Counting Minimal Semi-Sturmian Words

By: F. Blanchet-Sadri, Sean Simmons

Blanchet-Sadri, F., Simmons, S. (2013). Counting Minimal Semi-Sturmian Words. *Discrete Applied Mathematics, 161(18),* 2851-2861.doi: 10.1016/j.dam.2013.07.008

Abstract:

A finite *Sturmian* word w is a balanced word over the binary alphabet {a,b}, that is, for all subwords u andv of w of equal length, $||u|_a-|v|_a|\leq 1$, where $|u|_a$ and $|v|_a$ denote the number of occurrences of the lettera in u and v, respectively. There are several other characterizations, some leading to efficient algorithms for testing whether a finite word is Sturmian. These algorithms find important applications in areas such as pattern recognition, image processing, and computer graphics. Recently, Blanchet-Sadri and Lensmire considered finite *semi-Sturmian* words of minimal length and provided an algorithm for generating all of them using techniques from graph theory. In this paper, we exploit their approach in order to count the number of minimal semi-Sturmian words. We also present some other results that come from applying this graph theoretical framework to subword complexity.

Keywords: Combinatorics on words | Graph theory | Subword complexity | Semi-Sturmian words | Euler's totient function

Article:

1. Introduction

An infinite word w is an infinite sequence of letters from a finite alphabet. Any finite block of consecutive letters of w is a *factor* or *subword* of w. The word w is *Sturmian* if, for all non-negative integers n, there are exactly n+1 distinct subwords of w of length n. In other words, the *subword complexity* $p_w(n)$ of w, which counts the number of distinct subwords of length n of w, is equal to n+1. The fact $p_w(1)=1+1=2$implies that w is constructed from two distinct letters of the alphabet. Without loss of generality, we call these a and b. The well-known

Fibonacci word

abaababaabaababaababaabaababaabaab…

is Sturmian. It is defined by $F_{n+2}=F_{n+1}F_n$, where F_0=a and F_1=ab.

Sturmian words have been widely studied. Morse and Hedlund introduced the term “Sturmian trajectories” and did a first comprehensive study in 1940 in relation to symbolic dynamics [13]. Chapter 2 of Lothaire’s book “Algebraic Combinatorics on Words” provides a systematic exposition of Sturmian words, their numerous properties, and equivalent definitions [11]. Sturmian words appear in the literature under various names: rotation sequences, cutting sequences, Christoffel words, Beatty sequences, characteristic words, balanced words, nonhomogeneous spectra, billiard trajectories, etc. Application areas include linear filters [10], routing in networks [1], pattern recognition [5], image processing and computer graphics [6]. For example, counting the number of distinct digitized straight lines corresponds to counting the number of subwords of a given length in Sturmian words. A formula was conjectured by Dulucq and Gouyou-Beauchamps in [8] and later proved by Mignosi in [12].

A finite word w is Sturmian if it is a subword of an infinite Sturmian word. Linear-time algorithms have been provided for recognizing finite Sturmian words (see for example, Boshernitzan and Fraenkel [4] and de Luca and De Luca [7]). Berstel and Pocchiola also provided a linear probabilistic algorithm for generating randomly finite Sturmian words [2].

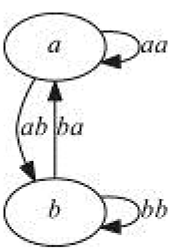
Now, a finite word w is *semi-Sturmian of order N* if $p_w(n)=n+1$ for $n=1,\ldots,N$. Note that the terminology *Sturmian of order N* was previously used by Blanchet-Sadri and Lensmire in [3] for such word, but we decided to adopt the terminology “semi-Sturmian” here to avoid confusion with finite Sturmian words. Not all semi-Sturmian words of order N are Sturmian, for instance, aabb is a semi-Sturmian word of order 2 but it is not a subword of any infinite Sturmian word. However every finite Sturmian word is semi-Sturmian of order Nfor some N. A semi-Sturmian word of order N is *minimal* if it has minimal length among all semi-Sturmian words of order N. Equivalently, it is minimal if it has length 2N. In [3], Blanchet-Sadri and Lensmire described an algorithm that generates all minimal semi-Sturmian words of each order $N\geq3$. Earlier in [14], it had been shown that the minimal length of a word w such that $p_w(n)=F_{n+2}$ for all n,$1\leq n\leq N$, isF_N+F_{N+2}, where $(F_n)_{n\geq1}$ is the Fibonacci sequence and N is a positive integer, and an algorithm had been given for generating such minimal words of each order $N\geq1$.
In this paper, our main result is to count the number of minimal semi-Sturmian words of order N for every integer N greater than 1. We show that this number is connected to Euler’s totient function ϕ from number theory, where the totient $\phi(n)$ of a positive integer n is the number of positive integers less than or equal ton that are coprime to n.

The contents of our paper is as follows. In Section 2, we review some basics on semi-Sturmian graphs and some graphs corresponding to given sets of words of a fixed length. We also recall conditions for the existence of Eulerian paths in graphs. In Section 3, we consider minimal words with subword complexity n+1, that is, minimal semi-Sturmian words. We count all minimal semi-Sturmian words of order N using a graph theoretical approach based on the above mentioned algorithm that generates all such words. We show that any graph produced by this algorithm belongs to one of three families of semi-Sturmian graphs that end up playing an important role in the counting. In Section 4, we use our techniques to extend our result further to include a lower bound on the number of minimal words with subword complexity n+k−1, where k is the alphabet size.

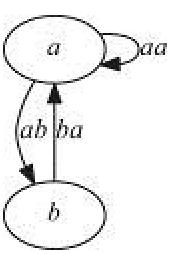
2. Preliminaries on graphs

We recall some graph theoretical concepts that will be useful. All graphs in this paper are assumed to be directed. The reader is referred to [9] for more information.

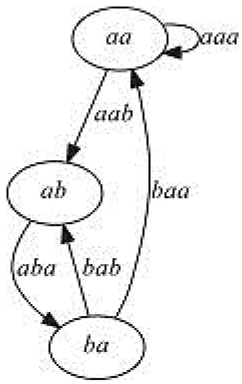
A graph G is said to be *semi-Sturmian of order* n if G has n vertices, n+1 edges, and contains an Eulerian path. The graph G is also said to be *semi-Sturmian* if it is semi-Sturmian of some order n. Moreover for any graph G=(V,E), we denote by L(G) its *line graph* which is the graph $G'=(V',E')$ where $V'=E$, and for all $v'_1, v'_2 \in V'$, $(v'_1, v'_2) \in E'$ if $v'_1 = (v_1, v_2)$ and $v'_2 = (v_2, v_3)$ for some $v_1,v_2,v_3 \in V$. Fig. 1(c) gives the line graph of Fig. 1(b).



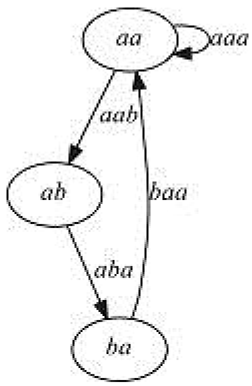
(a) G_2.



(b) G'_2.



(c) G_3.



(d) G'_3.

Fig. 1 Running Algorithm 1: this sequence of graphs produces, among others, the semi-Sturmian word aaabaa of order 3.

