

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Preliminaries</b>	<b>4</b>
2.1	Graph Theory . . . . .	4
2.2	Complexity Theory . . . . .	7
<b>3</b>	<b>Nonrepetitive graph colorings</b>	<b>9</b>
3.1	Definition of Nonrepetitiveness . . . . .	9
3.2	The complexity of nonrepetitive vertex-colorings . . . . .	10
3.3	The complexity of nonrepetitive edge-colorings . . . . .	14
<b>4</b>	<b>The complexity of the Thue Number problem</b>	<b>17</b>
4.1	The Thue number and Thue chromatic number . . . . .	17
4.2	Constructing the graph . . . . .	18
4.2.1	The structure of $H$ . . . . .	18
4.2.2	Subgraphs of $H$ . . . . .	20
4.3	The coloring of $H$ . . . . .	21
4.4	The final part of the reduction . . . . .	26
<b>5</b>	<b>The Thue chromatic number</b>	<b>31</b>
5.1	Algorithms for vertex-colorings . . . . .	31
5.2	Bounds on Planar graphs . . . . .	33
5.3	Colorings of Trees . . . . .	33
<b>6</b>	<b>Other Thue type problems</b>	<b>36</b>
<b>7</b>	<b>Conclusion</b>	<b>37</b>

# 1 Introduction

In 1906, Axel Thue showed the construction of an infinite word where no two adjacent parts are the same using only three letters [1]. This concept of nonrepetitiveness has since been applied to other fields, probably most interestingly in graph theory. The nonrepetitiveness of graph colorings and the number of colors needed to construct such nonrepetitive colorings have been a topic since Alon et. al. introduced the idea in [2]. They focused on coloring the vertices of graphs in ways to that no partial coloring of a path repeats itself immediately, similar to Thue's word. It's also possible to delve into the nonrepetitiveness of colorings of the edges instead of the vertices of graphs. This thesis aims to give an overview of some of the results, both in vertex- and edge-colorings with the main task being the analysis of Fedor Manin's proof of complexity for the Thue Number problem. The Thue Number  $\pi(G)$  of a graph  $G$  is the minimum of colors that is needed for a nonrepetitive edge-coloring of  $G$ . Manin showed this problem to be  $\Sigma_2^P$ -complete, but even though his proof has been referenced in [3] and [4], it hasn't been published. Most of it's correct, although there are some minor mistakes and many places in the proof where the motivation or reason for a particular construction is missing, even going as far as leaving out parts of the construction process. This thesis attempted to fix both the incorrect parts of the proof and add to the construction in such a way that it might become more accessible and easier to understand.

First, though, some general preliminaries in graph and complexity theory will be covered to give the reader a background in these fields.

Then, in the third section, the formal definition of nonrepetitiveness for graph colorings and the complexity of the corresponding decision problems will be given. These problems serve both as an introduction to the Thue Number proof, in fact one of them is its predecessor, and to nonrepetitive colorings in general. Furthermore, comparing the two might make clear why the Thue Number problem for edge-colorings has been solved, but the corresponding vertex-coloring problem hasn't yet.

The fourth section is about edge-colorings, more specifically Manin's proof that the Thue Number problem is  $\Sigma_2^P$ -complete [5]. This proof will be examined in detail as it's probably the most important result for nonrepetitive edge-colorings as well as generally applicable, which isn't the case for most results specific to vertex-colorings. In this proof, as well as Manin's proof in section three, some inconsistencies and missing links have been resolved and the proof has also been fleshed out a good bit more than in the original paper.

Section five, then, is intended to give a bit of an overview of some relevant results when it comes to nonrepetitive vertex-colorings, though it certainly doesn't aim to be complete. The first part introduces two randomized algorithms, one for finding repetitive paths of a specific length and the other constructing a nonrepetitive coloring. After that, bounds for two classes of graphs that have received particular attention will be presented: planar graphs and trees. The results here, with maybe the exception of trees, won't be presented in depth as a lot of them have been found using probabilistic methods which go beyond the scope of this thesis. They're mentioned to show how diverse methods and results for different colorings and graphs are, not as an exhaustive explanation of nonrepetitive vertex-colorings. Finally, section six will mention other applications and variants of nonrepetitiveness and Thue Number type problems in other mathematical fields.

## 2 Preliminaries

Before going into the specifics of nonrepetitive colorings, this section provides the background knowledge necessary for the two main fields of this thesis: graph theory and complexity. Basic as well as more specific terminology that will become important later will be defined here. Broader and more detailed explanation can be found in [6], [7] and [8] that were used for this section. Note that this section is focused on graph and complexity theory, but in places some general mathematical background might still be necessary to understand parts of this thesis.

### 2.1 Graph Theory

**Definition 1** *A graph  $G$  is defined as two finite sets  $V$  and  $E$ , representing the vertices and edges of  $G$  respectively, so that  $E$  is a subset of  $V \times V$ , meaning that each edge from  $E$  connects two vertices from  $V$  [6, p. 1].*

This is denoted as  $G = (V, E)$ . Hereafter, any graph  $G$  will refer to an undirected and unweighted graph unless mentioned otherwise. Thus, edges can be traversed in both directions. Furthermore, all graphs will be simple graphs, not containing multi-edges (more than one edge from a vertex  $v_i$  to another vertex  $v_j$ ) or loops (edges starting from and ending at the same vertex) [7, pp. 3-4].

**Definition 2** *A graph  $G' = (V', E')$  is called a subgraph of  $G$  if  $V'$  is a subset of  $V$  and  $E'$  is a subset of  $E$  [6, p. 2].*

**Definition 3** *The maximal degree  $\Delta$  of a graph is the maximal number of edges connected to a single vertex [6, p. 4].*

There is also a minimum degree, but it won't be relevant for the topics covered here, so degree will always refer to the maximum degree.

