

Pattern avoidance on graphs

Jarosław Grytczuk^{a, b}

^a*Faculty of Mathematics, Computer Science, and Econometrics, University of Zielona Góra, 65-516 Zielona Góra, Poland*

^b*Algorithmic Research Group, Faculty of Mathematics and Computer Science, Jagiellonian University, Kraków, Poland*

Received 5 December 2002; received in revised form 6 July 2004; accepted 18 November 2005

Available online 4 December 2006

Abstract

In this paper we consider colorings of graphs avoiding certain patterns on paths. Let X be a set of variables and let $p = x_1x_2 \dots x_r$, $x_i \in X$, be a pattern, that is, any sequence of variables. A finite sequence s is said to match a pattern p if s may be divided into non-empty blocks $s = B_1B_2 \dots B_r$, such that $x_i = x_j$ implies $B_i = B_j$, for all $i, j = 1, 2, \dots, r$. A coloring of vertices (or edges) of a graph G is said to be p -free if no path in G matches a pattern p . The pattern chromatic number $\pi_p(G)$ is the minimum number of colors used in a p -free coloring of G .

Extending the result of Alon et al. [Non-repetitive colorings of graphs, Random Struct. Alg. 21 (2002) 336–346] we prove that if each variable occurs in a pattern p at least $m \geq 2$ times then $\pi_p(G) \leq c\Delta^{m/(m-1)}$, where c is an absolute constant. The proof is probabilistic and uses the Lovász Local Lemma. We also provide some explicit p -free colorings giving stronger estimates for some simple classes of graphs. In particular, for some patterns p we show that $\pi_p(T)$ is absolutely bounded by a constant depending only on p , for all trees T .

© 2006 Elsevier B.V. All rights reserved.

MSC: primary 05C38 15A15; secondary 05A15 15A18

Keywords: Nonrepetitive graph coloring; Pattern chromatic number; Avoidable pattern

1. Introduction

In this paper we consider a more general form of graph colorings introduced in [2] as an analog of Thue's nonrepetitive sequences. Recall that a finite sequence $u = u_1u_2 \dots u_n$ of symbols from a set C is nonrepetitive if it does not contain a sequence of the form $yy = y_1 \dots y_my_1 \dots y_m$, $y_i \in C$, as a subsequence of consecutive terms. For instance the sequence $u = 1232132312321$ over the set $C = \{1, 2, 3\}$ is nonrepetitive, while $v = 12313231321$ is not. A well-known theorem of Thue [23] asserts that there exist arbitrarily long nonrepetitive sequences built of only three different symbols.

Nonrepetitive sequences have striking applications in many fields of mathematics. In fact, they were rediscovered independently by many authors, each time showing up new properties and unexpected relationships with seemingly distant problems (see [1,10,18,19,21]). A lot of various generalizations of Thue's sequences have been considered so far respecting the wealth of their unusual properties. The whole branch of Pattern Avoidance emerged, with many beautiful results, methods and applications intersecting such diverse areas as Combinatorics, Group Theory, Universal Algebra, Number Theory, Dynamical Systems, Formal Language Theory, etc. (see [4–16,18–24]).

E-mail address: J.Grytczuk@wmie.uz.zgora.pl.

The following graph theoretic version of nonrepetitiveness has been introduced recently in [2]. A coloring of vertices of a graph G is *nonrepetitive* if the sequence of colors on any simple path in G is nonrepetitive. The minimal number of colors needed is the *Thue number* of G , denoted by $\pi(G)$. For instance, Thue's theorem asserts that $\pi(P_n) = 3$, for all $n \geq 4$, where P_n is the simple path with n vertices. It has been proved in [2] that there exists an absolute constant c_1 such that $\pi(G) \leq c_1 \Delta^2$, for all graphs G with maximum degree at most Δ . Moreover, this bound is nearly tight, since for any $\Delta > 1$ there are graphs G with maximum degree Δ satisfying $\pi(G) \geq c_2 \Delta^2 (\log \Delta)^{-1}$, for some absolute constant $c_2 > 0$. The proofs rely on the probabilistic method and at the moment no constructive argument is known for any $\Delta \geq 3$.

