THUE CHOOSABILITY OF TREES

FRANCESCA FIORENZI AND PASCAL OCHEM

Laboratoire de Recherche en Informatique, UMR 8623, Université Paris-Sud, 91405 Orsay, France

PATRICE OSSONA DE MENDEZ[1]

Centre d'Analyse et de Mathématique Sociales, CNRS, UMR 8557, 54 Bd Raspail, 75006, Paris, France

XUDING ZHU[2]

Department of Mathematics, Zhejiang Normal University, China

Abstract. A vertex colouring of a graph G is *nonrepetitive* if for any path $P = (v_1, v_2, \dots, v_{2r})$ in G, the first half is coloured differently from the second half. The *Thue choice number* of G is the least integer ℓ such that for every ℓ-list assignment $\mathcal{L}$ of G, there exists a nonrepetitive $\mathcal{L}$-colouring of G. We prove that for any positive integer ℓ, there is a tree T with $\pi_{\mathrm{ch}}(T) > \ell$. On the other hand, it is proved that if G' is a graph of maximum degree Δ, and G is obtained from G' by attaching to each vertex v of G' a connected graph of tree-depth at most z rooted at v, then $\pi_{\mathrm{ch}}(G) \le c(\Delta, z)$ for some constant $c(\Delta, d)$ depending only on Δ and z.

1. Introduction

A sequence $S = (s_1, s_2, \dots,)$ is *nonrepetitive* if no two adjacent blocks are identical, i.e., for any positive integers i and r, one has $(s_i, s_{i+1}, \dots, s_{i+r-1}) \neq (s_{i+r}, s_{i+r+1}, \dots, s_{i+2r-1})$. It was proved by Thue in 1906 [13] that there is an infinite sequence on three symbols $\{0, 1, 2\}$ which is nonrepetitive. This

E-mail addresses: `fiorenzi@lri.fr, ochem@lri.fr, pom@ehess.fr, xudingzhu@gmail.com`.

[1]Research partially supported by Academia Sinica.

[2]Research partially supported by ZJNSF No. Z6110786.

result was rediscovered many times in different contexts. It has many generalizations and important applications in distinct fields of Mathematics and Computer Sciences (see [1], [3], [5], [7], [10]). In particular, it is the starting point of symbolic dynamics, a field with many deep results and challenging open problems.

Recently, connection between Thue sequence and graph theory has attracted much attention [6]. A *vertex colouring* of a graph G is a mapping c which assigns to each vertex v of G a color $c(v)$. A vertex colouring is *nonrepetitive* if for any path $P = (v_1, v_2, \ldots, v_{2r})$ in G with an even number of vertices, the first half is coloured differently from the second half. The *Thue chromatic number* of G, denoted by $\pi(G)$, is the least number of colours used in a nonrepetitive colouring of G. In this language, an infinite nonrepetitive sequence is just a nonrepetitive colouring of the infinite path, and the Thue Theorem is equivalent to say that the infinite path has Thue chromatic number 3.

There are many other classes of graphs whose Thue chromatic numbers have been studied. The Thue chromatic number of all cycles are determined [4]. It is proved in [2] that for any graph G, $\pi(G) \leqslant 16\Delta(G)^2$, and there are graphs with Thue chromatic number of order $\Delta(G)^2/\log\Delta(G)$. Graphs of treewidth k are known to have $\pi(G) \leqslant 4^k$ [9].

The list nonrepetitive colouring of graphs is studied in [8]. A *ℓ-list assignment* is a mapping $\mathcal{L}$ which assigns to each vertex v of G a set $\mathcal{L}(v)$ of ℓ permissible colours. A *nonrepetitive $\mathcal{L}$-colouring* of G is a nonrepetitive colouring c of G such that $c(v) \in \mathcal{L}(v)$ for each vertex v. The *Thue choice number* of G, denoted by $\pi_{\mathrm{ch}}(G)$, is the least integer ℓ such that for any ℓ-list assignment $\mathcal{L}$, there exists a nonrepetitive $\mathcal{L}$-colouring of G.