Now, let S be a set of words of length n. Combining ideas from de Bruijn and Rauzy graphs, the graph $G_S=(V,E)$, defined in [3], is as follows: V is the set of all factors of length $n-1$ of words in S, and E consists of all edges (x,x') so that there exists a word $y\in S$ with x as a prefix and x' as a suffix. The edge (x,x')can be identified (or labelled) with the word y. See Fig. 1(a) for an example where S={aa,ab,ba,bb}.

It is worth noting that every path in a graph of the form G_S corresponds to a word. More specifically, let $x_0,\dots,x_m$ be a path in G_S where $x_0,\dots,x_m$ are vertices. Then this path corresponds to the word w where$w[0\dots n-1]=x_0,w[1\dots n)=x_1,\dots,w[|w|-n+1\dots|w|)=x_m$ (here $m=|w|-n+1$). Moreover if pand q are different paths, then they correspond to different words. A similar construction allows us to view every path in the graphs $L(G_S),L(L(G_S)),\dots$ as a word. We say that if p is a path in some subgraph of$L(\cdots(L(G_S))\cdots)$, then p corresponds to a word.

We end this section with a well-known result on the existence of Eulerian paths. The notation $\mathtt{ideg}(v)$ refers to the indegree of vertex v and $\mathtt{odeg}(v)$ to its outdegree.

Lemma 1.

Let G *be a graph, and let* x*and* y*be vertices in* G.

- *If* x=y, *then there is an Eulerian path from* x*to* y*if and only if* G*is strongly connected and* $\mathtt{ideg}(v) = \mathtt{odeg}(v)$*for all* v.

- *If* x≠y, *then there is an Eulerian path from* x*to* y*if and only if* G*is weakly connected,* $\mathtt{ideg}(x) = \mathtt{odeg}(x) - 1, \mathtt{ideg}(y) = \mathtt{odeg}(y) + 1$, *and* $\mathtt{ideg}(v) = \mathtt{odeg}(v)$*for all other vertices* v.

3. Our main result

Our main goal is to prove the following result. Recall that the *Euler totient* ϕ(n) of a positive integer n is the number of positive integers less than or equal to n that are coprime to n. For example, ϕ(9)=6 since 1,2,4,5,7 and 8 are coprime to 9.

Theorem 1.

For N≥2, *the number of minimal semi-Sturmian words of order* N,S(N), *satisfies*

$$S(N) = 6 + 4\left(\sum_{n=3}^{N} \phi(n)\right) + (N-1)\phi(N+1)$$

where ϕ(n)*is the Euler totient function.*

We begin by recalling an algorithm due to Blanchet-Sadri and Lensmire [3], which we illustrate in Fig. 1.

Algorithm 1 Constructing a Minimal Semi-Sturmian Word of Order $N \geq 3$

Set $G_2 = G_{\{aa,ab,ba,bb\}}$
Create G'_2 by deleting an edge from G_2
for $i = 3$ to N **do**
  Build $G_i = L(G'_{i-1})$
  if G_i has $i + 1$ edges **then**
    Set $G'_i = G_i$
  else
    Create G'_i from G_i by deleting an edge from G_i so that G'_i has an Eulerian path
Find an Eulerian path p in G'_N
Find the word w associated to p
return w

Theorem 2 [3]**.**

Algorithm 1 *outputs a word* w*if and only if* w*is a minimal semi-Sturmian word of order* N.

To simplify our exposition, we introduce the following notation: if G=(V,E) is a semi-Sturmian graph, then we write $G \Rightarrow G'$ if one of the following holds:

1. L(G) has |E|+1 edges and G'=L(G);

2. L(G) has |E|+2 edges and G' is formed by removing an edge in L(G) so that G' has an Eulerian path.

In counting minimal semi-Sturmian words, there are some important families of semi-Sturmian graphs to consider.

Definition 1.

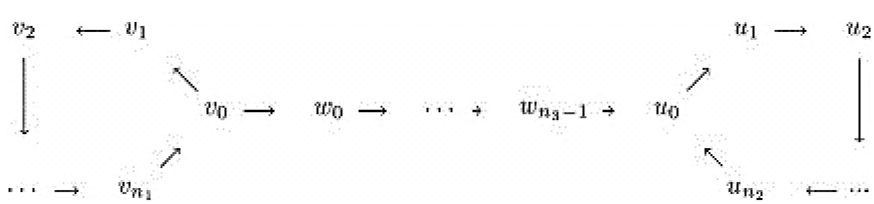
1. The graph $A_{n1,n2;n3}$ with vertices $v_0,\ldots,v_{n1},u_0,\ldots,u_{n2},w_0,\ldots,w_{n3-1}$ is composed of the cycles
$v_0,\ldots,v_{n1},v_0,$

$u_0,\ldots,u_{n2},u_0,$

and the path

$v_0,w_0,\ldots,w_{n3-1},u_0$

(see Fig. 2).

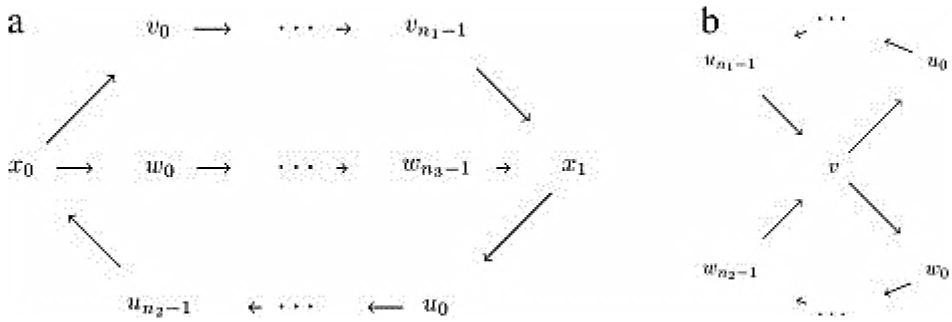


2. The graph $B_{n1,n2;n3}$ with vertices $x_0,x_1,v_0,\ldots,v_{n1-1},w_0,\ldots,w_{n3-1},u_0,\ldots,u_{n2-1}$ is composed of the paths
$x_0,v_0,\ldots,v_{n1-1},x_1,$

$x_0,w_0,\ldots,w_{n3-1},x_1,$

$x_1,u_0,\ldots,u_{n2-1},x_0$

(see Fig. 3(a)).



3.The graph $C_{n1,n2}$ with vertices $v,u_0,\ldots,u_{n1-1},w_0,\ldots,w_{n2-1}$ is composed of the cycles
$v,u_0,\ldots,u_{n1-1},v$

$v,w_0,\ldots,w_{n2-1},v$

(see Fig. 3(b)). This can be thought of as a cycle with n_1+1 edges attached at one point with a cycle with n_2+1 edges.

We first prove a few simple properties about the above graphs, since they are the cornerstone for the proof of our main result.

Lemma 2.

Let G *be a semi-Sturmian graph with more than one vertex.*

1.*If* G*is isomorphic to* $A_{n1,n2;n3}$*then there is a unique graph* H*so that* $G\Rightarrow H$. *Moreover,* H*is isomorphic to* $A_{n1,n2;n3+1}$.

2.*If* G*is isomorphic to* $B_{n1,n2;n3}$*then there is a unique graph* H*so that* $G\Rightarrow H$. *If* $n_2=0$*then* H*is isomorphic to* $C_{n1+1,n3+1}\cong C_{n3+1,n1+1}$. *Otherwise it is isomorphic to* $B_{n1+1,n2-1;n3+1}$.

