

Anagram-free Graph Colouring

Tim E. Wilson David R. Wood*

School of Mathematical Sciences
Monash University
Melbourne, Australia

{timothy.e.wilson, david.wood}@monash.edu

Submitted: Jul 1, 2016; Accepted: Mar 17, 2018; Published: May 11, 2018

© The authors. Released under the CC BY-ND license (International 4.0).

Abstract

An anagram is a word of the form WP where W is a non-empty word and P is a permutation of W . We study anagram-free graph colouring and give bounds on the chromatic number. Alon *et al.* [Random Structures & Algorithms 2002] asked whether anagram-free chromatic number is bounded by a function of the maximum degree. We answer this question in the negative by constructing graphs with maximum degree 3 and unbounded anagram-free chromatic number. We also prove upper and lower bounds on the anagram-free chromatic number of trees in terms of their radius and pathwidth. Finally, we explore extensions to edge colouring and k -anagram-free colouring.

Mathematics Subject Classifications: 05C15

1 Introduction

A *square* is a word of the form WW where W is a non-empty word. A word X is *square-free* if no subword of X is a square. Much of the study of squares involves avoiding squares or, equivalently, characterising the set of square-free words. An early result in this area is the construction, by Thue [22], of arbitrarily long square-free words on three symbols. More recently, Alon *et al.* [1] generalized the concept of square-free words to graph colouring by requiring colourings to avoid square sequences along subpaths of a graph. A *vertex colouring* of a graph G is a function $f : V(G) \rightarrow C$ where C is a set of colours and $V(G)$ is the vertex set of G . A vertex colouring of a graph G is *square-free* if the sequence of colours on each subpath of G is not a square. The *square-chromatic number* of a graph G is the minimum number of colours in a square-free colouring of G . Square-free graph colouring has been extensively studied and is often referred to as

*Research supported by the Australian Research Council.

nonrepetitive colouring [3, 11, 13]. The square-free chromatic number of G is denoted $\pi(G)$ and is also known as the *Thue number* or *nonrepetitive chromatic number*. With this notation, Thue's result is $\pi(P) \leq 3$ for every path P .

In this paper we introduce anagram-free graph colouring, which is a topic suggested for study by Alon *et al.* [1]. This definition follows naturally from the established notion of anagrams in combinatorics on words [4]. An *anagram* is a word of the form W_1W_2 where W_1 is a non-empty word and W_2 is a permutation of W_1 . A word is *anagram-free* if it contains no anagram as a subword. An anagram is also called an *abelian square* and an anagram-free word is also called an *abelian square-free word* or a *strongly non-repetitive sequence*. Our generalization to graph colouring follows the example set by square-free colouring. A vertex colouring of a graph G is *anagram-free* if the sequence of colours on each subpath of G is not an anagram. The *anagram-free chromatic number* of G is the minimum number of colours in an anagram-free colouring of G , denoted by $\phi(G)$.

Results analogous to Thue's exist for anagram-free words. Anagram-free words on three symbols have length at most 7, for example *abcabc* [5]. Ker an [16, 17] constructed anagram-free words of arbitrary length on four symbols. In the context of anagram-free graph colouring, Ker anen's result is $\phi(P) \leq 4$ for every path P . Throughout this paper we prove bounds on ϕ that demonstrate significant differences in the behaviour of ϕ and π . This is somewhat surprising considering the similarity of Ker anen's and Thue's results.

1.1 Bounding by maximum degree

An area of study central to square-free colouring is bounding $\pi(G)$ by functions of the maximum degree, $\Delta(G)$. Alon *et al.* [1] proved a result that implies $\pi(G) \leq c\Delta(G)^2$ for some constant c . Several subsequent works improved the value of c [9, 12, 14] with the best known value being $c = 1 + o(1)$ [8]. Earlier proofs use the Lov asz Local Lemma while the proof in [8] uses entropy compression.

Alon *et al.* [1] states as an open problem whether ϕ is bounded by a function of maximum degree. We answer this question, thus finding the first significant difference between π and ϕ . We prove that, unlike π , no function of maximum degree is an upper bound on ϕ .

Theorem 1. *Graphs of maximum degree 3 have unbounded anagram-free chromatic number.*

Theorem 1 complements the result of Richmond and Shallit [19] regarding enumeration of anagrams. They counted anagrams with the aim of using the Lov asz Local Lemma to prove that ϕ is bounded on paths. They concluded that the probability that a random word is an anagram is too large for this approach. A proof using the Lov asz Local Lemma that ϕ is bounded on paths would likely apply to graphs of maximum degree 3, which Theorem 1 shows is impossible.

We also consider variations of square-free colouring and apply them to anagram-free colouring. An *edge colouring* of a graph is an assignment of colours to the edges of the graph. An edge colouring of G is *square-free* if every subpath of G has a square-free colour sequence along its edges. Alon *et al.* [1] defined the *square-chromatic index*,

denoted $\pi'(G)$, as the minimum number of colours in a square-free edge colouring of G . We define anagram-free edge colouring similarly. An edge colouring of a graph G is *anagram-free* if every subpath of G has an anagram-free colour sequence along its edges. The corresponding *anagram-free chromatic index* is denoted $\phi'(G)$.

The obvious bound, $\phi'(G) \geq \Delta(G)$, follows from the observation that edges incident to a common vertex receive distinct colours in an anagram-free edge colouring. We prove a significant improvement on this bound with the following result.

Theorem 2. *Trees of maximum degree 3 have unbounded anagram-free chromatic index.*

Theorem 2 is for a more restricted class of graphs than Theorem 1, so Theorem 2 says more about ϕ' than Theorem 1 does about ϕ . The degree bound of 3 in Theorems 1 and 2 is best possible since $\phi(P) \leq 4$ for all paths P , which implies $\phi(C) \leq 5$ for all cycles C . Indeed, to anagram-free 5-colour a cycle assign one vertex a unique colour and then anagram-free 4-colour the remaining vertices. The proof that $\phi'(C) \leq 5$ is analogous. Whether this bound can be improved to 4 is an open problem. The analogous problem for square-free colouring was solved by Currie [6], who showed that $\pi(C) \leq 3$ for every cycle with the exception of order 5, 7, 9, 10, 14 and 17.

1.2 Bounding ϕ with given radius or pathwidth

Our initial bounds for ϕ and ϕ' motivate further study of ϕ for trees. First, note that Brešar [2] proved that $\pi(T) \leq 4$ for every tree T . In contrast, we prove the following:

Theorem 3. *Trees have unbounded anagram-free chromatic number.*

We also study ϕ for trees as a function of radius. The *radius* of a tree T is the minimum, taken over all vertices v in T , of the maximum distance of a vertex from v . Such a vertex v is called a *centre* of T . We obtain the following tight bound on ϕ as a function of radius.