**Definition 4** *Two vertices  $v_i$  and  $v_j$  are adjacent, if there is an edge  $e_k$  that connects the two. Two edges are adjacent if they share one vertex [6, p. 1-2].*

**Definition 5** *A path on a graph is a sequence of the form*

$$W = v_0 e_0 v_1 e_1 \dots e_{n-1} v_n$$

*where, starting from an initial vertex  $v_0$ , vertices and edges alternate until the final vertex  $v_n$  is reached.*

In a simple graph, a path can also be represented by either just its vertices [7, p. 10] or edges [6, p. 4]. The length  $l$  of a path is the number of edges it contains [6, p. 5].

**Definition 6** *An open path is a path that visits no vertex more than once.*

Thus cycles on the path are not possible (see 'trail' in [7, p. 10]). Open paths are sometimes called walks.

**Definition 7** *An open path through every vertex is called a Hamiltonian path [6, p. 14].*

This shouldn't be confused with the Hamiltonian circle, a Hamiltonian path visits every vertex exactly once whereas a Hamiltonian circle then goes on to visit the first vertex in the path again.

**Definition 8** *An edge-coloring of a graph  $G$  is a function that matches each edge to a color contained in  $C$ , the set of colors. A vertex-coloring is a function matching the vertices instead of the edges to the colors in  $C$  [7, pp. 7-8].*

The color of an edge  $e$  in a coloring  $C$  will be denoted as  $C(e)$ . An edge of color  $c$  whose other attributes aren't relevant will be referred to as  $[c]$ .

There are some special graph structures that will become important later on: the *Clique* and the *Hypercube*.

**Definition 9** *A Clique is a graph, in which every two vertices  $v_i, v_j$  are adjacent (see figure 1). A Clique containing  $k$  vertices is denoted as  $k$ -Clique [7, pp. 21, 111].*

As all vertices are adjacent in an  $n$ -Clique, there are  $n - 1$  edges per vertex. The degree of all vertices is therefore the same in this graph. An example of a Clique can be seen in fig. 1.

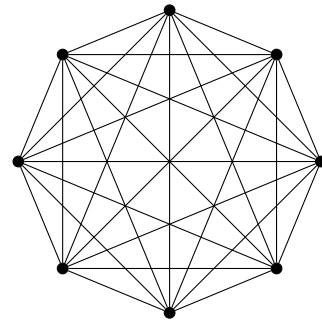


Fig. 1: 8-Clique

**Definition 10** A *Hypercube* is an  $n$ -dimensional Graph (denoted  $n$ -Hypercube) that can be constructed by labeling a set of vertices of size  $2^n$  with a different element of  $\{0,1\}^n$  each and connecting the vertices that differ in one position (see fig. 2) [6, p. 60].

If  $n > 1$ , the construction of the  $(n + 1)$ -Hypercube consist essentially of two  $n$ -Hypercubes being connected (see fig. 3). Both for the Hypercube and Clique, the number of edges connected to a vertex is the same for all vertices, e.g. 7 for a 8-Clique or 4 for a 4-Hypercube.

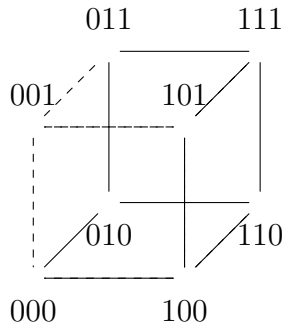


Fig. 2: 3-Hypercube

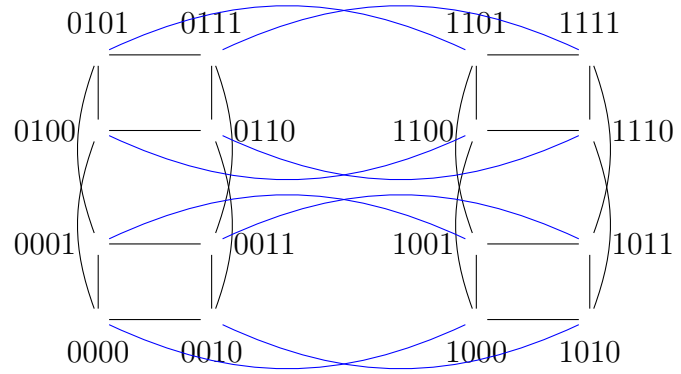


Fig. 3: 4-Hypercube constructed out of two 3-Hypercubes

## 2.2 Complexity Theory

To discuss the complexity of the following problems, some background in complexity theory, specifically the *Polynomial Hierarchy*, will be needed. First, however,  $P$  and  $NP$ , two specific complexity classes within the hierarchy will be introduced.

$P$  and  $NP$  are similarly defined using Turing machines.

**Definition 11**  $P$  contains all problems that can be solved in polynomial time, that means all problems a Deterministic Turing Machine can decide in  $p(x)$  computational steps,  $p(n)$  being any polynomial and  $x$  the length of the input.

**Definition 12**  $NP$  contains all problems that can be solved in polynomial time by a Non-deterministic Turing Machine [8].

An alternative way to define  $NP$  is that  $NP$  contains all problems for which a solution can be verified in polynomial time [9, p. 59]. Its counterpart, the complexity class of  $coNP$ , contains all problems where an instance that is not a solution can be verified as such in polynomial time [9, p. 154-155].

