Nonrepetitively 3-colorable subdivisions of graphs with a logarithmic number of subdivisions per edge

Matthieu Rosenfeld*

June 13, 2021

Abstract

We show that for every graph G and every graph H obtained by subdividing each edge of G at least $O(\log |V(G)|)$, H is nonrepetitively 3-colorable. In fact, we show that $O(\log \pi'(G))$ subdivisions per edge are enough, where $\pi'(G)$ is the nonrepetitive chromatic index of G. This answers a question of Wood and improves a similar result of Pezarski and Zmarz that stated the existence of at least one 3-colorable division with a linear number of subdivision vertices per edge.

1 Introduction

A sequence $s_1 \ldots s_{2n}$ is *a square* if $s_i = s_{i+n}$ for each $i \in \{1, \ldots, n\}$. A sequence is *repetitive* if it contains a consecutive subsequence that is a square and it is *nonrepetitive* (or *square-free*) otherwise. For instance, the words **hotshots**, **repetitive** and **alfalfa** are repetitive and the words **total** and **minimize** are nonrepetitive.

The work of Thue on nonrepetitive words is regarded as the starting point of combinatorics on words [14, 15] (see [4] for a translation in modern mathematical English). He showed that there are infinite square-free sequences over three elements. Many generalizations and variations of this notion have been studied. In particular, the notion of nonrepetitive coloring of graphs was introduced by Alon et al. [1] (see [16] for a recent survey on this topic). We say that a coloring (either of the vertices or of the edges) of a graph

*LIRMM, CNRS, Université de Montpellier

is *nonrepetitive* if the sequence of colors of any path is nonrepetitive. The *nonrepetitive chromatic number* (resp. *nonrepetitive chromatic index*) of a graph, denoted by $\pi(G)$ (resp. $\pi'(G)$) is the smallest number of colors in a nonrepetitive coloring of the vertices (resp. the edges) of the graph. Alon et al. showed that $\pi'(G)$ is in $O(\Delta^2)$ where Δ is the maximum degree of G [1]. Different authors successively improved the upper bounds on the nonrepetitive chromatic number and the nonrepetitive chromatic index and the best known bound for the nonrepetitive chromatic number is also in $O(\Delta^2)$ [5, 7, 9, 12].

Nonrepetitive coloring of subdivisions of graphs were also widely studied. We say that a graph G' is a *subdivision* of the graph G if G' is obtained by replacing each edge vw of G by a path P with endpoints vw, where the new paths are pairwise internally disjoints. If each edge is replaced by a path with at least d internal vertices then G' is a $(\geq d)$*-subdivision* of G. Barát and Wood proved that every graph has a nonrepetitively 4-colorable subdivision [2]. Pezarski and Zmarz reduced 4 to 3 [11] (solving a conjecture of Grytczuk [8]). This result is a strong generalization of Thue's result. In these two results, the number of division vertices per edge is $O(|V(G)|)$ or $O(|E(G)|)$. Djumović et al.showed that every graph has a nonrepetitively 5-colorable subdivision with $O(\log |V(G)|)$ division vertices per edge [5]. Their result is in fact stronger than that since it holds in the list-coloring setting and that their bound is in fact $O(\log \Delta)$ where Δ is the maximal degree of the graph. Finally, Wood proved that every graph has a nonrepetitively 5-colorable subdivision with $O(\log \pi(G))$ division vertices per edge [16]. It is slightly stronger since it implies the same bound of $O(\log \Delta)$, but it does not hold in the list-coloring setting and requires all the edges to be subdivided in the same amount of internal vertices. In a recent survey Wood asked the following question.

Open Problem 1. *[16, Open Question 6.21] Does every graph G have a nonrepetitively 3-colourable subdivision with $O(\log |V(G)|))$ or even $O(\log \pi(G))$ division vertices per edge?*

In this article, we give a positive answer to the first part of this question. In fact, we show that there exists a function $f(n) = O(\log n)$ such that any subdivision of any graph G with a least $f(\pi'(G))$ subdivision per edge is nonrepetitively 3-colorable. Since $\pi'(G)$ is in $O(\Delta^2)$, the quantity $O(\log \pi'(G)))$ is smaller than $O(\log \Delta))$ which is itself smaller than the suggested $O(\log |V(G)|))$. However, this is not clear how $O(\log \pi'(G)))$ compares to $O(\log \pi(G))$ so we are not able to solve the second part of this question. The number of subdivision vertices per edge that we require is $558 \log_2(n) + O(1)$. This result is optimal in the sense that $\Omega(\log n)$ division vertices are needed on some edges of any nonrepetitively $O(1)$-colorable subdivision of K_n[10]. However, we expect the optimal multiplicative coefficient in front of the log to be much smaller than 558.

Moreover, we show that any subdivision is nonrepetitively colorable as long as each edge is subdivided enough. This is much stronger than showing that there exists one nonrepetitively colorable subdivision. Remark that all the aforementioned results only showed the existence of one nonrepetitively colorable subdivision. Our proof becomes much simpler if we only care about the existence of one nonrepetitively 3-colorable subdivision.

The article is organized as follows. We first provide in Section 2 some definitions and recall some useful results from the literature. Then in Section 3, we show the existence of a set of words that is needed for our main construction. In particular, we show that there are sets of n-good words of exponential size. In Section 4, we use these sets of n-good words to show our main result. The main idea is to start from a nonrepetitive coloring of the edges of a graph and to "encode" each color by a well chosen n-good word. In the last section, we discuss possible ways to improve the bound on the number of needed subdivisions per edge.

