2 × N GRIDS HAVE UNBOUNDED ANAGRAM-FREE CHROMATIC NUMBER

Saman Bazarghani, Paz Carmi, Vida Dujmović, Pat Morin

Abstract. We show that anagram-free vertex colouring a 2 × n square grid requires a number of colours that increases with n. This answers an open question in Wilson’s thesis and shows that even graphs of pathwidth 2 do not have anagram-free colourings with a bounded number of colours.

1 Introduction

Two strings s and t are said to be anagrams of each other if s is a permutation of t. A single string s := s₁, ..., s₂r is anagramish if its first half s₁, ..., sᵣ and its second half sᵣ₊₁, ..., s₂r are anagrams of each other. A string is anagram-free if it does not contain an anagramish substring.

In 1961, Erdős [3] asked if there exists arbitrarily long strings over an alphabet of size 4 that are anagram-free.¹ In 1968 Evdokimov [4, 5] showed the existence of arbitrarily long anagram-free strings over an alphabet of size 25 and in 1971 Pleasants [10] showed an alphabet of size 5 is sufficient. Erdős’s question was not fully resolved until 1992, when Keränen [7] answered it in the affirmative.

A path v₁, ..., v₂r in a graph G is anagramish under a vertex c-colouring φ : V(G) → {1, ..., c} if φ(v₁), ..., φ(v₂r) is an anagramish string. The colouring φ is an anagram-free colouring of G if no path in G is anagramish under φ. The minimum integer c for which G has an anagram-free vertex c-colouring is called the anagram-free chromatic number of G and is denoted by χπ(G).

For a graph family ℱ, χπ(ℱ) := max{χπ(G) : G ∈ ℱ} or χπ(ℱ) := ∞ if the maximum is undefined. The results on anagram-free strings discussed in the preceding paragraph can be interpreted in terms of χπ(ℱ) where ℱ is the family of all paths. Slightly more complicated than paths are trees. Wilson and Wood [13] showed that χπ(T) = ∞ for the family T of trees and Kamčev, Łuczak, and Sudakov [6] showed that χπ(T₂) = ∞ even for the family T₂ of binary trees.

One positive result in this context is that of Wilson and Wood [13], who showed that every tree T of pathwidth p has χπ(T) ≤ 4p + 1. Since trees are graphs of treewidth 1 it is natural to ask if this result can be extended to show that every graph G of treewidth t

¹This was an incredibly prescient question since it is not at all obvious that there exist arbitrarily long anagram-free strings over any finite alphabet. The only justification for choosing the constant 4 is that a short case analysis rules out the possibility of length-8 anagram-free strings over an alphabet of size 3.
and pathwidth $p$ has $\chi_\pi(G) \leq f(t, p)$ for some $f : \mathbb{N}^2 \to \mathbb{N}$. Carmi, Dujmović, and Morin [2] showed that such a generalization is not possible for any $t \geq 3$ by giving examples of $n$-vertex graph of pathwidth 3 (and treewidth 3) with $\chi_\pi(G) \in \Omega(\log n)$. The obvious remaining gap left by these two works is graphs of treewidth 2. Our main result is to show that $G_n$, the $2 \times n$ square grid has $\chi_\pi(G_n) \in \omega_n(1)$. Since $G_n$ has pathwidth 2, we have:

**Theorem 1.** For every $c \in \mathbb{N}$, there exists a graph of pathwidth 2 that has no anagram-free vertex $c$-colouring.

Wilson [11, Section 7.1] conjectured that $\chi_\pi(G_n) \in \omega_n(1)$, so this work confirms this conjecture. Prior to the current work, it was not even known if the family of $n \times n$ square grids had anagram-free colourings using a constant number of colours.

In a larger context, this lower bound gives more evidence that, except for a few special cases (paths [4, 7, 10], trees of bounded pathwidth [13], and highly subdivided graphs [12]), the qualitative behaviour of anagram-free chromatic number is not much different than that of treedepth/centered colouring [9]. Very roughly: For most graph classes, every graph in the class has an anagram-free colouring using a bounded number of colors precisely when every graph in the class has a colouring using a bounded number of colours in which every path contains a colour that appears only once in the path.

The remainder of this paper is organized as follows: Section 2 gives some definitions and states a key lemma that shows that, under a certain periodicity condition, every sufficiently long string contains a substring that is $\epsilon$-close to being anagramish. In Section 3 we prove Theorem 1. In Section 4 we prove the key lemma. Section 5 concludes with some final remarks about the (non-)constructiveness of our proof technique.