Theorem 4. *Every tree T of radius h has $\phi(T) \leq h + 1$. Moreover, for every $h \geq 0$ there is a tree T of radius h such that $\phi(T) \geq h$.*

The bound in Theorem 4 is poor for paths since the radius of a path is roughly half its length. To address this, we prove a bound on ϕ for trees that is bounded for paths. To do so we use pathwidth, which is a well studied parameter in square-free colouring [10, 8] as well as more generally. Let $\text{pw}(T)$ denote the pathwidth of a tree T .

Theorem 5. *For every tree T , $\phi(T) \leq 4 \text{pw}(T) + 1$. Moreover, for every $p \geq 0$ there is a tree T such that $\phi(T) \geq p \geq \text{pw}(T)$.*

Note that since every tree T on n vertices has pathwidth $O(\log n)$, Theorem 5 implies that $\phi(T) \leq O(\log n)$. It is open whether $\phi(G) \leq 5$ for graphs of pathwidth 1 because not all trees of pathwidth 1 are paths.

1.3 k -anagram-free colouring

We also consider the k -power generalisation of square-free colouring and apply it to anagram-free colouring. A k -power is a word W^k where W is a non-empty word, for $k \geq 2$. In this setting a square is a 2-power. A k -power-free colouring is a colouring avoiding paths with k -power colour sequences. This definition can be applied for vertex or edge colouring. The corresponding k -power-free chromatic number is $\pi_k(G)$ and the k -power-free chromatic index is $\pi'_k(G)$ [11].

In this paper, we introduce k -anagram-free colouring. k -anagrams, often called abelian k -powers, are an established object of study in combinatorics on words [7]. An k -anagram is a word $W_1W_2 \dots W_k$ where each W_i is a permutation of a non-empty word W , for $k \geq 2$. A colouring of a graph G is k -anagram-free if the sequence of colours on each subpath of G is not a k -anagram. We apply this definition to both vertex and edge colourings. The corresponding k -anagram-free chromatic number is denoted by $\phi_k(G)$ and the k -anagram-free chromatic index is denoted by $\phi'_k(G)$.

Every k -anagram contains an $(k-1)$ -anagram so a k -anagram-free colouring it is also $(k+1)$ -anagram-free. Thus, for every graph G ,

$$\phi(G) = \phi_2(G) \geq \phi_3(G) \geq \phi_4(G) \geq \dots \quad (1)$$

with an analogous expression for ϕ'_k . Therefore we can immediately apply upper bounds for ϕ and ϕ' to ϕ_k and ϕ'_k respectively. With this in mind we first study lower bounds of ϕ_k and ϕ'_k and start by studying bounds as a function of maximum degree. We prove the following generalisation of Theorem 1.

Theorem 6. *For $k \geq 2$, k -anagram-free chromatic number is unbounded on graphs of maximum degree $k+1$.*

Theorem 1 is implied by Theorem 6 with $k=2$. Note that the degree bound in Theorem 6 depends on k and it is open whether such a result holds for a degree bound independent of k . We also investigate upper bounds for k -anagram-free colouring and prove the following contrasting result.

Theorem 7. *If $k \geq 4$ then $\phi_k(T) \leq 4$ and $\phi'_k(T) \leq 4$ for every tree T .*

This result is somewhat surprising given Theorems 1 and 3 which say that ϕ_2 and ϕ'_2 are unbounded on trees. Theorem 7 leaves a gap at $k=3$, which motivates the question of whether ϕ_3 and ϕ'_3 are bounded on trees. We have upper bounds for ϕ_3 on trees, in terms of radius and pathwidth, due to Equation (1) and Theorems 4 and 5. We prove a similar upper bound for ϕ'_3 in the following theorem.

Theorem 8. *For every tree T , $\phi'_3(T) \leq 4 \text{pw}(T)$.*

These bounds for ϕ_3 and ϕ'_3 depend on pathwidth or radius. The question of whether ϕ_3 and ϕ'_3 are bounded on trees is an open problem.

We also give tighter bounds than those in Theorem 7 for larger values of k . We improve the bound to 3 for $k \geq 6$ and to 2 for $k \geq 8$ by using results on 3-anagram-free and 4-anagram-free words [7].

2 Lower Bounds

In this section we prove lower bounds for ϕ and ϕ' . Most of these bounds depend upon counting the occurrence of colours in each half of a path, which motivates the following definition. A *colour multiset* of size n on c colours is a multiset of size n with entries from $[c]$ where $[c] := \{1, 2, \dots, c\}$. Let $\mathcal{M}_{n,c}$ be the set of colour multisets of size n on c colours. Note that $|\mathcal{M}_{n,c}|$ equals the number of ways to place n unlabelled balls in c labelled boxes, which is well known to equal $\binom{n+c-1}{c-1}$; see [20, Section 1.9]. We give a weaker bound by noting that the number of occurrences of each colour is in $\{0, 1, \dots, n\}$, therefore

$$|\mathcal{M}_{n,c}| \leq (n+1)^c. \quad (2)$$

This simple bound suffices for our needs because we only require that, once c is fixed, $|\mathcal{M}_{n,c}|$ is bounded by a polynomial in n . For a coloured graph G , let $M(G)$ be the multiset of colours that occur in G and call $M(G)$ the *colour multiset* of G . A c -*colouring* of a graph G is a colouring of G , either for vertices or edges, where the colour set has size c . Note that if G is vertex c -coloured then $M(G) \in \mathcal{M}_{|V(G)|,c}$, and if G is edge c -coloured then $M(G) \in \mathcal{M}_{|E(G)|,c}$. We intentionally allow the definition to apply to edge or vertex colouring and in each case the usage will be clear from the context.

The function M is useful for the analysis of anagram-free colouring. Note that a vertex coloured path v_1, \dots, v_{2i} is an anagram if and only if

$$M(v_1, \dots, v_i) = M(v_{i+1}, \dots, v_{2i}).$$

Similarly, for $i \geq 1$, an edge coloured path v_1, \dots, v_{2i+1} is an anagram if and only if

$$M(v_1v_2, \dots, v_iv_{i+1}) = M(v_{i+1}v_{i+2}, \dots, v_{2i}v_{2i+1}).$$

The indices in these equations highlight a distinction between ϕ and ϕ' . In a vertex colouring only the paths of even order can be anagrams. However, in an edge colouring only the paths of even length can be anagrams.