The main idea is implicitly to generalize the notion of square-free morphism. A *morphism* is a map $h : \Sigma^* \to \Sigma^*$ such that the image of each word is given by the concatenation of the image of the letters. A *square-free morphism* is a morphism such that the image of every square-free word is a square-free word. A directed graph is nonrepetitively colored if the sequence of colors of every directed path is nonrepetitive. Given a directed graph D, two sets of colors C and Σ, a nonrepetitive edge coloring $\phi : A(D) \to C$ of D and a square-free morphism $h : C \mapsto \Sigma$. If we can subdivide D in such a way that the sequence of colors of the subdivision of any edge e is $h(\phi(e))$, then this subdivision is nonrepetitively $|\Sigma|$-colorable[1]. We cannot use the same construction for undirected graphs. But if we can find a square-free morphism such that the image of every square-free word is square-free even if we replace the image of a letter by its mirror at some arbitrary positions, then we can use the same idea. There are extra technicalities, since we show that it works for any large enough subdivision. Naturally, the proof mostly relies on combinatorics on words.

2 Preliminaries

A *word* is a finite sequence over a finite set that we call the *alphabet*. A *factor* of a word is a contiguous subsequence of this word, that is, if there are two

[1] We might need one extra color to color the vertices from the original graph.

words p and s such that $w = pfs$ then f is a factor of w. If p (resp. s) is empty then f is also a *prefix* (resp. *a suffix*) of w. A prefix (resp. a suffix) of w is *proper* if it is not equal to w. The *length* of a word u is denoted by $|u|$. A word u *occurs* in v at position p, if the factor of length $|u|$ of v that starts at position p is exactly u. The *mirror image* $\overline{w}$ of a word w is the word obtained by reading w from right to left. We let $\overline{w}^1 = \overline{w}$ and $\overline{w}^0 = w$.

We recall the following result from Shur on the number of square-free ternary words.

Theorem 1 ([13]). *For all n, let $C_{sq}(n)$ be the number of ternary square-free words. Then*

$$\limsup_{n\to\infty} C_{sq}(n)^{\frac{1}{n}} \geq 1,30175907\,.$$

Since any factor of a square-free word is square-free, C_{sq} is a submultiplicative function (i.e., we have $C_{sq}(i+j) \leq C_{sq}(i)C_{sq}(j)$, for all i, j). By Fekete's Lemma we deduce the following Corollary of Theorem 1.

Corollary 1. *For all integer $n \geq 1$, $C_{sq}(n) > 1.3^n$.*

We now recall Turán's Theorem and a simple corollary.

Theorem 2 (Turán's Theorem). *Let G be any graph with n vertices, such that G is K_{r+1}-free. Then the number of edges in G is at most*

$$|E(G)| \leq \left(1 - \frac{1}{r}\right)\frac{n^2}{2}\,.$$

Corollary 2. *Any graph G contains independent set of size at least $\frac{n}{1+d(G)}$ where $d(G)$ is the average degree of G.*

Proof. Let G be a graph of average degree $d(G)$ and let H be the complement of G. Then the number of edges of H is

$$\begin{aligned}|E(H)| = \frac{n(n-1)}{2} - |E(G)| = \frac{n(n-1)}{2} - \frac{nd(G)}{2} &= \frac{n^2}{2}\left(1 - \frac{1+d(G)}{n}\right) \\ &> \frac{n^2}{2}\left(1 - \frac{1}{\frac{n}{1+d(G)} - 1}\right)\,.\end{aligned}$$

By Turán's Theorem, this implies that H cannot be $K_{\frac{n}{1+d(G)}}$-free. Hence, G contains an independent set of size $\frac{n}{1+d(G)}$ as desired. □

3 Good sets

Let $\sigma = 1202120121021201021$, $\overline{\sigma} = 1201021201210212021$ be the mirror image and let $\rho = \overline{\sigma}0\sigma$. By construction p is a palindrome, $|\sigma| = 19$ and $|\rho| = 39$.

A word v is *nice* if $\rho v \rho$ is square-free and contains only two occurrences of σ and two occurrences of $\overline{\sigma}$. The only occurrences of σ and $\overline{\sigma}$ are inside the two occurrences of ρ. Let $\mathcal{N}$ be the set of nice words and for any integer n, let $\mathcal{N}_n$ be the set of nice words of length n.

For all $n \geq 8750$, let l_n be the lexicographically least word of the set $\mathcal{N}_n$. We show in Lemma 7 that there exist nice words of every length at least 8750. So l_n is properly defined for any $n \geq 8750$.

A set of words S is *mirror-free* if for any word w from S such that $w \neq \overline{w}$, $\overline{w}$ is not in S. For any integer $n \geq 8750$, a set of words $\mathcal{S} \subseteq \mathcal{N}_n$ is *n-good* if it is mirror-free and if for all $i \in \{2n+100, \ldots, 7n\}$ and all $u, v \in \mathcal{S}$ such that $u \neq v$ the words $\rho u \rho v \rho l_i$, $\rho \overline{u} \rho v \rho l_i$, $\rho u \rho \overline{v} \rho \overline{l_i}$ and $\rho \overline{u} \rho \overline{v} \rho \overline{l_i}$ are square-free.

In Lemma 2, we show the central property of n-good set, but first we show the following Lemma as a warm-up exercise.

Lemma 1. *Let $n \geq 8750$ be an integer, S be an n-good set, $u, v \in S$ and $i \in \{2n+100, \ldots, 7n\}$ be an integer then $\rho u \rho l_i \rho v \rho$ is square-free.*

Proof. Suppose, for the sake of contradiction, that $\rho u \rho l_i \rho v \rho$ contains a square ww. First remark, that since S is an n-good set $\rho u \rho l_i \rho$ and $\rho l_i \rho v \rho$ are both square-free. Hence there exist a non-empty suffix x of ρu and a non-empty prefix y of $v\rho$ such that $x \rho l_i \rho y$ is a square.

