

Since $a_s > 1$ and $s > 0$ it is not difficult to prove by the Rauzy rules that there exists a word u such that $w_{n-1}{=}u^{a_s-1}x\$$, $w_n = u^{a_s}x\$$, for some x that is a prefix of u. By Lemma 23 we further have that $w_{n+1} = u^{a_s+1}x\$$, for some x prefix of u. Hence a suffix of w_{n+1} longer than w_n is of the form $hu^{a_s}x\$$, with h a proper suffix of u.

Suppose that there is an arc from state $w_i\$^{-1}$ to the final state labeled by v. Since $|w_i\$^{-1}v| > |w_n|$ we have that $w_i v = hu^{a_s}x\$$.

Consider the subcase $|w_i\$^{-1}| < |hu^{a_s-1}x|$. In this case $hu^{a_s-1}x = w_i\$^{-1}v''$, for some $v'' \neq \varepsilon$. We claim that v'' is a non-empty prefix of v and, more precisely, $v{=}v''yx$ with $xy{=}u$. Indeed $w_i\$^{-1}v''yx{=}hu^{a_s-1}xyx{=}hu^{a_s}x{=}w_i\$^{-1}v{=}$. But $w_i\$^{-1}v'' = hu^{a_s-1}x$ is a suffix of state $w_n\$^{-1}$ and therefore, by Proposition 25, an arc from state $w_i\$^{-1}$ to the final state labeled by v cannot exist.

Consider now the remaining subcase $|w_i\$^{-1}| \geqslant |hu^{a_s-1}x| > |w_{n-1}\$^{-1}|$. Since the only state longer than $w_{n-1}\$^{-1}$ is $w_n\$^{-1}$, then $w_i = w_n$, which is impossible since the integer i we have taken in exam is such that $1 \leqslant i < n$. Hence in the case $a_s > 1$ there are no other arcs added. Therefore, in this case the previous inductive characterization is correct.

Let us now examine the case $a_s = 1$. Since $s > 0$ the directive sequence involved is of the form $[a_0, \ldots, a_{s-1}, 1, 1]$.

We can suppose that $s - 1 \geqslant 0$ by the base of the induction.

If $s{-}1{=}0$, the directive sequence we are interested in is $[a_0, 1, 1]$ and $w_{n-1}{=}a^{a_0-1}\$$, $w_n{=}a^{a_0}\$$ and $w_{n+1}{=}a^{a_0}ba^{a_0}\$$. Therefore, for any $i < a_0$ there exists an arc from state $w_i\$^{-1}$ to the final state labeled by $v{=}ba^{a_0}\$$. Indeed $a^i ba^{a_0}\$$ is a suffix of w_{n+1} and has length greater than the length of $w_n\$^{-1}$, and there is no non-empty prefix v'' of v that is such that $a^i v''$ is a suffix of any other state $w_j\$^{-1} = a^j$ because any non-empty prefix of v must contain the letter b. Therefore, we have added a_{s-1} new arcs labeled by $v{=}ba^{a_0}\$$ that coincides with the word T'_n of the inductive characterization. In $CDAWG(w_n) = a^{a_0}\$$ any state $w_i\$^{-1}$, $1 < n$ has two outgoing arcs, one labeled \$ to the final state and another labeled a to $w_{i+1}\$^{-1}$ with the exception of $w_{n-1}\$^{-1} = a^{a_0-1}$ that has one arc labeled by \$ and the other labeled $a\$$ toward the final state. $CDAWG(w_{n+1})$with the arcs we have added, is such that any state $w_i\$^{-1}$, $1 \leqslant i < n$ has three outgoing arcs. Any state of the CDAWG of a word over an alphabet of three letters cannot have more than three outgoing arcs. Moreover, there are exactly two arcs from state $w_n\$^{-1} = a^{a_0}$ to the final state, one labeled \$ and the other T'_n and it is not difficult to see that there are no other arcs that satisfy the conditions in Proposition 25. Therefore, also in this case the previous inductive characterization is correct.

