Rose-Hulman Undergraduate Mathematics Journal

Volume 18
Issue 1

Article 15

On the Long-Repetition-Free 2-Colorability of Trees

Joseph Antonides
The Ohio State University

Claire Kiers
University of North Carolina, Chapel Hill

Nicole Yamzon
San Francisco State University

Rose-
Hulman
Undergraduate
Mathematics
Journal

on the long-repetition-free 2-colorability of trees

Joseph Antonides[a] Claire Kiers[b] Nicole Yamzon[c]

Volume 18, No. 1, Spring 2017

Sponsored by

Rose-Hulman Institute of Technology
Department of Mathematics
Terre Haute, IN 47803
mathjournal@rose-hulman.edu
scholar.rose-hulman.edu/rhumj

[a]The Ohio State University
[b]University of North Carolina, Chapel Hill
[c]San Francisco State University

Rose-Hulman Undergraduate Mathematics Journal
Volume 18, No. 1, Spring 2017

On the long-repetition-free 2-colorability of trees

Joseph Antonides   Claire Kiers   Nicole Yamzon

Abstract. A word $\bar{w} = \bar{u}\bar{u}$ is called a *long square* if $\bar{u}$ is of length at least 3; a word $\bar{w}$ is called *long-square-free* if $\bar{w}$ contains no sub-word that is a long square. If there exists a k-coloring of the vertices of a graph G such that, for any path P in G, the word generated by the coloring of P is long-square-free, then G is called *long-repetition-free k-colorable.* We show that every rooted tree of radius $r \leq 7$ is long-repetition-free 2-colorable. We also show that there exists a class of trees which are not long-repetition-free 2-colorable.

Acknowledgements: This research was conducted in part under the direction of David Milan at the 2013 Research Experience for Undergraduates program at the University of Texas at Tyler and was supported in part by the National Science Foundation (Grant DMS-1062740). This research was also conducted in part under the direction of Annika Miller at Susquehanna University.

1 Introduction

The aim of this paper is to introduce the concepts of "long square" and "long repetition," and to prove two main results regarding the long-repetition-free 2-colorability of trees. Most terminology relevant to this paper is provided, but the reader is encouraged to reference M. Lothaire [3] and Bondy and Murty [1] for more information regarding combinatorics on words and graph theory, respectively.

The investigation of words that avoid specific letter patterns (such as, for example, squares, cubes, or palindromes) has been a topic of interest in the mathematical sciences and the biological sciences since the early 20th century, beginning largely with the papers of Axel Thue [4, 5]. Thue proved that it is impossible to define an infinitely-long, square-free word over a binary alphabet; however, Thue proved that it *is* possible to define an infinitely-long, square-free word over a ternary alphabet, and Thue gave a definition for such a word (which today is called the Thue-Morse word). One can use such a word to color the vertices of a graph and, in so doing, create what we call repetition-free 3-colorings.

We say a word $\bar{w}$ is a "long square" if $\bar{w} = \bar{u}\bar{u}$ for some word $\bar{u}$ where $|\bar{u}| \geq 3$. Much is known regarding words that avoid squares generally (and graph colorings that avoid repetitions generally) - see Grytczuk [2] for a survey of such results; however, little is known about words that allow for squares of lengths 2 and 4.

In this article, we begin in Section 2 by defining terms used throughout the paper. In Section 3, we prove our first main result, Proposition 3.1, which asserts that every tree of radius $r \leq 7$ is long-repetition-free 2-colorable. The proof of this proposition relies on a computer program which inputs a word and returns whether or not the word is long-square-free. Finally, in Section 4, we prove two lemmas which lead to the proof of our second main result, Proposition 4.4, which asserts that there exists a class of trees which are not long-repetition-free 2-colorable.

2 Terminology and Notation

We begin by providing several useful definitions and detailing the notation that is used throughout this paper.