Now $l_i \rho$ is not a factor of w since it contains only one occurrence of ρ and that l_i is much longer than the gap between any other occurrences of ρ. For the same reason ρl_i is not a factor of w. There exists x', y' such that $l_i = x'y'$ and $w = x\rho x' = y'\rho y$. Since $|x'| + |y'| = |l_i| \geq 2n + 100$, assume, without loss of generality that $|x'| \geq n + 50$. Then $|x'| > |y|$ and xp is a proper prefix of $y'\rho$ which implies that there is a second occurrence of ρ in $y'\rho$. This contradicts the fact that l_i is a nice word. $\square$

Lemma 2. *Let k and n be two integers. Let S be an n-good set, Σ be an alphabet and $f : \Sigma \mapsto S$ be an injective map. Then, for any square-free word $w_1 \ldots w_k \in \Sigma^k$, any sequence of integers $(s_i)_{1 \leq i \leq k} \in \{2n+100, \ldots, 7n\}^k$ and any sequence $(r_i)_{1 \leq i \leq k} \in \{0, 1\}^k$ the word*

$$\rho \prod_{i=1}^{k} \overline{f(w_i)\rho l_{s_i} \rho f(w_i)}^{r_i} \rho = \rho \overline{f(w_1)\rho l_{s_1} \rho f(w_1)}^{r_1} \ldots \rho \overline{f(w_k)\rho l_{s_k} \rho f(w_k)}^{r_k} \rho$$

is square-free.

Proof. Suppose, for the sake of contradiction, that there is a square uu in $\rho\prod_{i=1}^{k}\overline{f(w_i)\rho l_{t_i}\rho f(w_i)}^{r_i}\rho$. There are $x_1,\ldots,x_{3k}\in\mathcal{N}_n$ such that for all i, $\overline{f(w_i)\rho l_{s_i}\rho f(w_i)}^{r_i}=x_{3i-2}\rho x_{3i-1}px_{3i}$. This implies that

$$\rho\prod_{i=1}^{k}\overline{f(w_i)\rho l_{s_i}\rho f(w_i)}^{r_i}\rho=\rho x_1\rho x_2\rho\ldots\rho x_{3k}\rho$$

and that there are no other occurrences of σ and $\overline{\sigma}$ than the $3k+1$ occurrences that are inside the occurrences of ρ.

Let $L=\{l_i : i\in\{2n+100,\ldots,7n\}\}$. Then for any j, $x_j\in L$ if and only if $j\equiv 2 \mod 3$. Since S is mirror-free, for any i, x_i and x_{i+1} are different. Thus, by the definition of n-good sets and by Lemma 1, for any i, $\rho x_i\rho x_{i+1}\rho x_{i+2}\rho$ is square-free. Thus there are at least 3 occurrences of ρ in uu. At least one of σ or $\overline{\sigma}$ appears twice in u. Assume, without loss of generality, that the middle of the square uu does not cut any occurrence of σ (if it is not the case we can use the same argument with $\overline{\sigma}$ instead). Then σ occurs at least twice in u.

Let $l\geq 2$ be the number of occurrences of σ in u. Since the explicit occurrences of σ are the only occurrences, and that there are as many occurences of σ in the two consecutives occurences of u, there exist an integer i and words y,y',z,z' such that y is a proper suffix of $\sigma x_{i-l}\overline{\sigma}0$, $y'z=x_i\overline{\sigma}0$, z' is a prefix of $x_{i+l}\overline{\sigma}0$ and

$$\begin{aligned}u&=y\sigma x_{i+1-l}\overline{\sigma}0\sigma x_{i+2-l}\overline{\sigma}0\sigma\ldots\overline{\sigma}0\sigma x_{i-1}\overline{\sigma}0\sigma y'\\&=z\sigma x_{i+1}\overline{\sigma}0\sigma x_{i+2}\overline{\sigma}0\sigma\ldots\overline{\sigma}0\sigma x_{i+l-1}\overline{\sigma}0\sigma z'\,.\end{aligned}$$

Moreover, the occurences of σ have to match each others and *synchronise* the rest of u, that is, $y=z$, $y'=z'$ and for all $j\in\{0,\ldots,l-2\}$,

$$x_{i+1-l+j}=x_{i+1+j}\,. \tag{1}$$

Recall that, for any j, $x_j\in L$ if and only if $j\equiv 2 \mod 3$.

If $l=2$, then $x_{i-1}=x_{i+1}$ which implies $i\equiv 2 \mod 3$. Moreover, $y=z$ and $y'=z'$ imply

$$|x_i\overline{\sigma}0|=|y|+|z'|<|\sigma x_{i-l}\overline{\sigma}0|+|x_{i+l}\overline{\sigma}0|=|x_{i-l}|+|x_{i+l}|+59\,.$$

But since, $i\equiv 2 \mod 3$, $|x_i|\geq 2n+100$ and $|x_{i+1}|=|x_{i-1}|=n$ which contradicts the previous equation.

Let us now take care of the case $l\geq 3$. We first show that $l\equiv 0 \mod 3$. For all $j\in\{0,\ldots,l-2\}$, $i+1-l+j\equiv 2 \mod 3$ if and only if $i+1+j\equiv 2 \mod 3$. Thus if $i+1-l\equiv 2 \mod 3$ or $i+2-l\equiv 2 \mod 3$ then l has to be divisible by 3. If it is not the case then $i+1-l \equiv 0 \mod 3$, $|x_{i+1-l}| = |x_{i+2-l}| = n$ which implies $|x_{i+1}| = |x_{i+2}| = n$ and $i+1 \equiv 0 \mod 3$ and finally $l \equiv 0 \mod 3$.

We have three different cases to consider.

Case $i+1-l \equiv 1 \mod 3$**:** First recall that, for every $j \equiv 1 \mod 3$, $x_j = \overline{f(w_{(j+2)/3})}^{(j+2)/3}$. Thus for all integer j, $x_{i+1+3j-l} = \overline{f(w_{(i-l)/3+1+j})}^{r_{(i-l)/3+1+j}}$ and $x_{i+1+3j} = \overline{f(w_{i/3+1+j})}^{r_{i/3+1+j}}$. By equation (1), for all $j \in \{0, \ldots, l/3-1\}$,

$$\overline{f(w_{(i-l)/3+1+j})}^{r_{(i-l)/3+1+j}} = \overline{f(w_{i/3+1+j})}^{r_{i/3+1+j}} .$$

The function f is an injective and maps to S which is mirror-free, so the previous equation implies that for all $j \in \{0, \ldots, l/3-1\}$,

$$w_{(i-l)/3+1+j} = w_{i/3+1+j} .$$

This implies that there is a square in w which is a contradiction.