The purpose of this paper is to consider more general *pattern avoiding* colorings of graphs defined as follows. Let X be an infinite set of symbols, called *variables*, and let $p = x_1 x_2 \dots x_r$, $x_i \in X$, be a *pattern*, i.e., any finite sequence over X . A finite sequence s is said to *match* a pattern p if s may be divided into non-empty blocks $s = B_1 \dots B_r$ such that $x_i = x_j$ implies $B_i = B_j$, for all $i, j = 1, 2, \dots, r$. In other words, there is a substitution of non-empty blocks into variables of p giving the sequence s . For instance, the sequence $s = 1231232121$ matches a pattern $p = xxyy$, as it may be obtained from p by substituting $x = 123$ and $y = 21$. We will write $s = M(p)$ to denote that s matches p under some assignment M of blocks to variables of p .

A coloring of vertices (or edges) of a graph G is said to *encounter* a pattern p if there exists a simple path P in G such that the sequence of colors on P matches p . Otherwise, we say that a coloring *avoids* p , or is *p -free*. The *pattern chromatic number* $\pi_p(G)$ is the minimal number of colors in a p -free coloring of G .

A general problem is to recognize those patterns p for which $\pi_p(G)$ is bounded by a constant in the class of graphs G with bounded maximum degree. In that case the function

$$\pi_p(n) = \max\{\pi_p(G) : \Delta(G) \leq n\}$$

is well defined for all $n \geq 1$, and we will say that p is *avoidable on graphs*. This concept extends a well known notion of avoidable patterns introduced independently by Zimin [24] and Bean et al. [5]. Recall that a pattern p is *avoidable* if there are arbitrarily long sequences avoiding p over some finite set of symbols. In our notation this is expressed by $\pi_p(P_n) \leq c_p$ for every $n \geq 1$, where c_p is a constant depending only on p .

Although avoidable patterns were completely characterized already in [5] and [24], a lot of important questions are still unanswered. For instance, it is not known if $\pi_p(P_n) \leq 5$ holds for all $n \geq 1$ and all avoidable patterns p . An excellent exposition of this fascinating branch of Combinatorics on Words may be found in [15].

The paper is organized as follows. In Section 2 we prove that there is an absolute constant c such that

$$\pi_p(n) \leq cn^{m/(m-1)}$$

for any pattern p containing each of its variables at least twice, where $m = m(p)$ is the minimum multiplicity of a variable occurring in p . The proof is a simple modification of a probabilistic argument from [2], which is based on the local lemma. The question whether any avoidable pattern is avoidable on graphs remains open. In Section 3 we consider p -free *edge* colorings of graphs. The same probabilistic estimate holds for the related *pattern chromatic index* $\pi'_p(G)$ of a graph G . However, explicit colorings are found giving slightly better bounds for a few special classes of graphs and patterns. The final section contains some of a variety of problems for future investigations.

2. The probabilistic bound on $\pi_p(G)$

In this section we give a probabilistic upper bound for $\pi_p(G)$ for some classes of patterns p . The result is based on the following equivalent formulation of the Lovász Local Lemma (see [3,17]), which was previously applied in [2] for the pattern $p = xx$.

Lemma 1. (The Local Lemma; multiple version) Let A_1, A_2, \dots, A_n be events in an arbitrary probability space and let $G = (V, E)$ be a related dependency graph, where $V = \{A_1, A_2, \dots, A_n\}$ and A_i is mutually independent of all the events $\{A_j : A_i A_j \notin E\}$, for each $1 \leq i \leq n$. Let $V = V_1 \cup V_2 \cup \dots \cup V_N$ be a partition such that all the events $A_i \in V_r$ have the same probability p_r , $r = 1, 2, \dots, N$. Suppose that there are real numbers $0 \leq a_1, a_2, \dots, a_N < 1$ and $\Delta_{rs} \geq 0$,

$r, s = 1, 2, \dots, N$ such that the following conditions hold:

- $p_r \leq a_r \prod_{s=1}^N (1 - a_s)^{\Delta_{rs}}$ for all $r = 1, 2, \dots, N$,
- for each $A_i \in V_r$ the size of the set $\{A_j \in V_s : A_i A_j \in E\}$ is at most Δ_{rs} , for all $r, s = 1, 2, \dots, N$.

