On free spectra of varieties of locally threshold testable semigroups

Igor Dolinka

Department of Mathematics and Informatics, University of Novi Sad
Trg Dositeja Obradovića 4, 21101 Novi Sad, Serbia
e-mail: dockie@dmi.uns.ac.rs

Abstract

A semigroup S is said to be ℓ-threshold k-testable if it satisfies all identities $\mathbf{u} = \mathbf{v}$ where $\mathbf{u}, \mathbf{v}$ is an arbitrary pair of words over a finite alphabet Σ such that they simultaneously belong or fail to belong to any ℓ-threshold k-testable (regular) language. We give an asymptotic formula for the free spectrum of the variety $\mathcal{T}_{k,\ell}$ of all ℓ-threshold k-testable semigroups, thereby providing an asymptotic upper bound on the size of an arbitrary finitely generated locally threshold testable semigroup. The combinatorial interpretation of this task yields an enumeration problem for particular edge labellings of de Bruijn graphs.

1 Introduction

For an alphabet Σ, let Σ^+ denote the free semigroup on Σ, consisting of all nonempty words (finite sequences) over Σ. An equivalence relation ρ on Σ^+ is said to *saturate* the language $L \subseteq \Sigma^+$ if L is a union of some ρ-classes; equivalently, $L = \bigcup_{\mathbf{w}\in L} \mathbf{w}/\rho$. Now for $k \geq 2$, $\ell \geq 1$ consider the equivalence $\rho_{k,\ell}$ defined so that $(\mathbf{u}, \mathbf{v}) \in \rho_{k,\ell}$ if and only if either $\mathbf{u} = \mathbf{v}$, or $|\mathbf{u}|, |\mathbf{v}| \geq k$ and the following conditions hold:

(1) the prefixes of $\mathbf{u}$ and $\mathbf{v}$ of length $k-1$ coincide;

(2) the suffixes of $\mathbf{u}$ and $\mathbf{v}$ of length $k-1$ coincide;

(3) for any $1 \leq j \leq \ell$ the sets of factors (subwords consisting of consecutive letters) of length k occurring at least j times in $\mathbf{u}$ and $\mathbf{v}$, respectively, coincide.

Clearly, the latter condition can be replaced by

(3') for any word $\mathbf{w}$ of length k the numbers of occurrences of $\mathbf{w}$ as a factor in $\mathbf{u}$ and $\mathbf{v}$ either coincide or are both greater than $\ell - 1$.

Key words and phrases: semigroup, threshold testable, free spectrum, de Bruijn digraph, edge labelling, strongly connected

2010 Mathematics Subject Classification: 20M07 (primary), 08A40, 05A16, 05C20, 05C40

A language L is said to be *ℓ-threshold k-testable* if it is saturated by $\rho_{k,\ell}$. A language is *locally threshold testable* if it is ℓ-threshold k-testable for some k, ℓ.

Locally threshold testable languages were introduced by Beauquier and Pin [3]. Today they have a variety of applications, for example in pattern recognition, speech recognition, natural language modelling, etc., see [14, 19]. Their algebraic and automata-theoretic properties were studied by Trahtman [16, 17, 18]. For $\ell = 1$ they reduce to a quite familiar notion in formal language theory, *locally testable languages*, see [4, 12].

Motivated by these definitions, a semigroup S is said to be *ℓ-threshold k-testable* [17, 18] if it satisfies all semigroup identities $\mathbf{u} = \mathbf{v}$ such that $(\mathbf{u}, \mathbf{v}) \in \rho_{k,\ell}$; similarly as above, S is *locally threshold testable* if it is ℓ-threshold k-testable for some k, ℓ. Since the class $\mathcal{T}_{k,\ell}$ of all ℓ-threshold k-testable semigroups is defined by identities, it forms a variety in the sense of universal algebra [5]. Also, finite semigroups from $\mathcal{T}_{k,\ell}$ arise as syntactic semigroups [13] of ℓ-threshold k-testable languages. A (nontrivial) equational axiomatisation of $\mathcal{T}_{k,\ell}$ was given in [17].

