RAIRO-Theor. Inf. Appl. 46 (2012) 123–130
DOI: 10.1051/ita/2011122

Available online at:
www.rairo-ita.org

REPETITION THRESHOLDS FOR SUBDIVIDED GRAPHS AND TREES

Pascal Ochem[1] and Elise Vaslet[2]

Abstract. The *repetition threshold* introduced by Dejean and Brandenburg is the smallest real number α such that there exists an infinite word over a k-letter alphabet that avoids β-powers for all $\beta > \alpha$. We extend this notion to colored graphs and obtain the value of the repetition thresholds of trees and "large enough" subdivisions of graphs for every alphabet size.

Mathematics Subject Classification. 68R15.

1. Introduction

A non-repetitive coloring f of a graph is a vertex coloring containing no square, that is, the graph does not contain a non-intersecting path $v_1, \ldots, v_{2r}$ such that $f(v_i) = f(v_{i+r})$ for all $i \in [1, r]$.

This notion can be extended by considering repetitions of fractional exponent, as it has been done in the framework of combinatorics on words.

Up to now, the most studied problem is the following: we fix the exponent (*i.e.*, exponent 2, corresponding to squares) to be avoided and minimize the number of colors, *i.e.*, the alphabet size. For large subdivisions of graphs, this so-called non-repetitive chromatic number is proved to be 3 [6]. The non-repetitive chromatic number of trees is 4 [4].

Aberkane and Currie [1] considered the problem the other way. They fix the alphabet size and study what exponents can be avoided. They show that there

Keywords and phrases. Combinatorics on words, repetition threshold, square-free coloring.

[1] CNRS, LRI, Université Paris-Sud 11, 91405 Orsay Cedex, France. ochem@lri.fr

[2] IML, UMR 6206, Université Aix-Marseille II, Campus de Luminy, Case 907, 13288 Marseille Cedex 9, France. vaslet@iml.univ-mrs.fr

Article published by EDP Sciences

© EDP Sciences 2011

https://doi.org/10.1051/ita/2011122 Published online by Cambridge University Press

exist binary circular words of every length avoiding exponents strictly greater than 5/2. The cycle of length 5 shows that this is the best possible bound.

A notion of repetition threshold (similar to Dejean's repetition threshold [3]) can then be defined for graphs. Aberkane and Currie's result sets it to 5/2 for cycles.

We settle the problem of finding the repetition thresholds of two graph classes, namely the subdivided graphs and the trees, for every alphabet size k. This improves earlier results on nonrepetitive coloring on these classes [4, 6].

2. Preliminaries

Let $k \in \mathbb{N}$ and let f be a k-coloring of a finite graph G. We call *factor* the sequence of colors on a non-intersecting path in a k-colored graph G. We recall the usual notions of period and exponent: a word w is a repetition with period p and excess e if $w = pe$, and e is a prefix of w. The exponent of the repetition w is the ratio $\exp(w) = \frac{|pe|}{|p|}$. A word or a colored graph is said to be (α^+, n)-free if it does not contain as a factor a repetition with exponent strictly greater than α and period of length at least n. We use α^+-free as a short notation for $(\alpha^+, 1)$-free. Let G be a graph. We define the repetition threshold $RT(k, G)$ by

$$\inf_{k\text{-coloring } f} \sup \{exp(w) \mid w \text{ is a factor in the } k\text{-coloring } f \text{ of } G\}.$$

For a graph class $\mathcal{G}$, we define $RT(k, \mathcal{G}) = \sup_{G \in \mathcal{G}} RT(k, G)$. Let us express some known results with this definition: $RT(2, \mathcal{C}) = 5/2$ [1] and $RT(4, \mathcal{T}) \leq 2$ [4], where $\mathcal{C}$ the class of cycles and by $\mathcal{T}$ the class of trees. A subdivision of a graph G is a graph obtained from G by a sequence of edge subdivisions. A subdivision of an edge $\{u, v\}$ consists in the addition of a new vertex w and the replacement of the edge $\{u, v\}$ by the edges $\{u, w\}$ and $\{w, v\}$. It has been shown that for every graph G, there exists a subdivision G_s of G such that $RT(3, G_s) < 2$ [6].