It follows from the definition that for any graph G, $\pi_{\mathrm{ch}}(G) \geq \pi(G)$. Some of the upper bounds for Thue chromatic number of graphs G are indeed upper bounds for their Thue choice number. For example, the proof in [2] actually shows that for any graph G, $\pi_{\mathrm{ch}}(G) \leqslant 16\Delta(G)^2$. However, the proofs of some other upper bounds for $\pi(G)$ do not work for $\pi_{\mathrm{ch}}(G)$. For example, the proofs for the upper bounds of the Thue chromatic number of paths, trees, graphs of bounded treewidth, etc., do not work for their Thue choice number. Using the probabilistic method, it is proved in [8] that the infinite path P_∞ has Thue choice number at most 4. As $\pi(P_\infty) = 3$, Thue choice number of the infinite path is either 3 or 4.

The question whether the Thue choice number of trees are bounded by a constant was asked in [8]. In this paper, we give a negative answer to this question, by proving that for any positive integer ℓ, there is a tree T with $\pi_{\mathrm{ch}}(T) > \ell$. So the tree-width of a graph does not provide an upper bound on its Thue choice number. On the other hand, it is easy to show that graphs of bounded tree-depth have bounded Thue choice number. We shall prove a more general class of graphs have bounded Thue choice number: for any integers Δ, z, there is a constant $c(\Delta, z)$ for which the following holds: if G is obtained from a graph G' of maximum degree at most Δ by attaching to each vertex v of G' a connected graph of tree-depth at most z, then $\pi_{\text{ch}}(G) \leq c(\Delta, z)$.

2. Thue choice number of trees are unbounded

First, we give some definitions and notations. Suppose T is a rooted tree with root v. The *level* of vertices of T is defined recursively: the level of v is 1, and if u has level k then the sons of u have level $k+1$. The *height* of T is the maximum level of a vertex of T.

A $[k]$*-tree* is a rooted tree such that

- all the leaves have the same level,
- each internal vertex has exactly k sons.

Let a be a positive integer. A vertex colouring c of a tree is a*-distinct* if

- c takes values in a finite set of colours $\{1, \ldots, a\}$,
- the sons of a vertex have all different colours.

Note that $[k]$-tree always admits an a-distinct colouring if $k \leq a$.

A *descending path* in a tree is a path starting from the root.

We can now state the main result of this section.

Theorem 1. *Let ℓ be a positive integer and let T be a $\left[(\ell+1)\binom{\ell^3}{\ell}\right]$-tree of height $\ell+1$. Then $\pi_{\text{ch}}(T) > \ell$.*

Notice that the tree above has $\sum_{h=0}^{\ell} \left((\ell+1)\binom{\ell^3}{\ell}\right)^h = \ell^{\Theta(\ell^2)}$ vertices. We thus obtain the following.

Corollary 2. *The maximum Thue choice number of trees of order n asymptotically satisfies*

$$\max_{|T|=n} \pi_{\text{ch}}(T) = \Omega\left(\left(\frac{\log n}{\log\log n}\right)^{1/2}\right).$$

The rest of this section is dedicated to the proof of Theorem 1.

We fix a set A of colors of cardinality $a \geq \ell$ and a $\left[(\ell+1)\binom{a}{\ell}\right]$-tree T. We use an ℓ-list assignment $\mathcal{L}$ on T such that every ℓ-subset of A is assigned to exactly $(\ell+1)$ of the sons of an internal vertex (the list assigned to the root could be arbitrarily chosen).

Consider a potential nonrepetitive $\mathcal{L}$-colouring of T. Among the $(\ell+1)$ sons of an internal vertex sharing the same list, at least two of them must get the same colour by the pigeon-hole principle. Moreover, for each internal vertex, at least $(a-\ell+1)$ distinct colours appear twice among its sons: select one son for each list such that its color appears also on another son and suppose for the sake of contradiction that (at least) ℓ distinct colours do not appear among the selected sons. Then the list consisting of these ℓ forbidden colours is assigned to one of the selected son, a contradiction.