Definition 2.1. An *alphabet* is a finite set of elements called *letters*, and we typically denote alphabets by a single capital letter (for example, A). The set of all words over alphabet A is denoted A^*. We typically denote words of arbitrary length by a single lower-case letter with an overhead bar (for example, $\bar{w}$), and we typically denote letters (words of length 1) by a single lower-case letter (for example, a). We denote the concatenation of words $\bar{w}$ and $\bar{v}$ by $\bar{w}\bar{v}$ or, if we wish to stress the significance of concatenation, by $\bar{w} \cdot \bar{v}$. If a word $\bar{u}$ contains sub-word $\bar{w}$, we say $\bar{w} \leq \bar{u}$.

Definition 2.2. For word $\bar{w} = a_1 \ldots a_n$ over alphabet A, we define the *reversal* of $\bar{w}$ to be $\overleftarrow{w} = a_n \ldots a_1$. Let $\bar{w}$ be a word. We say $\bar{w}$ is a *palindrome* if $\bar{w} = \overleftarrow{w}$. Similarly we say $\bar{w}$ is a *long palindrome* when $\bar{w}$ is of at least length 4. We call $\bar{w}$ a *square* if there exists some non-empty word $\bar{u}$ such that $\bar{w} = \bar{u}\bar{u}$. Furthermore, $\bar{w}$ is *square-free* if $\bar{w}$ contains no sub-word that is a square. Similarly, word $\bar{w} = \bar{u}\bar{u}$ is a *long square* if $|\bar{u}| \geq 3$. Word $\bar{w}$ is *long-square-free* if $\bar{w}$ contains no sub-word that is a long square.

Definition 2.3. Let $G = (V, E)$ be a simple, connected graph, and let $A = \{1, \ldots, k\} \subset \mathbb{N}$. Graph G is said to be *long-repetition-free k-colorable* if there exists a coloring $\chi : V \to A$ such that, for every path $P = u_1 e_1 \ldots e_{n-1} u_n \subseteq G$, the word $\chi(u_1) \ldots \chi(u_n)$ is long-square-free.

Definition 2.4. Let $G = (V, E)$ be a connected graph. The *distance* between vertices $u, v \in V$ is defined to be the number of edges contained in the shortest path from u to v. The *eccentricity* of vertex v is the maximum distance between v and any other vertex in G. The *radius* of graph G is the minimum eccentricity of its vertices, and the *center* of graph G is the sub-graph induced by the vertices of minimum eccentricity.

Definition 2.5. A tree $T = (V, E)$ is called a *rooted tree* if one vertex has been distinguished to be called the *root* of T. The edges of a rooted tree are described with an implicit orientation away from the root vertex. In a rooted tree, the *parent* of a vertex v is the vertex adjacent to v on the path to the root vertex. A *child* (plural, *children*) of vertex v is a vertex of which v is the parent. A *descendant* of a vertex v is any vertex which is either a child of v or is, recursively, the descendant of any of the children of v. A *sibling* to a vertex v is any other vertex in V which has the same parent as v. The k^{th} *generation* of root r of tree T, denoted $G_k(T)$, is the set of descendant vertices of distance k from r. The *height* of T is the number of generations in T from r, with the convention that $G_0(T) = \{r\}$.

In particular, it is of note that if $T = (V, E)$ is a rooted tree, then for root $r \in T$ and vertex $v \in V$ there exists a unique paths from r to v and from v to r. As a matter of notation, it is perhaps noteworthy that while the k^{th} generation from r is of course dependent on our choice of r, we simply notate this $G_k(T)$ since, for this paper, the particular vertex distinguished as the root is not particularly relevant and furthermore does not change once the root has been chosen.

Definition 2.6. The *Tyler tree of height n*, denoted $T_n = (V_n, E_n)$, is the rooted tree with root r in which each vertex in the j^{th} generation from r has $2^{n-j} + 1$ children, for every $0 \leq j \leq n - 1$. Computationally, $|V_n| = 1 + \displaystyle\sum_{j=0}^{n-1} \prod_{k=n-j}^{n} (2^k + 1)$. A *binary tree* is a rooted tree in which each vertex has at most two children. A full binary tree with height n is denoted B_n. The reader can verify that the Tyler tree of height n contains $2^n + 1$ Tyler trees of height $n - 1$, and it is of note that all vertices in generation n are leaves.