The Thue-Morse word $w_{TM} = 0110100110010110100101100110 10 \ldots$ is the fixed point of $0 \to 01,\ 1 \to 10$.

Let $w[i, \ldots, j]$ denote the subword $w_i w_{i+1} \ldots w_j$ of the word w.

Finally, we denote by Σ_k the k-letter alphabet $\{0, 1, \ldots, k-1\}$.

3. Repetition threshold for subdivided graphs

For a given k, we define the real number

$$\alpha_k = \sup_G \inf_{G_s \text{ subdivision of } G} RT(k, G_s).$$

By abuse of notation, we set $RT(k, \mathcal{S}) = \alpha_k$. We can see $\mathcal{S}$ as the pseudo-class of "large enough" subdivisions of graphs.

https://doi.org/10.1051/ita/2011122 Published online by Cambridge University Press

Theorem 3.1.

- $RT(2, \mathcal{S}) = \frac{7}{3}$,
- $RT(3, \mathcal{S}) = \frac{7}{4}$,
- $RT(k, \mathcal{S}) = \frac{3}{2}$, *for* $k \geq 4$.

Proof. We first prove the upper bounds. For any graph G, we construct a subdivision G_s of G and a suitable coloring of G_s. Without loss of generality, we can suppose that G is a (large) complete graph. Let e denote the number of edges of G. For each k, we consider a set W_k containing $2e$ words over the k-letter alphabet such that all words in W_k have a non-empty common prefix and a non-empty common suffix. An edge in G is replaced by a chain of vertices of degree 2 in G_s. A vertex of degree at least 3 in G_s thus corresponds to an original vertex in G and is said *big*. Such a chain consists of two half-chains. We color each half-chain with a distinct word $w \in W_k$ such that the first letter of w corresponds to the color of a big vertex. The last vertex of an half-chain is identified with the last vertex of the other half-chain of a same chain. This common vertex is called the *center* of the chain and is colored with the common last letter of all words in W_k. A vertex is said to be *special* if it is a big vertex or a center. Two half chains sharing a special vertex are said to be *opposite* since they have opposite reading directions.

A morphism $h : \Sigma_s^* \to \Sigma_e^*$ is *synchronizing* if for any $a, b, c \in \Sigma_s$ and $v, w \in \Sigma_e^*$, if $h(ab) = vh(c)w$, then either $v = \varepsilon$ and $a = c$, or $w = \varepsilon$ and $b = c$. In order to prove upper bounds on the exponent of repetitions, it is convenient to use synchronizing uniform morphisms together with the following lemma from [5]:

Lemma 3.2. *Let* $\alpha, \beta \in \mathbb{Q}$, $1 < \alpha < \beta < 2$ *and* $n \in \mathbb{N}^*$. *Let* $h : \Sigma_s^* \to \Sigma_e^*$ *be a synchronizing* q*-uniform morphism (with* $q \geq 1$*). If* $h(w)$ *is* (β^+, n)*-free for every* α^+*-free word* w *such that* $|w| < \max\left(\frac{2\beta}{\beta-\alpha}, \frac{2(q-1)(2\beta-1)}{q(\beta-1)}\right)$, *then* $h(t)$ *is* (β^+, n)*-free for every (finite or infinite)* α^+*-free word* t.

Case $k = 2$:
We consider a set W_2 of (distinct) factors of the Thue-Morse word with prefix 0110010 and suffix 1011001. Suppose that there exists a repetition r in G_s with exponent strictly greater than $\frac{7}{3}$. If r contains at most one big vertex (resp. at most one center), then either:

- r is a factor of w_{TM} and its exponent is at most 2.
- r has period at most 6 and we check that its exponent is at most $\frac{7}{3}$.
- The period of r contains exactly one occurrence of at least one factor in $\{01001101, 10110010\}$. Those factors appear near the big vertex or the center because of the prefix 0110010 and the suffix 1011001 in the half chains, and since they are not factors of w_{TM}, we have a contradiction.