A variety $\mathcal{V}$ of semigroups (and, more generally, of algebras) is *locally finite* if each finitely generated member of $\mathcal{V}$ is finite. An equivalent way of saying this is that the *free spectrum* [10] of $\mathcal{V}$, the sequence $f_n(\mathcal{V}) = |\mathfrak{F}_n(\mathcal{V})|$ of cardinalities of n-generated free objects of $\mathcal{V}$, consists entirely of finite numbers. In this short note we supply an asymptotic estimation of the free spectrum of $\mathcal{T}_{k,\ell}$ (showing, in passing, that it is locally finite).

Theorem 1.1. *We have*

$$f_n(\mathcal{T}_{k,\ell}) \sim n^{2(k-1)}(\ell + 1)^{n^k}$$

as $n \to \infty$.

The previous asymptotic formula was proved for the case $\ell = 1$ (that is, for locally testable semigroup varieties) in an earlier contribution [8] by the author, which was in turn a generalisation of the particular case $k = 2$ (the variety of 2-testable semigroups) computed in [11].

The interest in obtaining such an asymptotic formula is at least two-fold. First of all, if L is an ℓ-threshold k-testable language over an n-element alphabet X_n whose syntactic semigroup is S then the syntactic morphism $\varphi : X_n^+ \to S$ (such that $L = \varphi^{-1}(P)$ for an appropriate subset $P \subseteq S$) factorises through the natural homomorphism $X_n^+ \to \mathfrak{F}_n(\mathcal{V})$; thus the above theorem provides a rough upper bound on the size of S and so of the number of states of a DFA needed to accept L (the construction of which is a task with a number of practical applications). And secondly, it is valuable to the (algebraic) theory of free spectra of semigroups and monoids, since it provides yet another class of semigroups (and semigroup varieties) whose free spectra are *small*, which means that $\log f_n(\mathcal{V}) \in \mathcal{O}(n^c)$ for some $c \geq 0$. This is particularly interesting in the light of the currently principal driving force in this theory, the *Seif Conjecture* [15], which attempts to demarcate between small and log-exponential free spectra of finitely generated monoid varieties.

Even though the topic of free spectra of varieties has a very definite general-algebraic appearance, it is often the case with problems related to this topic that universal algebra is just a guise concealing their truly combinatorial nature, with some known structures from discrete mathematics popping up during the analysis that leads to the required computations. This will be no different in this note as well. Therefore, in the next section we review some necessary notions from (di)graph theory (see [2] for a standard background) and establish a connection to the initial formulation of the sequence $f_n(\mathcal{T}_{k,\ell})$ we aim to estimate. The proof of Theorem 1.1 in this new combinatorial setting is deferred to Sect. 3.

2 De Bruijn graphs and their ℓ-labellings

In this paper a *directed graph* (or *digraph*) is just a structure (V, E) such that $E \subseteq V \times V$; in other words, we allow loops and edges in both directions between vertices. An edge $e = (u, v)$, $u, v \in V$, is written uv for brevity. A *walk* is a sequence of edges $e_1 \dots e_m \in E^+$ such that for each $1 \leq i < m$, if $e_i = u_i v_i$ and $e_{i+1} = u_{i+1} v_{i+1}$ then $v_i = u_{i+1}$; that is, consecutive edges in the sequence are adjacent. A walk is *closed* if its initial and final vertices coincide. The fact that the considered walk (or path) starts at vertex u and ends at v will be denoted by $u \rightsquigarrow v$.