## 2 Periodicity in Strings

An *alphabet* $\Sigma$ is a finite non-empty set. A *string over* $\Sigma$ is a (possibly empty) sequence $s := s_1, \ldots, s_n$ with $s_i \in \Sigma$ for each $i \in \{1, \ldots, n\}$ The *length* of $s$ is the length, $n$, of the sequence. For an integer $k$, $\Sigma^k$ (the $k$-fold cartesian product of $\Sigma$ with itself) is the set of all length-$k$ strings over $\Sigma$. The *Kleene closure* $\Sigma^* := \bigcup_{k=0}^\infty \Sigma^k$ is the set of all strings over $\Sigma$. For each $1 \leq i \leq j \leq n + 1$, $s[i : j] := s_i, s_{i+1}, \ldots, s_{j-1}$ is called a *substring* of $s$. (Note the convention that $s[i : j]$ does not include $s_j$, so $s[i : j]$ has length $j - i$.)

Let $s := s_1, \ldots, s_n$ be a string over an alphabet $\Sigma := [s_i : i \in \{1, \ldots, n\}]$ and, for each $a \in \Sigma$, define $h_a(s) := |\{i \in \{1, \ldots, n\} : s_i = a\}|$. The *histogram* of $s$ is the integer-valued $|\Sigma|$-vector $h(s) := (h_a(s) : a \in \Sigma)$ indexed by elements of $\Sigma$. Observe that a string $s_1, \ldots, s_2r$ is anagramish if and only if $h(s_1, \ldots, s_r) = h(s_{r+1}, \ldots, s_{2r})$ or, equivalently, $h(s_1, \ldots, s_r) - h(s_{r+1}, \ldots, s_{2r}) = 0$. For each $a \in \Sigma$, let $\tau_a := |h_a(s_1, \ldots, s_r) - h_a(s_{r+1}, \ldots, s_{2r})|$. Then $\tau(s) := \sum_{a \in \Sigma} \tau_a(s)$ is a useful measure of how far a string is from being anagramish and $\tau(s) = 0$ if and only if $s$ is anagramish.

A string $s := s_1, \ldots, s_n$ is $\ell$-periodic if each length-$\ell$ substring of $s$ contains every character in $\Sigma := \bigcup_{i=1}^n s_i$. We make use of the following lemma, which states that every sufficiently long $\ell$-periodic string contains a long substring that is $\epsilon$-close to being anagramish.

**Lemma 2.** For each $r_0, \ell \in \mathbb{N}$ and each $\epsilon > 0$, there exists a positive integer $n$ such that each
The proof of Lemma 2 is deferred to Section 4. We now give some intuition as to how it is used. The process of checking if a string is anagramish is often viewed as finding common terms in the first and second halves and crossing them both out. If this results in a complete cancellation of all terms, then the string is an anagram. Lemma 2 tells us that we can always find a long substring $s$ where, after exhaustive cancellation, only an $\epsilon$-fraction of the original terms remain. Informally, the substring $s$ is $\epsilon$-close to being anagramish.

Lemma 2 says that, if up to $\epsilon r$ terms in each half of $s$ were each allowed to cancel two terms each in the other half of $s$, then it would be possible to complete the cancellation process. To achieve this type of one-versus-two cancellation in our setting, we decompose our coloured pathwidth-2 graph into pieces of constant size. The vertices in each piece can be covered with one path or partitioned into two paths. In this way an occurrence $H_z$ of a particular coloured piece in one half can be matched with two like-coloured pieces $H_x$ and $H_y$ in the other half. We construct a single path $P$ that contains all vertices in $H_z$ and only half the vertices in each of $H_x$ and $H_y$. In this way, the colours of vertices $P \cap H_z$ can cancel the colours of the vertices in $P \cap (H_x \cup H_y)$.

Since Lemma 2 requires that the string $s$ be $\ell$-periodic, the following lemma will be helpful in obtaining strings that can be used with Lemma 2.

**Lemma 3.** Let $\Sigma$ be an alphabet and let $P : \Sigma^* \rightarrow \{0, 1\}$ be a function such that,

1. (A1) if $P(s') = 0$ for some substring $s'$ of $s$ then $P(s) = 0$; and
2. (A2) for each $n \in \mathbb{N}$, there exists at least one $s \in \Sigma^n$ such that $P(s) = 1$.

Then there exists $\ell \in \mathbb{N}$ and $\Xi \subseteq \Sigma$ such that, for each $n \in \mathbb{N},$

1. (C1) there exists $s \in \Xi^n$ such that $P(s) = 1$; and
2. (C2) every string in $\{s \in \Xi^n : P(s) = 1\}$ is $\ell$-periodic.