The levels of the Polynomial Hierarchy contain three complexity classes each:  $\Sigma_i^p$ ,  $\Delta_i^p$  and  $\Pi_i^p$  [10]. At level 0, that is  $i = 0$ , all three of them are equal and  $P$ :

**Definition 13**  $\Sigma_0^p = \Delta_0^p = \Pi_0^p = P$  [11]

The further levels are recursively defined using *oracles*. If a Turing Machine can use an oracle, it means that in a single computational step, an oracle for a language  $L$  will decide if a string on the Turing Machine's tape is in  $L$ . With that in mind, the definitions go as follows:

**Definition 14**  $\Sigma_{i+1}^p = \{ L \mid L \text{ is accepted by a Non-deterministic Turing Machine with an } L' \text{ oracle for some } L' \in \Sigma_i^p \}$

**Definition 15**  $\Pi_{i+1}^p = \{ L \mid \text{the complement of } L \text{ is accepted by a Non-deterministic Turing Machine with an } L' \text{ oracle for some } L' \in \Sigma_i^p \}$

**Definition 16**  $\Delta_{i+1}^p = \{ L \mid L \text{ is accepted by a Deterministic Turing Machine with an } L' \text{ oracle for some } L' \in \Sigma_i^p \}$

$P$  and  $NP$  are of course included here,  $\Delta_1^p = P$  and  $\Sigma_1^p = NP$  [10].  $\Pi_1^p$  is called  $coNP$ .

$\Sigma_i^p$ ,  $\Pi_i^p$  and  $\Delta_i^p$  are related in the following way:

$$\Sigma_i^p \cup \Pi_i^p \subseteq \Delta_{i+1}^p \subseteq \Sigma_{i+1}^p \cap \Pi_{i+1}^p$$

It's not yet decided if these inclusions are proper, but they give an idea of what the Polynomial Hierarchy looks like [11].

One possible way to decide which complexity class a problem belongs to, is the reduction from a problem with known complexity.

**Definition 17** *Let  $L$  and  $Q$  be two languages, then  $Q$  is reducible to  $L$  if there is a function  $f(n)$  that runs in polynomial time so that for any  $x$  in  $Q$ ,  $f(x)$  is in  $L$ .*

This is denoted as  $Q \leq_p L$ . If such a function exists,  $L$  is at least as complex as  $Q$  [8].

Three problems that will be reduced from later on in this thesis are *SAT*, *3SAT* and  $\forall\exists 3SAT$ . The later two are special cases of *SAT*, the Boolean Satisfiability Problem.

**Definition 18**  *$SAT$  is the language containing all satisfiable boolean formulas in conjunctive normal form.*

*SAT* is in *NP* and in fact *NP – hard*, meaning that every problem in *NP* can be reduced to *SAT*. Both of those attributes combined make it *NP – complete* [8, p. 338-346].

**Definition 19**  *$3SAT$  is a subset of  $SAT$  with all formulas with 3 or less literals per clause.*

$3SAT$  is also *NP – complete* [8, p. 347].

**Definition 20**  *$\forall\exists 3SAT$  finally is the satisfiability problem for formulas  $\varphi$  with 3 or less literals per clause that are of the form*  
 $\forall(x_1, \dots, x_i)\exists(y_1, \dots, y_j)\varphi(x_1, \dots, x_i, y_1, \dots, y_j)$ .

$\forall\exists 3SAT$  is  $\Pi_2^P$ -complete [4].



### 3 Nonrepetitive graph colorings

This section will introduce the notion of nonrepetitiveness before delving into problems stemming from it. It includes two smaller proofs, both showing the complexity of deciding whether a coloring is nonrepetitive, one for vertex- and one for edge-colorings. The vertex-coloring proof is taken from [3] and not much, beside an additional example and showing why it's in *coNP*, has been added. There are hardly any proofs of complexity when it comes to vertex-colorings, so this one, as the most basic problem for Thue Number type problems, gives a good point of orientation of how more complicated proofs might look like. Furthermore, the second proof [5] was directly extended to proof the complexity of the Thue Number problem, so comparing the two also gives an idea of why the vertex-coloring version of Thue Number can't be proven as easily with the same idea. The second proof itself, of course, leads up to the bigger Thue Number proof, but has a few inconsistencies that have been remedied here. Its concept and graph construction will be the basis for the next section.

#### 3.1 Definition of Nonrepetitiveness

Before talking about Thue number type problems, it makes sense to look at non-repetitiveness first, as it is the defining concept behind the Thue number. Nonrepetitive here means square-free. This can be applied to different fields like graph theory, geometry or number theory [12], but it was first introduced in the process of constructing a square-free word by Axel Thue in 1906.

**Definition 21** *A word  $w$  is square-free if there is no part  $x$  of the word that occurs two times in a row.*

Let the words  $a$  and  $b$  each consist of three different substrings,  $v$ ,  $x$  and  $y$ , then  $a = vxx$  is not square-free, but  $b = vxy$  is [1].

In graph theory, nonrepetitiveness is relevant with respect to nonrepetitive vertex- [3] and edge-colorings of graphs.

**Definition 22** *A coloring  $c$  of a graph  $G$  is nonrepetitive if there is no open path whose vertex or edge colors contain a color sequence  $x$  that is repeated two or more times in a row.*

To put it differently: a coloring of a graph is nonrepetitive if the coloring of every possible open path on the graph is nonrepetitive [2]. To give a baseline for the complexity of further problems concerning nonrepetitive graph colorings, first the complexity of determining whether a vertex- or edge coloring is nonrepetitive will be shown.