2.1 Edge colouring

We start with the proof of Theorem 2, which says that ϕ' is unbounded on trees of maximum degree 3. The proof uses the fact that the number of leaves in a complete binary tree grows exponentially with height while, for fixed c , $|\mathcal{M}_{n,c}|$ is bounded by a polynomial in n . We then associate colour multisets to leaves and show that the tree contains an anagram if two leaves share a colour multiset. A *complete binary tree* is a rooted tree such that every non-leaf vertex has two children and the leaves have equal distance to the root.

Theorem 2. *Trees of maximum degree 3 have unbounded anagram-free chromatic index.*

Proof. Fix $c \geq 1$ and choose $h \in \mathbb{Z}^+$ so that $2^h > (h+1)^c$. Let T be the rooted complete binary tree of height h with root vertex r . Fix an arbitrary edge c -colouring of T . By our choice of h and Equation (2)

$$\#\text{leaves} = 2^h > (h+1)^c \geq |\mathcal{M}_{h,c}|.$$

Since each root-to-leaf path in T has h edges, the number of leaves in T is greater than the number of distinct colour multisets on root-to-leaf paths in T . Therefore there are two leaves, p and q , such that $M(P) = M(Q)$ where P is the rp -path and Q is the rq -path. As illustrated in Figure 1.

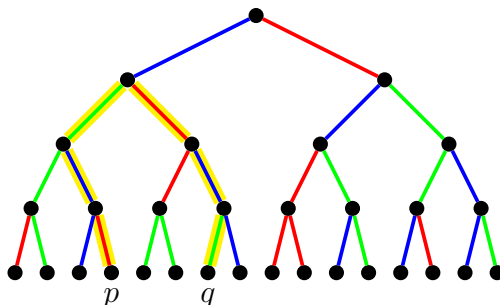


Figure 1: The complete binary tree of height 4 with an edge 3-colouring. The leaves p and q have the same associated colour multiset so the pq -path is an anagram.

Let v be the least common ancestor of p and q . Split these paths into three disjoint parts by defining R as the rv -path, P' as the vp -path and Q' as the vq -path. Note that $E(R)$ is exactly the set of edges shared by P and Q . Therefore

$$M(P) = M(P') \cup M(R) \text{ and } M(Q) = M(Q') \cup M(R).$$

Thus $M(P') = M(Q')$. Finally, note that $P' \cup Q'$ is a path and so T contains an anagram. \square

2.2 Vertex colouring

Using the line graph construction, ϕ is unbounded on graphs of maximum degree 4 as a corollary of Theorem 2. The *line graph* of a graph G , denoted $L(G)$, is the graph with $V(L(G)) = E(G)$ and an edge between vertices of $L(G)$ which are incident to a common vertex in G . It is well known that for ordinary graph colouring $\chi'(G) = \chi(L(G))$. We have a similar relation for anagram-free colouring, however, equality does necessarily not hold. Line graphs satisfy $\phi'(G) \leq \phi(L(G))$ because every paths edge sequence G correspond to a path $L(G)$. Theorem 2 says ϕ' is unbounded for trees of maximum degree 3 and, for these trees, $L(T)$ has maximum degree at most 4. Also, as we have just shown, $\phi'(T) \leq \phi(L(T))$. Therefore ϕ is unbounded on graphs of maximum degree 4. Keränen's result [16] implies $\phi(G) \leq 5$ for graphs of maximum degree 2, thus there is a gap which motivates study of ϕ on graphs of maximum degree 3.

We prove Theorem 1, that ϕ is unbounded on graphs of maximum degree 3, with a method similar to the proof of Theorem 2. We construct a graph with an exponential number of leaves, associate each leaf to a colour multiset and show that the graph contains an anagram if two leaves share a colour multiset.

Theorem 1. *Graphs of maximum degree 3 have unbounded anagram-free chromatic number.*

Proof. Let $c \geq 1$ and let $h \in \mathbb{Z}^+$ be odd such that $2^{(h+1)/2} > (h+2)^c$. Let T be the rooted tree, with root r , such that:

- vertices of depth h are leaves,
- vertices of even depth have two children,
- non-leaf vertices of odd depth have one child,

where the *depth* of a vertex is its distance from the root. Let G be the graph obtained from T by adding an edge between every pair of vertices in T that share a parent, as illustrated in Figure 2.

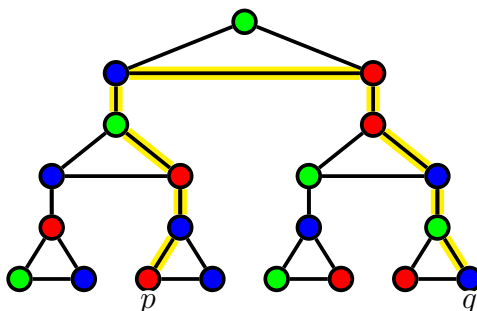


Figure 2: The graph G with $h = 5$, a vertex 3-colouring. Vertices p and q have the same colour multiset so the pq -path is an anagram.

Fix an arbitrary c -colouring of G . We now show that G contains an anagram. By Equation (2) and our choice of h ,

$$\#\text{leaves of } T = 2^{(h+1)/2} > (h+2)^c \geq |\mathcal{M}_{h+1,c}|.$$

So, because each root-to-leaf path in T has $h+1$ vertices, the number of leaves in T is greater than the number of distinct colour multisets on root-to-leaf paths in T . Therefore there are two leaves, p and q in T , such that $M(P) = M(Q)$ where P and Q are the rp -path in T and rq -path in T respectively. Split the two paths into three vertex-disjoint paths $R = P \cap Q$, $P' = P - V(R)$ and $Q' = Q - V(R)$. Note that

$$M(P) = M(P') \cup M(R) \text{ and } M(Q) = M(Q') \cup M(R).$$

Thus $M(P') = M(Q')$. By construction, $G[V(P') \cup V(Q')]$ is a path. Therefore G contains an anagram. \square

Theorem 1 only proves that ϕ is unbounded on graphs of maximum degree 3 without further reference to their structure. We prove below that ϕ is unbounded on trees. A particularly interesting question is whether there is a result analogous to Theorem 2: is ϕ bounded on trees of maximum degree 3? This motivates further investigation of ϕ on trees.

2.3 Vertex colouring trees

We now prove lower bounds for anagram-free vertex colourings of trees. These bounds follow from the following theorem. The *complete d -ary tree of height h* is the rooted tree such that every internal vertex has d children and all leaves are of distance h from the root.