Then $\Pr\left(\bigcap_{i=1}^n \bar{A}_i\right) > 0$.

The number of occurrences of a variable in a pattern p is called its *multiplicity*. Let $m(p)$ denote the minimum multiplicity among all variables occurring in p . For instance, $m(xyzxzyx) = 2$ and $m(xyxxzyx) = 1$. We will show below that any pattern with $m(p) \geq 2$ is avoidable on graphs.

Theorem 1. *There is an absolute constant c such that if p is a pattern with minimum multiplicity $m = m(p) \geq 2$ then*

$$\pi_p(G) \leq c \Delta^{m/(m-1)}$$

for all graphs G with maximum degree at most Δ .

Proof. Let G be a simple graph with maximum degree Δ and let p be a pattern of minimum multiplicity $m \geq 2$. Consider a random coloring of vertices of G with K colors, where K will be specified later. For a path R in G of length r (the number of vertices in R) denote by $A(R)$ the event that the sequence of colors of R matches p . Set $V_r = A(R) : R$ is a path of length r and note that $p_r \leq K^{-r(1-1/m)}$. Let $a_s = 2^{-s} \Delta^{-s}$ and note that $(1 - a_s) \geq e^{-2a_s}$, as $a_s \leq 2^{-1}$, for all $s \geq 1$. Since each path of length r shares a vertex with at most $rs \Delta^s$ paths of length s , we may take $\Delta_{rs} = rs \Delta^s$. Now, the Local Lemma applies provided

$$K^{-r(1-1/m)} \leq a_r \prod_s (1 - a_s)^{rs \Delta^s}.$$

Thus, it is enough to show that

$$K^{-r(1-1/m)} \leq 2^{-r} \Delta^{-r} \prod_s e^{-2a_s rs \Delta^s}.$$

Substituting $a_s = 2^{-s} \Delta^{-s}$ and reducing gives

$$K^{(m-1)/m} \geq 2\Delta \exp\left(2 \sum_s 2^{-s} s\right).$$

Since $\sum_{s=1}^{\infty} 2^{-s} s = 2$ the Local Lemma guarantees the existence of a p -free K -coloring if $K \geq (2\Delta e^4)^{m/(m-1)}$. This completes the proof. \square

A simple consequence of the above theorem extends the class of patterns avoidable on graphs.

Corollary 1. *Let p be a pattern of length l with d distinct variables. If $l \geq 2^d$ then p is avoidable on graphs.*

Proof. We use induction with respect to d to prove that if p is not avoidable on graphs then $l(p) < 2^d$. For $d = 1$ the assertion follows at once from Theorem 1. Assume that it holds up to some $d > 1$ and consider a pattern p with $d + 1$ distinct variables. If p is not avoidable on graphs then, according to Theorem 1, there is a variable x occurring only once in p . Then p can be written as $p = p_1 x p_2$, where p_1 and p_2 are patterns with at most d variables, none of which may be avoidable on graphs. Hence, by the induction hypothesis $l(p_i) \leq 2^d - 1$, $i = 1, 2$, which gives $l(p) \leq 2^{d+1} - 1$. \square

Unfortunately, not all avoidable patterns are covered by the above results. For instance, the pattern $p = xy t_1 y z t_2 z x t_3 y x t_4 x z$ with seven distinct variables is known to be avoidable on four symbols (see [4,15]), while $m(p) = 1$ and $l(p) = 14$. We do not know at the moment if p is avoidable on graphs with $\Delta = 3$.

3. Explicit colorings

Let us start with a simple construction of *edge* colorings of K_n avoiding patterns p , in which every variable occurs an even number of times. A definition of *p-free edge coloring* of a graph G is the same as for the vertex case, and the related *p-chromatic index* is denoted by $\pi'_p(G)$. Note however that full cycles in G are allowed to match a pattern p . Our constructions here are similar to those found in [2] for the pattern $p = xx$.