3.*If* G*is isomorphic to* $C_{n1,n2},n_1\leq n_2$, *then there exist exactly four distinct graphs,* H_0,H_1,H_2,H_3*so that* $G\Rightarrow H_i$*for each* i. *Moreover, up to ordering we get that*

(a)$H_0\cong A_{n1,n2;0}$*;*

(b)$H_1\cong A_{n2,n1;0}$*;*

(c)$H_2\cong B_{n1+1,n2-1;0}$*;*

(d)*If* $n_1 \neq 0$ *then* $H_3 \cong B_{n2+1,n1-1;0}$*; otherwise* $H_3 \cong C_{0,n2+1}$.

Proof.

Assume that $G \Rightarrow G'$,$G=(V,E)$. For Statement (1), it suffices to consider $G=A_{n1,n2;n3}$.
Then $V=\{v_0,\dots,v_{n1},u_0,\dots,u_{n2},w_0,\dots,w_{n3-1}\}$. Let $l_i=(v_i,v_{i+1})$ for $0 \le i<n_1$,$l_{n1}=(v_{n1},v_0)$. Then let $e_i=(u_i,u_{i+1})$ for $0 \le i<n_2$,$e_{n2}=(u_{n2},u_0)$. Finally let $f_0=(v_0,w_0)$,$f_{n3}=(w_{n3-1},u_0)$, and $f_{i+1}=(w_i,w_{i+1})$ for $0 \le i<n_3-1$. It follows that L(G) is the graph whose vertices are $l_0,\dots,l_{n1},e_0,\dots,e_{n2},f_0,\dots,f_{n3}$ consisting of the cycles $l_0,\dots,l_{n1},l_0,e_0,\dots,e_{n2},e_0$, and the path $l_{n1},f_0,\dots,f_{n3},e_0$. Thus L(G) has |E|+1 edges, so $G \Rightarrow G'$ if and only if $G'=L(G)$.
Moreover, $G'=L(G) \cong A_{n1,n2;n3+1}$.

For Statement (2), it suffices to consider $G=B_{n1,n2;n3}$. The result then follows by a similar argument to the above,
since $G \Rightarrow L(G)$ where $L(G) \cong C_{n1+1,n3+1}$ if $n_2=0$ and $L(G) \cong B_{n1+1,n2-1;n3+1}$ otherwise.

For Statement (3), it suffices to consider $G=C_{n1,n2}$, where $n_1 \le n_2$. Assume $n_1=0$, the other case being similar. Then $C_{n1,n2}$ is isomorphic to the graph with $V=\{w_0,\dots,w_{n2}\}$ and with $E=\{l_0,\dots,l_{n2},e\}$where $l_i=(w_i,w_{i+1})$ for $0 \le i<n_2$,$l_{n2}=(w_{n2},w_0)$, and $e=(w_0,w_0)$. Thus L(G) is the graph with vertices $\{l_0,\dots,l_{n2},e\}$ and
edges $e'=(e,e), l'_i=(l_i,l_{i+1})$ for $0 \le i<n_2, l'_{n_2}=(l_{n_2},l_0), f_0=(l_{n_2},e)$ and$f_1=(e,l_0)$. Let H_0 be the subgraph we get by removing f_0,H_1 the subgraph we get by removing f_1,$H_2$the subgraph we get by removing e', and H_3 the subgraph we get by removing l'_{n_2}. It is easy to see thatH_0,H_1,H_2 and H_3 are all distinct, and
that $H_0 \cong A_{n1,n2;0}$,$H_1 \cong A_{n2,n1;0}$,$H_2 \cong B_{n1+1,n2-1;0}$ and$H_3 \cong C_{0,n2+1}$. Moreover, by definition, $G \Rightarrow H_i$ for all i.

Therefore all that remains is to show that these are the only such graphs. To see this assume that $G \Rightarrow G'$. This implies G' is produced from L(G) be deleting one edge. If we delete l'_{n_2}, e', f_0 or f_1 then it is one of theH_i's. Therefore G' must be produced by deleting l'_i from L(G) for some i,$0 \le i<n_2$. If we remove l'_{n_2-1}then l_{n2} has two outgoing edges and no incoming ones, so G' cannot contain an Eulerian path, which is a contradiction. Therefore $0 \le i<n_2-1$. Removing l'_i means that l_{i+1} has one outgoing edge and no incoming edge, so any Eulerian path in G' must start at l_{i+1} by Lemma 1. However, since l_{n2} has two outgoing edges and one incoming edge, any Eulerian path must also start at l_{n2}. This is a contradiction. Therefore G' must be one of $H_0,\dots,H_3$, as we wanted. □

Lemma 3.

Let $G'_2,\dots,G'_N$ *be a sequence produced by Algorithm*

1. *Then there exist* n_1, n_2, n_3 *so that* G'_N *is isomorphic to one of* $C_{n1,n2}, B_{n1,n2;n3}$ *or* $A_{n1,n2;n3}$.

Proof.

We proceed by induction. It is easy to check that this holds for $N = 2$. Therefore consider $N > 2$. Then, for input $N-1$, Algorithm 1 can produce the sequence $G'_2, \ldots, G'_{N-1}$. Thus by induction G'_{N-1} is isomorphic to one of $C_{n1,n2}, B_{n1,n2;n3}$ or $A_{n1,n2;n3}$. Since $G'_{N-1} \Rightarrow G'_N$, we get by Lemma 2 that G'_N is isomorphic to one of $C_{n'_1,n'_2}$, $B_{n'_1,n'_2;n'_3}$ or $A_{n'_1,n'_2;n'_3}$ for some n'_1, n'_2, n'_3. □

It is the above lemma that makes $C_{n1,n2}, B_{n1,n2;n3}$ and $A_{n1,n2;n3}$ important for our purposes. Each of these families of graphs has numerous important quantities associated with it, introduced and denoted as follows, where $N \geq 2$:

$\mathbb{G}(N)$	$=$	$\{(G'_2, \ldots, G'_N) : G'_2, \ldots, G'_N$ are produced by Algorithm 1$\}$
$\mathbb{A}_{n_1,n_2;n_3}$	$=$	$\{(G'_2, \ldots, G'_{n_1+n_2+n_3+2}) \in \mathbb{G}(n_1+n_2+n_3+2) : G'_{n_1+n_2+n_3+2} \cong A_{n_1,n_2;n_3}\}$
$a_{n_1,n_2;n_3}$	$=$	$\lvert\mathbb{A}_{n_1,n_2;n_3}\rvert$
$\mathbb{A}_N$	$=$	$\{g \in \mathbb{G}(N) : g \in \mathbb{A}_{n_1,n_2;n_3}$ for some $n_1, n_2, n_3\}$
a_N	$=$	$\lvert\mathbb{A}_N\rvert$
$\mathbb{B}_{n_1,n_2;n_3}$	$=$	$\{(G'_2, \ldots, G'_{n_1+n_2+n_3+2}) \in \mathbb{G}(n_1+n_2+n_3+2) : G'_{n_1+n_2+n_3+2} \cong B_{n_1,n_2;n_3}\}$
$b_{n_1,n_2;n_3}$	$=$	$\lvert\mathbb{B}_{n_1,n_2;n_3}\rvert$ if $n_1 \geq n_3$, $b_{n_1,n_2;n_3} = 0$ otherwise
$\mathbb{B}_N$	$=$	$\{g \in \mathbb{G}(N) : g \in \mathbb{B}_{n_1,n_2;n_3}$ for some $n_1, n_2, n_3\}$
b_N	$=$	$\lvert\mathbb{B}_N\rvert$
$\mathbb{C}_{n_1,n_2}$	$=$	$\{(G'_2, \ldots, G'_{n_1+n_2+1}) \in \mathbb{G}(n_1+n_2+1) : G'_{n_1+n_2+1} \cong C_{n_1,n_2}\}$
c_{n_1,n_2}	$=$	$\lvert\mathbb{C}_{n_1,n_2}\rvert$
$\mathbb{C}_N$	$=$	$\{g \in \mathbb{G}(N) : g \in \mathbb{C}_{n_1,n_2}$ for some $n_1, n_2\}$
c_N	$=$	$\lvert\mathbb{C}_N\rvert$