If r contains at least two big vertices and at least two centers, then two half-chains should be matched in r. This is a contradiction because half-chains are distinct.

https://doi.org/10.1051/ita/2011122 Published online by Cambridge University Press

Case $k = 3$:
We consider the synchronizing 132-uniform morphism h in [5] such that for any $\frac{5}{4}^+$-free word $t \in \Sigma_6^*$, $h(t) \in \Sigma_4^*$ is $\frac{7}{5}^+$-free. From [2], there exists an infinite $\frac{5}{4}^+$-free word $w \in \Sigma_5^*$ such that one of the letter, say 4, appears every other six positions. For a fixed integer $i \geq 1$, we define the infinite word $w_{\langle i \rangle} \in \Sigma_6^*$ obtained from w by replacing every other i occurrences of the letter 4 with the letter 5, so that the frequency of the letter 5 in $w_{\langle i \rangle}$ is $\frac{1}{6i}$. We consider a set S of $2e$ words of length c with prefix 0123 and suffix 1230, such that the i^{th} element in S is a factor of $h(w_{\langle i \rangle})$, for $1 \leq i \leq 2e$. Notice that both 0123 and 1230 appear in every factor of length 100 of the h-image of any word. Words in S are in particular $\left(\frac{7}{5}^+\right)$-free words over Σ_4.

Consider also the following synchronizing 48-uniform morphism m:

$0 \mapsto 012010201202120121020102120121012010201210212021$
$1 \mapsto 012010201202120121012021020102120121012010212021$
$2 \mapsto 012010201202101210212012101202120102012021020121$
$3 \mapsto 012010201202101210201202120121012010212021020121$

The set W_3 contains the words of the form $w = m(x)01201020120$, with $x \in S$. So the words in W_3 have length $d = 48c + 11$. Since the morphisms h and m are synchronizing, we can take d large enough to ensure that the factors of length $d/100$ of distinct words in W_3 do not contain the same number of factors $m(h(5))$, and are thus distinct. The prefixes and suffixes of length 11 and the factors of length 52 of words in W_3 contain the factor 01201, whereas words in W_3 do not contain the mirror factor 10210. Using Lemma 3.2, we can check that the words in W_3 are $\left(\frac{7}{4}^+\right)$-free and $\left(\frac{3}{2}^+, 14\right)$-free.

Suppose that there exists a repetition $r = uvu$ in G_s with exponent strictly greater than $\frac{7}{4}$, that is, $|u| > 3|v|$. Obviously, r cannot be contained in one half-chain. A computer check shows that the period $p = |uv|$ of r must be at least 68, by looking at the neighborhood of a special vertex. We first consider the case where r is contained in two (opposite) half-chains. If the common special vertex belongs to v, then $|uv| \geq 68$ and $|u| > 3|v|$ gives $|u| \geq 52$. This is a contradiction since the factor 01201 appearing in one occurrence of u would induce a forbidden factor 10210 in the word in W_3 corresponding to the opposite chain containing the other ocurrence of u. If the common special vertex belongs to u, then the restriction of r to one of the half-chain is a repetition of period $p \geq 68$ and exponent at most $\frac{3}{2}$. To complete it and obtain the repetition r, there must exist a factor u' of u with size $|u'| > p\left(\frac{7}{4} - \frac{3}{2}\right) \geq 68 \times \frac{1}{4} > 11$. So this occurrence of u' contains the factor 01201 that is near the special vertex. Since the matching occurrence of u' is contained in the opposite half-chain, we have again a contradiction with the factors 01201 and 10210. Suppose now that r contains at least two special vertices. Since words in W_3 are $\left(\frac{3}{2}^+, 14\right)$-free, we have $\frac{3}{2}|uv| \geq d$. Recall that $|u| > 3|v|$, which gives $|u| > \frac{d}{2}$ and $|r| > \frac{7}{6}d$. Then, there exists a factor u' of u such that

https://doi.org/10.1051/ita/2011122 Published online by Cambridge University Press

either $|u'| \geq 52$ and both occurrences of u' are in half-chains with opposite reading directions, or $|u'| \geq d/100$ and both occurrences of u' are in distinct half-chains with the same reading direction. In both cases, we have a contradiction.