*Proof.* Take $\Xi$ to be a minimal subset of $\Sigma$ such that there exists $s \in \Xi^n$ with $P(s) = 1$, for each $n \in \mathbb{N}$. Such a $\Xi$ exists by (A2) and the fact that $\Sigma$ is finite. By definition, $\Xi$ satisfies (C1) so we need only show that it also satisfies (C2). If $|\Xi| = 1$ then we are done since every string over a 1-character alphabet is 1-periodic.

For $|\Xi| > 1$, the minimality of $\Xi$ implies that, for any $a \in \Xi$, there exist $\ell_a \in \mathbb{N}$ such that $P(s) = 0$ for each $s \in \Xi^{\ell_a}$. Therefore (A1) implies that, for each $\ell \geq \ell_a$, $P(s) = 0$ for each $s \in \Xi^\ell$. Set $\ell := \max\{\ell_a : a \in \Xi\}$. Now (A1) implies that, for every $s \in \Xi$, every length $\ell$-substring of $s$ contains every character in $\Xi$, so $s$ is $\ell$-periodic. \qed

### 3 Proof of Theorem 1

For each $n \in \mathbb{N}$, let $G_n$ be the $2 \times n$ square grid with top row $a_0, \ldots, a_{n-1}$ and bottom row $b_0, \ldots, b_{n-1}$ (see Figure 1). For convenience, we let $G := G_\infty$. For each $i, j \in \mathbb{N}$ with $i \leq j$, define $G[i : j] := G[\bigcup_{i \leq z \leq j} [a_z, b_z]]$ and we call $G[i : j]$ a $(j-i)$-block. Note that each $t$-block $G[i : i + t]$ is isomorphic to $G_t$ with the mapping $f : G_t \rightarrow V(G[i : i + t])$ given by $f_t(a_j) := a_{i+j}$ and $f_t(b_j) := b_{i+j}$ for each $j \in \{0, \ldots, t - 1\}$.
For each \( n \in \mathbb{N} \), let \( \Phi_{c,n} \) be the set of all \( c^{2n} \) functions \( \phi : V(G_n) \to \{1, \ldots, c\} \), i.e., all vertex \( c \)-colourings of \( G_n \). Given some \( \phi \in \Phi_{c,n} \), each \( t \)-block \( G[i : i + t] \) defines a vertex colouring \( \phi[i : i + t] \in \Phi_{c,i} \) of \( G_t \) defined as \( \phi[i : i + t](v) := \phi(f_i(v)) \) for each \( v \in V(G[i : i + t]) \).

Our strategy will be to break \( G_n \) up into small pieces using 4-blocks that all have the same colouring. Observe that any string \( s \in \Sigma^t \) defines a colouring of \( G_{4b} \) where \( \phi_s[4j : 4j + 4] := \phi_j \) for each \( j \in \{0, \ldots, b - 1\} \). Indeed, this is a bijection between \( c \)-colourings of \( G_{4b} \) and strings in \( \Phi_{c,4}^b \).

**Lemma 4.** If \( \chi_n(G_n) \leq c \) for each \( n \in \mathbb{N} \) then there exists \( \Xi_{c,4} \subseteq \Phi_{c,4} \) such that, for each \( b \in \mathbb{N} \), there exists \( s \in (\Xi_{c,4})^b \) such that \( \phi_s \) is an anagram-free vertex colouring of \( G_{4b} \) and every such \( s \) is \( \ell \)-periodic.

**Proof.** For any \( s \in (\Phi_{c,4})^* \), let \( P(s) := 1 \) if \( \phi_s \) is an anagram-free colouring of \( G_{4|s|} \) and let \( P(s) := 0 \) otherwise. Then \( P \) has property (A1) of Lemma 3 since any substring of \( s \) of \( s \) defines a colouring of \( G_{4|s|} \) that appears in the colouring of \( G_{4|s|} \); if the colouring \( \phi_s \) of \( G_{4|s|} \) is not anagrom-free then neither is the colouring \( \phi_s \) of \( G_{4|s|} \). By assumption, \( \chi_n(G) \leq c \), so \( G_{4b} \) has some anagram-free vertex \( c \)-colouring, so \( P \) also satisfies property (A2) of Lemma 3. The result now follows from Lemma 3. \( \square \)

Let \( \Xi_{c,4} \) and \( \ell \) be defined as in Lemma 4. Let \( \phi^* \) be an arbitrary element of \( \Xi_{c,4} \) and let \( \Sigma := \{(k, \phi) : k \in \{1, \ldots, \ell\}, \phi \in \Phi_{c,4k}\} \). For a string \( s := (k_1, \phi_1), \ldots, (k_r, \phi_r) \in \Sigma^* \), define \( n_{s,i} := 4(i + \sum_{j=1}^i k_j) \), define \( n_s := n_{s,r} \). Fix a vertex \( c \)-colouring \( \phi^* \in \Phi_{c,4} \) of \( G_4 \) and define the colouring \( \phi_s \) of \( G_n \) as follows:

1. for each \( i \in \{0, \ldots, r\} \), \( \phi[n_{s,i}, n_{s,i+4}] := \phi^* \); and
2. for each \( i \in \{1, \ldots, r\} \), \( \phi[n_{s,i-1} + 4, n_{s,i}] := \phi_i \).