Theorem 9. *The complete d -ary tree of height h does not have an anagram-free c -colouring when $d^c \leq (d/c)^h$.*

Proof. Let T be the complete d -ary tree of height h with root r . Let L be the leaves of T and fix an arbitrary c -colouring of T .

For each $v \in L$ let S_v be the sequence of colours on the rv -path. There are at most c^h such colour sequences since each has length $h + 1$ and they share an initial colour (the colour of the root). Since $|L| = d^h$ there is a set $C \subseteq L$ of size at least d^h/c^h such that $S_v = S_w$ for all $v, w \in C$. Thus C is a large set of leaves that have equal colour sequence on each root-to-leaf path. As illustrated in Figure 3.

Let R be the subtree of T induced by the set of all ancestors of leaves in C . The remainder of the proof is concerned with finding an anagram in R . Observe that if two vertices of R have the same depth, then they also have the same colour. Define a *level* as a maximal set of vertices in R that all have equal depth. R is coloured by level. Denote the levels l_0, l_1, \dots, l_h where $l_0 = \{r\}$.

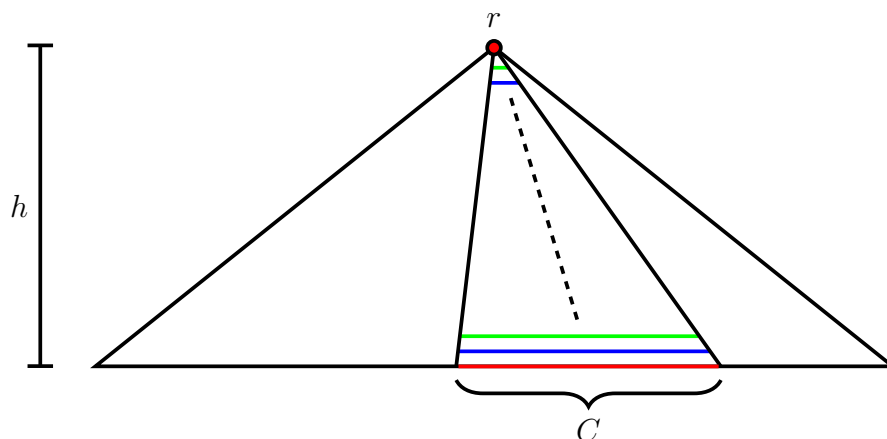


Figure 3: The complete d -ary tree of height h with a large set C of leaves which have the same root-to-leaf path colour sequence.

A level of R is *bad* if every vertex in the level has exactly one child in R . A level is *good* if it is not bad. Note that only level l_h contains vertices with no children. Let g be the number of good levels and b be the number of bad levels. Then $h + 1 = g + b$. We now prove that there are at least $c + 1$ good levels and so at least two good levels share a colour.

We bound the number of bad levels by considering the number of good levels required to obtain at least $(d/c)^h$ leaves. If l_i is bad then $|l_i| = |l_{i+1}|$ and if l_i is good then $|l_i| < |l_{i+1}| \leq d|l_i|$.

$$|l_i| \leq d^{\#\text{preceding good levels}}$$

since l_h is the final good level, it is preceded by $g - 1$ good levels. Thus

$$(d/c)^h \leq |l_h| \leq d^{g-1}$$

By assumption, $d^{c+1} \leq d^{h+1}c^{-h}$ which is at most d^g . Thus $c + 1 \leq g$, there are two good levels with the same colour.

Let l_a and l_b be two good levels that have the same colour and, without loss of generality, let $a < b$. Let $v \in l_a$ be a vertex with at least two children. All vertices in the levels between l_a and l_b have at least one child so there are two vertices $u, w \in B$ such that v is their least common ancestor.

Let p_0, p_1, \dots, p_n be the uv -path and q_0, q_1, \dots, q_n be the wv -path. Since R is coloured by level,

$$M(p_0, p_1, \dots, p_n) = M(q_0, q_1, \dots, q_n).$$

Vertices p_0 and q_n are in levels l_a and l_b respectively so they have the same colour. Thus

$$M(p_1, p_2, \dots, p_n) = M(q_0, q_1, \dots, q_{n-1}).$$

Therefore $p_0, p_1, \dots, p_{n-1}, p_n, q_{n-1}, \dots, q_1$ is an anagram. □

Theorem 9 implies Theorem 3, that ϕ is unbounded on trees, but it has uses beyond this result because it gives us control over the radius and maximum degree of the resulting trees. In particular, it implies the following two theorems each of which correspond to an extreme of radius or maximum degree.

Theorem 10. *For every integer $h \geq 2$ the complete $(h - 1)^h$ -ary tree of height h has $\phi(T) \geq h$.*

Proof. Let $h := c + 1$ and $d := c^{c+1}$. The conditions of Theorem 9 are satisfied because

$$d^c = c^{c(c+1)} = (c^{c(c+1)}c^{c+1})c^{-(c+1)} = d^{c+1}c^{-(c+1)} = (d/c)^h.$$

Therefore the complete c^{c+1} -ary tree of height $c + 1$ does not have an anagram-free c -colouring. □

Theorem 11. *For every integer $d' \geq 1$ there exists a tree T with maximum degree d' such that $\phi(T) \geq d' - 1$.*

Proof. Fix c and choose h so that

$$(c + 1)^c \leq \left(1 + \frac{1}{c}\right)^h = \left(\frac{c + 1}{c}\right)^h.$$

Thus Theorem 9 is satisfied with $d = c + 1$. Therefore the complete $(c + 1)$ -ary tree of height h does not have an anagram-free c -colouring. This tree has maximum degree $d' = c + 2$ and so $\phi(T) \geq c + 1 = d' - 1$. □

3 Upper bounds for ϕ on trees

In this section we complement the results of the previous section with some upper bounds for ϕ on trees. Our first bound comes from centred colouring. A vertex colouring of a graph G is *centred* if every subtree T of G contains a vertex whose colour appears exactly once in T . Centred colourings are anagram-free since every anagram contains an even number of occurrences of each colour and in a centred colouring every path contains a colour that occurs exactly once. Therefore, for every graph G , the centred chromatic number of G is an upper bound on $\phi(G)$. It is easily seen that every tree T of radius h has centred chromatic number at most $h + 1$, see [18, Section 6.5], thus $\phi(T) \leq h + 1$. This bound completes the first half of Theorem 4 and is attained by colouring each vertex by its distance from a centre of T . The second half of Theorem 4 follows from Theorem 10 and an inspection of the trees of radius 0 or 1.

We now work towards proving Theorem 5, which bounds ϕ by a function of pathwidth. While pathwidth is defined in terms of path decompositions, the only property we require is the following lemma.