We have considered up to now all cases except the one where

(i) $|w_i\$^{-1}v| > |w_n|$ for any arc we have to add from $w_i\$^{-1}$ toward the final state of $CDAWG(w_{n+1})$ labeled by v,
(ii) $a_s = 1$,
(iii) $s - 1 > 0$.

We deal this case in a similar way to the previous one when $s - 1 = 0$, as follows:

The directive sequence of $w_{n+1}\$^{-1}$ is $[a_0, \ldots, a_{s-1}, 1, 1]$. Let us suppose that $s - 1$ is even; the case $s - 1$ odd is absolutely analogous. Since $s - 1$ is even then we have applied rule $R_2 a_{s-1}$ times to get the $(n - 1)$th pair of standard words (Uba, Xab), where $Uba = (Xab)^{a_{s-1}Y}$ for some words $U, X \in \{a, b\}^*$ and $Y \in \{a, b\}^+$. Notice that this representation is valid only because $s - 1 > 0$. Therefore $|U| > |X|$ and $w_{n-1}\$^{-1} = U$.

Let us follow the directive sequence and apply rule R_1 once to get the nth pair of standard words $(UbaXab, UbaXabUba)$. We have that $w_n\$^{-1} = UbaX$.

Let us follow the directive sequence and apply rule R_2 once to get the $(n+1)$th pair of standard words $(UbaXabUba, UbaXab)$. We have that $w_{n+1}\$^{-1} = UbaXabU = w_n\$^{-1}abU$. By Lemma 23 we know that $w_{n+1}\$^{-1}$ is also equal to $UbaUbaX$ that, in turn, is equal to $Ubaw_n\$^{-1}$.

By the equality $w_{n+1}\$^{-1} = w_n\$^{-1}abU$ and since any $w_i\$^{-1}$, $i < n$ is a suffix of $w_n\$^{-1}$ we have that $w_i\$^{-1}abU\$$ is a suffix of w_{n+1}. If we set $v = abU\$$ then v is a label of a new arc from state $w_i\$^{-1}$ to the final state w_{n+1} only if $|w_i\$^{-1}v| > |w_n|$. But this is equivalent to $|w_i\$^{-1}abU| > |UbaX|$, that is equivalent to requiring that $|w_i\$^{-1}| > |X|$. This last equality is verified for all i for $i = n-1$ down to $n - a_{s-1}$ because $w_i\$^{-1} = (Xab)^{i-n+a_{s-1}+1}$. Hence we add a_{s-1} arcs to the same number of states, each labeled by $v = abU\$$ that coincides with the word T_n of previous inductive characterization.

Let us evaluate how many arcs we have now in $CDAWG(w_{n+1})$.

Any state of the CDAWG of a word over an alphabet of three letters cannot have more than three outgoing arcs. Moreover, there are exactly two arcs from state $wn\$^{-1} = a^{a_0}$ to the final state, one labeled \$ and the other T'_n and it is not difficult to see that there are no other arcs that satisfy the conditions in Proposition 25. Therefore $CDAWG(w_{n+1})$, that has n states, cannot have more than $3(n-1)+2 = 3n-1$ arcs. At this point, by induction, one can easily prove that $CDAWG(w_{n+1})$ have at least $3(n-1)+2$ arcs and therefore no other arc can be added and the previous inductive characterizations is completely proved. □

We are now ready to prove Theorem 19.

Proof. The proof follows trivially by observing that the length of T_n and T'_n in the inductive characterization of $CDAWG(w_n)$ is equal to $l_s + 1$ in the inductive definition of Sturmian graph $G([a_0, \ldots, a_s, 1])$. □

Remark 27. Notice that if $P/Q = [a_0, \ldots, a_s, 1]$, then the length of w_n defined in Theorem 19 is $P + Q$.

4. Infinite graphs

In analogy with finite and infinite words, we can define a convergence of semi-normalized weighted DAWGs.