For $\ell \geq 1$ let $\mathbb{N}_\ell^\infty = \{0, \dots, \ell-1\} \cup \{\infty\}$, where ∞ is just a symbol not belonging to the set $\mathbb{N}$ of natural numbers. An ℓ-labelling of a digraph $\mathcal{D} = (V, E)$ is a colouring of its edges in $\ell + 1$ colours, that is a function $\lambda : E \to \mathbb{N}_\ell^\infty$. Given an ℓ-labelling λ of $\mathcal{D}$, by $\mathcal{D}_\infty^{(\lambda)}$ we denote the digraph (V, E') that is obtained from $\mathcal{D}$ by retaining all the edges e such that $\lambda(e) = \infty$ and deleting all the others (and keeping the same set of vertices). Also, a walk $e_1 \dots e_m \in E^+$ in $\mathcal{D}$ is said to be a *covering walk* for λ if the following two conditions hold for an arbitrary edge $e \in E$:

(a) if $\lambda(e) = j < \ell$ then e occurs in the sequence $e_1 \dots e_m$ precisely j times;

(b) if $\lambda(e) = \infty$ then e occurs in the sequence $e_1 \dots e_m$ at least ℓ times.

In other words, the labelling λ happens to record the number of occurrences of each edge of $\mathcal{D}$ in the considered walk whenever that number is $< \ell$, and outputs ∞ if an edge occurs in the walk $\geq \ell$ times.

It is easy to check that given an n-element alphabet X_n, the equivalence $\rho_{k,\ell}$ on the free semigroup X_n^+ is in fact a semigroup congruence and, moreover, a *fully invariant* one, which means that it is stable under substitutions. By elementary universal-algebraic results (cf. [5, Lemma II.14.7]) we now have the following connection.

Lemma 2.1. *$X_n^+/\rho_{k,\ell}$ is isomorphic to the free object of $\mathcal{T}_{k,\ell}$ freely generated by X_n (i.e. by the copy of X_n consisting of one-letter words).*

Therefore, $f_n(\mathcal{T}_{k,\ell})$ is simply $|X_n^+/\rho_{k,\ell}|$, the index of the relation $\rho_{k,\ell}$ in X_n^+. The aim of computing such an index motivates to consider the following sequence of well-known digraphs, called *de Bruijn (di)graphs* [2, 6, 7, 9] $\mathcal{B}(n,r)$, $n,r \geq 1$. The vertex set of $\mathcal{B}(n,r)$ is X_n^r, the set of all words over X_n of length r. For $\mathbf{u},\mathbf{v} \in X_n^r$, there is an edge from $\mathbf{u}$ to $\mathbf{v}$ if and only if there is a word $\mathbf{w} \in X_n^+$ with $|\mathbf{w}| = r+1$ whose prefix of length r is $\mathbf{u}$, while its suffix of length r is $\mathbf{v}$; in other words, the suffix of $\mathbf{u}$ of length $r-1$ coincides with the prefix of $\mathbf{v}$ of the same length, $\mathbf{u} = xy_1 \dots y_{r-1}$, $\mathbf{v} = y_1 \dots y_{r-1}z$ (for some $x, y_1, \dots, y_{r-1}, z \in X_n$). Clearly, $\mathcal{B}(n,1)$ is the complete digraph on n vertices, while $\mathcal{B}(n,r+1)$ is the line digraph of $\mathcal{B}(n,r)$ for all $r \geq 1$.

Lemma 2.2. *Let $n,\ell \geq 1$ and $k \geq 2$. There is a bijection between the set of $\rho_{k,\ell}$-classes on X_n^+ containing words of length $\geq k$ and the set of all triples $(\mathbf{u},\mathbf{v},\lambda)$, where $\mathbf{u},\mathbf{v} \in X_n^{k-1}$ and λ is an ℓ-labelling of $\mathcal{B}(n,k-1)$ possessing a covering walk $\mathbf{u} \rightsquigarrow \mathbf{v}$.*