We would like to establish relationships between these quantities. We begin with a lemma that tells us that all words produced by Algorithm 1 are produced by a unique choice of graphs G'_i and path p.

Lemma 4.

Let w *be a minimal semi-Sturmian word of order* N>1. *Then the following hold:*

1. *Let* $G'_2, \ldots, G'_N$ *be a chain of graphs produced by Algorithm* 1. *Moreover, let* p *be an Eulerian path in* G'_N *that corresponds to a word* w. *Then* p *is the only path in* G'_N *that corresponds to* w.

2. *Let* $H'_2, \ldots, H'_N$ *be a chain produced by Algorithm* 1, *where some Eulerian path in* H'_N

corresponds to the word w. *Then* $H'_i = G'_i$ *for all* i *such that* $1<i\le N$.

Proof.

For Statement (1), assume that q is another Eulerian path in G'_N that corresponds to the semi-Sturmian wordw. Then the first vertex in both p and q must be the vertex corresponding to the word w[0…N−1). Then the next vertex of both must correspond to the word w[1…N), and so on. Therefore the ith vertex in p is the same as the ith vertex in q, so p=q.

For Statement (2), consider i such that $1<i\le N$. We proceed by induction on i. For i=2 the result follows trivially. Therefore assume i>2, and assume that the result holds for i−1. Then H'_i must be the subgraph of $L(H'_{i-1}) = L(G'_{i-1})$ whose vertices correspond to all subwords of w of length i−1 and whose edges correspond to all subwords of w of length i. However G'_i must also equal said graph, so $H'_i = G'_i$, and the result follows inductively. □

Lemma 5.

If $N>1$, *then* $|G(N)|=a_N+b_N+c_N$ *and the following equalities hold:*

$$a_N = \sum_{\substack{n_1,n_2,n_3\ge 0,\\ n_1+n_2+n_3+2=N}} a_{n_1,n_2;n_3}$$

$$b_N = \sum_{\substack{n_1,n_2,n_3\ge 0,\\ n_1\ge n_3,\\ n_1+n_2+n_3+2=N}} b_{n_1,n_2;n_3}$$

$$c_N = \sum_{\substack{n_1,n_2\ge 0,\\ n_1+n_2+1=N,\\ n_1\ge n_2}} c_{n_1,n_2}.$$

Proof.

Note that if $A_{n1,n2;n3}\cong A_{m1,m2;m3}$ then $n_i=m_i$ for each i. Therefore if $(n_1,n_2,n_3)\ne(m_1,m_2,m_3)$ then $A_{n1,n2;n3}$ is disjoint from $A_{m1,m2;m3}$. Moreover, we have that

$$\mathbb{A}_N = \bigcup_{\substack{n_1,n_2,n_3\ge 0,\\ A_{n_1,n_2;n_3}\text{ has } N \text{ vertices}}} \mathbb{A}_{n_1,n_2;n_3} = \bigcup_{\substack{n_1,n_2,n_3\ge 0,\\ n_1+n_2+n_3+2=N}} \mathbb{A}_{n_1,n_2;n_3}.$$

Thus the collection of all $A_{n1,n2;n3}$ with $n_1+n_2+n_3+2=N$ form a partition of A_N, so

$$a_N = |\mathbb{A}_N| = \sum_{\substack{n_1+n_2+n_3+2=N,\\ n_i\ge 0}} |\mathbb{A}_{n_1,n_2;n_3}| = \sum_{\substack{n_1+n_2+n_3+2=N,\\ n_i\ge 0}} a_{n_1,n_2;n_3}.$$

The proof is identical for b_N, except with $A_{n1,n2;n3}$ replaced by $B_{n1,n2;n3}$. Also, we need to take into account the fact that $B_{n1,n2;n3} \cong B_{n3,n2;n1}$. The proof is almost identical for c_N, except we need to take account of the fact that $C_{n1,n2} \cong C_{n2,n1}$.

To see that $|G(N)|=a_N+b_N+c_N$, we begin by noting that $A_{m1,m2;m3}$ is not isomorphic to $B_{n1,n2;n3}$ for any choice of parameters. Similarly $A_{m1,m2;m3}$ is not isomorphic to $C_{n1,n2}$ and $C_{m1,m2}$ is not isomorphic to $B_{n1,n2;n3}$ (the fact that none of these are isomorphic is a simple consequence of Lemma 2). Thus A_N, B_N, C_N is a partition of G(N), so
$|G(N)|=|A_N|+|B_N|+|C_N|=a_N+b_N+c_N$.
□

Lemma 6.

If N>1, *then* $S(N) = \sum_{(G'_2,\ldots,G'_N)\in\mathbb{G}(N)} \#(G'_N)$, *where* $\#(G'_N)$ *denotes the number of Eulerian paths in* G'_N.

Proof.
Consider the map ψ from the set
$\{(g,\, p) : (G'_2, \ldots, G'_N) = g \in \mathbb{G}(N),\ p \text{ is an Eulerian path in } G'_N\}$

to the set {w:w is a minimal semi-Sturmian word of order N}, that maps (g,p) to the word corresponding to p. We know that ψ is well defined and onto by Theorem 2. Moreover, by Lemma 4it is one-to-one, so ψ is a bijection. Then note that since

$$\sum_{(G'_2,\ldots,G'_N)\in\mathbb{G}(N)} \#(G'_N) = |\{(g,\, p) : (G'_2, \ldots, G'_N) = g \in \mathbb{G}(N),\ p \text{ is an Eulerian path in } G'_N\}|$$

and S(N)=|{w:w is a minimal semi-Sturmian word of order N}|, the fact that ψ is a bijection of finite sets implies our equality. □

To use Lemma 6, we need the following.

Lemma 7.

1.$A_{n1,n2;n3}$*has exactly one Eulerian path.*

2.$B_{n1,n2;n3}$*has exactly two Eulerian paths.*

3.$C_{n1,n2}$*has exactly* n_1+n_2+2*Eulerian paths.*

Proof.

We use the same notation as in Definition 1. For Statement (1), $A_{n1,n2;n3}$ has exactly one Eulerian path, namely $v_0,\ldots,v_{n1},v_0,w_0,\ldots,w_{n3-1},u_0,\ldots,u_{n2},u_0$.