Let us begin with some definitions. The first one concerns isomorphic (finite and infinite) graphs, that are, roughly speaking, graphs which contain the same number of graph vertices connected in the same way.

Definition 28. Two graphs $G = (V, E)$ and $H = (V', E')$ are said to be isomorphic if there is a bijection f from V to V' such that (u, v) is an edge in G if and only if $(f(u), f(v))$ is an edge in H. Moreover, if G and H are weighted graphs, arcs (u, v) and $(f(u), f(v))$ must be labeled by the same weight.

Now let us define infinite graphs. More precisely, we have the following definition that uses the notion of distance between two states, that is, the size of the smallest path between them, i.e., the minimal number of arcs in a path that connect them.

Definition 29. A sequence $\{G_m\}_{m=0\ldots\infty}$, of semi-normalized weighted DAWGs with positive weights, converges to the infinite weighted DAG G if for any constant $K \geqslant 0$ there exists a number $\hat{m}$ such that for any $m \geqslant \hat{m}$ the restriction of G_m and G to the set X_K of states having distance from the initial state smaller than K are isomorphic with isomorphism f_K. Moreover, if $K_1 > K$, then the restriction of f_{K_1} to X_K coincides with f_K.

Proposition 30. *Let G and H be two different convergence limits of sequence $\{G_m\}_{m=0\ldots\infty}$. Then the graphs G and H are isomorphic.*

Proof. Let x be a state in G and d be its distance from the initial state. Since G is a convergence limit of $\{G_m\}_{m=0\ldots\infty}$, if $K = d + 2$ then there exists a number $\hat{m}$ such that for any $m \geqslant \hat{m}$ the restriction of G_m and G to states having distance from the initial state smaller than K are isomorphic with the same isomorphism f_K. Therefore, for any $m \geqslant \hat{m}$ there exists a state y in G_m such that $f_K(x) = y$. Moreover, by hypothesis we know that H is another convergence

limit of $\{G_m\}_{m=0\ldots\infty}$. Therefore, there exists also a number $\check{m}$ such that for any $m \geqslant \check{m}$ the restriction of G_m and H to states having distance from the initial state smaller than K are isomorphic with the same isomorphism f'_K. Therefore for any $m \geqslant \check{m}$ there exists a state z in H such that ${f'_K}^{-1}(y) = z$. Now define $\bar{m} = \max(\hat{m}, \check{m})$. We know that for any $m \geqslant \bar{m}$ the restriction of G_m and G to states having distance from the initial state smaller than K are isomorphic with isomorphism f_K and that the restriction of G_m and H to states having distance from the initial state smaller than K are isomorphic with the same isomorphism f'_K. State x in G is mapped by $g_K = f_K \circ {f'_K}^{-1}$ on state z in H. Now let us recall that by definition if $K_1 > K$, then the restriction of f_{K_1} to X_K coincides with f_K and the restriction of f'_{K_1} to X_K coincides with f'_K. It is not difficult then to prove that the mapping $g : G \to H$ for any $K > 0$ coincides with g_K is an isomorphism. □

Definition 31. For any irrational number $\alpha > 0$, we define the Sturmian graph $G(\alpha)$ as the unique limit, up to isomorphism, of the sequence of graphs $\{G(P_m/Q_m)\}$, $m = 0 \ldots \infty$, where $\{P_m/Q_m\}_{m=0\ldots\infty}$, is the sequence of convergents to α.

Given a Sturmian graph $G(P_m/Q_m)$ there is a natural way of numbering the states, that is the order in which they have been created with the inductive construction, except the final state that we still call F and that we consider as the first being created. From now on, this will be the numbering of the states of any Sturmian graph.