Proof. Define a mapping φ which assigns to $\mathbf{w}/\rho_{k,\ell}$ (for a word $\mathbf{w}$ such that $|\mathbf{w}| \geq k$) the triple $(\mathbf{u},\mathbf{v},\lambda)$, where $\mathbf{u}$ is the prefix of $\mathbf{w}$ of length $k-1$, $\mathbf{v}$ is the suffix of $\mathbf{w}$ of length $k-1$, while for an edge e directed from $\mathbf{u} = xy_1 \dots y_{r-1}$ to $\mathbf{v} = y_1 \dots y_{r-1}z$ we set $\lambda(e) = j$ ($j \geq 0$) if $xy_1 \dots y_{r-1}z$ appears in $\mathbf{w}$ as a factor $j < \ell$ times, and otherwise (if it appears at least ℓ times in $\mathbf{w}$) we define $\lambda(e) = \infty$. By the very definition of the relation $\rho_{k,\ell}$ we have $\mathbf{w}/\rho_{k,\ell} = \mathbf{w}'/\rho_{k,\ell}$ if and only if $\varphi(\mathbf{w}/\rho_{k,\ell}) = \varphi(\mathbf{w}'/\rho_{k,\ell})$; this shows that φ is both well-defined and injective.

It remains to see that φ is surjective. Indeed, let $e_1 \dots e_m$ be a covering walk $\mathbf{u} \rightsquigarrow \mathbf{v}$ for λ. Then there are words $\mathbf{u}_1, \dots, \mathbf{u}_{m-1}$ over X_n, all of length $k-1$, such that $e_1 = \mathbf{u}\mathbf{u}_1$, $e_i = \mathbf{u}_{i-1}\mathbf{u}_i$ for all $1 < i < m$, and $e_m = \mathbf{u}_{m-1}\mathbf{v}$. Let $x^{(i)}$ be the rightmost letter of $\mathbf{u}_i$, and let x' be the rightmost letter of $\mathbf{v}$. Then upon defining $\mathbf{w}_0 = \mathbf{u}x^{(1)} \dots x^{(m-1)}x'$ it is straightforward to show that $\varphi(\mathbf{w}_0/\rho_{k,\ell}) = (\mathbf{u},\mathbf{v},\lambda)$. □

Thus we instantly obtain the following upper bound.

Lemma 2.3.

$$f_n(\mathcal{T}_{k,\ell}) \leq n^{2(k-1)}(\ell+1)^{n^k} + \frac{n^k - n}{n-1} .$$

Proof. The second term on the right-hand side represents the number of all non-empty words over X_n of length at most $k-1$. On the other hand, by the previous lemma to each element of $\mathfrak{F}_n(\mathcal{T}_{k,\ell}) \cong X_n^+/\rho_{k,\ell}$ represented by a word of length $\geq k$ uniquely corresponds a triple $(\mathbf{u},\mathbf{v},\lambda)$ consisting of two vertices $\mathbf{u},\mathbf{v}$ of $\mathcal{B}(n,k-1)$ and an ℓ-labelling λ of its edges with a covering walk $\mathbf{u} \rightsquigarrow \mathbf{v}$. There are $(n^{k-1})^2$ ways to choose the pair $(\mathbf{u},\mathbf{v})$, while $(\ell+1)^{n^k}$ is the total number of all ℓ-labellings of the edges of $\mathcal{B}(n,k-1)$, so the inequality follows immediately. □

Therefore, to prove Theorem 1.1 it suffices to exhibit a lower bound on the number of triples indicated in Lemma 2.2 that is asymptotically equivalent to $n^{2(k-1)}(\ell+1)^{n^k}$ as $n \to \infty$. This is the form of our main result that we are going to prove in the next section.

3 Proof of Theorem 1.1

In the following, we restrict our attention only to those ℓ-labellings λ of $\mathcal{B}(n, k-1)$ that have a *closed* covering walk including all of its vertices; later we will argue that these suffice to attain the required lower bound. In fact, our aim will be to show that a random ℓ-labelling of a de Bruijn graph has a walk with the desired properties with high probability [1] as $n \to \infty$. To this end, the following sufficient condition comes handy.