For Statement (2), x_0 has two outgoing edges and one incoming edge, x_1 has two incoming edges and one outgoing edge, and every other vertex has exactly one incoming and one outgoing edge. It is easy to see that this implies that $B_{n1,n2;n3}$ has at most two Eulerian paths. We are able to show that two such paths exist:

$x_0,v_0,\ldots,v_{n1-1},x_1,u_0,\ldots,u_{n2-1},x_0,w_0,\ldots,w_{n3-1},x_1$

$x_0,w_0,\ldots,w_{n3-1},x_1,u_0,\ldots,u_{n2-1},x_0,v_0,\ldots,v_{n1-1},x_1$.

For Statement (3), the proof is very similar to the above. There are exactly two Eulerian paths starting at v, and exactly one Eulerian path starting at each of the n_1+n_2 other vertices ($C_{n1,n2}$ has n_1+n_2+1vertices). Therefore the total number of Eulerian paths is $2(1)+1(n_1+n_2)=n_1+n_2+2$. □

Lemma 8.
If $N>2$, *then* $S(N)=a_N+2b_N+(N+1)c_N$.

Proof.
Denote by $\#(G'_N)$ the number of Eulerian paths in G'_N. By Lemma 6, we have

$$\begin{aligned} S(N) &= \sum_{(G'_2,\ldots,G'_N)\in\mathbb{G}(N)} \#(G'_N) \\ &= \sum_{(G'_2,\ldots,G'_N)\in\mathbb{A}_N} \#(G'_N) + \sum_{(G'_2,\ldots,G'_N)\in\mathbb{B}_N} \#(G'_N) + \sum_{(G'_2,\ldots,G'_N)\in\mathbb{C}_N} \#(G'_N). \end{aligned}$$

By Lemma 7, this equals

$$\sum_{(G'_2,\ldots,G'_N)\in\mathbb{A}_N} 1 + \sum_{(G'_2,\ldots,G'_N)\in\mathbb{B}_N} 2 + \sum_{(G'_2,\ldots,G'_N)\in\mathbb{C}_N} (N+1) = a_N + 2b_N + (N+1)c_N.$$

□

Lemma 9.

If $N>2$, *then* $2c_{N-1}=a_N-a_{N-1}$.

Proof.

Note that

$$a_N = \sum_{(G'_2,\ldots,G'_N)\in\mathbb{A}_N} 1 = \sum_{(G'_2,\ldots,G'_{N-1})\in\mathbb{G}(N-1)} \left(\sum_{\substack{G'_{N-1}\to G'_N,\\ G'_N\cong A_{n_1,n_2;n_3} \text{ for some } n_1,n_2,n_3}} 1 \right).$$

However, we know from Lemma 2 that

$$\sum_{\substack{G'_{N-1}\to G'_N,\\ G'_N\cong A_{n_1,n_2;n_3} \text{ for some } n_1,n_2,n_3}} 1$$

is equal to 1 if $(G'_2,\ldots,G'_{N-1})\in\mathbb{A}_{N-1}$, is equal to 2 if $(G'_2,\ldots,G'_{N-1})\in\mathbb{C}_{N-1}$, and is equal to 0 otherwise. Putting the above together, we get that $a_N=a_{N-1}+2c_{N-1}$. □

The proofs of the next three lemmas use almost identical techniques to those in Lemma 9.

Lemma 10.

If $N>2$, *then* $|G(N)|=a_{N-1}+b_{N-1}+4c_{N-1}$.

Proof.

Note that

$$|\mathbb{G}(N)| = \sum_{g\in\mathbb{G}(N)} 1 = \sum_{\{g=(G'_2,\ldots,G'_{N-1})\in\mathbb{G}(N-1)\}} \left(\sum_{\{G'_N : G'_{N-1}\Rightarrow G'_N\}} 1 \right).$$

Also note that if $g=(G'_2,\ldots,G'_{N-1})\in\mathbb{A}_{N-1}$ or $g=(G'_2,\ldots,G'_{N-1})\in\mathbb{B}_{N-1}$, then by Lemma 2 there is exactly one G'_N such that $G'_{N-1}\Rightarrow G'_N$. If, on the other hand, $g=(G'_2,\ldots,G'_{N-1})\in\mathbb{C}_{N-1}$ then there are four G'_N's such that $G'_{N-1}\Rightarrow G'_N$. Plugging this into the above, we get that

$$\begin{aligned}
|\mathbb{G}(N)| &= \sum_{\{g=(G'_2,\ldots,G'_{N-1})\in\mathbb{A}_{N-1}\}} \left(\sum_{\{G'_N : G'_{N-1}\Rightarrow G'_N\}} 1 \right) + \sum_{\{g=(G'_2,\ldots,G'_{N-1})\in\mathbb{B}_{N-1}\}} \left(\sum_{\{G'_N : G'_{N-1}\Rightarrow G'_N\}} 1 \right) \\
&\quad + \sum_{\{g=(G'_2,\ldots,G'_{N-1})\in\mathbb{C}_{N-1}\}} \left(\sum_{\{G'_N : G'_{N-1}\Rightarrow G'_N\}} 1 \right) \\
&= \sum_{\{g=(G'_2,\ldots,G'_{N-1})\in\mathbb{A}_{N-1}\}} 1 + \sum_{\{g=(G'_2,\ldots,G'_{N-1})\in\mathbb{B}_{N-1}\}} 1 + \sum_{\{g=(G'_2,\ldots,G'_{N-1})\in\mathbb{C}_{N-1}\}} 4 \\
&= a_{N-1} + b_{N-1} + 4c_{N-1}.
\end{aligned}$$

□

Lemma 11.

1.*If* $n>1$, *then* $c_{n,0}=c_{0,n}=c_{0,n-1}$;

2.*If* $n_1,n_2>0,n_1\neq n_2$, *then* $c_{n1,n2}=b_{n1-1,0;n2-1}+b_{n2-1,0;n1-1}$;

3.*If* $n_1,n_2>0,n_1=n_2$, *then* $c_{n1,n2}=b_{n1-1,0;n2-1}$.

Proof.

We prove that for $n>1,c_{n,0}=c_{n-1,0}$, the other proofs are similar. Let $N=n+1$. Note that

$$c_{n,0} = |\mathbb{C}_{n,0}| = \sum_{g\in\mathbb{C}_{n,0}} 1 = \sum_{\substack{g=(G'_2,\ldots,G'_N)\in\mathbb{G}(N),\\ G'_N\cong C_{n,0}}} 1$$

$$= \sum_{g=(G'_2,\ldots,G'_{N-1})\in\mathbb{G}(N-1)} \left(\sum_{\{G'_N : G'_{N-1}\Rightarrow G'_N,\, G'_N\cong C_{n,0}\}} 1 \right).$$

By Lemma 2, we see that if $(G'_2,\ldots,G'_{N-1})\in\mathbb{C}_{n-1,0}$ then

$$\sum_{\{G'_N : G'_{N-1}\Rightarrow G'_N,\, G'_N\cong C_{n,0}\}} 1$$

is equal to 1 and otherwise it is equal to 0. Plugging this into the above equation, we get

$$c_{n,0} = \sum_{g=(G'_2,\ldots,G'_{N-1})\in\mathbb{C}_{n-1,0}} 1 = c_{n-1,0}.$$

□

Lemma 12.