### 3.2 The complexity of nonrepetitive vertex-colorings

**Theorem 1** *Deciding whether a vertex-coloring of a graph is nonrepetitive is  $coNP$ -complete [3].*

The problem of deciding whether a vertex-coloring of a graph  $G(V, E)$  is nonrepetitive has been proven to be  $coNP$  – *hard* by reducing from the Hamiltonian path problem, which is  $NP$  – *complete*. The idea behind Marx’s and Schäfer’s proof is to construct a graph  $H$  from  $G$  with two parts, the first part representing the vertices of  $G$  with a subgraph each and the second one having a path for every possible path in  $G$  that is as long as a Hamiltonian path would be ( $n$  vertices with  $n = |E|$ ).  $H$  is colored in such a way that a repetitively colored path would have to go through both parts with the coloring from the first path repeating itself in the second path. If a path through the second half of  $H$ , however, corresponds to a path in  $G$  that is not a Hamiltonian path but a path that visits at least one vertex more than once, the subgraph for that vertex in the first half of  $H$  would have to also be visited more than once - meaning that the path through the first half is not an open path. Thus  $H$  is colored nonrepetitively only if  $G$  doesn’t have a Hamiltonian path. This implies that the problem of deciding whether a vertex-coloring is nonrepetitive is  $coNP$  – *hard*.

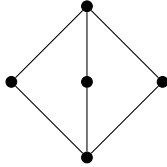


Fig. 4: A  $K_{2,3}$  Graph

The construction of  $H$  goes as follows: for the first part, for every vertex  $i \leq n$  ( $n = |E|$ ) of  $G$ , a  $K_{2,n}$  graph (fig. 4) is added. It consists of two vertices, colored  $a$  and  $b$  respectively, with  $n$  vertices between them, each connected to both the  $a$ - and  $b$ -vertex. As one vertex is added for each vertex in  $G$ , the vertices are colored  $c_{i,j}$ ,  $i$  being the vertex in  $G$  for which the subgraph was added and  $j$  ranging from 1 to  $n$ . The  $[b]$  vertices are now connected to  $[a]$  vertex,  $z$ , colored  $c$ , that marks the transition between the two parts of  $H$ . Lastly, to connect the different subgraphs, for every  $[b]$  vertex,  $d_{i,j}$ -colored vertices are added. A  $d_{i,j}$  vertex is connected to the  $[b]$  vertex of the subgraph with the same value for  $i$  and to all  $[a]$  vertices for  $j$  between 1 and  $n$  where  $j \neq i$ . So now all  $[a]$  and  $[b]$  vertices are connected to all other  $K_{2,n}$  subgraphs of the first part of  $H$ . An example of this construction for a simple graph (fig. 5) is shown in fig. 6.

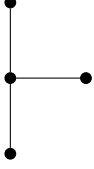


Fig. 5:  
Graph  $F$

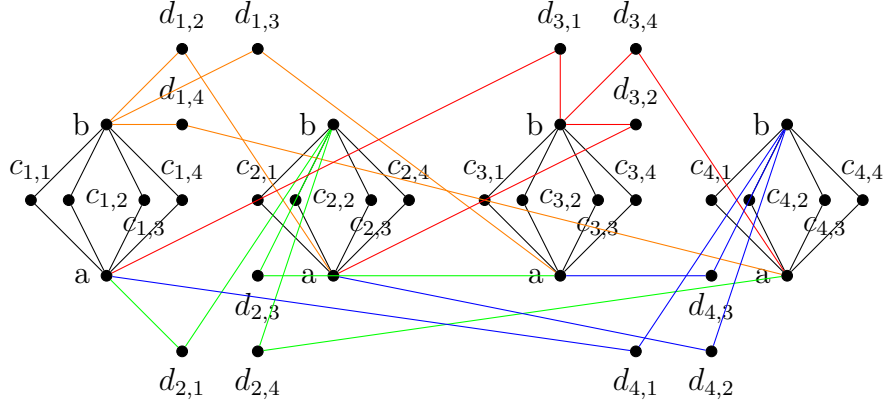


Fig. 6: The first part of  $H$  constructed for  $F$ ;  
The marked edges are connected to a  $d_{i,j}$ -colored vertex

For the second part, for every combination of  $i$  and  $j$  ( $1 \leq i, j \leq n$ ), a path containing three vertices colored  $a$ ,  $c_{i,j}$  and  $b$  in this order is added to  $H$ . Where  $j = 1$ , the  $[a]$  vertex is connected to the middle vertex  $z$ . Between the  $[b]$  vertices for  $j = 1$  and the  $[a]$  vertices for  $j = 2$ ,  $d_{i,j}$ -colored vertices are added. The value for  $i$  is the same as in the three vertex path to which the new vertex becomes connected via the  $[b]$  vertex and  $j$  takes on the values of every vertex the  $i$ -th vertex in  $G$  is connected to by an edge in  $E$ . Those  $[d]$  vertices are then connected to the  $[a]$  of the three vertices path corresponding to the  $j$  value in their color. This way, the connections in  $G$  are mimicked in the structure of  $H$ . This is repeated for every  $j$  until all the three vertex subgraphs have been connected. The last step is to connect the  $[b]$  vertices of the paths where  $j = n$  to a  $[c]$  end vertex  $z'$ . Fig. 7 shows the completed example construction. Now  $H$  has a first part that allows any path through the graph to be mimicked by the coloring, a vertex of a different color separating the two parts of  $H$ , and a second part whose connections are dependent on the edges of  $G$ . In the first and second part respectively,  $a$  and  $b$  are the only non-unique colors and vertices colored  $a$  or  $b$  are never adjacent. Thus a repetitive path has to run through both parts, that means it needs to pass the  $c$ -colored vertex  $z$ . Because of this, it also needs to reach the end vertex, as it is the only other vertex colored with  $c$ . So the only possible square path would look like this:  $w = QzQ'z'$  with  $Q$  and  $Q'$  and of course  $z$  and  $z'$  having the same coloring. To get from vertex  $z$  to vertex  $z'$ ,  $w$  has to pass through  $n$  vertices colored  $c_{i,j}$  for some  $i, j$ , visiting the same colors in the first part of  $H$ . As all vertices colored  $c_{i,j}$  with the same value for  $i$  are contained in the same subgraph of the first half of  $H$ ,  $w$  can only visit one  $c_{i,j}$  colored vertex for each value of  $i$ . This is only possible if  $G$  has