Case $i+1-l \equiv 0 \mod 3$**:** We use the same idea as in the previous case with the fact that for every $j \equiv 0 \mod 3$, $x_j = \overline{f(w_{j/3})}^{r_{j/3}}$. In this case we obtain that for all $j \in \{0, \ldots, l/3-1\}$,

$$w_{(i+1-l)/3+j} = w_{(i+1)/3+j} .$$

This implies that there is a square in w which is a contradiction.

Case $i+1-l \equiv 2 \mod 3$**:** This case is almost identical to the previous ones. We know that for every $j \equiv 0 \mod 3$, $x_j = \overline{f(w_{j/3})}^{r_{j/3}}$. Moreover for all j, $i+2-l+3t \equiv 0 \mod 3$. By the same argument, for all $j \in \{0, \ldots, l/3-1\}$,

$$w_{(i+1-l)/3+j} = w_{(i+1)/3+j} .$$

This implies that there is a square in w which is a contradiction. □

This property is essential to construct the nonrepetitive coloring of a subdivided graph. The idea is to encode the colors of the edges of a nonrepetitive edge coloring of the initial graph. Any vertex from the initial graph will be colored by 0 and the path corresponding to any edge colored c in the original graph should receive the color sequence $\sigma f(c)\rho l_i \rho f(c)\overline{\sigma}$ (with the l_i of the right length). The fact that we can choose a different l_i for every edge means that we can find a right encoding as long as the edge is subdivided enough. The fact that we can replace the encoding of each w_i by the mirror image of the encoding means that we can take an arbitrary orientation of the subdivided edge to apply the encoding to the corresponding path.

We also need a variant of this property. This variant will be useful for paths that start and end in the subdivision of the same edge (i.e., paths that appear in the subdivision of a cycle).

Lemma 3. *Let k and n be two integers. Let S be an n-good set, Σ be an alphabet and $f : \Sigma \mapsto S$ be an injective map. Let $w_1 \dots w_k \in \Sigma^k$ be a square-free word such that $w_2 w_3 \dots w_k w_1$ is also square-free. Let $(t_i)_{1\le i\le k} \in \{2n + 100, \dots, 7n\}^k$ be a sequence of integers, and $(r_i)_{1\le i\le k} \in \{0,1\}^k$ be a sequence of 0 and 1. Let a and b be a pair of words such that $\overline{f(w_1)\rho l_{t_1} \rho f(w_1)}^{r_1} \rho = ab$. Then the word*

$$
\begin{aligned}
&b \left(\prod_{i=2}^{k} \overline{f(w_i)\rho l_{t_i} \rho f(w_i)}^{r_i} \rho \right) a \\
&= b\rho \overline{f(w_2)\rho l_{t_2} \rho f(w_2)}^{r_2} \rho \overline{f(w_2)\rho l_{t_2} \rho f(w_2)}^{r_2} \rho \dots \rho \overline{f(w_k)\rho l_{t_k} \rho f(w_k)}^{r_k} \rho a
\end{aligned}
$$

is square-free.

Proof. Suppose, for the sake of contradiction, that there is a square uu in $b \left(\prod_{i=2}^{k} \overline{f(w_i)\rho l_{t_i} \rho f(w_i)}^{r_i} \rho \right) a$.

Since $w_1 w_2 \dots w_k$ and $w_2 w_3 \dots w_k w_1$ are both square-free, Lemma 2 implies that $b \left(\prod_{i=2}^{k} \overline{f(w_i)\rho l_{t_i} \rho f(w_i)}^{r_i} \rho \right)$ and $\left(\prod_{i=2}^{k} \overline{f(w_i)\rho l_{t_i} \rho f(w_i)}^{r_i} \rho \right) a$ are also square-free. Hence the first occurrence of u starts in b and the second occurrence ends in a. Assume, without loss of generality, that the middle of the square does not cut an occurrence of σ. Since $k > 3$, we know that uu contains enough occurrences of σ to be synchronized by the occurrences of σ and this implies that the number of occurrences of σ in u is a multiple of 3. With the same argument as in the proof of Lemma 2 one easily shows that for every $i \in \left\{1, \dots, \frac{k}{2} - 1\right\}$,

$$
w_{i+1} = w_{k/2+i+1} \, .
$$

We also easily verify that $\overline{f(w_{k/2+1})\rho l_{t_{k/2+1}} \rho f(w_{k/2+1})}^{r_{k/2+1}}$ is a factor of $ab = \overline{f(w_1)\rho l_{t_1} \rho f(w_1)}^{r_1} \rho$. This is only possible if $w_1 = w_{k/2+1}$. Thus, for every $i \in \left\{0, \dots, \frac{k}{2} - 1\right\}$,

$$
w_{i+1} = w_{k/2+i+1} \, .
$$

This is a contradiction since $w_1 \dots w_k$ is square-free. $\square$

3.1 Exponentially large good sets

We show in this subsection that there are exponentially many nice words of any length and we use that to show that there are exponentially large n good sets.

Let $h : \Sigma \mapsto \Sigma^*$, be the map such that

$$\begin{aligned}
h(0) =& \{012102120210201021201210, 0121021202102012021201210\}\,,\\
h(1) =& \{120210201021012102012021, 1202102010210120102012021\},\\
h(2) =& \{201021012102120210120102, 2010210121021201210120102\}\,.
\end{aligned}$$

A word v is an *image* of w by h if v can be obtained by replacing each occurrence of any letter i of w by any word of the corresponding set $h(i)$. The set of images of w by h is denoted by $h(w)$. The authors of [3] introduced h and showed the following property.[2]

Theorem 3 ([3, Theorem 20])**.** *For any square-free word $w \in \{0,1,2\}^*$ and any $v \in h(w)$, v is square-free.*

We will use h to show that the set of nice words has exponential growth. First, we need a few simple facts about h, σ and ρ.