Lemma 3.1. *For $\ell \geq 1$, let λ be an ℓ-labelling of edges of a digraph $\mathcal{D} = (V, E)$ such that $\mathcal{D}_\infty^{(\lambda)}$ is strongly connected. Then λ has a closed covering walk that includes all the vertices of $\mathcal{D}$.*

Proof. By the given conditions, $\mathcal{D}_\infty^{(\lambda)}$ does not have isolated vertices, so any covering walk of λ must include all the vertices of $\mathcal{D}$ (as it must traverse each edge of $\mathcal{D}_\infty^{(\lambda)}$ at least ℓ times). Now we describe a covering walk for λ. Let u_1 be an arbitrary vertex of $\mathcal{D}$ and let

$$e_1 = u_1 v_1, \ldots, e_i = u_i v_i, \ldots, e_M$$

be an enumeration of the edges e of $\mathcal{D}$ such that $\lambda(e) > 0$. Let E_∞ be the set of edges of $\mathcal{D}_\infty^{(\lambda)}$. Since the latter digraph is strongly connected by assumption, for any $1 \leq i \leq M$ there is a word $U_i \in E_\infty^+$ representing a path $v_i \rightsquigarrow u_i$ in $\mathcal{D}_\infty^{(\lambda)}$. Similarly, for each $1 \leq i < M$ there is a word $W_i \in E_\infty^+$ representing a path $v_i \rightsquigarrow u_{i+1}$ in $\mathcal{D}_\infty^{(\lambda)}$, and also a word W_M representing a path $v_M \rightsquigarrow u_1$.

Now let the function $\lambda' : E \to \{0, 1, \ldots, \ell - 1, \ell\}$ be defined by

$$\lambda'(e) = \begin{cases} \lambda(e) & \text{if } \lambda(e) \neq \infty, \\ \ell & \text{if } \lambda(e) = \infty, \end{cases}$$

and consider the following word from E^+:

$$\prod_{i=1}^{M} (e_i U_i)^{\lambda'(e_i) - 1} e_i W_i.$$

We claim that this word represents a (closed) covering walk for λ. Indeed, the walk represented by $e_i U_i$ is a cycle that traverses $e_i = u_i v_i$ and then returns from v_i to u_i only along edges e such that $\lambda(e) = \infty$. This cycle is traversed $\lambda'(e_i) - 1$ times and then e_i is traversed for the $\lambda'(e_i)$th time, followed by a path along the edges of $\mathcal{D}_\infty^{(\lambda)}$ from v_i to the starting vertex u_{i+1} of e_{i+1} (or back to u_1 if $i = M$). This ensures that each edge f such that $0 < \lambda(f) \leq \ell - 1$ is traversed *precisely* $\lambda'(f) = \lambda(f)$ times, while each edge e with $\lambda(e) = \infty$ is traversed *at least* $\lambda'(e) = \ell$ times, just as wanted. The walk is closed because it both starts and ends at u_1. □

So, to obtain a closed walk with the required properties, it suffices to have the 'backbone' $\mathcal{D}_\infty^{(\lambda)}$ (of a labelling λ of $\mathcal{D}$) strongly connected. We now show that this is an almost sure event for a (reasonable) random labelling in case when $\mathcal{D}$ is $\mathcal{B}(n, k-1)$.

Proposition 3.2. *Let λ be a random ℓ-labelling ($\ell \geq 1$) of the edges of the de Bruijn graph $\mathcal{B}(n, k-1)$ (with an arbitrary distribution for the labels such that the probability of assigning the label ∞ to an edge is $p > 0$). Then $\mathcal{B}(n, k-1)_{\infty}^{(\lambda)}$ is strongly connected with high probability as $n \to \infty$.*