Lemma 32. *Given an integer $\tilde{m}$, let $[a_0, \ldots, a_{\tilde{m}}]$ be the simple continued fraction expansion of $P_{\tilde{m}}/Q_{\tilde{m}}$. There exists a graph $G_{\tilde{m}}$ such that for any $m \geqslant \tilde{m} + 3$ the restriction of $G(P_m/Q_m)$ to the set $X_{\tilde{m}}$ of the first $\sum_{i=0}^{\tilde{m}} a_i$ states is isomorphic to $G_{\tilde{m}}$.*

Proof. Let $G([a_0, \ldots, a_{\tilde{m}}])$ be the Sturmian graph of directive sequence $[a_0, \ldots, a_{\tilde{m}}]$. By the inductive definition of Sturmian graphs, we know that from it we can obtain the Sturmian graph $G([a_0, \ldots, a_{\tilde{m}}, 1])$. In fact, this new graph is obtained by only splitting each arc of maximal weight in $G([a_0, \ldots, a_{\tilde{m}}])$ in one arc of that weight minus 1 from the same outgoing state to a new state (the same for each arc) and two arcs from this new state towards the final one, one labeled 1 and the other labeled $P_{\tilde{m}-1} + Q_{\tilde{m}-1} + 1$. Moreover, if $a_{\tilde{m}} = 1$, then for each state of out-degree 2, except the new one, one must add a new outgoing arc labeled $P_{\tilde{m}-1} + Q_{\tilde{m}-1} + 1$ towards the final state, with the exception of the new state that has already one such arc. By Proposition 1 this new Sturmian graph $G([a_0, \ldots, a_{\tilde{m}}, 1])$ contains exactly $n = \sum_{i=0}^{\tilde{m}} a_i + 1$ states, including $a_{\tilde{m}}$ of out-degree 2 and $(n - a_{\tilde{m}})$ of out-degree 3, and $3(n-1) - a_{\tilde{m}}$ arcs.

Now let us construct $G([a_0, \ldots, a_{\tilde{m}}, 1, 1]) = G([a_0, \ldots, a_{\tilde{m}}, 2])$. This new graph is obtained by splitting each arc of maximal weight $(P_{\tilde{m}-1} + Q_{\tilde{m}-1} + 1)$ in $G([a_0, \ldots, a_{\tilde{m}}, 1])$ in one arc of weight $(P_{\tilde{m}-1} + Q_{\tilde{m}-1})$ from the same outgoing state to a new state (the same for each arc) and two arcs from this new state towards the final one, one labeled 1 and the other labeled $P_{\tilde{m}} + Q_{\tilde{m}} + 1$. Moreover, since the penultimate partial quotient is 1, then for each state of out-degree 2, except the new one, one must add a new outgoing arc labeled $P_{\tilde{m}} + Q_{\tilde{m}} + 1$ towards the final state, with the exception of the new state that has already one such arc.

Finally, let us construct $G([a_0, \ldots, a_{\tilde{m}}, 1, 1, 1])$ and see what happens to the arcs of weight $(P_{\tilde{m}} + Q_{\tilde{m}} + 1)$. Each of them is split into one arc of weight $(P_{\tilde{m}} + Q_{\tilde{m}})$ from the same outgoing state to a new state (the same for each arc) and two arcs from this new state towards the final one.

Let us examine what happens in all but the final states of $G([a_0, a_1, \ldots, a_{\tilde{m}}, 1]$, i.e., the first $\sum_{i=0}^{\tilde{m}} a_i$ states, if we examine $G([a_0, \ldots, a_{\tilde{m}}, 1, 1, 1])$. Each of them has out-degree 3 and neither their outgoing arcs nor their labels will ever change further in every successive graph in the converging sequence.