3 A Class of Trees that is Long-Repetition-Free 2-Colorable

The goal of the present section is to prove the first of our two main results: Proposition 3.1, which asserts that every rooted tree of radius less than or equal to 7 can be long-repetition-free 2-colored.

Proposition 3.1. *Every rooted tree of radius less than or equal to 7 is long-repetition-free 2-colorable.*

Proof. Let $T = (V, E)$ be a tree of radius $R \leq 7$, and choose a root vertex r in the center of T; that is, choose r to have minimal eccentricity. Let $g = \max\{k \in \mathbb{N} \mid G_k(r) \neq \varnothing\}$ be the number of generations in T from vertex r. First, note that $g \leq 7$. Indeed, if $g \geq 8$, then there exists a vertex $w \in G_8(r)$, so $d(r, w) = 8$, and so the eccentricity of r is at least 8. However, since T has radius $R \leq 7$, there exist vertices of eccentricity less than 8, contradicting our choice of r; therefore, $g \leq 7$. It suffices to show the proposition holds for $g = 7$, so assume without loss of generality that $g = 7$. Assume also that at least two vertices belong to each generation $G_k(r)$ for all $1 \leq k \leq 7$.

Color the vertices of T in the following way. Let $\bar{a} = 00010111$, and define $\chi : V \to \{0, 1\}$ by $\chi(v) = a_j$ whenever $v \in G_j(r)$, for $0 \leq j \leq 7$. See Figure 3.2 for an example of a tree of radius 7 whose vertices have been colored according to χ.

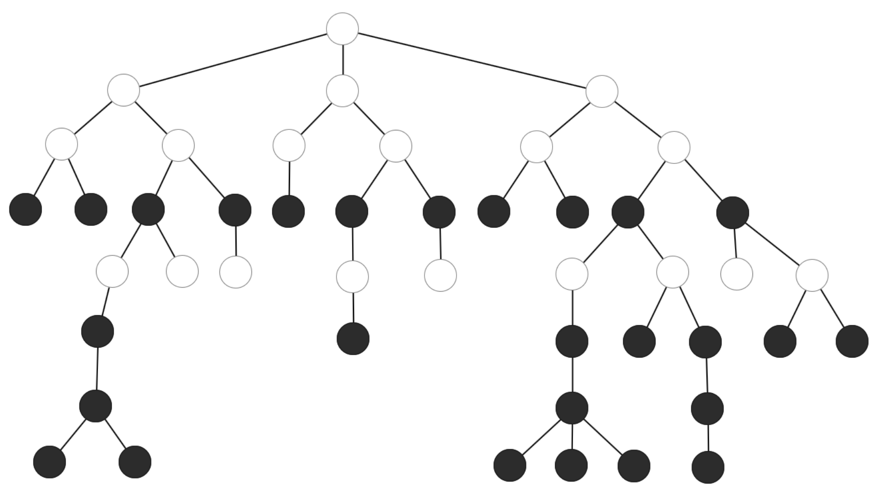

Figure 3.2. *The coloring χ as defined in Proposition 3.1 applied to a tree of radius 7.*