Before we take on proving this proposition, we summarise several known facts and notions (discussed in [8]) we need in the following proof. A *cut* in a digraph $\mathcal{D} = (V, G)$ is a partition $(S, V \setminus S)$ of its vertex set such that $S \notin \{\varnothing, V\}$. The *edge boundary* ∂S with respect to this cut is the set of all edges of $\mathcal{D}$ crossing the cut *from* S *to* $V \setminus S$, that is, of all edges $e = uv$ such that $u \in S$, $v \notin S$. If $\mathcal{P}_{1/2}(V)$ denotes the set of all subsets S of V such that $|S| \leq |V|/2$, then the *isoperimetric number* of $\mathcal{D}$ is defined by

$$i(\mathcal{D}) = \min_{S \in \mathcal{P}_{1/2}(V)} \frac{|\partial S|}{|S|} .$$

The isoperimetric number of a (di)graph is a spectral index, meaning that it is closely related to the eigenvalues of the (di)graph in question. The paper of Delorme and Tillich [7] provides a lower bound on the isoperimetric number of de Bruijn digraphs (in fact, [7] considers the undirected version of de Bruijn graphs, but the directed case is easy to derive from this, see [8]). Their result is the following inequality:

$$i(\mathcal{B}(n, r)) \geq \frac{n}{4(r-1)} ,$$

holding for $r \geq 2$.

Therefore, if $k \geq 3$ and $S \in \mathcal{P}_{1/2}(V_{n,k-1})$, where $V_{n,k-1}$ is the vertex set of the de Bruijn digraph $\mathcal{B}(n, k-1)$, then

$$|\partial S| \geq \frac{n|S|}{4(k-2)} .$$

On the other hand, $\mathcal{B}(n, 1)$ is the complete digraph on n vertices (the vertex set is X_n), so it trivially holds that

$$|\partial S| \geq |S| \cdot |X_n \setminus S| \geq \frac{n|S|}{2} .$$

Hence, we have just proved the following assertion.

Lemma 3.3. *Let S be an arbitrary subset of the vertex set of $\mathcal{B}(n, k-1)$, $k \geq 2$, containing at most $n^{k-1}/2$ words. Then there exists a positive number C (depending only on k but not on n and S) such that*

$$|\partial S| \geq Cn|S|$$

holds.

Proof of Proposition 3.2. Let $P_{n,k,\ell}$ denote the probability that a random ℓ-labelling λ of $\mathcal{B}(n, k-1)$ (with the probability distribution on assigning labels to edges as specified in the formulation of the proposition) has the property that $\mathcal{B}(n, k-1)^{(\lambda)}_{\infty}$ is strongly connected. If the latter is *not* the case then there exists a cut $(S, V_{n,k-1} \setminus S)$ crossed by no edge e such that $\lambda(e) = \infty$; in other words, all edges crossing the considered cut have a finite λ-label. For this particular cut, the probability of such an event is $P(S) = (1-p)^{|\partial S|}$. If $|S| \leq n^{k-1}/2$ a lower bound for $|\partial S|$ is provided by Lemma 3.3; otherwise, if $|S| > n^{k-1}/2$, since any de Bruijn digraph is Eulerian (the out-degree and the in-degree coincide at any vertex), we must have $|\partial S| = |\partial \overline{S}|$ (where $\overline{S} = V_{n,k-1} \setminus S$), so Lemma 3.3 applies again. As $1 - p < 1$, we thus have

$$P(S) \leq (1-p)^{Cn\mu(S)}$$

for some $C > 0$, where $\mu(S) = \min\{|S|, n^{k-1} - |S|\}$.

It is now clear that

$$1 - P_{n,k,\ell} \leq \sum_{S} P(S),$$

where the sum $\mathcal{S}$ on the right-hand side ranges over all cuts of $\mathcal{B}(n, k-1)$. By denoting $s = |S|$ and $\mu(s) = \min\{s, n^{k-1} - s\}$, we obtain

$$\begin{aligned}
\mathcal{S} &\leq \sum_{s=1}^{n^{k-1}-1} \binom{n^{k-1}}{s} (1-p)^{Cn\mu(s)} = \sum_{s=1}^{n^{k-1}-1} \binom{n^{k-1}}{\mu(s)} (1-p)^{Cn\mu(s)} \leq \\
&\leq 2 \sum_{s=1}^{\left\lfloor \frac{n^{k-1}+1}{2} \right\rfloor} \binom{n^{k-1}}{s} (1-p)^{Cns} \leq 2 \sum_{s=1}^{n^{k-1}} \binom{n^{k-1}}{s} (1-p)^{Cns} = \\
&= 2\left[(1 + (1-p)^{Cn})^{n^{k-1}} - 1 \right].
\end{aligned}$$