No matter how many more times we apply the inductive constructive step, the first $\sum_{i=0}^{\tilde{m}} a_i$ states have out-degree 3 and neither their outgoing arcs nor their labels will ever change further in every successive graph in the converging sequence. Moreover, the reader can see that no matter what the values of $a_{\tilde{m}+1}$, $a_{\tilde{m}+2}$ and $a_{\tilde{m}+3}$ are, i.e., not only in the special case $a_{\tilde{m}+1} = a_{\tilde{m}+2} = a_{\tilde{m}+3} = 1$ that is the one we have examined, $G([a_0, \ldots, a_{\tilde{m}+3}])$ will have the characteristics seen above. Indeed, when $a_{\tilde{m}+1} > 1$ or $a_{\tilde{m}+2} > 1$ the first $\sum_{i=0}^{\tilde{m}} a_i$ states have out-degree 3 and neither their outgoing arcs nor their labels will ever change further, already in $G([a_0, \ldots, a_{\tilde{m}+2}])$. Hence, it would be sufficient to consider the graph $G_{\tilde{m}}$that is the restriction of $G(P_{\tilde{m}+3}/Q_{\tilde{m}+3})$ to the set $X_{\tilde{m}}$ of the first $\sum_{i=0}^{\tilde{m}} a_i$ states to prove the claim. □

The previous lemma says, roughly speaking, that the sequence of Sturmian graphs $\{G(P_m/Q_m)\}_{m=0\ldots\infty}$, where $\{P_m/Q_m\}_{m=0\ldots\infty}$, is the sequence of convergents to α, has a larger and larger common initial part, i.e., this sequence converges, as formally proved in the following proposition.

Proposition 33. *The graph $G(\alpha)$ exists.*

Proof. Let $\{P_m/Q_m\}_{m=0\ldots\infty}$ be the sequence of convergents to α. By Lemma 32, we know that for any positive integer $\tilde{m}$ there exists a graph $G_{\tilde{m}}$ such that the restriction of any graph of the sequence $\{G(P_m/Q_m)\}_{m=\tilde{m}+3\ldots\infty}$ and of $G_{\tilde{m}}$ to the first $\sum_{i=0}^{\tilde{m}} a_i$ states are isomorphic. Therefore, it is sufficient to define $G(\alpha)$ as the graph that for any $\tilde{m}$ coincides in the first $\sum_{i=0}^{\tilde{m}} a_i$ states with the graph $G_{\tilde{m}}$ to prove the claim.

The graph $G(\alpha)$ is exactly the unique limit, up to isomorphism, of the sequence of graphs $\{G(P_m/Q_m)\}$, $m=0\ldots\infty$.

The graph $G(\alpha)$ turns out to be a normalized weighted DAG such that each state, except the final one which has no outgoing arcs, has out-degree 3. It is worth noticing the behavior of the final state F. Its distance from the initial state is 1, and, consequently, it belongs to $G(\alpha)$. Moreover, eventually any state will have an arc toward F of weight 1. Therefore, in $G(\alpha)$ F has no outgoing arcs but infinitely many ingoing arcs, each weighted by 1.

As Sturmian words represent rays geometrically, the final state F in $G(\alpha)$ can be thought as the analogy of the vanishing point in projective geometry. In analogy with the finite case, we call F the final state of the DAWG. □

Before going on, we give the extension of Definition 3.

Definition 34. An infinite normalized weighted DAG G has the (h, ∞)-counting property, or, said differently, it counts from h to ∞, if any path from the initial state to the final one has weight in the range $h\ldots\infty$ and for any i, $i \geqslant h$, there exists just one unique path from the initial state to the final one having weight i. An infinite semi-normalized weighted DAG G' has the (h, ∞)-counting property, or, in short, it counts from h to ∞ if any non-empty path from the initial state has weight in the range $h\ldots\infty$ and for any i, $i \geqslant h$, there exists just one unique path that starts from the initial state and has weight i.

Theorem 35. *For any positive irrational α, $G(\alpha)$ can count from 1 up to infinity.*

Proof. Let i be any number in the range $1\ldots\infty$. We have to prove that there exists a path from the initial state of $G(\alpha)$ to the final one whose weight is i and that it is unique.