Case $k = 4$:
We consider the set S defined in the case $k = 3$ and the following 11-uniform synchronizing morphism m:

$0 \mapsto 01320213032$
$1 \mapsto 01312023203$
$2 \mapsto 01232120323$
$3 \mapsto 01213231302$

The set W_4 contains the words of the form $w = m(x)01$, with $x \in S$. As in the case $k = 3$, we can suppose that the factors of length $d/100$ of distinct words in W_4 are distinct, where $d = 11c + 2$ is the length of words in D. The prefixes and suffixes of length 2 and the factors of length 12 of words in W_4 contain the factor 01, whereas words in W_4 do not contain the mirror factor 10. Using Lemma 3.2, we can check that the words in W_4 are $\left(\frac{3}{2}^+\right)$-free and $\left(\frac{10}{7}^+, 34\right)$-free.

Suppose that there exists a repetition $r = uvu$ in G_s with exponent strictly greater than $\frac{3}{2}$, that is, $|u| > |v|$. Obviously, r cannot be contained in an half-chain We check that the period $p = |uv|$ of r must be at least 22, by looking at the neighborhood of a special vertex. We first consider the case where r is contained in two (opposite) half-chains. If the common special vertex belongs to v, then $|uv| \geq 22$ and $|u| > |v|$ gives $|u| \geq 12$. This is a contradiction since the factor 01 appearing in one occurrence of u would induce a forbidden factor 10 in the word in W_4 corresponding to the opposite chain containing the other ocurrence of u. If the common special vertex belongs to u, then u must contain a factor 01 or 10 that cannot be matched in the other occurrence of u. Suppose now that r contains at least two special vertices. Since words in W_4 are $\left(\frac{10}{7}^+, 34\right)$-free, we have $\frac{10}{7}|uv| \geq d$. Recall that $|u| > |v|$, which gives $|u| > \frac{7}{20}d$ and $|r| > \frac{21}{20}d$. Then, there exists a factor u' of u such that either $|u'| \geq 12$ and both occurrences of u' are in half-chains with opposite reading directions, or $|u'| \geq d/100$ and both occurrences of u' are in distinct half-chains with the same reading direction. In both cases, we have a contradiction.

Now, we prove the lower bounds.

Let us call *spider* a tree with at most one vertex of degree strictly greater than two. Notice that every subdivision G_s of a spider G contains G as a subgraph. For each k, we give a spider G such that any k-coloring of G contains a repetition of exponent at least the value of $RT(k, \mathcal{S})$ given in the theorem. Every subdivision G_s of this spider G contains G as a subgraph and thus a repetition of exponent at least $RT(k, \mathcal{S})$, which proves the lower bound.

https://doi.org/10.1051/ita/2011122 Published online by Cambridge University Press

Case $k = 2$:
Every $\left(\frac{7}{3}^+\right)$-free binary word of length 4 has a prefix in $F = \{00, 11, 010, 101, 0110, 1001\}$. Consider the spider G on 13 vertices containing a vertex v of degree 4 incident to 4 paths of length 3. Without loss of generality, v gets color 0. Since there are 4 paths to color in G and only 3 words in F starting with 0, two paths are colored with the same word in F. Now, for any factor $f[1, d]$ in F, the factor $f_d f_{d-1} \dots f_2 f_1 f_2 \dots f_{d-1} f_d$ is a repetition of exponent at least $\frac{7}{3}$.

Case $k = 3$:
Dejean [3] proved that the repetition threshold of a path on 39 vertices is $\frac{7}{4}$. Such a path is a spider.

Case $k \geq 4$:
We consider the spider $G = K_{1,k}$, that is, G contains a vertex v with degree k and its neighbors. We have to color $k + 1$ vertices, v and its neighbors, using k colors only, so at least two of them get the same color. If v has the same color as one of its neighbors then we get a square, and if two neighbors of v have the same color then we get a repetition of exponent $\frac{3}{2}$. □

4. Repetition threshold for trees

Theorem 4.1.