Lemma 4.
1. *For any letters $a, b, c \in \{0,1,2\}$, $v \in h(ab)$ and $v' \in h(c)$, the word v' cannot occur as an internal factor of v (i.e., the only possible occurrences of v' are as suffix or prefix).*
2. *For any $a, b \in \{0,1,2\}$ and $v \in h(ab)$, neither σ or $\overline{\sigma}$ are factor of v.*
3. *If $a \in \{1,2\}$, $b \in \{0,1,2\} \setminus \{a\}$ and $v \in h(0ab)$, then the word ρv is square-free and contains exactly one occurrence of σ and $\overline{\sigma}$.*
4. *If $a \in \{1,2\}$, $b \in \{0,1,2\} \setminus \{a\}$ and $v \in h(ba0)$, then the word $v\rho$ is square-free and contains exactly one occurrence of σ and $\overline{\sigma}$.*

It is a bit tedious to verify the four claims of this Lemma by hand so we provide a simple computer program that verifies Lemma 4[3]. The first of this 4 facts implies that images by h synchronize, that is, as long as a factor of an image is of length at least 50 (the length of the images of two letters), there is a unique way to split it into different images by h. This kind of properties are really useful to establish that an image of a square-free word by h does

[2]In fact, the map that they used contains 4 words in each set and here we used only 2 of them.

[3]See the ancillary file `verifying_lemma4.cpp`.

not contains any large squares (i.e., squares that are large enough to allow us to use this synchronization property). On the other hand, the other facts are quit useful to establish that there are no short squares. With this lemma in hand it is relatively simple to provide a proof of Theorem 3 (which we will not do). We can use these facts to show that there are exponentially many nice words.

Lemma 5. *Let $w \in \{0,1,2\}^*$ be a word such that $|w| \geq 4$ and $10w01$ is square-free. Then any $v \in h(0w0)$ is nice.*

Proof. Let w and v be as in the Theorem statement. Let $w_1 \dots w_n = w$ with $w_1, \dots, w_n \in \{0,1,2\}$ and for all i let $v_i \in h(w_i)$ such that $v = v_1 \dots v_n$. Recall that we need to show that $\rho v \rho$ is square-free and contains only two occurrences of σ and two occurrences of $\overline{\sigma}$. By 2., 3. and 4. of Lemma 4, any occurrence of σ or $\overline{\sigma}$ is in ρ, so it only remains to show that $\rho v \rho$ is square-free.

Suppose, for the sake of contradiction, that there is a square in $\rho v \rho$, that is, there are words $x, y \in \{0,1,2\}^*$ and $u \in \{0,1,2\}^+$ such that $\rho v \rho = xuuy$. Theorem 3 implies that uu cannot be a factor of v. One easily verifies that ρ is square-free. Thus x is a proper prefix of ρ or y is a proper suffix of ρ. Assume, without loss of generality, that x is a proper suffix of ρ. Let r be the nonempty suffix of ρ such that $xr = \rho$. Fact 3. of Lemma 4 implies that the square uu is not a factor of $\rho v_1 v_2 v_3$, thus $|uu| \geq |v_1 v_2 v_3| + 2$ and v_1 is a factor of u. Thus rv_1 is a prefix of u. We have to distinguish between two different cases depending on the length of r.

Case $|r| \geq 4$: By hypothesis $|w| \geq 4$ and $|v_2 \dots v_n| \geq 5 \times 24 > |v_1| + 2|\rho| \geq |v_1 r \rho|$ which gives $|rv_1| + \frac{|r+v+\rho|}{2} < |rv|$. We deduce that the $|rv_1|$ first letters of the second occurrence of u do not overlap with the final occurrence of ρ. This implies that rv_1 is a factor of v. From Fact 1. of Lemma 4, v_1 can only appear as the image of 0 and thus r must appear as the suffix of the image of a letter. However, it is easy to verify that r is not the suffix of any image of a letter if $|r| \geq 4$ (it is enough to verify this with $|r| = 4$), which is a contradiction.

Case $|r| \leq 3$: Then r is also the suffix of an image of 1 by h (since 021 is suffix of any image of 1). There is $v_0 \in h(1)$ such that uu is also a factor of $v_0 v \rho$. By hypothesis w was chosen such that $10w01$ is square-free and Theorem 3 implies that there is no square in $v_0 v$. Thus the square in $v_0 v \rho$ overlaps with the final occurrence of ρ. By symmetry of the previous case, the square overlaps by at most 3 letters with the final occurence of ρ which implies that the square is also an image of $10w01$ which is a contradiction since any image of $10w01$ by h is square-free by Theorem 3. □

We can deduce an exponential lower bound on the size of $\mathcal{N}_n$.

Theorem 4. *For all $n > 8750$,*

$$|\mathcal{N}_n| > \frac{1.01^n}{7.8}\,.$$

Proof. Let P be the set of words over $\{0,1,2\}$ such that for any $w \in P$, $|w| \geq 4$, $10w01$ is square-free. Lemma 5 implies that for all n

$$|\mathcal{N}_n| \geq |\{v \in \{0,1,2\}^n | \exists w \in P, v \in h(0w0)\}|\ .$$

The set of images by h of two different words are distinct, hence

$$|\mathcal{N}_n| \geq |\{w \in P | h(0w0) \cap \{0,1,2\}^n \neq \emptyset\}|\ . \tag{2}$$

Every letter has an image of length 24 and an image of length 25 by h. Thus for any integer $n \geq 14 \times 25^2 = 8750$ and any word w of length $\frac{n}{24} - 13 \leq |w| \leq \frac{n}{24}$, w admits at least an image of size n by h. That is, for any $n \geq 8750$,

$$\{w \in P | h(0w0) \cap \{0,1,2\}^n \neq \emptyset\ \}| \geq \left|\left\{w \in P | \frac{n}{24} - 13 \leq |w| \leq \frac{n}{24}\right\}\right|\ .$$

By symmetry, there are exactly $\frac{C_{sq}(n)}{6}$ square-free words over $\{0,1,2\}$ of length n starting by 10. Moreover, one easily verifies that every square-free word over $\{0,1,2\}$ of length 14 contains at least one occurrence of 01. Thus for every integer n, there are at least $\frac{C_{sq}(n)}{6}$ ternary square-free words of length between $n-13$ and n starting with 10 and ending with 01.