By definition, we know that the Sturmian graph $G(\alpha)$ is the unique limit, up to isomorphism, of the sequence of graphs $\{G(P_m/Q_m)\}_{m=0\ldots\infty}$, where $\{P_m/Q_m\}_{m=0\ldots\infty}$ is the sequence of convergents to α. Furthermore, by Theorem 5 we have that every graph $G(P_m/Q_m)$ has the $(1, P_m + Q_m - 1)$-counting property.

For any i in the range $1\ldots\infty$, let $\tilde{m}$ be an integer such that $G(P_{\tilde{m}}/Q_{\tilde{m}})$ contains a unique path weighted i from its initial state to its final state and that the greatest arc label l in $G(P_{\tilde{m}}/Q_{\tilde{m}})$ is greater than $i+1$. In such a way at the next step this label is transformed to $l - 1 > i$ and represents the weight of an arc towards a new state that has two outgoing arcs, one labeled 1 and the other one labeled $l' > l$. Hence, for any $m > \tilde{m}$ the path labeled i in $G(P_m/Q_m)$ goes only through the first $\sum_{i=0}^{\tilde{m}} a_i$ states. By Lemma 32 this implies that the path weighted i in $G(P_{\tilde{m}+3}/Q_{\tilde{m}+3})$ remains fixed and it is the same also in $G(\alpha)$. Therefore, $G(\alpha)$ contains a path p from its initial state to its final state weighted i.

What we have to prove now is that this path is unique. Let us suppose, in order to obtain a contradiction, that there is another such path; call it p_1. Let K_1 be a constant such that the set X_{K_1} of Definition 29 contains all states in this new path. Let $\bar{K} = \max(K, K_1)$. By definition, there would exist an integer $\bar{m}$ such that $G(P_{\bar{m}}/Q_{\bar{m}})$ contains two paths labeled i from its initial state to its final state, which is a contradiction, because of the counting property of $G(P_{\bar{m}}/Q_{\bar{m}})$.

If $\alpha = (\sqrt{5}+1)/2$, that is, the golden ratio, we call $G(\alpha)$ the golden graph.

Since any state reaches the vanishing state with an arc of weight 1, we can eliminate the vanishing state and the arcs going to it and the new graph $G'(\alpha)$ can count from 0 to infinity, supposing each state "terminal". Indeed $G'(\alpha)$ is the limit graph of the sequence $G'(P_n/Q_n)$, $n=0\ldots\infty$, where P_n/Q_n, $n=0\ldots\infty$, is the sequence of convergents to α and $G'(P_n/Q_n)$ is defined in Remark 6 (cf. Fig. 7). □

Proposition 36. *For every $n \geqslant 1$ the golden graph uses $\mathrm{O}(\log_\varphi n)$ states to count from one to n.*

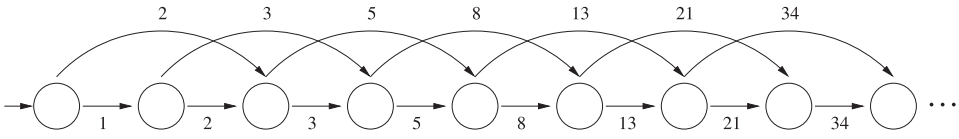

Fig. 7. The semi-normalized graph obtained by eliminating the vanishing state and the arcs going to it in the golden graph $G(\sqrt{5}+1/2)$.

Proof. Let f_s be the sth Fibonacci number. Then $G(f_{s+2}/f_{s+1})$ has $s+1$ states, because $f_{s+2}/f_{s+1} = [a_0, a_1, \ldots, a_s]$ with, for any j, $0 \leqslant j \leqslant s$, $a_j = 1$. By Theorem 5, $G(f_{s+2}/f_{s+1})$ can count from 1 up to $n = f_{s+2} + f_{s+1} - 1$.