- $RT(2, \mathcal{T}) = \frac{7}{2}$,
- $RT(3, \mathcal{T}) = 3$,
- $RT(k, \mathcal{T}) = \frac{3}{2}$, *for* $k \geq 4$.

Proof. We define the family $U_n,\ n \geq 0$ of rooted trees as follows: U_0 is a vertex, U_{n+1} is obtained from $k^n + 1$ copies of U_n by adding a root vertex adjacent to the root of each copy. The family U_n is universal, that is, every (rooted) tree is the subgraph of U_n for some n. We thus have $RT(k, \mathcal{T}) = \lim_{n\to\infty} RT(k, U_n)$. A tree is *level-colored* if it is colored such that all vertices at the same depth have the same color.

We prove now that in any k-coloring of U_n, there exists in U_n a level-colored complete binary subtree of height n. Suppose that the property holds for U_n and consider any coloring of U_{n+1}. We have $k^n + 1$ copies of U_n, each containing a level-colored binary tree, and there are at most k^n ways to color one of these binary trees. So two level-colored binary trees have the same coloring, and this creates the expected level-colored binary tree in U_{n+1}.

The repetition threshold of a k-coloring of U_n is at least the largest exponent of a repetition in a level-colored complete binary tree of height n that it contains. Moreover, the largest exponent of a repetition in a level-coloring is the same for U_n and for a binary complete tree of height n, because the set of factors is the same in these two trees.

https://doi.org/10.1051/ita/2011122 Published online by Cambridge University Press

The study of $RT(k, \mathcal{T})$ is thus reduced to that of level-colorings of complete binary trees, so we can consider the word $w = w_1 w_2 \dots$ such that w_i is the color appearing at depth $i-1$. The factors of the level-colored complete binary tree are exactly those of the form $w_{m+l} w_{m+l-1} \dots w_{m+1} w_m w_{m+1} \dots w_{m+r-1} w_{m+r}$.

We first prove the upper bounds. For each k, we consider an infinite word w over the k-letter alphabet such that for all l, r, and m, the exponent of $w_{m+l} w_{m+l-1} \dots w_{m+1} w_m w_{m+1} \dots w_{m+r-1} w_{m+r}$ is at most the expected value of $RT(k, \mathcal{T})$. Without loss of generality, we suppose that $l \leq r$.

Case $k = 2$:
We choose w to be the Thue-Morse word w_{TM}. Notice that the factor $w_{m+2} w_{m+1} w_m w_{m+1} w_{m+2}$ is in $\{10001, 01010, 11011, 00100, 10101, 01110\}$, so it is not a factor of w_{TM}. Suppose that $w_{m+l} w_{m+l-1} \dots w_{m+1} w_m w_{m+1} \dots w_{m+r-1} w_{m+r}$ is a repetition with a period of length $p \geq 2$ and exponent strictly greater than $\frac{7}{2}$ (a repetition with period of length 1 has exponent at most 3). If $l = 1$, then the exponent is $2 + \frac{1}{p} \leq \frac{5}{2} < \frac{7}{2}$. If $l \geq 2$, then the factor $w_{m+2} w_{m+1} w_m w_{m+1} w_{m+2}$ must appear in $w[m, m+r]$, which is a factor of w_{TM}. A contradiction.

Case $k = 3$:
We choose w to be the image of any ternary square-free word by the morphism $0 \mapsto 00$, $1 \mapsto 11$, $2 \mapsto 22$ which doubles every letter. The tree thus contains the cubes 000, 111, and 222, but no cube of period at least two. To see this, consider a potential cube uuu of period at least two as two overlapping occurrences of the square uu. Since w contains no square of period at least two, the leftmost occurrence of uu cannot be contained in the left branch of the cube factor. So the leftmost factor uu contains the factor bab where $a, b \in \Sigma_3$ and a is the color of the root. This is a contradiction, since the matched occurrence of bab in the rightmost factor uu is contained in the right branch of uuu but bab is not a factor of w.