Theorem 2. *Let p be a pattern such that each variable occurring in p occurs an even number of times. Then $\pi'_p(K_{2^k}) \leq 2^k - 1$ for all $k \geq 1$. In consequence, $\pi'_p(K_n) \leq 2n - 3$ for every $n \geq 2$.*

Proof. Our argument is based on addition mod 2. Let a positive integer k be chosen so that $2^{k-1} < n \leq 2^k$. Label the vertices of K_n by different elements of the additive group \mathbb{Z}_2^k , the direct product of k copies of \mathbb{Z}_2 . Next, color the edges by non-zero elements of \mathbb{Z}_2^k , such that an edge with vertices labeled by x and y gets the color $x + y$. It is seen easily that, by the mod 2 property, for every path P in K_n there is a color class intersecting P in an odd number of edges; otherwise P would be a cycle. Hence, P cannot match p , which completes the proof. \square

Corollary 2. *Let p be a pattern of length $l(p) \geq 2^d$, where d is a number of distinct variables appearing in p . Then $\pi'_p(K_n) \leq 2n - 3$ for every $n \geq 2$.*

Proof. We will show that a pattern $p = x_1x_2 \dots x_l$, $l \geq 2^d$, must contain a subpattern satisfying the assumption of Theorem 2. Let y_1, y_2, \dots, y_d denote all distinct variables of p . For $1 \leq i \leq l$ and $1 \leq j \leq d$ define $v_j^{(i)}$ as a residue mod 2 of the number of occurrences of y_j among the first i terms of p . Consider a related sequence of binary vectors $v^{(i)} = (v_1^{(i)}, \dots, v_d^{(i)})$, $i = 1, \dots, l$. Since $l \geq 2^d$, either $v^{(i)} = (0, \dots, 0)$ is a zero vector, for some i , or there are i_1 and i_2 for which $v^{(i_1)} = v^{(i_2)}$. In the first case a subpattern $p_1 = x_1 \dots x_{i_1}$ does the job. In the second case, one may take $p_2 = x_{i_1+1} \dots x_{i_2}$. \square

In the next two examples we use pattern avoiding sequences with additional constraints for colorings of trees. Let S be a finite set of avoidable patterns. It is not hard to see that there exists a finite set C such that all patterns of S are *simultaneously* avoidable over C , i.e., there is an infinite sequence $w(S)$ over C avoiding each pattern of S (see [15]). Denote by $\mu(S)$ the smallest size of such a set C . Consider the class of patterns \mathcal{D} satisfying the following condition: for every $p \in \mathcal{D}$, in any decomposition $p = p_1p_2$ at least one of subpatterns p_i is avoidable. For instance, the set of patterns from Corollary 1 satisfies this condition.

A *reflection* of a sequence $u = u_1u_2 \dots u_n$ is a sequence $\tilde{u} = u_nu_{n-1} \dots u_1$ that is, u written backward. In the following theorem we will make use of a sequence avoiding simultaneously not only all avoidable subpatterns of p , but also their reflections.

Theorem 3. *Let $p \in \mathcal{D}$ be a fixed pattern with $m(p) \geq 2$ and let S be the set containing all avoidable subpatterns of p together with their reflections. Let T be any tree with $\Delta(T) \geq 2$. Then $\pi'_p(T) \leq 2\mu(S)(\Delta(T) - 1)$.*

Proof. Let T be a tree of maximum degree $\Delta \geq 2$ and let $p = x_1x_2 \dots x_n$. Choose a vertex of degree strictly less than Δ as a root of T and arrange the rest of vertices by their distance from the root. Then the edges of T can be partitioned into levels, L_1, L_2, \dots consisting of disjoint stars. Let $w(S) = w_1w_2 \dots$ be an infinite sequence over the set $C = \{1, 2, \dots, \mu(S)\}$ avoiding all patterns of S . That is, if $p = rqs$, where q is avoidable and r, s are possibly empty, then $w(S)$ avoids q and \tilde{q} . Define another sequence $u(S) = u_1u_2 \dots$ over the set of pairs $\{(t, j) : t \in C, j = 0, 1\}$ by $u_t = (w_t, j)$, where $t \equiv j \pmod{2}$ for all $t \geq 1$. Clearly, $u(S)$ consists of at most $2\mu(S)$ symbols, still avoids all patterns of S , but satisfies additionally $u_t \neq u_{t+1}$ for all $t \geq 1$. Next, take $2\mu(S)$ disjoint sets of colors $C_i = \{i', i'', \dots, i^{(\Delta-1)}\}$, $i = 1, 2, \dots, 2\mu(S)$ and color each star on level L_t by distinct colors from the set C_{u_t} . We claim that this coloring is p -free. In fact, suppose that $M(p) = M(x_1) \dots M(x_n)$ is a sequence of colors on some path $P \subseteq T$ matching a pattern p under some substitution M . Clearly P cannot be monotone, hence it must have two edges e, f of the same star with colors from the same set C_i . This situation happens only once in P . Moreover, since $m(p) \geq 2$, the common vertex of edges e and f must separate two adjacent blocks of $M(p)$; otherwise, the colors of e and f would appear consecutively