See Figure 2. In words, \( G_n \) is decomposed into blocks each of whose length is a multiple of 4. There are colourful blocks of lengths \( 4k_1, 4k_2, \ldots, 4k_{|s|} \leq 4\ell \) and these are interleaved with boring blocks, each of length 4. The colourful blocks have their vertex colours determined by \( s \). The boring blocks are all coloured the same way, by \( \phi^* \).
Define the function $P : \Sigma^* \rightarrow \{0, 1\}$ so that $P(s) := 1$ if $\phi_s$ is an anagram-free colouring of $G_n$, and $P(s) := 0$ otherwise. Observe that any substring $s'$ of $s$ defines a colouring $\phi_{s'}$ of $G_{n'}$ that appears in the colouring $\phi_s$ of $G_n$. Therefore $P$ satisfies property (A1) of Lemma 3. Furthermore, if $\chi_n(G_n) \leq c$ for each $n \in \mathbb{N}$, then Lemma 4 implies that there exists a string $s \in \Sigma^b$ with $P(s) = 1$ for each $b \in \mathbb{N}$. Therefore $P$ satisfies property (A2) of Lemma 3. Therefore, Lemma 3 implies the following result:

**Lemma 5.** If $\chi_n(G_n) \leq c$ for each $n \in \mathbb{N}$ then there exists $\ell \in \mathbb{N}$ such that, for each $k \in \mathbb{N}$ there exists an $\ell$-periodic string $s \in \Sigma^k$ such that $\phi_s$ is an anagram-free vertex colouring of $G_n$.

Lemma 2 and 5 immediately imply:

**Lemma 6.** If $\chi_n(G_n) \leq c$ for each $n \in \mathbb{N}$ then, for each $r_0 \in \mathbb{N}$ and $\varepsilon > 0$, there exists $r \geq r_0$ and a string $s := s_1, \ldots, s_{2r} \in \Sigma^{2r}$ with $\tau(s) \leq \varepsilon r$ such that $\phi_s$ is an anagram-free vertex colouring of $G_n$.

**Proof.** By Lemma 5, for each $k \in \mathbb{N}$ there exists an $\ell$-periodic string $s' \in \Sigma^k$ such that $\phi_{s'}$ is an anagram-free vertex colouring of $G_n$. Applying Lemma 2 to $s'$ proves the existence of the desired string $s$. $\square$

Lemma 6 shows the existence of colourings of $G_n$ for arbitrarily large values of $n$ that are defined by strings that are $\varepsilon$-close to being anagramish. The last step, done in the next lemma, is to show that there is enough flexibility when constructing a path in $G_n$ that we can find a path that has an anagramish colour sequence.

**Lemma 7.** For any $\ell \in \mathbb{N}$, there exists $\varepsilon > 0$ and $r_0 \in \mathbb{N}$ such that, for any integer $r \geq r_0$ and any $\ell$-periodic $s := s_1, \ldots, s_{2r} \in \Sigma^{2r}$ with $\tau(s) \leq \varepsilon r$, the graph $G_n$ contains a path $P = v_1, \ldots, v_{2m}$ such that $\phi_s(v_1), \ldots, \phi_s(v_{2m})$ is anagramish.

**Proof.** For each $a \in \Sigma$ define $\delta_a := h_a(s_1, \ldots, s_r) - h_a(s_{r+1}, \ldots, s_{2r})$ and define sets $A_a \subseteq \{i \in \{1, \ldots, r - 1\} : s_i = a\}$ and $B_a \subseteq \{i \in \{r + 1, \ldots, 2r - 1\} : s_i = a\}$ as follows:

1. If $\delta_a = 0$ then $A_a = B_a = \emptyset$.
2. If $\delta_a > 0$ then $|A_a| = 2\delta_a = 2|B_a|$.
3. If $\delta_a < 0$ then $|A_a| = |B_a| = \frac{1}{2}|B_a|$.