We claim that χ is a long-repetition-free 2-coloring of T. Suppose, by way of contradiction, that there exists a path $P \subseteq T$ such that $\chi(P)$ is a long repetition. Let $P = v_1 e_1 v_2 e_2 \ldots v_n e_n v_{n+1} \ldots v_{2n-1} e_{2n-1} v_{2n}$. Clearly $n \geq 3$ in order for P to be a long repetition, and since $g \leq 7$, it follows that $n \leq 7$. Since T is a tree, two adjacent vertices in P cannot belong to the same generation; moreover, if $v_i, v_{i+1} \in P$ are adjacent and if $v_i \in G_j(r)$ (for some $0 \leq j \leq 7$), then either $v_{i+1} \in G_{j+1}(r)$ or $v_{i+1} \in G_{j-1}(r)$. Furthermore, since T is a tree, the sequence of generations $g(v_1), \ldots, g(v_{2n})$ of vertices $v_1, \ldots, v_{2n}$ is either strictly increasing, strictly decreasing, increases and then decreases, or decreases and then increases. Consequently, it follows that P is a subgraph of some path P'_ℓ, where P'_ℓ is a path starting at a vertex in $G_7(r)$, ascending to a vertex in $G_\ell(r)$, and descending to another vertex in $G_7(r)$, for some $0 \leq \ell \leq 4$. Hence there are five cases to consider for $\chi(P)$: (1) $\chi(P) \leq 1110111$, (2) $\chi(P) \leq 111010111$, (3) $\chi(P) \leq 11101010111$, (4) $\chi(P) \leq 1110100010111$, and (5) $\chi(P) \leq 111010000010111$.

To prove the statement of the proposition, it suffices to check that the words 1110111, 111010111, 11101010111, 1110100010111, and 111010000010111 are long-square-free. For this, we use a computer program (written in the language Python, given in the Appendix). In brief, the computer program inputs a word and checks each subword of length 6 or more contained in the word. If a long square is found, then the program prints the detected long square and returns 'True'; if not, then the program returns 'False'. Using this program, we find the words 1110111, 111010111, 11101010111, 1110100010111, and 111010000010111 to be long-square-free.

Thus, we find the coloring $\chi(P)$ of path P to be long-repetition-free, and so the coloring of every path in T is long-repetition-free, as desired. □

4 A Class of Trees that is Not Long-Repetition-Free 2-Colorable

One may ask whether or not every tree is long-repetition-free 2-colorable. In this section, we show that a certain class of trees require a minimum of 3 colors for a long-repetition-free coloring, making 3 colors the lowest bound for long-repetition-free colorings of trees. In order to gain this result, we first prove Lemma 4.1 and Lemma 4.2.

Lemma 4.1. *Every binary word of length at least* 9 *contains a long palindrome.*

Proof. It is clear that we need only consider binary words of length 9, for any binary words of length greater than 9 contains a binary word of length 9. Let $a_1a_2\ldots a_9$ be a word over alphabet $A = \{0, 1\}$. Define a_i^c to be the binary complement of a_i; more precisely, $a_i^c \equiv a_i + 1 \mod 2$ for every $i \leq 9$. The reader can verify that for $a_1 \ldots a_9$ to be long-palindrome-free, the following conditions must hold:

(i) If $a_i = a_{i+3}$, then $a_{i+1} = a_{i+2}^c$, for $1 \leq i \leq 6$;

(ii) If $a_i = a_{i+4}$, then $a_{i+1} = a_{i+3}^c$, for $1 \leq i \leq 5$;

(iii) If $a_i = a_{i+1}$, then $a_{i+1} = a_{i+2}^c$, for $2 \leq i \leq 6$.

Suppose without loss of generality that $a_1 = 0$. We consider four cases. First, suppose $a_1 = a_4 = a_5 = 0$. By (i), it follows that $a_2 = a_3^c$. Then $a_2 = 1$ and $a_3 = 0$, for otherwise we have long palindrome 00100. However, we now have no choice for a_6, for $a_6 = 0$ yields long palindrome 0000, and $a_6 = 1$ yields long palindrome 10001.

Suppose next that $a_1 = a_4 = a_5^c = 0$. By (i), it follows that $a_2 = a_3^c$. Then $a_2 = 0$ and $a_3 = 1$, for otherwise we have long palindrome 1001. Then $a_6 = 1$, for otherwise we have long palindrome 01010, and so $a_7 = 1$, for otherwise we have long palindrome 0110. However, we now have no choice for a_8, for $a_8 = 0$ yields long palindrome 01110, and $a_8 = 1$ yields long palindrome 1111.

Suppose that $a_1 = a_4^c = a_5^c = 0$. By (iii), $a_6 = 0$, and it follows that $a_3 = 1$ for otherwise we have long palindrome 0110. However, we now have no choice for a_2, for $a_2 = 0$ yields long palindrome 01110, and $a_2 = 1$ yields long palindrome 1111.