on a monotone part of P , which contradicts the mod 2 property of $u(S)$. Hence, p can be written as $p = p_1 p_2$ and both blocks $M(\tilde{p}_1)$ and $M(p_2)$ appear on monotone paths of T . This contradicts the assumption that $p \in \mathcal{D}$ and the definitions of S and $u(S)$. \square

The following theorem concerns vertex colorings of trees and shows that $\pi_p(T)$ is bounded by a constant depending only on p (but not on Δ), if $p \in \mathcal{D}$ and $m(p) \geq 2$. The proof goes along similar lines.

Theorem 4. *Let $p \in \mathcal{D}$ be a fixed pattern with $m(p) \geq 2$. Let S be the set containing all avoidable subpatterns of p and all their reflections. Then $\pi_p(T) \leq 3\mu(S)$, for any tree T .*

Proof. Let $w(S)$ be as above and let $u(S) = u_1 u_2 \dots$ be defined by $u_t = (w_t, j)$, where $j \in \{0, 1, 2\}$ and $t \equiv j \pmod{3}$, for all $t \geq 1$. Arrange a tree T into levels and color all vertices of level L_t with the same color u_t . Suppose that some path P in T matches a pattern p under a substitution M . Since P cannot be monotone it must contain a subpath $v_1 v_2 v_3$ where v_2 is the highest vertex of P . Let $c_i = (a_i, j_i)$ be the color of v_i , $i = 1, 2, 3$. Since $j_1 = j_3$, a triple $c_1 c_2 c_3$ cannot appear on monotone part of P , by the mod 3 property of u . It follows that v_2 separates blocks of $M(p)$, which leads to a contradiction as in the preceding proof. \square

4. Final discussion

We would like to conclude with a few of a range of problems that arise naturally on the intersection of two topics mixed in this paper—Pattern Avoidance and Graph Coloring. Certainly, the main problem that is left open is to characterize those patterns p that are avoidable on graphs. As we mentioned in the Introduction, a complete characterization is known for patterns avoidable in the usual sense. It might be the case that the following conjecture is true.

Conjecture 1. Any avoidable pattern is avoidable on graphs.

By Theorem 1 and Corollary 1, the only patterns left to be considered are those with $m(p) = 1$ and $l(p) < 2^d$, where d is the number of distinct variables of p . Note also that validity of this statement for vertex colorings implies the edge case, but not conversely.

Recall from the Introduction that $\pi_p(n)$ denotes $\max\{\pi_p(G) : \Delta(G) \leq n\}$. If p is avoidable on graphs then $\pi_p(n)$ is finite for all $n \geq 1$ and there is a question of its asymptotics.

Conjecture 2. For any pattern p avoidable on graphs there is a constant k such that $\pi_p(n) = O(n^k)$.

Again, by Theorem 1, only the case of $m(p) = 1$ is not clear.

The case of edge colorings is somewhat different and here one may expect lower order of magnitude for the analogous function $\pi'_p(n)$. A conjecture generalizing the one proposed in [2] looks as follows.

Conjecture 3. For any pattern p avoidable on graphs $\pi'_p(n) = O(n)$.

Our next problem is inspired by one of the major problems in Pattern Avoidance—the question whether some absolute constant number of symbols suffice to avoid any avoidable pattern (see [4,7,15]). We will take a risk and propose the following more general conjecture.

Conjecture 4. For every $n \geq 1$ there is a constant c depending only on n such that $\pi_p(n) \leq c$ for all patterns p avoidable on graphs.