Let $A := \bigcup_{a \in \Sigma} A_a$ and $B := \bigcup_{a \in \Sigma} B_a$. The sets $A_a$ and $B_a$ are chosen so that they satisfy the following **global independence constraint**: There is no pair $i, j \in A \cup B$ such that $i - j = 1$. To see that this is possible, first observe that, because $r$ is not in $A$ or in $B$ we need only concern ourselves with pairs where both $i, j \in \{1, \ldots, r - 1\}$ or pairs where both $i, j \in \{r + 1, \ldots, 2r - 1\}$. Thus, we can choose the elements of $A_a$, for each $a \in \Sigma$ and then independently choose the elements of $B_a$, for each $a \in \Sigma$.

We show how to choose the elements of $A_a$ for each $a \in \Sigma$. The same method works for choosing the elements in $B_a$. Observe that, because $s$ is $\ell$-periodic, $|\{i \in \{1, \ldots, r - 1\} : s_i = a\}| \geq \lceil (r - 1)/\ell \rceil$ for each $a \in \Sigma$. This allows us to greedily choose the elements in $A_a$ for each $a \in \Sigma$. At each step we simply avoid choosing $i$ if $i - 1$ or $i + 1$ have already been chosen in some previous step. At any step in the process, at most $\varepsilon r$ elements have already been
Figure 3: Constructing the anagramish path $P$: (1) top and bottom paths are matched with zig-zag paths; (2) all remaining colourful blocks receive top paths; (3) all boring blocks receive top, updown, or downup paths.

Figure 4: Subpaths of $P$ through colourful blocks: A top path and a bottom path contribute the same amount as a single zig-zag path.

chosen in previous steps and each of these eliminates at most 2 options. Therefore, there will always be an element available to choose, provided that $2\epsilon r < \lfloor (r-1)/\ell \rfloor$. In particular, for any $r \geq r_0 \geq 2\ell$, $\epsilon < 1/(4\ell)$ works.

We now construct the path $P$ in a piecewise fashion. Refer to Figure 3. For each $i \in \{1, \ldots, 2r\}$, let $H_i := G[n_{5,i} : n_{5,i+1} - 4]$. The subgraph $H_1, \ldots, H_{2r}$ are what is referred to above as colourful blocks. The colouring of $V(H_i)$ by $\phi_s$ is defined by $s_i$.

1. For each $a \in \Sigma$ such that $\delta_a > 0$, group the elements of $A_a$ into pairs. For each pair $(i, j)$, $P$ contains the path through the top row of $H_i$ and the path through the bottom row of $H_j$. For each element $i \in B_a$, $P$ contains the zig-zag path with both endpoints in the top row of $H_i$ and that contains every vertex of $H_j$. (Note that the zig-zag path begins at the top and ends at the bottom row because $H_i$ is a $t$-block for $t$ a multiple of 4; in particular, $t$ is even.)

2. For each $a \in \Sigma$ such that $\delta_a < 0$ we proceed symmetrically to the previous case, but reversing the roles of $A_a$ and $B_a$. Specifically, we group the elements of $B_a$ into pairs. For each pair $(i, j)$, $P$ contains the path through the top row of $H_i$ and the path through the bottom row of $B_j$. For each element $i \in A_a$, $P$ contains the zig-zag
path with both endpoints in the top row of $H_i$ and that contains every vertex of $H_i$.

3. For each $i \in \{1, \ldots, 2r\} \setminus \bigcup_{a \in \Sigma} (I_a \cup B_a)$, $P$ contains the top row of $H_i$.

The rules above define the intersection, $P_i$, of $P$ with each colourful block $H_i$ of $G_{ns}$. If $P_i$ is the path through the bottom (top) row of $H_i$ then we call $H_i$ a bottom (top) block. If $P_i$ is the zig-zag path that contains every vertex of $H_i$ then we call $H_i$ a zig-zag block. Note that $\sum_{a \in \Sigma} |\delta_a| = 0$ and this implies that the number of bottom blocks among $H_1, \ldots, H_{r-1}$ is the same as the number of bottom blocks among $H_{r+1}, \ldots, H_{2r}$. Indeed, this number is exactly $\beta := \frac{1}{2} \sum_{a \in \Sigma} |\delta_a| = \frac{1}{2} \tau(s)$.

We now define how $P$ behaves for the boring blocks, that we name $Q_0, \ldots, Q_{2r-1}$. The first boring block $Q_0$ comes immediately before $H_1$. Each boring block $Q_j$, for $j \in \{1, \ldots, 2r-1\}$ comes immediately after $H_j$ and immediately before $H_{j+1}$. In almost every case, $P$ uses the path through the top row of $Q_j$. The only exceptions are when $H_j$ or $H_{j+1}$ are bottom blocks. Note that, because of the global independence constraint, these two cases are mutually exclusive. See Figure 5.

1. When $H_j$ is a bottom block $P$ uses a path that begins at the bottom row of $Q_j$ but moves immediately to the top row of $Q_j$ and uses the entire path along the top row. We call this a downup path.