Finally, suppose that $a_1 = a_4^c = a_5 = 0$. By (ii), $a_2 = 0$, and it follows that $a_3 = 0$ for otherwise we have long palindrome 0110. Then $a_6 = 1$, for otherwise we have long palindrome 00100; $a_7 = 1$, for otherwise we have long palindrome 01010; and $a_8 = 1$, for otherwise we have 0110. However, we now have no choice for a_9, for $a_9 = 0$ yields long palindrome 01110, and $a_9 = 1$ yields long palindrome 1111. □

Lemma 4.2. *Let T_n denote the Tyler tree of height n with root r, and apply an arbitrary 2-coloring ϕ to T_n. Then T_n contains a binary sub-tree of height n, B_n, with root r such that, for every $u, v \in G_i(B_n)$, $\phi(u) = \phi(v)$, for all $0 \le i \le n$. In other words, the vertices in the same generation of B_n share the same color.*

Proof. We prove the result by induction on the height, n, of the Tyler tree.

Tyler tree T_1 contains exactly four vertices: root r and 3 children vertices adjacent to r. By the pigeonhole principle, at least 2 of these 3 children vertices must be colored with the same color. Thus, T_1 contains a full binary tree of height 1 whose root is r and where all the vertices on the same level have the same color. See Figure 4.3 for an example of T_1 and T_2 each with an embedded full binary tree colored by generations.

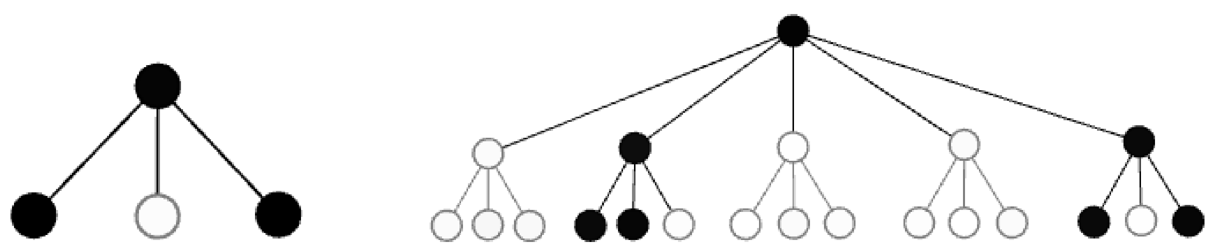

Figure 4.3. *The Tyler tree of height 1 (left) and the Tyler tree of height 2 (right) each contain a binary sub-tree in which every vertex in each generation from the root has the same color.*

Suppose that every Tyler tree of height $k \ge 1$ contains a binary sub-tree such that, for every $u, v \in G_i(B_k)$, we have $\phi(u) = \phi(v)$, for every $0 \le i \le k$. Let T_{k+1} denote the Tyler tree of height $k+1$ with root r, and apply arbitrary 2-coloring ϕ to T_{k+1}. By definition, r has $2^{k+1}+1$ children, and each of these children are the root of a Tyler tree of height k. Therefore, T_{k+1} contains $2^{k+1}+1$ Tyler trees of height k as sub-trees, and, by the inductive hypothesis, it follows that each of these Tyler trees of height k contain a binary sub-tree of height k such that, for every $u, v \in G_i(B_k)$, we have $\phi(u) = \phi(v)$, for every $0 \le i \le k$.