For any i in the range $1 \ldots n$, let $\tilde{s}$ be an integer such that $G(f_{\tilde{s}+2}/f_{\tilde{s}+1})$ contains a unique path weighted i from its initial state to its final state and that the greatest arc label $l = f_{\tilde{s}+1} + f_{\tilde{s}} + 1$ in $G(f_{\tilde{s}+2}/f_{\tilde{s}+1})$ is greater than $i+1$. In such a way at the next step this label is transformed in $l - 1 = f_{\tilde{s}+1} + f_{\tilde{s}} > i$ and represents the weight of an arc towards a new state that has two outgoing arcs, one labeled 1 and the other one labeled $f_{\tilde{s}+2} + f_{\tilde{s}+1} + 1 > l$. Hence, for any $s > \tilde{s}$ the path labeled i in $G(f_{\tilde{s}+2}/f_{\tilde{s}+1})$ goes through the first $\sum_{i=0}^{\tilde{s}} a_i = \tilde{s} + 1$ states only. By Lemma 32 and reasoning similar to that in the proof of Theorem 35, we obtain that the golden graph uses $O(\log_\varphi n)$ states to count from one to n. □

Remark 37. In the Fibonacci numeration system, every number has only one binary representation that does not contain two consecutive 0s. There exists a bijection between the set of these representations and the set of paths in the golden graph. This bijection associates the representation of natural number n with the path weighted by n in the golden graph.

Definition 38. An infinite graph having a countable number of states *has the local property* or *is local* with constant k, if there exists a way of numbering states such that for any state i all outgoing arcs (i, j) are such that $i - k \leqslant j \leqslant i + k$.

The next proposition connects the structure of a Sturmian graph with the continued fraction expansion of α.

Proposition 39. *$G'(\alpha)$ is local if and only if α has bounded partial quotients in its continued fraction expansion.*

Proof. Suppose that $\alpha = [a_0, a_1, \ldots, a_s, \ldots]$. First of all notice (more formally by induction) that all arcs in Sturmian graphs are of the form (i, j) with $j > i$ except the ones that point to the final state. Therefore, to prove the locality of G' we have to prove that there exists a constant k such that for any state i all outgoing arcs (i, j) are such that $j \leqslant i + k$. Notice also that any state in Sturmian graph has out-degree at least 2 except the final state.

Let us consider the graph $G([a_0, a_1, \ldots, a_s, 1, 1] = [a_0, a_1, \ldots, a_s, 2])$. No matter what the value of a_{s+1} is, the above graph is one of the sequence of graphs that has $G(\alpha)$ as a limit. Since the penultimate partial quotient is 1 then in the inductive constructive step, for each state of out-degree 2 a new outgoing arc was added labeled $l_s + 1$ towards the final state, with the exception of the last created state that has already one such arc. The label $l_s + 1$ is the latest and the largest label, because once a label is created it can only decrease in inductive steps and because the sequence of l_s, $s = 0, 1, \ldots$ is strictly increasing. Therefore, in the next inductive constructive step all states that in $G([a_0, a_1, \ldots, a_s, 1]$ have out-degree 2 points with a new arc toward the latest created state (with label l_s) and neither the outgoing arcs nor their labels ever change further in every succeeding graph in the converging sequence.

The situation above holds for any $s \geqslant 0$. By Proposition 1 we have that the number of states that in $G([a_0, a_1, \ldots, a_s, 1]$ have out-degree 2 is a_s. Let i be one of such states. The fact above, in turns, implies that for any arc (i, j), $j \neq F$ in the limit graph is such that $j - 1 \leqslant a_s$. If $\alpha = [a_0, a_1, \ldots, a_s, \ldots]$ has partial quotients bounded by K, then $G'(\alpha)$ is local with constant $K + 1$.

Conversely, by same argument used above, it is possible to prove that there are states with in-degree a_s. If α has unbounded partial quotients then $G'(\alpha)$ has unbounded in-degree. It is not difficult to prove that graphs having unbounded in-degree cannot have the local property. □

Acknowledgments

We want to thank both the anonymous Reviewers, and especially the one that pointed out a flow in the statement of Proposition 25.