Assume that $n_1+n_2+n_3+2>2,n_i\geq 0$. *Then the following hold:*

1.*If* $n_1=0$, *then* $b_{n1,n2;n3}=0$;

2.*If* $n_1,n_3>0$, *then* $b_{n1,n2;n3}=b_{n1-1,n2+1;n3-1}$;

3.*If* $n_3=0,n_1>0$*and* $n_1-1\neq n_2+1$, *then* $b_{n1,n2;0}=c_{n1-1,n2+1}$;

4.*If* $n_3=0,n_1>0$*and* $n_1-1=n_2+1$, *then* $b_{n1,n2;0}=2c_{n1-1,n2+1}$.

Proof.

We prove Statements 2 and 4, the other statements are similar. For Statement 2, let $N=n_1+n_2+2$. Moreover assume $n_1\geq n_3$, the other case being trivial. Note that

$$b_{n_1,n_2;n_3} = |\mathbb{B}_{n_1,n_2;n_3}| = \sum_{g\in \mathbb{B}_{n_1,n_2;n_3}} 1 = \sum_{\substack{g=(G'_2,\ldots,G'_N)\in \mathbb{G}(N),\\ G'_N \cong B_{n_1,n_2;n_3}}} 1$$

$$= \sum_{g=(G'_2,\ldots,G'_{N-1})\in \mathbb{G}(N-1)} \left(\sum_{\{G'_N : G'_{N-1} \Rightarrow G'_N,\, G'_N \cong B_{n_1,n_2;n_3}\}} 1 \right).$$

By Lemma 2, we see that if $(G'_2, \ldots, G'_{N-1}) \in \mathbb{B}_{n_1-1,n_2+1;n_3-1}$ then

$$\sum_{\{G'_N : G'_{N-1} \Rightarrow G'_N,\, G'_N \cong B_{n_1,n_2;n_3}\}} 1$$

is equal to 1 and otherwise it is equal to 0. Plugging this into the above equation, we get

$$b_{n_1,n_2;n_3} = \sum_{g=(G'_2,\ldots,G'_{N-1})\in \mathbb{B}_{n_1-1,n_2+1;n_3-1}} 1 = b_{n_1-1,n_2+1;n_3-1}.$$

For Statement 4, let $N=n_1+n_2+2$. Note that

$$b_{n_1,n_2;0} = |\mathbb{B}_{n_1,n_2;0}| = \sum_{g\in \mathbb{B}_{n_1,n_2;0}} 1 = \sum_{\substack{g=(G'_2,\ldots,G'_N)\in \mathbb{G}(N),\\ G'_N \cong B_{n_1,n_2;0}}} 1$$

$$= \sum_{g=(G'_2,\ldots,G'_{N-1})\in \mathbb{G}(N-1)} \left(\sum_{\{G'_N : G'_{N-1} \Rightarrow G'_N,\, G'_N \cong B_{n_1,n_2;0}\}} 1 \right).$$

By Lemma 2, $C_{n1-1,n2+1} \Rightarrow B_{n1,n2;0}$ when $n_1-1 \le n_2+1$ and $C_{n2+1,n1-1} \Rightarrow B_{n1,n2;0}$ when$n_2+1 \le n_1-1$. We see that if $(G'_2, \ldots, G'_{N-1}) \in \mathbb{C}_{n_1-1,n_2+1}$ then
equation(1)

$$\sum_{\{G'_N : G'_{N-1} \Rightarrow G'_N,\, G'_N \cong B_{n_1,n_2;0}\}} 1$$

is equal to 1, if $(G'_2, \ldots, G'_{N-1}) \in \mathbb{C}_{n_2+1,n_1-1}$ then the sum (1) is also equal to 1, and otherwise it is equal to 0. Plugging these into the above equation, we get

$$\begin{aligned} b_{n_1,n_2;0} &= \sum_{g=(G'_2,\ldots,G'_{N-1})\in \mathbb{C}_{n_1-1,n_2+1}} 1 + \sum_{g=(G'_2,\ldots,G'_{N-1})\in \mathbb{C}_{n_2+1,n_1-1}} 1 \\ &= c_{n_1-1,n_2+1} + c_{n_2+1,n_1-1} \\ &= 2c_{n_1-1,n_2+1}. \end{aligned}$$

□

Lemma 13.

1. *If* $n_1+n_2+1 \ge 2, n_1, n_2 \ge 0$*, then*

$$c_{n_1,n_2} = \begin{cases} 2 & \text{if } \gcd(n_1+1, n_1+n_2+2) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

2. *If* $m_1+m_2+m_3+2 \ge 2, m_1, m_2, m_3 \ge 0$*, then*

$$b_{m_1,m_2;m_3} = \begin{cases} 2 & \textit{if } m_1 > m_3 \textit{ and } \gcd(m_1 - m_3, m_2 + m_1 + 2) = 1 \\ 0 & \textit{otherwise.} \end{cases}$$

Proof.

We prove this by induction on $N=n_1+n_2+1=m_1+m_2+m_3+2$. Note that this holds for the case $N=2$. Therefore assume that $N>2$.

For Statement (1), first consider the case $n_1=0$. Then $c_{n1,n2}=c_{0,n2}=c_{0,n2-1}$ by Lemma 11(1). Since $\gcd(0+1,0+n_2-1+2)=1$ we know that $c_{0,n2}=c_{0,n2-1}=2$. Since $\gcd(0+1,0+n_2+2)=1$ it follows that the claim holds in this case. The case with $n_2=0$ is almost identical. So assume that $n_1,n_2>0$. If $n_1=n_2$ then $c_{n1,n2}=b_{n1-1,0;n2-1}$ by Lemma 11(3); since $n_1-1=n_2-1$ it follows by induction that $c_{n1,n2}=0$. Therefore assume that $n_1\neq n_2$. Consider the case $n_1<n_2$, the other being similar. Then by Lemma 11(2), $c_{n1,n2}=b_{n1-1,0;n2-1}+b_{n2-1,0;n1-1}$. Since $n_1<n_2$ it follows that $b_{n1-1,0;n2-1}=0$, so $c_{n1,n2}=b_{n2-1,0;n1-1}$. Then note that $\gcd(n_2-n_1,n_2+1)=\gcd(n_1+1,n_2+1)=\gcd(n_1+1,n_1+n_2+2)$, so it follows by induction that

$$c_{n_1,n_2} = b_{n_2-1,0;n_1-1} = \begin{cases} 2 & \text{if } \gcd(n_2 - n_1, n_2 + 1) = \gcd(n_1 + 1, n_1 + n_2 + 2) = 1 \\ 0 & \text{otherwise} \end{cases}$$

which is what we wanted.
For Statement (2), assume $m_1\leq m_3$. Then by a simple inductive argument on m_1 and Lemma 12, we get that $b_{m1,m2;m3}=b_{0,m2+m1;m3-m1}=0$. If $m_1>m_3$, a similar argument gives us that $b_{m1,m2;m3}=b_{m1-m3,m2+m3;0}$. There are two cases to consider. First if $m_1-m_3-1\neq m_2+m_3+1$, we obtain by induction that

$$\begin{aligned} b_{m_1-m_3,m_2+m_3;0} &= c_{m_1-m_3-1,m_2+m_3+1} \\ &= \begin{cases} 2 & \text{if } m_1 > m_3 \text{ and } \gcd(m_1 - m_3, m_2 + m_1 + 2) = 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

as desired.