2. When $H_{j+1}$ is a bottom block, $P$ uses a path that begins at the top row of $Q_j$ and moves immediately to the bottom row of $Q_j$ and uses the entire path along the bottom row. We call this a updown path.

This completely defines the path $P := v_1, \ldots, v_{2m}$. All that remains is to argue that $\rho := \phi_s(v_1), \ldots, \phi_s(v_{2m})$ is anagramish.

Observe that the number of downup paths and the number of updown paths in $Q_0, \ldots, Q_{r-1}$ is exactly the same as the number of bottom blocks among $H_1, \ldots, H_{r-1}$ which is exactly $\beta$. Similarly, the number of updown paths and downup paths in $Q_{r+1}, \ldots, Q_{2r-1}$ is exactly $\beta$. Now every path that is neither downup nor updown uses the top row. This implies that the sequence of colours contributed to $\rho$ by the intersection of $P$ with $Q_0, \ldots, Q_{r-1}$ is a permutation of the sequence of colours contributed to $\rho$ by the intersection of $P$ with $Q_{r+1}, \ldots, Q_{2r-1}$.

Finally, by construction, each pair of top and bottom blocks in $H_1, \ldots, H_{r-1}$ contributes exactly the same amount as a single matching zig-zag block in $H_{r+1}, \ldots, H_{2r-1}$. Specifically, if $x, y \in A_a$, and $z \in B_a$, $H_x$ is a top block, $H_y$ is a bottom block and $H_z$ is a zig-zag block, then the contributions of $P_x$ and $P_y$ to $\rho$ cancels out the contribution of $P_z$. After doing this cancellation exhaustively, all that remains are top blocks, which also cancel each other perfectly. This completes the proof. □
Proof of Lemma 2. Define an even-length string $t$ over the alphabet $\Sigma$ to be a-unbalanced if $\tau_a(t) > \ell|t|/\ell$ and a-balanced otherwise. If $t$ is a-balanced for each $a \in \Sigma$ then $t$ is balanced. Observe that, if $t$ is balanced then $\tau(t) \leq \ell|t|$. A string is everywhere unbalanced if it contains no balanced substring of length $r \geq r_0$. Our goal therefore is to show that there is an upper bound $n := n(\ell, \epsilon, r_0)$ on the length of any $\ell$-periodic everywhere unbalanced string.

Let $h$ be a positive integer (that determines $n$ and whose value will be discussed later), and let $n := r_02^h$. Let $s$ be an $\ell$-periodic everywhere unbalanced string of length $n$ over the alphabet $\Sigma$. The fact that $s$ is $\ell$-periodic, implies that the $|\Sigma| \leq \ell$. Assume, without loss of generality, that $r_0$ is a multiple of $\ell$.

Consider the complete binary tree $T$ of height $h$ whose leaves, in order, are length-$r_0$ strings whose concatenation is $s$ and for which each internal node is the substring obtained by concatenating the node’s left and right child. Note that for each $v \in V(T)$ and each $a \in \Sigma$, the fact that $s$ is $\ell$-periodic and $r_0$ is multiple of $\ell$ implies that $h_a(v) \geq |v|/\ell$.

For each $a \in \Sigma$, let $S_a := \{v \in V(T) : v$ is $a$-unbalanced$\}$. Since $s$ is everywhere unbalanced, $\bigcup_{a \in \Sigma} S_a = V(T)$. Therefore,

$$(h + 1)n = \sum_{v \in V(T)} |v| \leq \sum_{a \in \Sigma} \sum_{v \in S_a} |v|$$

and therefore, there exists some $a^* \in \Sigma$ such that $\sum_{v \in S_{a^*}} |v| \geq (h + 1)n/|\Sigma| \geq (h + 1)n/\ell$. At this point we are primarily concerned with appearances of $a^*$, so let $X := S_{a^*}$, and, for each node $v \in V(T)$, let $W(v) := h_{a^*}(v)$.

For each non-leaf node $v$ of $T$, let $R(v)$ denote a child of $v$ such that $W(R(v)) \leq \frac{1}{2}W(v)$. (It is helpful to think of $T$ as being ordered so that each right child $y$ with sibling $x$ has $W(y) \leq W(x)$.) For a non-leaf node $v \in X$ the fact that $v$ is $a^*$-unbalanced implies that

$$W(R(v)) \leq \frac{1}{2}W(v) - \frac{\epsilon}{2\ell} \cdot |v| \leq (\frac{1}{2} - \frac{\epsilon}{2\ell})W(v)$$

From this point on we use the following shorthands. For any $S \subseteq V(T)$, $L(S) := \sum_{v \in S} |v|$, $W(S) := \sum_{v \in S} W(v)$, and $R(S) = \{R(v) : v \in S\}$. Summarizing, we have a complete binary tree $T$ of height $h$ and $X \subseteq V(T)$ with the following properties:
1. For each $v \in V(T)$, $W(v) \geq |v|/\ell$.
2. $L(X) \geq (h+1)n/\ell$.
3. For each non-leaf node $v \in X$, $W(R(v)) \leq (1 - \frac{\ell}{2})W(v)$.