From the $\mathcal{L}$-colouring of T we keep $a-\ell+1$ sons for each internal vertex v, such that they have different colors and the $a-\ell+1$ different colours appear twice among the sons of v in T. This way, we obtain an a-distinct colouring of some subtree T' of T, and T' is an $[a-\ell+1]$-tree.

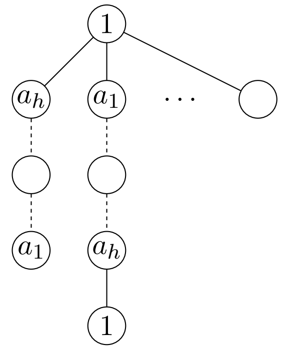
The color of a vertex in T' must be distinct from the color of its father to avoid repetitions of length two. It is also distinct from the color of its grand-father to avoid repetitions of length four, since there exists an uncle in T with the same color as the father. This means that T' contains no palindrome.

We consider now the tree T' only, and we suppose that its a-distinct colouring is nonrepetitive and avoids palindromes. We will eventually obtain a contradiction, for a large enough.

Lemma 3. *Let T' be an $[a-\ell+1]$-tree of height $\ell+1$ as above. For any nonrepetitive a-distinct colouring on T' that avoids palindromes, the number of vertices at level ℓ having a son with the same colour as one of its ancestor, is at most*

$$\sum_{h=2}^{\ell-1}(a-\ell+1)^{\ell-1-h}\left(a^h-(a-\ell+1)^h\right).$$

Proof. Let h be an integer such that $2\le h\le \ell-1$. Consider an $[a-\ell+1]$-tree R of height $h+2$. For any nonrepetitive a-distinct colouring on R that avoids palindromes, the number of vertices at level $h+1$ having a son with the same colour as the root, is at most $a^h-(a-\ell+1)^h$. Indeed, suppose that a vertex at level $h+1$ has a son with the same colour as the root. Call *nostalgic* such a vertex. Say that the common colour of the root and of the leaf is 1. Consider the descending path coloured $1a_1\dots a_h$ ending a nostalgic vertex in R. The inverse descending path coloured $1a_h\dots a_1$ is distinct from the previous one since R does not contain palindromes. Moreover, it cannot exist in R since it would create the repetition $(a_1\dots a_h1)^2$.



Now, there are $a^h - (a - \ell + 1)^h$ words of length h over $\{1, 2, \ldots, a\}$ that do not correspond to one of $(a - \ell + 1)^h$ descending paths in R. For every such missing word, there is at most one nostalgic vertex in R by considering the inverse descending path.

Since T' avoids palindromes, a vertex does not have the same colour as its father (resp. grand-father). The bound in the statement is obtained by considering nostalgic vertices of subtrees rooted at each internal vertex of T'. □

Let T' be an $[a-\ell+1]$-tree of height $\ell+1$ on which is defined an a-distinct vertex colouring. Observe that, among the $a-\ell+1$ sons of a vertex, at most ℓ of them have the same colour as one of their ancestors. Consider a vertex at level ℓ such that its colour and the colours of its ancestors are pairwise distinct. The number of such vertices is at least

$$(a - 2\ell + 1)^{\ell-1}.$$

A vertex at level ℓ that simultaneously verifies this latter condition and has no son coloured as one of its own ancestor is called *bad*.

By Lemma 3, the number of bad vertices is at least

$$(a - 2\ell + 1)^{\ell-1} - \sum_{h=2}^{\ell-1} (a - \ell + 1)^{\ell-1-h} \left(a^h - (a - \ell + 1)^h \right). \tag{1}$$

Notice that (1) is a polynomial in a with leading term $a^{\ell-1}$. It is thus positive for a large enough, say $a = \ell^3$. Hence T' contains a bad vertex. By considering the ℓ vertices on the descending path to the bad vertex, and the $(a-\ell+1)$ colours on its sons, we obtain a contradiction: $\ell+(a-\ell+1) = a+1$ vertices should get distinct colours but we have only a colours.

3. Graph classes with bounded Thue choice number

In order to find classes of graphs whose Thue choice number is bounded, we need to restrict to those classes which do not contain the class of trees as a subclass. One such class consists of graphs of bounded maximum degree.