Therefore assume that $m_1-m_3-1=m_2+m_3+1$. Then we get that $b_{m1-m3,m2+m3;0}=2c_{m1-m3-1,m2+m3+1}=0$, where the last equality follows since $\gcd(m_1-m_3-1+1,m_1-m_3-1+m_2+m_3+1+2)=\gcd(m_1-m_3,2(m_1-m_3-1)+2)=\gcd(m_1-m_3,2(m_1-m_3))\geq m_1-m_3=m_2+m_3+2>1$, so $c_{m1-m3-1,m2+m3+1}=0$ by the inductive hypothesis. However, note that since $\gcd(m_1-m_3,m_1+m_2+2)=\gcd(m_1-m_3,m_3+m_2+2)=\gcd(m_1-m_3,m_1-m_3)=m_1-m_3=m_2+m_3+2>1$, the result follows. □

Lemma 14.

If $N\geq 2$, *then the following equalities hold:*

1. $c_N=\phi(N+1)$;

2. $a_N = 2 + 2\sum_{n=3}^{N}\phi(n)$;
3. $b_N = 2 + \left(\sum_{n=3}^{N}\phi(n)\right) - \phi(N+1)$.

Proof.

Consider N≥2. For Statement (1),

$$c_N = \sum_{\substack{0\le n_1 \le n_2,\\ n_1+n_2+1=N}} c_{n_1,n_2}.$$

By Lemma 13(1), this equals

$$2|\{(n_1, n_2) : \gcd(n_1+1, n_1+n_2+1) = 1, n_1+n_2+1 = N, 0 \le n_1 \le n_2\}|$$
$$= 2\left|\left\{n_1 : \gcd(n_1+1, N+1) = 1, 1 \le n_1+1 \le \frac{N}{2}\right\}\right|$$

where the above equality follows since the map ξ from the set
$\{(n_1,n_2):\gcd(n_1+1,n_1+n_2+1)=1,n_1+n_2+1=N,0\le n_1\le n_2\}$

to the set

$$\left\{n_1 : \gcd(n_1+1, N+1) = 1, 1 \le n_1+1 \le \frac{N}{2}\right\}$$

defined by $\xi(n_1,n_2)=n_1$ is a bijection. Setting $m=n_1+1$ tells us that the above equals

$$2\left|\left\{m : \gcd(m, N+1) = 1, 1 \le m \le \frac{N}{2}\right\}\right| = 2\left|\left\{m : \gcd(m, N+1) = 1, 1 \le m < \frac{N+1}{2}\right\}\right|$$
$$= \left|\left\{m : \gcd(m, N+1) = 1, 1 \le m < \frac{N+1}{2}\right\}\right|$$
$$+ \left|\left\{m : \gcd(m, N+1) = 1, 1 \le m < \frac{N+1}{2}\right\}\right|.$$

By noting that gcd(m,N+1)=1 if and only if gcd(N+1−m,N+1)=1, we get that

$$\left|\left\{m : \gcd(m, N+1) = 1, 1 \le m < \frac{N+1}{2}\right\}\right| = \left|\left\{m : \gcd(m, N+1) = 1, N+1 > m > \frac{N+1}{2}\right\}\right|.$$

So plugging this into the above chain of equalities, we deduce that

$$c_N = \left|\left\{m : \gcd(m, N+1) = 1, 1 \le m < \frac{N+1}{2}\right\}\right| + \left|\left\{m : \gcd(m, N+1) = 1, N+1 > m > \frac{N+1}{2}\right\}\right|$$
$$= \left|\left\{m : \gcd(m, N+1) = 1, m \ne \frac{N+1}{2}, 0 < m < N+1\right\}\right|$$
$$= |\{m : \gcd(m, N+1) = 1, 0 < m < N+1\}|$$
$$= \phi(N+1)$$

where the second to last equality follows since, if $m = \frac{N+1}{2}$ then $\gcd(m, N+1) = m = \frac{N+1}{2} \ne 1$.

Therefore we get the result we wanted.

For Statement (2), we prove this by induction on N. It holds when N=2, so assume N>2. Then by the previous part of the proof and Lemma 9, we get that

$$a_N = a_{N-1} + 2c_{N-1} = 2 + 2\left(\sum_{n=3}^{N-1}\phi(n)\right) + 2\phi(N) = 2 + 2\sum_{n=3}^{N}\phi(n).$$

For Statement (3), we proceed by induction on N. We know that the claim holds when N=2, so assume thatN>2. By Lemma 10, $|G(N)|=a_{N-1}+b_{N-1}+4c_{N-1}$. Moreover, $|G(N)|=a_N+b_N+c_N$. Thus $a_{N-1}+b_{N-1}+4c_{N-1}=a_N+b_N+c_N$. Solving for b_N, it follows by induction that

$$\begin{aligned} b_N &= a_{N-1} + b_{N-1} + 4c_{N-1} - a_N - c_N = -2\phi(N) + 4\phi(N) - \phi(N+1) + b_{N-1} \\ &= 2\phi(N) - \phi(N+1) + b_{N-1} = 2 + \left(\sum_{n=3}^{N}\phi(n)\right) - \phi(N+1). \end{aligned}$$

□

We conclude this section with the proof of Theorem 1:

Proof.

For N=2, we can easily check that this holds. For N>2, this follows from Lemma 8 and Lemma 14. □

4. Conclusion

Our techniques can help count minimal words with subword complexity other than f(n)=n+1. In particular they can help with those with complexity f(n)=n+k−1, where k is the size of the alphabet. For N>0, how many words w of minimal length exist such that $p_w(n)=n+k-1$ for n=1,…,N? Theorem 3 gives a lower bound on this number.

Lemma 15.

Let N>2, *let* A*be a* k-*letter alphabet, let* H'_2*be any spanning subgraph of* $G_{\{ab:a,b\in A\}}$*such that* H'_2*has an Eulerian path and* k+1*edges, and let* $H'_2,\ldots,H'_N$*be any sequence such that* $H'_i \Rightarrow H'_{i+1}$*for* 2≤i<N. *If* w*is a word corresponding to an Eulerian path in* H'_N, *then* w*is a minimal word of order* N*with subword complexity* f(n)=n+k−1.

Proof.
The proof is identical to that of Theorem 2. □

Lemma 16.
Let A*be a* k-*letter alphabet.*

1.*There are at least* $\frac{k(k-1)}{2}k!$*spanning subgraphs of* $G_{\{ab:a,b\in A\}}$*isomorphic to some graph of the form* $A_{n1,n2;n3}$.
2.*There are at least* $k!$*spanning subgraphs of* $G_{\{ab:a,b\in A\}}$*isomorphic to some graph of the form* $C_{0,n}$.

Proof.

For Statement (1), let A_1 consist of all subsets of A containing two elements and A_2 consist of all orderings of the elements in A. Then $A_1\times A_2$ has $\frac{k(k-1)}{2}k!$ elements. Therefore to prove our result it suffices to produce a subgraph of $G_{\{ab:a,b\in A\}}$ isomorphic to some $A_{n1,n2;n3}$ for each element in $A_1\times A_2$. So consider the element $(\{a_r,a_s\},(a_0,\ldots,a_{k-1}))$ in $A_1\times A_2$. Without loss of generality we can assume that $r<s$. Then consider the graph which contains the path $a_0,\ldots,a_{k-1}$, and so contains the edges(a_i,a_{i+1}) for all i such that $0\leq i<N$, as well as the edges (a_r,a_0) and (a_{k-1},a_s). Then this is a spanning subgraph of $G_{\{ab:a,b\in A\}}$ isomorphic to $A_{r,k-1-s;s-r-1}$. Moreover, each element in $A_1\times A_2$produces a different graph, so our result follows.