In the end we restrict ourselves to one concrete situation. Theorem 4 shows that for certain patterns the number $\pi_p(T)$ is absolutely bounded from above for all trees T , no matter how large is $\Delta(T)$. In particular, for $p = xx$ one may show that $\pi(T) \leq 4$. So, one naturally wonders whether similar phenomenon could hold for other classes of graphs, at least for some specific patterns.

Conjecture 5. There is a natural number N such that $\pi(G) \leq N$, for any planar graph G .

Acknowledgement

This work was partially supported by Grant KBN 1P03A 017 27.

References

- [1] J.-P. Allouche, J. Shallit, The ubiquitous Prouhet–Thue–Morse sequence, Sequences and their applications (Singapore, 1998), Springer Series in Discrete Mathematics and Theoretical Computer Science, Springer, London, 1999, pp. 1–16.
- [2] N. Alon, J. Grytczuk, M. Hałuszczak, O. Riordan, Non-repetitive colorings of graphs, *Random Struct. Alg.* 21 (2002) 336–346.
- [3] N. Alon, J.H. Spencer, The Probabilistic Method, Second ed., Wiley, New York, 2000.
- [4] K.A. Baker, G.F. McNulty, W. Taylor, Growth problems for avoidable words, *Theoret. Comput. Sci.*, 69 (1989) 319–345.
- [5] D.R. Bean, A. Ehrenfeucht, G.F. McNulty, Avoidable patterns in strings of symbols, *Pacific J. Math.* 85 (1979) 261–294.
- [6] Ch. Choffrut, J. Karhumäki, Combinatorics of Words, in: G. Rozenberg, A. Salomaa (Eds.), *Handbook of Formal Languages*, Springer, Berlin, Heidelberg, 1997, pp. 329–438.
- [7] J.D. Currie, Open problems in pattern avoidance, *Amer. Math. Monthly* 100 (1993) 790–793.
- [8] J.D. Currie, Words avoiding patterns: open problems, manuscript.
- [9] F.M. Dekking, Strongly non-repetitive sequences and progression free sets, *J. Combin. Theory Ser. A* 16 (1974) 159–164.
- [10] W.H. Gottschalk, G.A. Hedlund, A characterization of the Morse minimal set, *Proc. Amer. Math. Soc.* 15 (1964) 70–74.
- [11] J. Grytczuk, Pattern avoiding colorings of Euclidean spaces, *Ars Combin.* 64 (2002) 109–116.
- [12] J. Grytczuk, W. Śliwa, Non-repetitive colorings of infinite sets, *Discrete Math.* 265 (2003) 365–373.
- [13] J. Larson, R. Laver, G. McNulty, Square-free and cube-free colorings of the ordinals, *Pacific J. Math.* 89 (1980) 137–141.
- [14] M. Lothaire, *Combinatorics on Words*, Addison-Wesley, Reading, MA, 1983.
- [15] M. Lothaire, *Algebraic Combinatoric on Words*, Cambridge, 2001.
- [16] C. Mauduit, Multiplicative properties of the Thue–Morse sequence, *Periodica Math. Hungar* 43 (2001) 137–153.
- [17] M. Molloy, B. Reed, Graph Colouring and the Probabilistic Method, *Algorithms and Combinatorics*, vol. 23, Springer, Berlin, 2002.
- [18] M. Morse, A one-to-one representation of geodesics on a surface of negative curvature, *Amer. J. Math.* 43 (1921) 35–51.
- [19] P.S. Novikov, On periodic groups, *Dokl. Akad. Nauk SSSR* 127 (1959) 749–752.
- [20] P.A.B. Pleasants, Non-repetitive sequences, *Proc. Cambridge Philos. Soc.* 68 (1970) 267–274.
- [21] E. Prouhet, Memoire sur quelques relations entre les puissances des nombres, *C. R. Acad. Sci. Paris* 33 (1851) 31.
- [22] M. Queffelec, Substitutions dynamical systems - Spectral analysis, *Lecture Notes in Mathematics*, vol. 1294, Springer, Berlin, 1987.
- [23] A. Thue, Über unendliche Zeichenreihen, *Norske Vid Selsk. Skr. I. Mat. Nat. Kl. Christiana*, 7 (1906) 1–22.
- [24] A.I. Zimin, Blocking sets of terms, *Mat. Sb. (NS)* 119 (1982) 363–375.