As mentioned in Section 1, the proof in [2] shows that for any graph G, $\pi_{\mathrm{ch}}(G) \leqslant 16\Delta(G)^2$. Another such class consists of graphs of bounded tree depth defined below.

The *closure* of a rooted tree (T, r), is defined as the graph $\mathrm{clos}(T, r)$ in which $V(\mathrm{clos}(T, r)) = V(T)$ and $v_1v_2 \in E(\mathrm{clos}(T, r))$ if and only if v_1 is an ancestor of v_2 or v_2 is an ancestor of v_1 in (T, r). For a connected graph G, the *tree-depth* of G [11] is the least integer h such that there is a rooted tree (T, r) of height h such that G is a subgraph of $\mathrm{clos}(T, r)$. For a disconnected graph G, its tree-depth is the maximum of the tree-depth of its connected components.

Lemma 4. *For every positive integer ℓ, the maximum Thue choice number of graphs of tree-depth ℓ is equal to ℓ.*

Proof. The proof of Theorem 1 gives a rooted tree of height $\ell + 1$, which is then a graph of tree-depth at most $\ell + 1$, whose Thue choice number is at least $\ell + 1$.

On the other hand, let G be a connected graph of tree-depth at most ℓ. Let (T, r) be a rooted tree of height ℓ such that G is a subgraph of $\mathrm{clos}(T, r)$.

Assume that to each vertex v is assigned a list of ℓ colours $\mathcal{L}(v)$. As a vertex in (T, r) has at most $\ell - 1$ ancestors, it is possible to find a colour assignment c such that for every vertex v, $c(v) \in \mathcal{L}(v)$ and $c(v)$ is different from the colours of its ancestors in (T, r). Now consider any path P in G. There exists a vertex $v \in V(P)$ which is an ancestor of all the other vertices in $V(P)$. Then the colour $c(v)$ cannot appear twice on P and hence the colour sequence of P is nonrepetitive. It follows that the Thue choice number of T is at most ℓ. $\square$

The following result shows that a super class of graphs obtained by gluing together graphs of bounded maximum degree and bounded tree-depth also have bounded Thue choice number.

Theorem 5. *Let G be a graph of maximum degree Δ and G' be a graph obtained from G by attaching to each vertex v of G a connected graph H_v of tree-depth at most z, i.e., identify v with some vertex of H_v, then G' has Thue choice number at most $\lceil (2^z - 4)\Delta^{2^z-4} e^{4(2^z-3)(2^z-2)} \rceil$.*

The proof of Theorem 5 uses Lovász Local Lemma.

Lemma 6 (Lovász Local Lemma)**.** *Let $A_1, A_2, \ldots, A_n$ be events in a probability space. Let $D = (V, E)$ be a graph with vertex set $V = \{A_1, A_2, \ldots, A_n\}$. For each $A \in V$, $\Gamma(A)$ (respectively, $\Gamma[A]$) is the open (respectively, close)* *neighborhood of A in D. If each event A is independent from any collection $\mathcal{A}$ of events such that $\mathcal{A} \cap \Gamma[A] = \emptyset$, and there exists an assignment of reals $x : V \to (0,1)$ to the events such that for any event A,* $\Pr[A] \leq x_A \prod_{B\in\Gamma(A)}(1-x_B)$, *then the probability of avoiding all events in V is positive, i.e.,* $\Pr[\overline{A}_1 \wedge \overline{A}_2 \wedge \ldots \wedge \overline{A}_n] > 0$.