Lemma 12 ([21] Lemma 5). *Every tree T with at least one edge contains a path P such $\text{pw}(T - V(P)) \leq \text{pw}(T) - 1$.*

We also require two trivial properties of pathwidth. The first is that edgeless graphs have pathwidth 0 and the second is that the pathwidth of a disconnected graph is the maximum pathwidth over each of its components.

Theorem 13. *Every tree of pathwidth m has an anagram-free vertex $(4m + 1)$ -colouring.*

Proof. The proof is an induction on m . The base case holds because every tree T of pathwidth 0 is edgeless and thus anagram-free 1-colourable. Now assume that every tree with pathwidth $p \leq m$ is anagram-free vertex $(4p + 1)$ -colourable.

Let T be a tree of pathwidth $m + 1$. By Lemma 12, there exists a path $P \subseteq T$ such that $\text{pw}(T - V(P)) \leq m$. Each component of $T - V(P)$ has pathwidth at most m . By induction we may anagram-free colour each component of $T - V(P)$ with a common set of $4m + 1$ colours. Then use four additional colours to anagram-free colour P , by [17].

We now show that this colouring is anagram-free. Let Q be a path in T . If Q is entirely contained within a component of $T - V(P)$, then by induction, Q is not an anagram. Otherwise, Q intersects P . The intersection of Q and P is an anagram-free subpath of P and the colours in $Q \cap P$ occur nowhere else in Q . Therefore Q is not an anagram. \square

We now use Theorems 13 and 10 to prove Theorem 5.

Theorem 5. *For every tree T , $\phi(T) \leq 4\text{pw}(T) + 1$. Moreover, for every $p \geq 0$ there is a tree T such that $\phi(T) \geq p \geq \text{pw}(T)$.*

Proof. The first part follows directly from Theorem 13. For the second part it is well known, and easily proved, that the pathwidth of a tree is at most its radius. Therefore, by Theorem 10, there exists a tree T with $\phi(T) \geq p \geq \text{pw}(T)$ for all $p \geq 2$. For the remaining cases use the path of order 2 for $p = 1$ and the empty graph for $p = 0$. \square

The main open problem that arises from the above results is whether ϕ is bounded for trees of maximum degree 3. The complete binary tree of height h is the key example. Centred colourings provides a trivial upper bound of $h + 1$ (Theorem 4). We now give a non-trivial colouring of the complete binary tree to demonstrate that the bound from centred colouring is not best possible.

Theorem 14. *If T is the complete binary tree of height h , then*

$$\phi(T) \leq \frac{h}{2} + \frac{1}{2} \log_2(h + 1) + 1. \quad (3)$$

Proof. We proceed by induction on h . The base case is satisfied as follows:

$$\phi(\text{singleton vertex}) = 1 \leq \frac{0}{2} + \frac{1}{2} \log_2(0 + 1) + 1 = 1.$$

Now assume the result holds up to $h - 1$. Let T be the complete binary tree of height h with root r . Let $t := 2 \lfloor \frac{h+2}{4} \rfloor$ and $T_t \subseteq T$ be the complete binary tree of height t with root r . T_t is the top half of T , which we colour directly. Let $b := h - t - 1$, this is the height of each subtree which will be coloured by induction. Colour T_t as follows:

- All vertices with even depth receive the same colour. Call this colour c .
- Each odd level is allocated distinct set of two colours. Vertices of odd depth are coloured with one of the two colours allocated to their level, so that each vertex receives a colour different from their sibling.

Note that the leaves of T_t have even depth so have colour c . This colouring is shown for $h = 8$ in Figure 4.

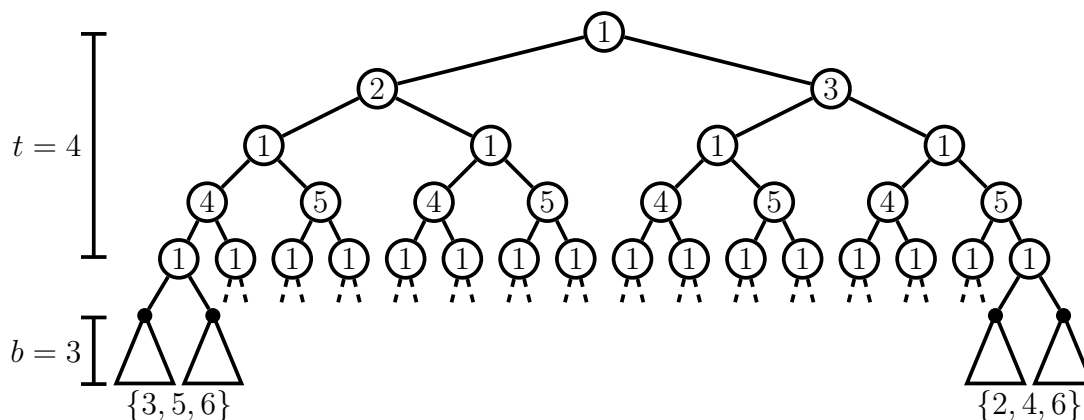


Figure 4: Schematic of colouring the complete binary tree with $h = 8$. The set of colours on shown subtrees of height b are shown below the trees.

Colour each remaining subtree of T by induction, avoiding colours that occur on their ancestors in T_t .

Claim 15. *This colouring of T is anagram-free.*

Proof. Let P be a path in T with even order at least 2. Let u be the shallowest vertex in P . If $u \notin V(T_t)$ then, by induction, P is not an anagram. Now consider the case where $u \in V(T_t)$. If u has odd depth then its colour is unique in P . Indeed, in T_t the colour of u only occurs in the level of u and in $T - V(T_t)$ the ancestors of u avoid its colour. Similarly, if u has even depth and u is an endpoint of P then the child of u in P is uniquely coloured in P .

The remainder of the proof is concerned with the case where $u \in V(T_t)$, u has even depth and neither endpoint of P is u . Let $x_1, x_2 \in V(P)$ be the endpoints of P and let $v_1, v_2 \in V(P)$ be the children of u such that v_i is an ancestor of x_i for $i \in [2]$. If both $x_1, x_2 \notin V(T_t)$ then v_1 (and indeed v_2) is uniquely coloured in P . In the remaining cases at least one of x_1 and x_2 are in $V(T_t)$.