For Statement (2), since there are k! orderings of the elements in A, it suffices to construct a subgraph of$G_{\{ab:a,b\in A\}}$ isomorphic to some $C_{0,n}$ for each ordering. So let $a_0,\ldots,a_{k-1}$ be such an ordering. Then consider the graph whose vertices are $a_0,\ldots,a_{k-1}$, and where (a_i,a_{i+1}) is an edge for each i such that$0\leq i<N$. Moreover, let (a_{k-1},a_0) and (a_0,a_0) be edges. Then this graph is uniquely determined by the ordering of the letters in A and is isomorphic to $C_{0,k-1}$. Moreover, it is a spanning subgraph of$G_{\{ab:a,b\in A\}}$. □

Lemma 17.

Let A*be a* k-*letter alphabet.*

1.*There are at least* $k!$*sequences* $H_2',\ldots,H_N'$*such that* H_2'*is a subgraph of* $G_{\{ab:a,b\in A\}}$*having an Eulerian path and* $k+1$*vertices,* $H_i'\Rightarrow H_{i+1}'$*for all* i*such that* $2\leq i<N$, *and* $H_N'\cong C_{0,n}$*for some* n.

2.*There are at least* $\frac{k(k-1)}{2}k!+2k!(N-2)$*sequences* $H_2',\ldots,H_N'$*such that* H_2'*is a subgraph of* $G_{\{ab:a,b\in A\}}$*having an Eulerian path and* $k+1$*vertices,* $H_i'\Rightarrow H_{i+1}'$*for all* i *such that* $2\leq i<N$, *and* $H_N'\cong A_{n1,n2;n3}$*for some* n_1,n_2,n_3.

Proof.

For Statement (1), we proceed by induction on N. We know that this holds for N=2. Therefore assume thatN>2 and that the claim holds for N−1. Then for each sequence $H'_2, \ldots, H'_{N-1}$ with $H'_{N-1} \cong C_{0,n}$ for somen, we know that there exists a unique graph H'_N so that $H'_{N-1} \Rightarrow H'_N$ and H'_N is isomorphic to $C_{0,n}$ for some n (see Lemma 2(3)(d)). Therefore the claim follows by induction.

For Statement (2), we also proceed by induction on N. This holds for N=2. Therefore assume that N>2 and that the claim holds for N−1. For each sequence $H'_2, \ldots, H'_{N-1}$ with $H'_{N-1} \cong C_{0,n}$ for some n, there exist two graphs H'_N and G'_N so that $H'_{N-1} \Rightarrow H'_N$, $H'_{N-1} \Rightarrow G'_N$, and both H'_N and G'_N are isomorphic to some$A_{n1,n2;n3}$ (see Lemma 2(3)(a,b)). Also for each sequence $H'_2, \ldots, H'_{N-1}$ with $H'_{N-1} \cong A_{n1,n2;n3}$ for somen_1,n_2,n_3, there exists a unique H'_N so that $H'_{N-1} \Rightarrow H'_N$ and so that $H'_N \cong A_{n1,n2;n3}$ for some n_1,n_2,n_3(see Lemma 2(1)). Therefore the claim follows inductively. □

We now prove our bound.

Theorem 3.

For N≥2*, the number of minimal words of order* N*over a* k*-letter alphabet with subword complexity* f(n)=n+k−1,S(N,k)*, satisfies*

$$S(N,k) \geq k!(k+N-1) + 2k!(N-2) + \frac{k(k-1)}{2}k!$$

Proof.

By Lemma 17 there are at least k! sequences $H'_2, \ldots, H'_N$, where H'_2 is a subgraph of $G_{\{ab:a,b\in A\}}$having an Eulerian path and k+1 vertices, $H'_i \Rightarrow H'_{i+1}$ for 2≤i<N, and $H'_N \cong C_{0,n}$ for some n. Moreover, each such H'_N has k+N−1 Eulerian paths (since $H'_N \cong C_{0,n}$ where n=N+k−3), so these graphs contribute k!(k+N−1) to the total. Moreover, there are at least $\frac{k(k-1)}{2}k! + 2k!(N-2)$ sequences $H'_2, \ldots, H'_N$, where H'_2 is a subgraph of $G_{\{ab:a,b\in A\}}$ having an Eulerian path and k+1 vertices, $H'_i \Rightarrow H'_{i+1}$ for 2≤i<N, and $H'_N \cong A_{n1,n2;n3}$ for some n_1,n_2,n_3. Each of these sequences contributes one word to the total (since $A_{n1,n2;n3}$ has exactly one Eulerian path). Thus adding everything up gives us our result. □
Note that the above bound can be improved by including more families of graphs (as opposed to just$A_{n1,n2;n3}$ and $C_{0,n}$), but the proof becomes trickier.

References

[1] E. Altman, B. Gaujal, A. Hordijk, Balanced sequences and optimal routing, Journal of the Association for Computing Machinery 47 (2000) 752–775.

[2] J. Berstel, M. Pocchiola, Random generation of finite Sturmian words, Discrete Mathematics 153 (1996) 29–39.

[3] F. Blanchet-Sadri, J. Lensmire, On minimal Sturmian partial words, Discrete Applied Mathematics 159 (2011) 733–745.

[4] M. Boshernitzan, A.S. Fraenkel, A linear algorithm for nonhomogeneous spectra of numbers, Journal of Algorithms 5 (1984) 187–198.

[5] A.M. Bruckstein, The self-similarity of digital straight lines, in: 10th International Conference on Pattern Recognition, Vol. I, IEEE Press, 1990, pp. 485–490.

[6] A.M. Bruckstein, Self-similarity properties of digitized straight lines, in: P. Bhattacharya, R.A. Melter, A. Rosenfeld (Eds.), Vision Geometry, in: Contemporary Mathematics, vol. 119, American Mathematical Society, Providence, RI, 1991, pp. 1–20.

[7] A. de Luca, A. De Luca, Some characterizations of finite Sturmian words, Theoretical Computer Science 356 (2006) 118–125.

[8] S. Dulucq, D. Gouyou-Beauchamps, Sur les facteurs des suites de Sturm, Theoretical Computer Science 71 (1990) 381–400.

[9] J.L. Gross, J. Yellen, Handbook of Graph Theory, CRC Press, 2004.

[10] J.C. Kieffer, Sturmian minimal systems associated with the iterates of certain functions on an interval, in: J.C. Alexander (Ed.), Dynamical Systems, in: Lecture Notes in Mathematics, vol. 1342, Springer-Verlag, Berlin, 1988, pp. 354–360.

[11] M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, Cambridge, 2002.

[12] F. Mignosi, On the number of factors of Sturmiam words, Theoretical Computer Science 82 (1991) 71–84.

[13] M. Morse, G.A Hedlund, Symbolic dynamics II. Sturmian trajectories, American Journal of Mathematics 62 (1940) 1–42.

[14] M.-w. Wang, J. Shallit, On minimal words with given subword complexity, Electronic Journal of Combinatorics 5 (1998) 1–16.