Write $\alpha = (1-p)^{-C} > 1$. Since the sequence α^n grows asymptotically faster than n^k, it follows that

$$\begin{aligned}
0 &\leq \lim_{n\to\infty} ((1 + \alpha^{-n})^{n^{k-1}} - 1) \leq \lim_{n\to\infty} ((1 + n^{-k})^{n^{k-1}} - 1) = \\
&= \lim_{n\to\infty} \left(\left(\left(1 + \frac{1}{n^k} \right)^{n^k} \right)^{1/n} - 1 \right) = \lim_{n\to\infty} (e^{1/n} - 1) = 0.
\end{aligned}$$

Hence, $\mathcal{S} \to 0$ as $n \to \infty$, proving that $\lim_{n\to\infty} P_{n,k,\ell} = 1$, just as wanted. □

An immediate consequence is the following.

Corollary 3.4. *Let $\Lambda_{n,k,\ell}$ be the number of all ℓ-labellings of $\mathcal{B}(n, k-1)$ with the property that $\mathcal{B}(n, k-1)^{(\lambda)}_{\infty}$ is strongly connected. Then $\Lambda_{n,k,\ell} \sim (\ell+1)^{n^k}$ as $n \to \infty$.*

It remains to 'exploit' the existence of a large closed covering walk of a labelling to produce sufficiently many triples of the form described in Lemma 2.2.

Lemma 3.5. *Let λ be an ℓ-labelling of $\mathcal{B}(n,k-1)$ such that $\mathcal{B}(n,k-1)^{(\lambda)}_{\infty}$ is strongly connected (having, consequently, a closed covering walk that includes all the vertices of $\mathcal{B}(n,k-1)$). Then for each $\mathbf{u},\mathbf{v}\in X_n^{k-1}$ there is a covering walk $\mathbf{u}\rightsquigarrow\mathbf{v}$ for λ.*

Proof. Let $e_1\dots e_K$ be a closed covering walk for λ visiting all the vertices of the digraph $\mathcal{B}(n,k-1)$. The edges in this walk can be cyclically permuted so that the closed walk $\mathbf{w}=e_{i_0}\dots e_K e_1\dots e_{i_0-1}$ starts and ends at $\mathbf{u}$. On the other hand, since $\mathcal{B}(n,k-1)^{(\lambda)}_{\infty}$ is strongly connected, there is a walk $\mathbf{u}\rightsquigarrow\mathbf{v}$ along the edges e satisfying $\lambda(e)=\infty$ represented by a word $\mathbf{w}'\in E^{+}_{\infty}$. Now it is straightforward to see that the walk represented by the word $\mathbf{w}\mathbf{w}'$ has the required properties. □

Proof of Theorem 1.1. By the previous lemma, for each ℓ-labelling λ of $\mathcal{B}(n,k-1)$ such that $\mathcal{B}(n,k-1)^{(\lambda)}_{\infty}$ is strongly connected there are at least $|X_n^{k-1}|^2=n^{2(k-1)}$ covering walks for λ such that for any two of them either their initial vertices are different, or their final vertices are different. By Lemma 2.2, these walks give rise to a set of $n^{2(k-1)}$ words lying in different $\rho_{k,\ell}$-classes. The combined effect of Lemmas 2.2 and 3.1 is that the inequality

$$f_n(\mathcal{T}_{k,\ell})\geq n^{2(k-1)}\Lambda_{n,k,\ell}$$

holds, where $\Lambda_{n,k,\ell}$ has the same meaning as in Corollary 3.4. It is this corollary that supplies the desired lower bound for $f_n(\mathcal{T}_{k,\ell})$, which, along with the inequality contained in Lemma 2.3, proves the theorem. □

Acknowledgement. The research of the author is supported by the Ministry of Education and Science of the Republic of Serbia through Grant No.174019, and by a grant (Contract 114–451–2002/2011) of the Secretariat of Science and Technological Development of the Autonomous Province of Vojvodina.