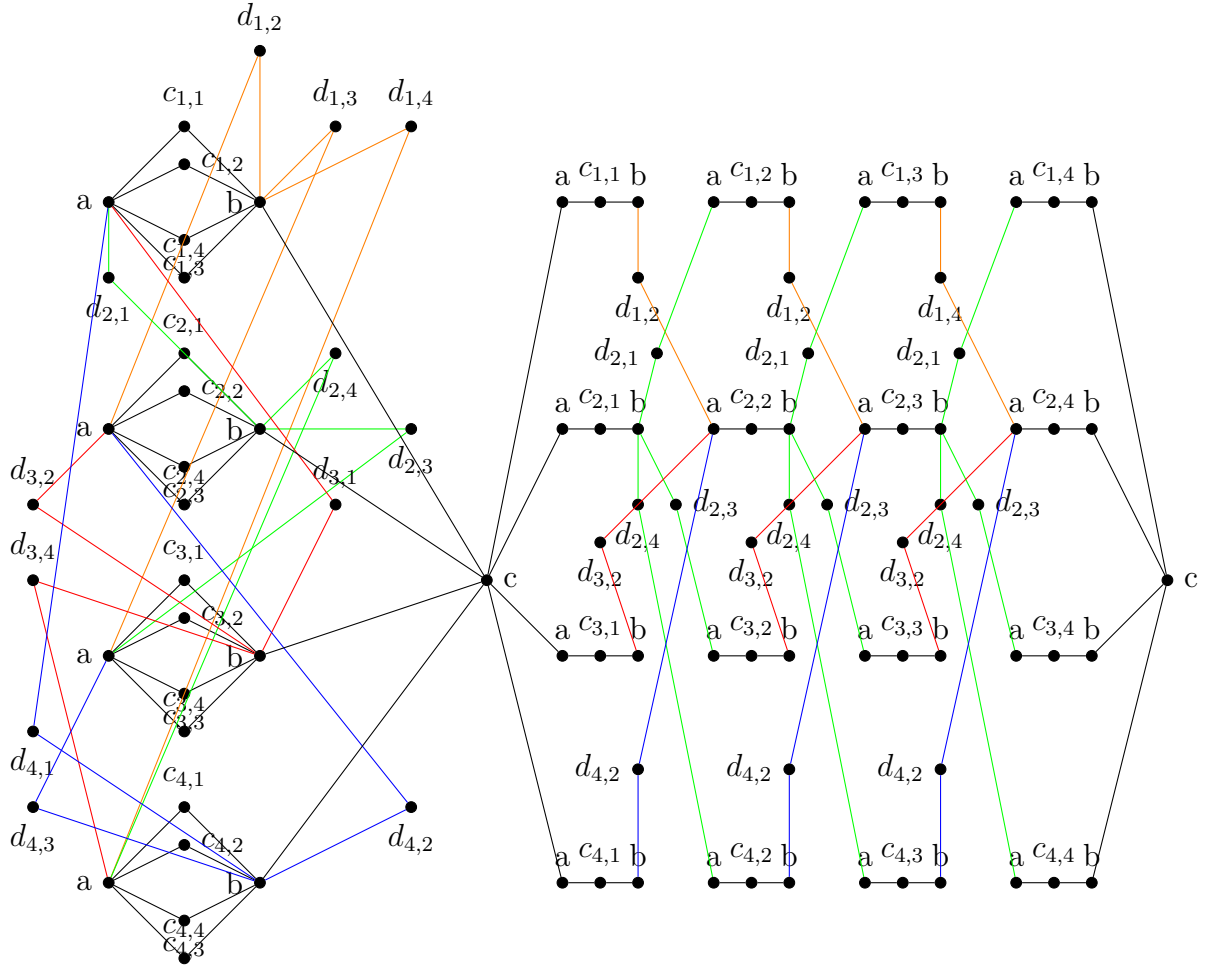


Fig. 7: The whole of  $H$  constructed for  $F$   
Edges connected to vertices colored  $d_{i,j}$  are marked

a Hamiltonian path, because then, by construction of the second part, there is a path through the second half of  $H$  that doesn't visit any two vertices with the same value for  $i$ . The coloring of this path can simply be copied for a path through the first part. If  $G$  doesn't have a Hamiltonian path, there is no path through the second part of  $H$  without having to visit two different vertices with the same value for  $i$ . To get the same coloring for the first part, however, the path there would have to pass through one of the  $K_{2,n}$  subgraphs twice. Therefore it is not an open path.

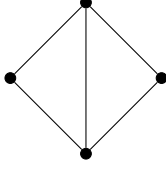


Fig. 8:  $F'$

Fig. 9: Example construction for  $F'$  [3]

$F$  is an example of a graph without a Hamiltonian path, so it will not be possible to find a path of the form  $w = QzQ'z'$ . The example given in fig. 9 (based on the graph from fig. 8) from Marx's and Schäfer's paper, though, has a square path.

Obviously, the construction of  $H$  takes a polynomial number of steps depending on the number of vertices in  $G$ . So the reduction works, making the problem of deciding the nonrepetitiveness of a vertex-coloring *coNP* – *hard*.