References

[1] A. Blumer, J. Blumer, D. Haussler, A. Ehrenfeucht, M.T. Chen, J. Seiferas, The smallest automaton recognizing the subwords of a text, Theoret. Comput. Sci. 40 (1) (1985) 31–55.

[2] A. Blumer, J. Blumer, D. Haussler, R. McConnell, A. Ehrenfeucht, Complete inverted files for efficient text retrieval and analysis, J. ACM 34 (3) (1987) 578–595.

[3] A. Blumer, D. Haussler, A. Ehrenfeucht, Average sizes of suffix trees and dawgs, Discrete Appl. Math. 24 (1989) 37–45.

[4] I. Borosh, H. Niederreiter, Optimal multipliers for pseudo-random number generation by the linear congruential method, BIT 23 (1983) 65–74.

[5] C. Brezinski, History of continued fractions and Padé approximants, Springer Series in Computational Mathematics, vol. 12, Springer, Berlin, 1991.

[6] M. Crochemore, Reducing space for index implementation, Theoret. Comput. Sci. 292 (1) (2003) 185–197.

[7] M. Crochemore, R. Vérin, Direct construction of compact directed acyclic word graphs, CPM97, in: A. Apostolico, J. Hein (Eds.), Lecture Notes in Computer Science, vol. 1264, Springer, Berlin, 1997, pp. 116–129.

[8] A. de Luca, F. Mignosi, Some combinatorial properties of Sturmian words, Theoret. Comput. Sci. 136 (1994) 361–385.

[9] G.H. Hardy, E.M. Wright, An Introduction to the Theory of Numbers, fifth ed., Oxford University Press, Oxford, 1989.

[10] J. Holub, M. Crochemore, On the implementation of compact DAWG's, in: Proceedings of the Seventh Conference on Implementation and Application of Automata, University of Tours, Tours, France, July 2002, Lecture Notes in Computer Science, vol. 2608, Springer, Berlin, 2003, pp. 289–294.

[11] S. Inenaga, H. Hoshino, A. Shinohara, M. Takeda, S. Arikawa, G. Mauri, G. Pavesi, On-line construction of compact directed acyclic word graphs, in: Proceedings of CPM 2001, Lecture Notes in Computer Science, vol. 2089, Springer, Berlin, 2001, pp. 169–180.

[12] S. Inenaga, H. Hoshino, A. Shinohara, M. Takeda, S. Arikawa, G. Mauri, G. Pavesi, On-line construction of compact directed acyclic word graphs, Discrete Appl. Math. 146 (2) (2005) 156–179.

[13] R. Klette, A. Rosenfeld, Digital straightness—a review, Discrete Appl. Math. 139 (1–3) (2004) 197–230.

[15] G. Larcher, On the distribution of sequences connected with good lattice points, Monatshefte Math. 101 (1986) 135–150.

[17] M. Lothaire, Algebraic combinatorics on words, Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge, 2002.

[18] F. Mignosi, Infinite words with linear subword complexity, Theoret. Comput. Sci. 65 (1989) 221–242.

[19] O. Perron, Die Lehre von den Kettenbrüchen, B.G. Teubner, Stuttgart, 1954.

[20] M. Raffinot, On maximal repeats in strings, Inform. Process. Lett. 83 (2001) 165–169.

[21] G. Rauzy, Mots infinis en arithmétique, in: M. Nivat, D. Perrin (Eds.), Automata on Infinite Words, Lecture Notes in Computer Science, vol. 192, Springer, Berlin, 1985, pp. 165–171.

[22] J. Shallit, Real numbers with bounded partial quotients, Enseignement Math. 38 (1992) 151–187.

[23] S.K. Zaremba, La méthode de "bons treillis" pour le calcul des intégrales multiples, in: S.K. Zaremba (Ed.), Applications of Number Theory to Numerical Analysis, Academic Press, New York, 1972, pp. 39–119.