We can now apply Corollary 1,

$$|\{w \in P | h(0w0) \cap \{0,1,2\}^n \neq \emptyset\ \}| \geq \frac{C_{sq}(\lfloor \frac{n}{24} \rfloor)}{6} > \frac{1.3^{\lfloor \frac{n}{24} \rfloor}}{6} > \frac{1.01^n}{7.8}\,.$$

Together with equation (2), we conclude $|\mathcal{N}_n| > \frac{1.01^n}{7.8}$. □

We use the exponential lower bound to establish the existence of exponentially large n-good sets, but first we show one more property of nice words.

Lemma 6. *Let n be a positive integer and $u, v \in \mathcal{N}$. If the word $\rho u \rho v \rho$ is not square-free then u is a prefix of v or v is a suffix of u. In particular, if $|u| = |v|$ then $u = v$.*

Proof. Let n be a positive integer and $u, v \in \mathcal{N}_n$ such that $u \neq v$. Let ww be a square in $\rho u \rho v \rho$. Since u and v are nice, $\rho u \rho$ and $\rho v \rho$ are square-free. Hence the second occurrence of ρ is a factor of ww. We also know that the only occurrences of σ in $\rho u \rho v \rho$ (resp. of $\overline{\sigma}$) are the three occurrences inside each occurrence of ρ.

Suppose, for the sake of contradiction, that there are two non-empty words u_1 and u_2 such that $u = u_1 u_2$, w is a suffix of ρu_1 and a prefix of $u_2 \rho v \rho$. Since w contains ρ as a factors and that there are exactly three occurrences of ρ this implies that $w = \rho u_1$ which is a contradiction with the fact that $u_2 \rho v \rho$ does not starts with ρ. By symmetry, we reach a similar contradiction if we try to split v in $v_1 v_2$.

Hence there exists ρ_1, ρ_2 such that $\rho_1 \rho_2 = \rho = \overline{\sigma} 0 \sigma$ and w is a suffix of $\rho u \rho_1$ and a prefix of $\rho_2 v \rho$. Assume, without loss of generality, that σ is a suffix of ρ_2 (otherwise $\overline{\sigma}$ is a prefix of ρ_1 and the rest of the argument is symmetric). Since the only occurrence of σ in $\rho u \rho_1$ is inside ρ, we deduce that $\rho_2 u \rho_1 = w$. Since w is a prefix of $\rho_2 v \rho$, u is a prefix of v or v is a prefix of u. $\square$

We can finally show the existence of exponentially large n-good sets.

Lemma 7. *For any $n > 8750$, there exists an n-good set S of size at least*

$$|S| \geq \frac{1.01^n}{16(60n^2 + 1)} .$$

Proof. Let $\mathcal{N}'_n$ be the set obtained by removing from $\mathcal{N}_n$ any prefix or suffix of every l_i with $i \in \{2n + 100, \dots, 7n\}$ and by keeping for each pair of mirror images only the lexicographically smallest of the two. Each l_i is responsible for removing at most two words from $\mathcal{N}_n$ so $|\mathcal{N}'_n| \geq \frac{|\mathcal{N}_n| - 10n}{2} \geq \frac{1.01^n}{15.6} - 5n$. Since $n > 8750$, we can simplify the bound

$$|\mathcal{N}'_n| \geq \frac{1.01^n}{16} .$$

For any $u, v \in \mathcal{N}'_n$, we say that u forbids v if $u \neq v$ and for some $i \in \{n + 1, n + 2, \dots, 5n\}$, $\rho u \rho v \rho l_i \rho$, $\rho \overline{u} \rho v \rho l_i$, $\rho u \rho \overline{v} \rho \overline{l_i} \rho$ or $\rho \overline{u} \rho \overline{v} \rho \overline{l_i} \rho$ contains a square.

We now count how many words v are forbidden by a given u.

Let $u, v \in \mathcal{N}'_n$ be such that $u \neq v$ and $\rho u \rho v \rho l_i \rho$ is not square-free. Lemma 6 implies that both $\rho u \rho v \rho$ and $\rho v \rho l_i \rho$ are square-free (since $u \neq v$ and v is not a prefix of l_i). There is a non-empty suffix u' of ρu and a non-empty prefix l' of $l_i \rho$ such that the square $ww = u' \rho v \rho l'$. Moreover, there exist two non-empty words v_1 and v_2 such that $\sigma v \overline{\sigma} = v_1 v_2$ and $w = u' \overline{\sigma} 0 v_1 = v_2 0 \sigma l'$.

Indeed, the middle of the square cannot be located outside of $\sigma v\overline{\sigma}$ since there would be too many occurences of σ or $\overline{\sigma}$ on one side of the square. Finally, remark that either v_1 contains σ as a prefix or v_2 contains $\overline{\sigma}$ as a suffix (both could be true). In both cases, using the fact that there are only two other occurrences of σ and $\overline{\sigma}$, we deduce that $|w| = |v\rho|$. Thus v_1 is a prefix of $0\sigma l_i\rho$ and v_2 is a suffix of $\rho u\overline{\sigma}0$ and v is uniquely determined by u, l_i and the position of the square. There are $|v\rho| = n + 39$ possible positions, less than $5n$ possibles values for l_i, so u forbids at most $(n + 39) \times 5n$ words because of $\rho u\rho v\rho l_i$. The count is similar for $\rho\overline{u}\rho v\rho l_i$, $\rho u\rho\overline{v}\rho\overline{l_i}\rho$ and $\rho\overline{u}\rho\overline{v}\rho\overline{l_i}\rho$, so u forbids at most $(n+39)\times 5n\times 4$ words. This is upper bounded by $30n^2$ since $n > 8750$.