**Claim 1** *Deciding whether a vertex-coloring of a graph is nonrepetitive is in coNP.*

Let  $G'$  be a graph with an arbitrary vertex-coloring. A Turing machine can non-deterministically pick one path on  $G'$  and simply test every possible pattern from every vertex to see if it repeats itself. There are  $\frac{n-i}{2}$  possible patterns for a path of length  $n$ , starting from the  $i$ -th vertex that could be repeated. Going through them all would take  $n \cdot \frac{n-i}{2}$  steps. Thus, a Turing machine can non-deterministically decide if the given coloring is a negative instance of the problem, containing a path that is repetitively colored, in  $O(|V|^2)$  steps and that is precisely the definition of *coNP*.

So the problem is *coNP* – *complete*.

### 3.3 The complexity of nonrepetitive edge-colorings

**Theorem 2** *Deciding whether an edge-coloring is nonrepetitive is coNP-complete [5].*

The corresponding problem for edge-colorings is in the same complexity class and the proof is somewhat similar. They are, however, not directly related and one can't be inferred from the other [3]. The reason Manin's proof might seem familiar is that it's also based on the idea of a path with two parts that are colored identically to form the only possible repetitive colored path. The problem he reduces from is *3SAT*, corresponding to which a graph  $G$  is constructed.  $G$  has a so called snout, a sequence of vertices and edges with the sole purpose of providing the colors for the second part to mimic. The second part of  $G$  then consist of a variable gadget for each variable that splits into a positive and negative path variant and a clause gadget for each clause. The clause gadgets are connected to the variable gadgets in the same way the variables occur in the clause; if a variable is negated, the clause gadget is connected to a vertex in the negative path of the variable gadget, if it isn't, the clause gadget is connected to the positive path. As a result, the only way to pass through the second half of  $G$  with an open path, is to find an assignment satisfying the formula, traversing the variable gadgets opposite to that assignment and then the clause gadgets corresponding to it. If there is no satisfying assignment, at least one of the edges from the clause gadgets will need to pass through a vertex in the variable gadget that has already been visited. Thus a coloring in  $G$  is only repetitive if the *3SAT* problem it's constructed from is satisfiable - and nonrepetitive if it isn't.

To construct  $H$ , a few numbers from the *3SAT* problem  $f$  are needed. Let  $M$  be the maximum of the number of instances of all variables, meaning the number of instances of the most common variable with  $x$  and  $\neg x$  being counted separately,  $n$  the number of variables and  $m$  the number of clauses in  $f$ . First,  $n$  variable gadgets with the following structure are added: each variable gadget begins with a vertex  $b_i$  and then splits into a positive and negative path, both of them containing  $M$  vertices. The paths are both connected to the starting vertex of the next variable gadget,  $b_{i+1}$ , except for the last gadget which is connected to a vertex  $c$ . Now the clause gadgets are added, one vertex for each clause in  $f$  and one vertex  $d$ . The first clause gadget is connected to  $c$  and to a vertex in every variable gadget corresponding to a variable the clause contains. If the clause is  $(x \vee y \vee \neg z)$ , the clause gadget will be connected to one vertex in the positive path of  $x$ , one vertex in the positive path of  $y$  and one vertex in the negative path of  $z$ . Those vertices in turn are connected to the next clause gadget, unless there are no further clause gadgets, then they're connected to  $d$ . A vertex in the

variable gadgets can't be connected to more than two clause gadgets. This is possible because the paths in the variable gadgets each have  $M$  vertices, ensuring that no variable has more instances than distinct vertices in its corresponding path. The snout is fairly simple now, it's merely a sequence of  $(M + 1)n + 2m + 1$ <sup>1</sup> vertices, with every vertex  $a_i$  connected to  $a_{i+1}$  and  $a_{(M+1)n+2m+1}$  connected to the first vertex of the first variable gadget,  $b_1$ . An example for  $f(x, y, z) = (x \vee y) \wedge \neg x \wedge \neg y$  is given in fig. 10.

$H$  now is colored as follows: The path from the beginning to the end of the snout is given a sequence of distinct colors, the edge from  $a_1$  to  $a_2$  being colored  $w_1$ , the next edge being colored  $w_2$  and so on until the edge to  $b_1$  is colored  $w_{(M+1)n+2m+1}$ <sup>2</sup>. This will be the first part of the potentially repetitively colored path. Thus, the coloring of the variable gadget starts again with  $w_1$ , the  $k$ -th edge in variable gadget  $i$  being colored  $w_{(M+1)(i-1)+k}$ <sup>3</sup>. Now only the clause gadgets through to vertex  $d$  are left. Starting from  $c$ , the edge between  $c$  and the first clause gadget is colored  $w_{(M+1)n}$ , then the edges from the clause gadget to the variable gadgets are colored  $w_{Mn+1}$  and so forth until the edges from the variable gadgets to  $d$  are colored  $w_{(M+1)n+2m+1}$ .

The construction is not yet complete, however, because for every  $b_i$  with  $1 \leq n$ , there are paths containing the last edge of the positive path of variable gadget  $i - 1$  followed by the last edge of the negative path of variable gadget  $i - 1$ . Both of course have the same color, so there would always be a very simple repetitively colored path in  $G$ . The solution here is to simulate a directed graph for the coloring by using "*direction-determining colors*" [5, p. 6]. As every vertex is connected to at most 3 other vertices (in the clause gadgets), the edges connected to that vertex need to have 3 different colors. So we need 3 sets of direction-determining colors. Every edge outside the snout colored  $w_i$  is now replaced with a path  $a_i^j b_i^j w_i c_k^i d_k^i$  with  $i$  and  $k$  in  $\{1, 2, 3\}$ . This way, the coloring can be changed so that no two edges with the same color share a vertex. The colors of the snout can be chosen to match the rest of the path by introducing new subgraphs replacing an edge, so the way the coloring works is not affected at all. These new subgraphs look like the one in figure 11.