Case $k = 4$:
We choose w to be any $\left(\frac{3}{2}\right)$-free word over Σ_4. Notice that w has no factors of the form aa nor aba where a and b are letters. So w contains no palindrome of size at least two. Suppose that uvu is a repetition of exponent strictly greater than $\frac{3}{2}$, that is, $|u| > |v|$, and suppose w.l.o.g that 01 is a prefix of u. The first factor u cannot contain a vertex and two of its sons, since this creates a factor aba that cannot be matched in the other factor u. So the first occurrence of u is a prefix of the left branch $w_{m+l} w_{m+l-1} \dots w_{m+1} w_m$.

Without loss of generality, w_m is the last letter of u, since otherwise the factor $w_{m+l} w_{m+l-1} \dots w_{m+2} w_{m+1} w_{m+2} \dots w_{m+r-1} w_{m+r}$ rooted at level $m+1$ is the repetition $uv'u$ where v' is the suffix of v of length $|v|-2$, so that the exponent of $uv'u$ is greater than the exponent of uvu. Thus $|u| = l+1$, and since $|v| < |u|$, we can consider the following cases on the right branch:

- $|v| = l$: the factor $w_{m+l} w_{m+l+1}$ should be 00, a contradiction.
- $|v| = l-1$: the factor $w_{m+l-1} w_{m+l} w_{m+l+1}$ should be 101, a contradiction.

https://doi.org/10.1051/ita/2011122 Published online by Cambridge University Press

- $|v| = l - t$, $t \geq 2$: the factor $w_{m+l-t+1} \dots w_{m+l}$ should be a palindrome of size at least two, a contradiction.

Now, we prove the lower bounds.

For each k, we prove that for any infinite word w over the k-letter alphabet, there exist l, r and m, such that the exponent of $w_{m+l}w_{m+l-1} \dots w_{m+1}w_m w_{m+1} \dots w_{m+r-1}w_{m+r}$ is at least $RT(k, \mathcal{T})$.

Case $k = 2$:
It is easy to check that every long enough binary word either has a factor in $F = \{000, 111, 0101, 1010, 0110110, 1001001\}$ or contains $(0011)^4$ as a factor. Moreover, for any factor $f[1, d]$ in F, the factor $f_d f_{d-1} \dots f_2 f_1 f_2 \dots f_{d-1} f_d$ is a repetition of exponent at least $\frac{7}{2}$.

Case $k = 3$:
Let F be the set of words obtained by permutations of the letter alphabet of the words in $\{00, 0101, 0121012, 012012012\}$. A computer check shows that every ternary word w of length 58 contains a factor $w[b, e]$ such that the factor $w_e w_{e-1} \dots w_{b+1} w_b w_{b+1} \dots w_{e-1} w_e$ contains a cube.
The following word of length 57 does not contain such a factor:
010201210212012021021012012010210210120120210212012102010.

Case $k \geq 4$:
It is clear that any level-coloring of the complete binary tree of height 1 contains a repetition of exponent at least $\frac{3}{2}$. □

Acknowledgements. We would like to thank an anonymous referee whose comments helped us improve the readability of the paper.

References

[1] A. Aberkane and J. Currie, There exist binary circular 5/2+ power free words of every length. *Electron. J. Comb.* **11** (2004) R10.
[2] J. Chalopin and P. Ochem, Dejean's conjecture and letter frequency. *RAIRO-Theor. Inf. Appl.* **42** (2008) 477–480.
[3] F. Dejean, Sur un théorème de Thue. *J. Combin. Theory. Ser. A* **13** (1972) 90–99.
[4] J. Grytczuk, Nonrepetitive colorings of graphs – a survey. *Int. J. Math. Math. Sci.* (2007) doi:10.1155/2007/74639.
[5] P. Ochem, A generator of morphisms for infinite words. *RAIRO-Theor. Inf. Appl.* **40** (2006) 427–441.
[6] A. Pezarski and M. Zmarz, Non-repetitive 3-Coloring of subdivided graphs. *Electron. J. Comb.* **16** (2009) N15.

Communicated by G. Richomme.
Received November 2, 2010. Accepted August 5, 2011.

https://doi.org/10.1051/ita/2011122 Published online by Cambridge University Press