For each $i \in \{0, \ldots, h\}$, let $X_i \subseteq X$ denote the set of nodes $v \in X$ for which the path from the root of $T$ to $v$ contains exactly $i$ nodes in $X$, excluding $v$. See Figure 6. Observe that, since each node in $X_i$ has an ancestor in $X_{i-1}$,

$$n \geq L(X_0) \geq L(X_1) \geq \cdots \geq L(X_h).$$

We will show that there exists an integer $t := t(\epsilon, \ell, r_0)$ such that, for each $i \in \{0, \ldots, h-t\}$,

$$L(X_{i+t}) \leq (1 - (1/2)^{t+1})L(X_i). \quad (1)$$

In this way,

$$(h+1)n/\ell \leq L(X) = \sum_{i=0}^{h} L(X_i) \leq \sum_{i=0}^{h} L(X_{t[i/t]}) \quad \text{(since $t[i/t] \leq i$)}
= t \cdot \sum_{i=0}^{h/t} L(X_{it}) \quad \text{(for $h$ a multiple of $t$)}
\leq t \cdot \sum_{i=0}^{\infty} (1 - (1/2)^{t+1})^i L(X_0) \quad \text{(by Equation (1))}
\leq tn \cdot \sum_{i=0}^{\infty} (1 - (1/2)^{t+1})^i \quad \text{(since $|X_0| \leq n$)}
= tn2^{t+1},$$

which is a contradiction for sufficiently large $h$; in particular, for $h > \ell t 2^{t+1} - 1$.

It remains to establish Equation (1), which we do now. Define $A_0 := X_i$ and, for each $j \geq 1$, define $A_j$ to be the subset of $X_{i+j}$ that are descendants of some node in $R(A_{j-1})$. See Figure 6. To upper bound $L(X_{i+j})$ observe that $X_{i+j}$ can be split into two sets $A'_j$ and $B$ defined as follows: The nodes $A'_j$ do not have an ancestor in $R(A_0)$ and therefore $L(A'_0) \leq (1/2)L(A_0)$. The nodes in $B$ do have an ancestor in $R(A_0)$ and therefore have an ancestor in $A_1$. Iterating this argument, we obtain

$$L(X_{i+j}) \leq (1/2)L(A_0) + (1/2)L(A_1) + \cdots + (1/2)^{j}L(A_{j-1}) + L(A_j) \leq (1/2)L(A_0) + (1/4)L(A_0) + \cdots + (1/2)^{j}L(A_0) + L(A_j) = (1 - (1/2)^{j})L(A_0) + L(A_j) = (1 - (1/2)^{j})L(X_i) + L(A_j).$$

So all that remains to establish 1 is to prove that $L(A_j) \leq (1/2)^{j+1}L(X_i)$. 

Figure 6: The partitioning of $X$ into $X_0, \ldots, X_h$. Shaded nodes are in $X$ and all nodes in $X_i$ are shaded with the same colour. Starting with $A_0 = X_0$, the elements of $A_0, \ldots, A_h$ are highlighted (in orange). The elements of $R(A_0), \ldots, R(A_h)$ are also highlighted (in pink).

To to this, observe that, for each $j \in \{1, \ldots, t\}$,

$$W(A_j) \leq W(R(A_{j-1})) \leq \left(\frac{1}{2} - \frac{\epsilon}{2\ell}\right) \cdot W(A_{j-1})$$

which implies

$$W(A_t) \leq \left(\frac{1}{2} - \frac{\epsilon}{2\ell}\right)^t W(A_0) \leq \left(\frac{1}{2} - \frac{\epsilon}{2\ell}\right)^t \cdot L(A_0) = \left(\frac{1}{2} - \frac{\epsilon}{2\ell}\right)^t \cdot L(X_i)$$

Since $s$ is $\ell$-periodic,

$$L(A_t) \leq \ell \cdot W(A_t) \leq \ell \cdot \left(\frac{1}{2} - \frac{\epsilon}{2\ell}\right)^t \cdot L(X_i) \leq L(X_i)/2^{t+1}$$

for $t = \lceil \log(2\ell)/\log((1/(1 - \frac{\epsilon}{\ell})) \rceil$.