As the direction-determining colors don't change the sequence of colors in the original coloring and can be copied by the snout to equal the coloring of any path through the variable and edge gadgets, so the coloring of the second part of  $H$  will be examined with just the original colors  $w_i$ . Every path's coloring through the whole second part can be matched by the snout

<sup>1</sup>In [5] the number is  $Mn + 2m$ , however, the snout is too small this way

<sup>2</sup>The number of edges in the snout equals the number of vertices, not more as assumed in [5]

<sup>3</sup>Again the numbering is off in [5]

and as the path traversing the snout from beginning to end is an open path, the variable- and edge gadgets are the deciding factor. Each variable gadget has to be traversed in order to get to the edge gadgets and ultimately to the vertex  $d$ . Because of that, one path of each variable gadget can't be visited again by any edges from the clause gadgets. To put it differently, for each variable, every open path from  $c$  to  $d$  can only visit either the positive or the negative path of each variable gadget. This means that the assignment for all variables is the same in every clause gadget. If there is a path like this actually connecting  $b_1$  to  $d$ , it must be a satisfying assignment for  $f$ , because the clause gadgets are only connected to the paths in the variable gadgets satisfying them. If  $f$  is not satisfiable, at least two clauses require each a different variable assignment, which is reflected in  $H$  as two edges from clause gadgets to different parts of the same variable gadget both needing to be traversed - and thus forming a path that is no longer open. A square path can have no different form; limiting the path to the snout doesn't work because every edge there has a different color, limiting the path to the variable gadgets doesn't work because with the direction determining colors in mind, they can only be traversed in one direction and again, the colors in the variable gadgets won't repeat themselves. The problem remains when using only the edge gadgets or combining edge and variable gadgets. Just the snout and variable gadgets also can't form a square path, as the colors in the variable gadgets are at the beginning of the snout and to reach the variable gadgets, all the colors in the edge gadgets would have to be traversed first. Thus, there is no other possible square path than the one through all three elements of  $H$ , so  $H$  contains a square path only if  $f$  is satisfiable. Again, the construction of  $G$  only takes a polynomial number of steps, therefore the reduction proves that the problem of deciding the nonrepetitiveness of an edge-coloring is *coNP - hard*.

**Claim 2** *The problem of deciding whether an edge-coloring is nonrepetitive is in coNP.*

The reasoning here is the same as in section 3.2., thus making the problem *coNP - complete*.



## 4 The complexity of the Thue Number problem

Now that the nonrepetitive coloring problem has been examined, Thue Number type problem which rely on it will be introduced. Thue Number type problems ask for the number of colors that are necessary for a nonrepetitive coloring to be possible. Again, there is an edge-coloring problem, Thue Number, and also a vertex-coloring variant, Thue chromatic Number. The Thue chromatic Number will be the topic of the next section, first the complexity of the Thue Number will be shown. The proof is the extension of the nonrepetitive edge-coloring problem and also from [5]. As before, this proof hasn't been published so there were some missing mistakes in the original paper that have been corrected here. Furthermore, parts with sparse explanation have been fleshed out in the following to give a more thorough understanding of the matter.

### 4.1 The Thue number and Thue chromatic number

**Definition 23** *In general, the Thue Number in respect to any sequence of elements is the minimum number of elements needed to make that sequence nonrepetitive.*

For words, this number is three [1]. For graph colorings, it's not as easy as to name a single number, since a coloring always refers to one specific graph. So every graph has a its own Thue Number: the least number of colors needed to color the graph nonrepetitively [2]. The problem explicitly named Thue Number, however, refers to only edge-colorings, the counterpart for vertex-colorings being Thue Chromatic Number [4].

**Definition 24** *The number of colors needed to construct a nonrepetitive edge-coloring of a graph  $G$  is called  $\pi(G)$ , the Thue Number of  $G$ . The number of colors needed to construct a nonrepetitive vertex-coloring of  $G$  is called  $\sigma(G)$ , its Thue Chromatic Number.*

In both cases, the problem is rather hard to solve. Manin proved Thue Number to be  $\Sigma_2^P$ -complete [5], which will be examined in the following section. The complexity of Thue Chromatic Number is still an open problem [4].

## 4.2 Constructing the graph

### 4.2.1 The structure of $H$

**Theorem 3** *The problem of deciding whether a graph  $G$  can be colored non-repetitively with  $k$  colors is  $\Sigma_2^P$ -complete*

The proof of complexity for  $\pi(G)$  in this section is a continuation from the one for the complexity of deciding the nonrepetitiveness of an edge-coloring of a graph. The key difference is the number of colors as an added factor. The task at hand is not just to prove that the graph  $H$  that will be constructed for the proof has a repetitive coloring only under a certain condition, but that also that this coloring consists of a fixed number of colors. The structure of  $H$  is similar to the one in section 3.3, in fact the graph from there can be extended to  $H$ . The problem reduced from here is  $\forall\exists SAT$  instead of  $3SAT$ , so the first step in this extension would of course be to differentiate between the universally and existentially quantified variables. This is rather easy, the first edge in every variable gadget's path corresponding to a universally quantified variable is colored  $w_i^1$  in the positive path and  $w_i^0$  in the negative path, keeping the original color  $w_i$  but forcing any repetitively colored path to take a specific path through the variable gadget depending on whether  $w_i^1$  or  $w_i^0$  is used in the snout. This graph will be referred to as  $G$ .