Proof of Theorem 5. Let $d = 2^z - 3$ and let $N = \lceil 2(d+1)\Delta^{2d+2}e^{4d^2+4d}\rceil$. Assume $\mathcal{L}$ is a list assignment with $|\mathcal{L}(v)| = N$ for each vertex v of G'. First, for each vertex v of G, let $c(v) \in \mathcal{L}(v)$ be a colour randomly chosen from $\mathcal{L}(v)$. For integers $0 \leq s \leq 2d$ and $k \geq 1$ and for a path $P = (v_1, v_2, \ldots, v_{2k+s})$ in G of length $2k+s-1$, let $A_{P,k,s}$ be the event that the first k vertices of P are coloured as the last k vertices, i.e., $c(v_i) = c(v_{k+s+i})$ for $i = 1, 2, \ldots, k$. Let $\mathcal{A}_k$ be the set of all events $A_{P,k,s}$, where $0 \leq s \leq 2d$, and let $\mathcal{A} = \cup_{k=1}^{\infty}\mathcal{A}_k$. Define a dependency graph with vertex set $\mathcal{A}$ so that $A_{P,k,s}$ and $A_{P',k',s'}$ are adjacent if and only if P and P' have a common vertex. For each vertex v, for each positive integer k and for each $0 \leq s \leq 2d$, there are at most $(k+s/2)\Delta^{2k+s} \leq (k+d)\Delta^{2k+2d}$ paths P of length $2k+s-1$ containing v. So a path P with $2k+s$ vertices intersects at most $(2k+s)(k'+d)\Delta^{2k'+2d} \leq (2k+2d)(k'+d)\Delta^{2k'+2d}$ paths of length $2k'+s'-1$ for any $0 \leq s' \leq 2d$. Therefore in the dependency graph, an event $A \in \mathcal{A}_k$ is adjacent to at most $2d(2k+2d)(k'+d)\Delta^{2k'+2d}$ events in $\mathcal{A}_{k'}$ if $k' \neq k$, and it is adjacent to at most $2d(2k+2d)(k+d)\Delta^{2k+2d} - 1$ events in $\mathcal{A}_k$.

For any event $A \in \mathcal{A}_k$, let

$$\begin{aligned} n_k &= (k+d)\Delta^{2(k+d)}, \\ x_A &= x_k = \frac{1}{n_k 2^k + 1}. \end{aligned}$$

Then

$$\frac{x_k n_k}{1-x_k} = \frac{1}{2^k}, \quad \sum_{k=1}^{\infty}\frac{x_k n_k}{1-x_k} = 1, \quad N^k \geq n_k 2^k \exp(4d(k+d)).$$

Therefore

$$\begin{aligned} x_A \prod_{B\leftarrow A}(1-x_B) &\geq x_k(1-x_k)^{-1}\prod_{k'=1}^{\infty}(1-x_{k'})^{4d(k+d)n_{k'}} \\ &\geq \frac{1}{n_k 2^k}\prod_{k'=1}^{\infty}\exp\left(\frac{-x_{k'}4d(k+d)n_{k'}}{1-x_{k'}}\right) \\ &= \frac{1}{n_k 2^k}\exp\left(-4d(k+d)\sum_{k'=1}^{\infty}\frac{x_{k'}n_{k'}}{1-x_{k'}}\right) \\ &= \frac{1}{n_k 2^k}\exp\left(-4d(k+d)\right) \\ &\geq N^{-k}. \end{aligned}$$

By Lovász Local Lemma, the probability of avoiding all the $A_{P,k,s}$ is positive. So there is an $\mathcal{L}$-colouring c of the vertices of G so that for any $k \geq 1$, for any $0 \leq s \leq 2d$, for any path P with $2k+s$ vertices, the first k vertices are not coloured in the same way as the last k vertices.

Now for each vertex v of G, let H_v be a connected graph of tree-depth at most d attached to G by identifying one of its vertices with v. We colour each vertex x of H_v with a colour in $\mathcal{L}(x)$ which is different from the colour of any ancestor of x, and also different from the colour of any vertex v' of G for which $\mathrm{dist}_G(v, v') \leq 2d$. Such a colouring exists, because $|\mathcal{L}(x)| = N = \lceil 2(d+1)\Delta^{2d+2} e^{4d^2+4d} \rceil \geq \Delta^{2d} + d + 1$.