5 Reflections

Although an explicit upper bound on $n := n(\epsilon, \ell, r_0)$ could be extracted from the proof of Lemma 2 it would likely be far from tight. We suspect that there is a Fourier analytic proof that would give better quantitative bounds. We have not pursued this, because we have no idea how to explicitly upper bound $\ell$, for reasons discussed in the next paragraph.

Lemma 3 and its proof give absolutely no clues to help find a concrete bound on $\ell$ or to find a minimal set $\Xi$. Indeed, for some choices of $P$, doing so can be a difficult problem. Consider the example where $|\Sigma| = 5$ and $P$ is predicate that tells whether or not its input is anagram-free. It is easy to see that this predicate $P$ satisfies (A1) and the result of Pleasants [10], published in 1970, shows that this $P$ satisfies (A2). The question of whether $|\Xi| = 4$ or
$|\Xi|=5$ is then the question of determining whether there exist arbitrarily long anagram-free strings on an alphabet of size 4. This was the open problem posed by Erdős [3] in 1961 and again by Brown [1] in 1971 and not resolved until 1992 when Keränen [7, 8] showed that the answer, in this case, is that $|\Xi|=4$. However, if this were not the case, then determining $\ell$ would be the question of determining the length of the longest anagram-free string over an alphabet of size 4.

Our proof uses Lemma 3 twice and each application uses a predicate $P$ that is considerably more complicated than asking if the input string is anagram-free. It seems unlikely that we will obtain concrete bounds upper bounds on $\ell$ as a function of $c$ except, possibly, through the use of computer search. The resulting value $\ell$ is used in the application of Lemma 2 and also within the proof of Lemma 7.

References

[1] T. C. Brown. Is there a sequence on four symbols in which no two adjacent segments are permutations of one another? *The American Mathematical Monthly*, 78(8):886–888, 1971. doi:10.1080/00029890.1971.11992892.

[2] Paz Carmi, Vida Dujmović, and Pat Morin. Anagram-free chromatic number is not pathwidth-bounded. In Andreas Brandstädt, Ekkehard Kühler, and Klaus Meer, editors, *Graph-Theoretic Concepts in Computer Science - 44th International Workshop, WG 2018, Cottbus, Germany, June 27-29, 2018, Proceedings*, volume 11159 of *Lecture Notes in Computer Science*, pages 91–99. Springer, 2018. doi:10.1007/978-3-030-00256-5_8.

[3] P. Erdős. Some unsolved problems. *Magyar Tud Akad Mat Kutato Int Kozl*, 6:221–254, 1961.

[4] A. A. Evdokimov. Strongly asymmetric sequences generated by finite number of symbols. *Doklady Akademii Nauk SSSR*, 179:1268–1271, 1968.

[5] A. A. Evdokimov. Strongly asymmetric sequences generated by finite number of symbols. *Soviet Mathematics Doklady*, 9:536–539, 1968.

[6] Nina Kamčev, Tomasz Łuczak, and Benny Sudakov. Anagram-free colourings of graphs. *Comb. Probab. Comput.*, 27(4):623–642, 2018. doi:10.1017/S096354831700027X.

[7] Veikko Keränen. Abelian squares are avoidable on 4 letters. In Werner Kuich, editor, *Automata, Languages and Programming, 19th International Colloquium, ICALP92, Vienna, Austria, July 13-17, 1992, Proceedings*, volume 623 of *Lecture Notes in Computer Science*, pages 41–52. Springer, 1992. doi:10.1007/3-540-55719-9_62.

[8] Veikko Keränen. A powerful abelian square-free substitution over 4 letters. *Theor. Comput. Sci.*, 410(38-40):3893–3900, 2009. doi:10.1016/j.tcs.2009.05.027.

[9] Jaroslav Nešetřil and Patrice Ossona de Mendez. Tree-depth, subgraph coloring and homomorphism bounds. *Eur. J. Comb.*, 27(6):1022–1041, 2006. doi:10.1016/j.ejc.2005.01.010.
[10] P. A. B. Pleasants. Non-repetitive sequences. *Proceedings of the Cambridge Philosophical Society*, 68:267–274, 1970. doi:10.1017/S0305004100046077.

[11] Tim E. Wilson. *Anagram-free Graph Colouring and Colour Schemes*. Ph.D. thesis, Monash University, 2019. doi:10.26180/5c72eca26d5c7.

[12] Tim E. Wilson and David R. Wood. Anagram-free colorings of graph subdivisions. *SIAM J. Discret. Math.*, 32(3):2346–2360, 2018. doi:10.1137/17M1145574.

[13] Tim E. Wilson and David R. Wood. Anagram-free graph colouring. *Electron. J. Comb.*, 25(2):P2.20, 2018. doi:10.37236/6267.