Each of these $2^{k+1}+1$ binary sub-trees has k generations, and, given the coloring constraints provided in the inductive hypothesis, there are exactly 2^k possible 2-colorings of these binary sub-trees. Attach to these binary sub-trees the root r. As there are 2 possible 2-colorings of r, there are a total of 2^{k+1} possible 2-colorings of this binary sub-tree of height $k+1$ in which each generation is similarly colored. Since there are $2^{k+1}+1$ binary sub-trees of height k and only 2^{k+1} possible 2-colorings in which each generation is similarly colored, it follows that at least two binary sub-trees of height k must be colored identically. The sub-tree induced by these two identically-colored full binary trees of height k, when attached to r, form a full binary tree of height $k+1$ in which, for every $u, v \in G_i(B_{k+1})$, we have $\phi(u) = \phi(v)$, for all $0 \leq i \leq k+1$, thereby completing the proof. □

Proposition 4.4. *If $n \geq 8$, then T_n is not long-repetition-free 2-colorable.*

Proof. Let T_8 denote the Tyler tree of height 8 with root r. We need only consider T_8 to prove the result. Apply an abritrary 2-coloring ϕ to T_8. By Lemma 4.2, T_8 contains a binary sub-tree B_8 in which, for each $u, v \in G_i(B_8)$, we have $\phi(u) = \phi(v)$, for all $0 \leq i \leq 8$.

Pick vertices $u_0, \ldots, u_8$ such that $u_i \in G_i(B_8)$ for each $0 \leq i \leq 8$, and such that $u_0, \ldots, u_8$ is a simple path. Let b_i denote the letter associated with $\phi(u_i)$, and let $\bar{b} = b_0 \ldots b_8$. By Lemma 4.1, $\bar{b}$ contains a palindrome $\bar{p}$ of length 4 or of length 5. Suppose $\bar{p} = b_j \ldots b_{j+m}$, where $0 \leq j \leq 4$ and where $m = 3$ or $m = 4$. Consider the path $P = u_{j+m} \ldots u_{j+1} u_j u'_{j+1} \ldots u'_{j+m}$, where $u'_i \in G_i(B_8)$ for each $0 \leq i \leq j+m$.

Let $\bar{c}$ be the word associated with $\phi(P) = \phi(u_{j+m} \ldots u_{j+1} u_j u'_{j+1} \ldots u'_{j+m})$. If $|\bar{p}| = 4$, then either $\bar{c} = c_1c_1c_1c_1c_1c_1c_1$, or $\bar{c} = c_1c_2c_2c_1c_2c_2c_1$, where $c_i \in \{0, 1\}$. In either case, $\bar{c}$ contains a long square. If $|\bar{p}| = 5$, then there are four possibilities: (1) $\bar{c} = c_1(c_1c_1c_1c_1)(c_1c_1c_1c_1)$, (2) $\bar{c} = c_1(c_2c_1c_2c_1)(c_2c_1c_2c_1)$, (3) $\bar{c} = c_1(c_1c_2c_1c_1)(c_1c_2c_1c_1)$ or (4) $\bar{c} = c_1(c_2c_2c_2c_1)(c_2c_2c_2c_1)$. In any of these four cases, $\bar{c}$ contains a long square. Since $\bar{c}$ contains a long square regardless of the length of $\bar{p}$, it follows that P contains a long repetition, and thus T_8 is not long-repetition-free 2-colorable. □

5 Appendix

The following computer program, written in the language Python, inputs a word (or a string) and checks every subword of length 6 or more contained in the given word. For each long square contained in the word, the program prints the long square and returns 'True', indicating that the word does contain a long square. If no long square is found in the word, then the program returns 'False', indicating that the word does not contain a long square and is therefore long-square-free.

```
def longSquareChecker(string):
    for length in range(3,len(string)//2+1):
        for index in range(0,len(string)-length*2+1):
            if string[index:index+length]...
            ==string[index+length:index+2*length]:
                print(string[index:index+length],index)
                return True
    return False
```

References

[1] J.A. Bondy and U.S.R. Murty. *Graph Theory*. Springer, 2008.

[2] Grytczuk. Nonrepetitive colorings of graphs - a survey. 2006.

[3] M. Lothaire. *Combinatorics on Words*. Cambridge University Press, 1997.

[4] A. Thue. Uber unendliche zeichenreihen. 7:1–22, 1906.

[5] A. Thue. Uber die gegenseitige lage gleicher teile gewisser zeichenreihen. 10:1–67, 1912.