Let G be the graph whose vertices are the words from $\mathcal{N}'_n$ and such that two words share an edge if one of them forbids the other one. The set of words corresponding to any independent set of G is an n-good set. Let S be the set of words corrsponding to the largest independent set of G. Since every word forbids at most $30n^2$ words, the average degree of the vertices of G is at most $60n^2$. By Corollary 2, there is an independent set of size at least $\frac{|\mathcal{N}'_n|}{60n^2+1}$. Thus $|S| \geq \frac{1.01^n}{16(60n^2+1)}$. □

4 The final construction

A graph G' is a $(\geq a, \leq b)$-*subdivision* of a graph G if G' can be obtained by subdividing each edge of G in at least a and at most b division vertices.

In Lemma 8, we use our results on n-goods sets to show that, if each edge of the graph is subdivided enough, but not too much, then we can nonrepetitively 3-color the resulting graph. To obtain Theorem 5, we then show that we can easily handle the edges that have too many subdivision vertices.

Lemma 8. *Let G be a graph and $n \geq 8750$ an integer such that $\pi'(G) \leq \frac{1.01^n}{16(60n^2+1)}$. Then for any $(\geq 4n + 216, \leq 9n)$-subdivision G' of G,*

$$\pi(G') = 3\,.$$

Proof. Let n, G and G' be as in the lemma statement. Let C be a set of colors of size $\pi'(G)$ and ϕ be a nonrepetitive edge C-coloring of G.

By Lemma 7, there is an n-good set S such that $|S| \geq \pi'(G)$. Let f be an injective map from C to S. Let $\vec{\sigma}$ be an arbitrary orientation of the edges of G.

Let $\phi' : V(G') \mapsto \{0, 1, 2\}$ be the 3-coloring of the vertices of G' such that

- the color of every vertex of G' that corresponds to an original vertex of G has color 0,
- for any edge e from G subdivided in $(v_1, \ldots, v_k)$ in G' with the v_i ordered according to $\vec{o}(e)$, the sequence of colors $(\phi'(v_1), \ldots, \phi'(v_n))$ is equal to $\sigma f(\phi(e))\rho l_{k-116-2n}\rho f(\phi(e))\overline{\sigma}$.

Remark that $|\sigma f(\phi(e))\rho| + |pf(\phi(e))\overline{\sigma}| = 2|\rho| + 2|\sigma| + 2n = 2n + 116$ and thus $|\sigma f(\phi(e))\rho l_{k-116-2n}\rho f(\phi(e))\overline{\sigma}| = k$. Thus ϕ' is well-defined. Our goal is now to show that ϕ' is nonrepetitive.

First remark that since every edge of G is subdivided at least $4n + 216$ times and at most $9n$ times this implies that for each edge of G subdivided into k vertices, $2n + 100 \leq k - 116 - 2n \leq 7n$. So the length of the l_i allows us to apply Lemma 2 and Lemma 3.

Let $\mathbf{p}$ be a path in G' whose two extremities do not belong to the subdivision of the same edge of G. Then it is a subpath of the subdivision of some path in G. Let $e_1, \ldots, e_k$ be this path of G. For all $i \in \{1, \ldots, k\}$, let $w_i = \phi(e_i)$, let r_i be 0 if $\vec{o}(e_i)$ goes in the same direction as the orientation of the path and $r_i = 1$ otherwise. For all $i \in \{1, \ldots, k\}$, let d_i be the integer such that e_i is subdivided into d_i vertices in G' and let $t_i = d_i - 116 - 2n$. Then by definition the sequence of colors of the path $\mathbf{p}$ from G' is a factor of

$$\rho \prod_{i=1}^{k} \overline{f(w_i)\rho l_{t_i}\rho f(w_i)}^{r_i} \rho \,.$$

Moreover, since ϕ is nonrepetitive, $w_1 \ldots w_p$ is square-free. By Lemma 2, $\mathbf{p}$ is nonrepetitively colored by ϕ'.

Now we need to show that the same property holds if the two extremities of a path $\mathbf{p}$ of G' belong to the subdivision of the same edge. If the path is short and completely contained in an edge then this is in fact solved as the previous case. Then the remaining case is that $\mathbf{p}$ starts in the subdivision of an edge e_1 of G, then leaves this subdivision, and comes back to it by the other side. Let $e_1, e_2, \ldots, e_n, e_1$ be the edges of G whose subdivision contains $\mathbf{p}$. For all $i \in \{1, \ldots, k\}$, let $w_i = \phi(e_i)$. For all $i \in \{1, \ldots, k\}$, define r_i and t_i as in the previous case. Then there are two words a and b such that $\overline{f(w_1)\rho l_{t_1}\rho f(w_1)}^{r_1} \rho = ab$ and such that the sequence of colors of $\mathbf{p}$ is

$$b \left(\prod_{i=2}^{k} \overline{f(w_i)\rho l_{t_i}\rho f(w_i)}^{r_i} \rho \right) a \,.$$

By Lemma 3, $\mathbf{p}$ is nonrepetitively colored by ϕ'.

We showed that every possible path of G' is nonrepetitively colored by ϕ' which implies that ϕ' is a nonrepetitive 3-coloring of G'. $\square$

Lemma 9. *Let G be a graph and H be a subdivision of G then $\pi'(H) \leq 2\pi'(G)+3$.*

Proof. Let ϕ be a nonrepetitive edge coloring of G over the set of colors C of size $\pi(G)$. Let C' be the set of colors obtained by adding three new colors α, β, γ and for each color $c \in C$ a new color c'.

Let ϕ' be an edge coloring of H such that for each edge e of G:

- if e is not subdivided in H then it has the same color in H and in G,
- if e is subdivided into two edges e_1 and e_2 then $\phi'(e_1) = \phi(e)$ and $\phi'(e_2) = \phi(e)'$,
- if e is subdivided in $k \geq 3$ edges $e_1, \ldots, e_k$, then $\phi'(e_1) = \phi'(e_k) = \phi(e)$ and the sequence $\phi'(e_2) \ldots \phi'(e_{k-1})$ is a square-free word over $\{\alpha, \beta, \gamma\}$.