Without loss of generality let $x_1 \in V(T_t)$. If $x_2 \notin V(T_t)$ then x_1 is uniquely coloured in P . If $x_2 \in V(T_t)$ then the colour c occurs an odd number of times in P . Indeed, u has colour c and $x_1, x_2 \in V(T_t)$ implies that both the v_1x_1 -path and the v_2x_2 -path contain the same number of vertices with colour c . \square

To complete the proof we show that this colouring satisfies Equation (3). Our colouring of T_t uses $t + 1$ colours as the even levels share a colour and there are $\frac{t}{2}$ odd levels which each use 2 colours.

Let v be a child of a leaf of T_t and T_b be the subtree of T rooted at v . Recall that colours on the vr -path do not occur in T_b . The number of distinct colours on the vr -path is $\frac{t}{2} + 1$ because the path contains $t + 1$ vertices and $\frac{t}{2} + 1$ of them have even depth so share colour c . Therefore the colouring of T_b can reuse $\frac{t}{2}$ of the colours used to colour T_t . So our colouring of T requires $\phi(T_b) - \frac{t}{2}$ colours in addition to those used to colour T_t . So

$$\phi(T) \leq t + 1 + \left(\phi(T_b) - \frac{t}{2} \right) = \frac{t}{2} + 1 + \phi(T_{h-t-1}).$$

By induction, since T_b is a complete binary tree of height $h - t - 1$,

$$\begin{aligned} \phi(T) &\leq \frac{t}{2} + 1 + \left(\frac{h-t-1}{2} + \frac{1}{2} \log_2(h-t) + 1 \right) \\ &= \frac{h}{2} + \frac{1}{2} + \frac{1}{2} \log_2(h-t) + 1 \\ &= \frac{h}{2} + \frac{1}{2} + \frac{1}{2} \log_2 \left(h - 2 \left\lfloor \frac{h+2}{4} \right\rfloor \right) + 1 \\ &\leq \frac{h}{2} + \frac{1}{2} + \frac{1}{2} \log_2 \left(h - \frac{h-1}{2} \right) + 1 \\ &= \frac{h}{2} + \frac{1}{2} + \frac{1}{2} \log_2 \left(\frac{h+1}{2} \right) + 1 \\ &= \frac{h}{2} + \frac{1}{2} \log_2(h+1) + 1. \end{aligned} \quad \square$$

4 k -anagram-free colourings

Recall that a k -anagram consists of k independently permuted copies of a word. In terms of colour multisets, a word $W_1W_2 \dots W_k$ is a k -anagram if for $i, j \in [k]$,

$$M(W_i) = M(W_j).$$

We defined k -anagram-free colouring as a generalisation of anagram-free colouring. Recall that $\phi_k(G)$ and $\phi'_k(G)$ are the k -anagram-free chromatic number and index respectively. In this section we show that ϕ_k is unbounded on graphs of maximum degree $k + 1$ and that ϕ_4 and ϕ'_4 are bounded on trees. The first result is a generalisation of Theorem 1, that ϕ is unbounded on graphs of maximum degree 3. The second result contrasts with Theorems 2 and 3 because it shows a shift in behaviour from unbounded to bounded as k increases.

4.1 Lower bounds

We now prove Theorem 6, which says that ϕ_k is not bounded by a function of maximum degree. The method is similar to that used in Theorem 1 to prove ϕ is not bounded by maximum degree. For each k and c we recursively construct a graph G such that every c -colouring of G contains a k -anagram. The proof generalises the $k = 2$ case in the sense that Theorem 1 is implied by Theorem 6.

Theorem 6. *For $k \geq 2$, the k -anagram-free chromatic number is unbounded on graphs of maximum degree $k + 1$.*

Proof. Let $S(t)$ be the statement that there exists a graph G with $\Delta(G) \leq k + 1$ and special vertices u and v with $\deg(u) = \deg(v) = 1$ such that for every vertex colouring of G at least one of the following holds:

- G contains a k -anagram, or
- $|D| \geq \left(\frac{k}{k-1}\right)^t$, where D is the set of colour multisets on uv -paths with length $4t$.

Claim 16. *$S(t)$ is true for all $t \geq 1$.*

Proof. We proceed by induction on t . First we prove $S(1)$. Let G be the graph obtained from a path P of order k by adding vertices a and b , each adjacent to every vertex in P . Add vertices u and v to G as well as edges ua and vb . Note that G satisfies the degree requirements of $S(1)$. Fix a colouring of G . If P is monochromatic then it is a k -anagram. If P is not monochromatic, then there are two paths u, a, p_1, b, v and u, a, p_2, b, v with distinct colour multisets. Therefore $|D| \geq 2 \geq \frac{k}{k-1}$ and so $S(1)$ is true. This graph is shown in Figure 5 for $k = 4$.

Assume $S(t)$ is true for some $t \geq 1$. Let G_1, \dots, G_k be copies of the graph guaranteed to exist by $S(t)$. Denote the two special vertices of G_i by u_i and v_i . Let G be the graph with $V(G) = \{u, v, a, b\}$ and $E(G) = \{ua, vb\}$. Add G_1, \dots, G_k to G as disjoint components. Finally, add the following edges to G :

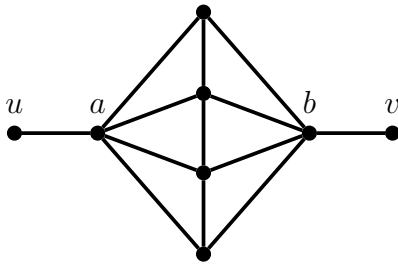


Figure 5: A graph satisfying $S(1)$ for $k = 4$.

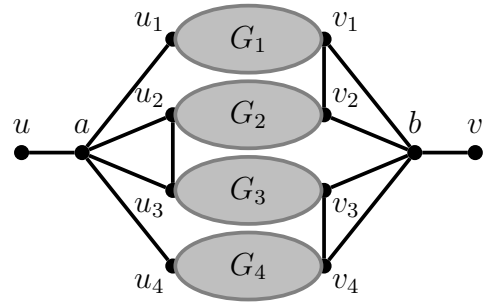


Figure 6: A graph satisfying $S(t)$ for $k = 4$. G_1, G_2, G_3 and G_4 are graphs satisfying $S(t - 1)$. Vertices v_i and u_i are the special vertices of degree 1 in G_i .

1. $u_i a$ for all $i \in [k]$.
2. $v_i b$ for all $i \in [k]$.
3. $u_i u_{i+1}$ for all even $i \in [k - 1]$.
4. $v_i v_{i+1}$ for all odd $i \in [k - 1]$.

This construction is shown in Figure 6 for $k = 4$.