References

[1] N. Alon and J. H. Spencer, *The Probabilistic Method*, Wiley, New York, 1992.

[2] J. Bang-Jensen and G. Z. Gutin, *Digraphs: Theory, Algorithms and Applications*, 2nd ed., Springer-Verlag, New York, 2010.

[3] D. Beauquier and J.-E. Pin, Factors of words, *Lecture Notes Comput. Sci.*, Vol. 372, Springer-Verlag, Berlin, 1989, pp. 63–79.

[4] J. A. Brzozowski and I. Simon, Characterizations of locally testable events, *Discrete Math.* **4** (1973), 243–271.

[5] S. Burris and H. P. Sankappanavar, *A Course in Universal Algebra*, Graduate Texts in Mathematics, Springer-Verlag, New York, 1981.

[6] N. G. de Bruijn, A combinatorial problem, *Nederl. Akad. Wetensch., Proc.* **49** (1946), 758–764.

[7] C. Delorme and J.-P. Tillich, The spectrum of de Bruijn and Kautz graphs, *Eur. J. Combin.* **19** (1998), 307–319.

[8] I. Dolinka, On free spectra of locally testable semigroup varieties, *Glasgow Math. J.* **53** (2011), 623–629.

[9] I. J. Good, Normal recurring decimals, *J. London Math. Soc.* **21** (1946), 167–169.

[10] G. Grätzer and A. Kisielewicz, A survey of some open problems on p_n-sequences and free spectra of algebras and varieties, in: *Universal Algebra and Quasigroup Theory* (Jadwisin, 1989), Heldermann Verlag, Berlin, 1992, pp. 57–88.

[11] K. Kátai-Urbán and Cs. Szabó, On the free spectrum of the variety generated by the combinatorial completely 0-simple semigroups, *Glasgow Math. J.* **49** (2007), 93–98.

[12] R. McNaughton and S. Papert, *Counter-free Automata*, M.I.T. Press, Cambridge, MA, 1971.

[13] J.-E. Pin, Finite semigroups and recognizable languages: an introduction, in: (ed. J. Fountain) *Semigroups, Formal Languages and Groups*, NATO Advanced Study Institute & Kluwer, Dordrecht, 1995, pp. 1–32.

[14] J. Ruiz, S. España, and P. Garcia, Locally threshold testable languages in strict sense: application to the inference problem, *Lecture Notes Comput. Sci.*, Vol. 1433, Springer-Verlag, Berlin, 1998, pp. 150–161.

[15] S. W. Seif, Monoids with sub-log-exponential free spectra, *J. Pure Appl. Algebra* **212** (2008), 1162–1174.

[16] A. N. Trahtman, Piecewise and local threshold testability of DFA, *Lecture Notes Comput. Sci.*, Vol. 2138, Springer-Verlag, Berlin, 2001, pp. 347–358.

[17] A. N. Trahtman, The identities of local threshold testability, in: (eds. C. Martin-Vide and V. Mitrana) *Grammars and Automata for String Processing: from Mathematics and Computer Science to Biology, and Back*, Taylor & Francis, London, 2003, pp. 347–356.

[18] A. N. Trahtman, Reducing the time complexity of testing for local threshold testability, *Theoret. Comput. Sci.* **328** (2004), 151–160.

[19] E. Vidal, F. Casacuberta, and P. Garcia, Grammatical inference and automatic speech recognition, in: *Speech Recognition and Coding*, Springer-Verlag, Berlin, 1995, pp. 175–191.