It is easy to verify that if there is a square in ϕ' then the colors inherited from ϕ form a square on G. □

Theorem 5. *Let G be a graph and $c = \max\left\{35216, 8^{\frac{\log(2\pi'(G)+3)}{\log(1.01)}} + 216\right\}$. Then for any $(\geq c)$-subdivision H of G,*

$$\pi(H) = 3\,.$$

Proof. Let $n = \max\left\{8750, 2^{\frac{\log(2\pi'(G)+3)}{\log(1.01)}}\right\}$. This implies $c = 4n + 216$.

Let G' be a subdivision of G such that H is a $(\geq 4n+216, \leq 9n)$-subdivision of G'. Let us first show that there exists such a graph H. Since $4n+216 < 2 \times (9n)$, for any integer $x \geq 4n+216$ there exists an integer $\gamma(x)$ such that $4n+216 \leq \frac{x}{\gamma(x)} \leq 9n$. Thus, for any edge e of G that is subdivided k times in H, we can choose e to be subdivided $\frac{k}{\gamma(k)}$ times in G'.

By Lemma 9, $2\pi'(G)+3 \geq \pi'(G')$. Let us now show that we can apply Lemma 8 to G' and H. We can verify by simple computation that $16(60n^2+1) < 1.01^{n/2}$ for any $n \geq 8750$. Hence, by definition of n.

$$\frac{1.01^n}{16(60n^2+1)} \geq 1.01^{n/2} \geq 2\pi'(G)+3 \geq \pi'(G')\,.$$

So n verifies the conditions of Lemma 8. Since H is a $(\geq 4n+216, \leq 9n)$-subdivision of G' we can apply Lemma 8 and we conclude that

$$\pi(H) = 3\,.$$

□

5 Improving the coefficient

We showed in Theorem 5 that for any graph G as long as there are at least $558\log_2(\pi'(G)) + O(1)$ division vertices per edge the resulting graph is 3-colorable. This result is optimal in the sense that $\Omega(\log n)$ division vertices are needed on some edges of any nonrepetitively $O(1)$-colorable subdivision of K_n[10]. However, we can try to reduce the multiplicative constant 558. By being more careful on the computations, we can replace 558 by $\frac{4}{\log_2(\gamma(\mathcal{N}))}$ where $\gamma(\mathcal{N})$ is the growth rate of the set of nice words. By using our lower bound of 1.01 on the growth rate of the number of nice words (Theorem 4) we obtain the bound $279\log_2(n) + O(1)$. But we expect the growth rate to be much closer to 1.3.

In fact, if instead of $\rho = \overline{\sigma}0\sigma$ we take any long enough square-free palindrome $\rho' = \overline{\sigma}'0\sigma'$ then it is easy to adapt the proof from [13] to show that the growth rate of the set of square-free words that avoids σ' and $\overline{\sigma}'$ can be arbitrarily close to 1.3. It probably does not change the growth rate to add the constraint that $\rho' w \rho'$ be square-free for every element w. However, we do not know how to prove this second point, but if it holds we can then replace 1.01 by 1.3. Our lower bound on the number of required division vertices per edge becomes $10.56\log_2(n) + O(1)$. We suspect that this coefficient would still be far from optimal.

References

[1] N. Alon, J. Grytczuk, M. Haluszcza, and O. Riordan. Nonrepetitive colorings of graphs. *Random Structures & Algorithms*, 21:336–346, 2002.

[2] J. Barát and D. R. Wood. Notes on nonrepetitive graph colour-ing. *Electronic Journal of Combinatorics*, 15:R99, 2008.

[3] J. Currie, T. Harju, P. Ochem, N. Rampersad, Some further results on squarefree arithmetic progressions in infinite words, *Theoretical Computer Science*, 799:140-148, 2019.

[4] J. Berstel. Axel Thue's papers on repetitions in words: a translation. *Publications du LaCIM 20, Université du Québec à Montréall*, 1995.

[5] V. Dujmović, G. Joret, J. Kozik, and D. R. Wood. Nonrepetitive colouring via entropy compression. *Combinatorica*, 36(6):661–686, Dec 2016.

[6] P.Erdős and L. Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In *Infinite and Finite Sets*, vol. 10 of *Colloq. Math. Soc. János Bolyai*, pp. 609–627. North-Holland, 1975

[7] D. Gonçalves, M. Montassier, and A. Pinlou. Entropy compression method applied to graph colorings. *arXiv e-prints*, arXiv:1406.4380, 2014.

[8] J. Grytczuk. Nonrepetitive colorings of graphs—a survey. *Int.J. Math. Math. Sci.*, 74639, 2007.

[9] J. Harant and S. Jendrol. Nonrepetitive vertex colorings of graphs. *Discrete Mathematics*, 312(2):374–380, 2012.

[10] J. Nešetřil, P. Ossona de Mendez, and D. R. Wood. Characterisations and examples of graph classes with bounded expansion. *Euro-pean J. Combin.*, 33(3):350–373, 2011.

[11] A. Pezarski and M. Zmarz. Non-repetitive 3-coloring of subdivided graphs. *Electronic Journal of Combinatorics*, 16(1), 2009.

[12] M. Rosenfeld. Another approach to non-repetitive colorings of graphs of bounded degree. *Electronic Journal of Combinatorics*, 27(3), 2020.

[13] A. M. Shur. Two-Sided Bounds for the Growth Rates of Power-Free Languages. In: Diekert V., Nowotka D. (eds)*Developments in Language Theory*, DLT 2009. Lecture Notes in Computer Science, vol 5583.

[14] A. Thue. Über unendliche Zeichenreihen. *'Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania*, 7:1–22, 1906.

[15] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania,*, 10:1–67, 1912.

[16] D. R. Wood. Nonrepetitive Graph Colouring. *arXiv e-prints*, arXiv:2009.02001, 2020.