We shall show that the colouring c is nonrepetitive. Assume to the contrary that $P = (x_1, x_2, \ldots, x_{2r})$ is a repetitive path in G'. If both end vertices x_1 and x_{2r} of P are vertices of G, then P is a path in G. By the first part of this proof, P cannot be repetitive. Thus we may assume that the initial vertex of P belong to H_v for some v of G. As proved in Lemma 4, P cannot be contained in H_v. So x_{2r} is either a vertex of G, or a vertex of $H_{v'}$ for some vertex v' of G which is distinct from v. In particular, $v \in V(P)$. As the colour $c(v)$ is not used by any other vertex of H_v, we conclude that $V(P) \cap V(H_v)$ is contained in the first half of P, i.e., $x_{r+1} \notin H_v$. Assume $|V(P) \cap V(H_v)| = t$. As noticed in [11], tree-depth is minor monotone and the tree-depth of a path of length L is $\lceil \log_2(L+2) \rceil$. It follows that the longest path H_v has length at most $2^z - 2$ hence $t \leq 2^z - 1$. If the other end vertex of P is a vertex of G, then $P' = (x_t, x_{t+1}, \ldots, x_{2r})$ is a path of G in which the first $r-t$ vertices of P' are coloured in the same way as the last $r-t$ vertices of P', contrary to the first part of this proof.

Assume $x_{2r} \in V(H_{v'})$ for some vertex v' of G distinct from v. By symmetry, we may assume that $|V(P) \cap V(H_{v'})| = t' \leq t$. Then $x_{r+1} \in V(G)$. Since $c(x_1) = c(x_{r+1})$, by the colouring defined above, we conclude that $\mathrm{dist}_G(v, x_{r+1})$ is at least $2d+1$. In particular, $r \geq 2d+1 \geq 2t$. Now the path $P' = (x_t, x_{t+1}, \ldots, x_r, x_{r+1}, \ldots, x_{2r-t+1})$ is in G, and the first $r-2t+2$ vertices are coloured as the last $r-2t+2$ vertices, contrary to the first part of this proof (where we consider $k = r-2t+2$ and $s = 2t-4 \leq 2^z-6 = 2d$). $\square$

References

[1] J-P. Allouche, J. Shallit, Automatic sequences. Theory, applications, generalizations, Cambridge University Press, Cambridge, 2003.

[2] N. Alon, J. Grytczuk, M. Hałuszczak, O. Riordan, Non-repetitive colourings of graphs, Random Struct. Alg. 21 (2002), 336-346.

[3] D. R. Bean, A. Ehrenfeucht, G. F. McNulty, Avoidable patterns in strings of symbols, Pacific J. Math. 85 (1979), 261-294.

[4] J.D. Currie, There are ternary circular square-free words of length n for $n \geqslant 18$, Electron. J. Combin. 9 (2002) #N10, 7pp.

[5] J.D. Currie, Pattern avoidance; themes and variations, Theor. Comput. Sci., 339 (2005), 7-18.

[6] J. Grytczuk, Nonrepetitive colourings of graphs - a survey, Int. J. Math. Math. Sci. (2007), Art. ID 74639, 10 pp.

[7] J. Grytczuk, Thue type problems for graphs, points, and numbers, Discrete Math. 308(19) (2008), Simonovits '06, pp. 4419–4429.

[8] J. Grytczuk, J. Przybyło and X. Zhu, Thue choice number of paths, Random Structures Algorithms 38(2011), no. 1-2, 162-173.

[9] A. Kündgen, M. J. Pelsmajer, Nonrepetitive colourings of graphs of bounded treewidth, Discrete Math. 308(19) (2008), Simonovits '06, pp. 4473–4478.

[10] M. Lothaire, Combinatorics on Words, Addison-Wesley, Reading MA, 1983.

[11] J. Nešetřil, P. Ossona de Mendez, Tree depth, subgraph coloring and homomorphism bounds, European J. Combin. 27(6) (2006), pp. 1022–1041.

[12] W. Pegden, Highly nonrepetitive sequences: winning strategies from the Local Lemma, Random Structures Algorithms 38(2011), no. 1-2, 140-161.

[13] A. Thue, Über unendliche Zeichenreichen, Norske Vid. Selsk. Skr., I Mat. Nat. Kl., Christiania, 7 (1906), 1-22.