First we show that G satisfies the degree requirements of $S(t + 1)$. Clearly, $\deg(u) = \deg(v) = 1$ and $\deg(a) = \deg(b) = k + 1$. Each u_i has degree 1 in G_i so $\deg(u_i) \leq 3 \leq k + 1$. Similarly, $\deg(v_i) \leq 3 \leq k + 1$. Every other vertex has the same degree as in G_i , which is at most $k + 1$.

Now fix a colouring of G . If some G_i contains a k -anagram then $S(t + 1)$ is satisfied so assume that each G_i is k -anagram-free. Let D_i be the set of colour multisets on paths of length $4t$ in G_i with endpoints u_i and v_i . By $S(t)$, we have $|D_i| \geq (k/(k - 1))^t$ for all $i \in [k]$. We now split the proof into two cases.

In the first case there exists a colour multiset M such that $M \in D_i$ for all $i \in [k]$. This means that each G_i contains a $u_i v_i$ -path P_i with $M(P_i) = M$. Type (iii) and (iv) edges between special vertices of G_i and G_{i+1} mean that the subgraph induced by $P_1 \cup P_2 \cup \dots \cup P_k$ is a path. This path is a k -anagram and so the colouring satisfies $S(t + 1)$.

In the second case there is no colour multiset that occurs in every D_i . Define the union of colour multisets as $\mathcal{U} := \bigcup_{i \in [k]} D_i$. For a colour multiset $M \in \mathcal{U}$ let $f(M)$ be the number of sets from $\{D_1, \dots, D_k\}$ that contain M . No colour multiset occurs in every D_i so $f(M) \leq k - 1$ and therefore

$$|\mathcal{U}|(k - 1) \geq \sum_{M \in \mathcal{U}} f(M) = \sum_{i \in [k]} |D_i| \geq k \left(\frac{k}{k - 1} \right)^t$$

Thus $|\mathcal{U}| \geq \left(\frac{k}{k - 1} \right)^{t+1}$. There is a bijection from \mathcal{U} to D because every uv -path of length $4t + 4$ shares vertices a, b, u and v . Therefore $|D| = |\mathcal{U}|$ and so $S(t + 1)$ is satisfied. \square

Let $c \geq 1$ be a number of colours and let t be sufficiently large so that

$$\left(\frac{k}{k-1}\right)^t > (4t+2)^c.$$

Let G be the graph guaranteed to exist by $S(t)$ and fix an arbitrary c -colouring of G . Let D be the set of colour multisets as defined previously. By (2) there are at most $(4t+2)^c$ colour multisets of size $4t+1$. Therefore $|D| \leq (4t+2)^c < \left(\frac{k}{k-1}\right)^t$. Thus G contains a k -anagram by $S(t)$. \square

The following natural question arises: Does there exist a d such that ϕ_k is unbounded on graphs of maximum degree d for all $k \geq 3$? Also, the analogous problem for edge colouring is open. We know of no family of graphs for which ϕ'_k is unbounded, except in the case of $k = 2$.

4.2 Upper bounds on trees

In this section we prove an upper bound on k -anagram-free colouring that contrasts with the results for anagram-free colouring.

Theorem 7. *If T is a tree and $k \geq 4$ then $\phi_k(T) \leq 4$ and $\phi'_k(T) \leq 4$.*

Proof. Root T at an arbitrary vertex $r \in V(T)$ and let h be the height of the resulting rooted tree. Let $C = c_0 \dots c_h$ be an anagram-free word on four symbols. Colour each vertex $u \in V(T)$ by c_x where x is the distance between u and r .

Let $P = P_1 \dots P_k$ be a path in T such that, for some $n \geq 1$, $|V(P_i)| = n$ for all $i \in [k]$. Note that P is a k -anagram if and only if $M(P_1) = M(P_2) = \dots = M(P_k)$. We now show that P is not a k -anagram.

P contains a unique vertex v closest to r . If $v \in V(P_i)$ with $i \geq 3$ then the colour sequence along $P_1 \cup P_2$ appears in C , so $M(P_1) \neq M(P_2)$. In the other case, $v \in V(P_i)$ with $i \leq 2$. Then the colour sequence along $P_3 \cup P_4$ appears in C , so $M(P_3) \neq M(P_4)$. In each case, P is not a k -anagram. Hence $\phi_k(T) \leq 4$.

The proof can be repeated for ϕ'_k by modifying the definition of P . In this case P has length nk and its subpaths P_1, P_2, \dots, P_k are edge-disjoint such that P is a k -anagram if and only if $M(P_1) = M(P_2) = \dots = M(P_k)$. \square

Theorem 7 demonstrates a qualitative change in behaviour as k increases. The case of $k = 3$ is an open problem that sits between bounded and unbounded behaviour. For ϕ_3 on trees we have upper bounds due to Equation (1) and Theorems 4 and 5. We prove similar upper bounds for ϕ'_3 using Dekking's [7] result $\phi_3(P) = 3$.

Theorem 8. *For every tree T , $\phi'_3(T) \leq 4 \text{pw}(T)$.*

Proof. The proof is by induction on m . The base case is satisfied because trees of pathwidth 0 are edgeless. Now assume that every tree T with pathwidth at most m has $\phi'_3(T) \leq 4m$.

Let T be a tree of pathwidth $m + 1$. By Lemma 12 there exists a path $P \subseteq T$ such that $\text{pw}(T - V(P)) \leq m$. Each component of $T - V(P)$ has pathwidth at most m , so can be 3-anagram-free edge-coloured with the same set of $4m$ colours, by induction. We now use four additional colours to colour the remaining edges. Dekking [7] proves $\phi_3(P) = 3$ so we can 3-anagram-free edge-colour P with three colours. The fourth extra colour is used to colour the edges between P and $T - V(P)$.

We now show that this colouring is 3-anagram-free. Let Q be a path in T . If Q is entirely contained within a component of $T - V(P)$ then, by induction, Q is not a 3-anagram. Otherwise Q intersects P . The intersection of Q and P is a 3-anagram-free subpath of P and the colours in $Q \cap P$ occur nowhere else in Q . Therefore Q is not a 3-anagram. \square

Note that no similar bound exists for ϕ'_2 because stars have pathwidth 1 and ϕ'_2 is unbounded on stars. A bound on $\phi'_3(T)$ as a function of radius follows from the relation between pathwidth and radius in trees. Note that since $\phi_3(P) \leq 3$ we are also able to prove that $\phi_3(T) \leq 3 \text{pw}(T) + 1$ with a proof similar to the proof of Theorem 13.

Dekking also proves $\phi_4(P) = 2$ and we use both results to improve upon Theorem 7 for larger k .

Theorem 17. *For all $z \geq 1$ and $k \geq 2z$, if $\phi_z(P) \leq y$ for all paths P then $\phi_k(T) \leq y$ for all trees T .*

Proof. Let T be a tree with root r and height h . Let $C = c_0 \dots c_h$ be a z -anagram-free word on y symbols. Colour each vertex $u \in V(T)$ by c_x where x is the distance between u and r .

Let $P = P_1 \dots P_k$ be a path in T such that $|V(P_i)| = n$ for some $n \geq 1$. Note that P is a k -anagram if and only if $M(P_1) = M(P_2) = \dots = M(P_k)$. We now show that P is not a k -anagram.

P contains a unique vertex v closest to r . If $v \in V(P_i)$ with $i > z$ then the colour sequence along $P_1 \cup P_2 \cup \dots \cup P_z$ appears in C . In the other case, $v \in V(P_i)$ with $i \leq z$, which implies the colour sequence along $P_{z+1} \cup P_{z+2} \cup \dots \cup P_{2z}$ appears in C . In each case there exist a, b such that $M(P_a) \neq M(P_b)$ because C is z -anagram-free. Therefore P is not a k -anagram. \square

Dekking [7] proves $\phi_3(P) = 3$ and $\phi_4(P) = 2$, so Theorem 17 implies $\phi_6(T) \leq 3$ and $\phi_8(T) \leq 2$. The corresponding results for edge colouring, $\phi'_6(T) \leq 3$ and $\phi'_8(T) \leq 2$, are achieved with modifications similar to those at the end of Theorem 7.

5 Open Problems

Throughout this paper we have posed many conjectures and open problems. In this section we provide a summary as well as some further questions.

The results of Sections 2 and 3 motivate further study of ϕ on trees. Whether ϕ is bounded on the complete binary tree is a particularly interesting question. We also

conjecture the result analogous to [6], that $\phi(C) \leq 4$ for cycles with only finitely many exceptions. The tight bounds given in Theorems 4 and 5 motivate further investigation of pathwidth and radius. We ask whether pathwidth is tied to ϕ on trees, that is, whether there exists a function f such that $\text{pw}(T) \leq f(\phi(T))$ for every tree T . Pathwidth is unbounded on complete binary trees so the two questions are related.

Section 4 contains two open problems. The first is whether ϕ'_k is bounded by maximum degree for some $k \geq 3$. The second is whether ϕ_3 and ϕ'_3 are bounded on trees.

Acknowledgements

Thanks to Gwenaël Joret for stimulating discussions.

Note

At the same time as the present paper was completed, Kamčev, Łuczak, and Sudakov [15] posted a paper on the arXiv that independently introduces *anagram-free* graph colouring. Each paper independently proves some of the results in the other paper. Note that Kamčev, Łuczak, and Sudakov proved that ϕ is unbounded on complete binary trees, thus solving the above-mentioned open problem.

References

- [1] N. Alon, J. Grytczuk, M. Hałuszczak, and O. Riordan. Nonrepetitive colorings of graphs. *Random Structures & Algorithms*, 21(3-4):336–346, 2002.
- [2] B. Brešar, J. Grytczuk, S. Klavžar, S. Niwczyk, and I. Peterin. Nonrepetitive colorings of trees. *Discrete Mathematics*, 307(2):163–172, 2007.
- [3] R. J. Britten and E. H. Davidson. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Quarterly Review of Biology*, 46(2):111–138, 1971.
- [4] R. Cori and M. R. Formisano. Partially abelian squarefree words. *Informatique théorique et applications*, 24(6):509–520, 1990.
- [5] L. J. Cummings. Strongly square-free strings on three letters. *Australasian J. Combinatorics*, 14:259–266, 1996.
- [6] J. D. Currie. There are ternary circular square-free words of length n for $n \geq 18$. *Electron. J. Combin*, 9(1):#N10, 2002.
- [7] F. M. Dekking. Strongly non-repetitive sequences and progression-free sets. *J. Combinatorial Theory, Series A*, 27(2):181–185, 1979.
- [8] V. Dujmović, G. Joret, J. Kozik, and D. R. Wood. Nonrepetitive colouring via entropy compression. *Combinatorica*, 36:661–686, 2016.
- [9] G. Fertin, A. Raspaud, and B. Reed. Star coloring of graphs. *J. Graph Theory*, 47(3):163–182, 2004.

- [10] A. Gagol, G. Joret, J. Kozik, and P. Micek. Pathwidth and nonrepetitive list coloring. *Electron. J. Combin.*, 23(4):#P4.40, 2016
- [11] J. Grytczuk. Nonrepetitive colorings of graphs a survey. *Int. J. Math. Math. Sci.*, 2007:Art. ID 74639.
- [12] J. Grytczuk. Nonrepetitive graph coloring. In *Graph Theory in Paris*, Trends in Mathematics, 209–218. Birkhauser, 2007.
- [13] J. Grytczuk, J. Kozik, and P. Micek. New approach to nonrepetitive sequences. *Random Structures & Algorithms*, 42(2):214–225, 2013.
- [14] J. Harant and S. Jendrol'. Nonrepetitive vertex colorings of graphs. *Discrete Math.*, 312(2):374–380, 2012.
- [15] N. Kamčev, T. Łuczak, and B. Sudakov. Anagram-free colorings of graphs. *Combinatorics, Probability and Computing*, 2017. doi:10.1017/S096354831700027X arXiv:1606.09062.
- [16] V. Keränen. Abelian squares are avoidable on 4 letters. In *Automata, languages and programming*, volume 623 of *Lecture Notes in Comput. Sci.*, pages 41–52. Springer, 1992.
- [17] V. Keränen. A powerful abelian square-free substitution over 4 letters. *Theoret. Comput. Sci.*, 410(38-40):3893–3900, 2009.
- [18] J. Nešetřil and P. O. De Mendez. *Sparsity: Graphs, Structures, and Algorithms*. Springer, 2012.
- [19] L. B. Richmond and Jeffrey Shallit. Counting abelian squares. *Electron. J. Combin.*, 16(1):#72, 2009.
- [20] R. P. Stanley. *Enumerative Combinatorics*, volume 1. Cambridge University Press, Cambridge, second edition, 2011.
- [21] M. Suderman. Pathwidth and layered drawings of trees. *Internat. J. Comput. Geom. Appl.*, 14(3):203–225, 2004.
- [22] A. Thue. Probleme über veränderungen von zeichenreihen nach gegebenen regeln. pages I. Math. naturv. Klasse, 10. Christiana Videnskabs-Selskabs Skrifte, 1914.