Then, various subgraphs, new gadgets, are added to  $G$  to form  $H$ , making the overall structure more complex and introducing the idea of bottlenecks, the places where nonrepetitively colored subgraphs are connected. Those are especially important, of course. The exact structure and coloring of the different gadgets will be covered in the next section, first the rules of when and where they're added will be given. Let  $n$  be the number of universally quantified variables,  $u$  the number of existentially quantified variables and  $c$  the number of colors used to color  $G$ . Further, let  $l$  be the smallest integer to fulfill the inequality  $2^l \geq c + u + 1$  and  $m = 4l + 3$ .  $H$  will be colored with  $2^m + 7c + 6$ <sup>4</sup> colors. This is a lot of colors and should give an indication as to how complex  $H$  will be in the end.

The gadgets that will now be added are called  $C$ -,  $N$ -,  $P$ - and  $Q$ -gadgets plus 7-Hypercubes that have no specific name. Any two opposing vertices of the 7-Hypercubes, however, are called edge-gadget vertices, with the vertex closer to the tip of the snout called  $v_0$  and the other one called  $v_1$ . The  $C$ -,  $N$ - and  $P$ -gadgets are closely related, they each replace an edge colored  $w_i^j$ . For the  $i$ -th universally quantified variable, the  $w_i^j$ -colored edge in the snout becomes a  $C$ -gadget (called  $C^i$ ), the edge in the negative path of the variable

---

<sup>4</sup>In [5], only  $2^m + 6$  colors are mentioned, the unique versions of some of these colors added for each consistency gadget, however, raise this number by the  $7c$  added here

gadget an  $N$ -gadget (or  $N^i$ ) and the edge in the positive path of the variable gadget a  $P$ -gadget ( $P^i$ ). As the gadgets have vertices on both ends, they are connected to the preexisting neighboring vertices with edges colored  $w_i^I$  before  $C^i$ ,  $N^i$  and  $P^i$  and  $w_i^F$  after. All three gadgets each have a  $v_0$  and a  $v_1$  edge-gadget vertex like the 7-Hypercube does. The  $C$ - and  $N$ -gadget additionally have a central vertex  $u$ . For every edge not carrying a true or false assignment in the coloring (so every edge that hasn't been replaced in the last step), a 7-Hypercube is added.

Lastly, the  $Q$ - or consistency gadgets: they are  $2^m$ -cliques and one is added for each universally quantified variable (called  $Q^{X_i}$  for the  $i$ -th variable) and for each color that is not associated with a universally quantified variable, that is every color but  $w_i^0$  and  $w_i^1$  (then referred to as  $Q^{e_i}$  or the plume of a Hypercube or  $P$ -gadget). They are used to connect all the other gadgets, as all of their vertices are merged with either edge-gadget vertices, central vertices of the two lower layers of Hypercubes and  $P$ -gadgets or the  $u$  vertices of  $C$ - and  $N$ -gadgets. The  $u$  vertex of  $C^i$  and  $N^i$  respectively becomes each one vertex of  $Q^{X_i}$ . Two vertices of  $P^i$  are merged with each one vertex of  $Q^{X_i}$ , one with distance three (called  $u_0$ ) and the other with distance four (called  $u_1$ ) from  $v_0$ . The same goes for the Hypercubes. Additionally, either the central vertex of the first two Hypercube and  $P$ -gadget layers or an edge-gadget vertex become part of the consistency gadget. In the former case, the consistency gadget is then called the plume of the vertex.

The edge gadgets are also connected at their  $v$  vertices, with a  $v_0$  vertex of one gadget becoming the  $v_1$  of the second. All edge gadgets of a consistency gadget  $Q$  are connected to edge gadgets of the same consistency gadget  $Q^i$  by their  $v_0$  vertices. The same goes for the edge gadgets connected to the  $v_1$  vertices of  $Q$ 's edge gadgets. In the first case,  $Q^i$  corresponds to the edge preceding the edge in  $G$  that  $Q$  represents, in the second case it's the edge after  $Q$ 's edge in  $G$  <sup>5</sup>.

Now  $H$  has become quite big, with a lot of 7-Hypercubes added and them being heavily connected by the consistency gadgets. The gadgets added for universally quantified variables are the only points in the original graph  $G$  that are also connected to this structure of Hypercubes. Before describing a coloring for these rather complex connections, the structure of the gadgets themselves will be shown and they will be colored nonrepetitively.

---

<sup>5</sup>This is never explicitly mentioned in [5], but important for the final part of the reduction

### 4.2.2 Subgraphs of $H$

All of the gadgets used, except for the consistency gadgets, are based on Hypercubes although the  $C$ -,  $N$ - and  $P$ -gadgets have been modified. The  $C$ - and  $N$ -gadgets are each a sequence of Hypercubes connected to form one structure. The  $C$ -gadget (fig. 12) is made up of a 2-Hypercube, two single vertices that can be viewed as a 1-Hypercube each and a 3-Hypercube.

Some vertices have added edges that would not exist in a normal Hypercube. Their purpose is to elevate the degree of these vertices. The  $N$ -gadget is somewhat similar, but consists of a 3-Hypercube connected to a 4-Hypercube (fig. 13). These plumes in The  $C$ - and  $N$ -gadgets will be used later to restrict the possible colorings in these gadgets, as they will be for the  $P$ -gadget.

The  $P$ -gadget is a 7-Hypercube with several modifications. First, its dimensions are labeled: the dimensions  $a$ ,  $b$  and  $ab$  cover the distance from the vertex  $u_0$  to  $v_0$ . The dimensions  $c$ ,  $d$ ,  $ac$  and  $ad$  then span the distance between the vertex  $u_0$  and the vertex  $v_1$  as they have been defined above. The edges from  $v_1$  in the dimensions  $c$ ,  $d$ ,  $ac$  and  $ad$  are removed. Then, two vertices,  $u_{-1}$  and  $u_{-2}$ , are defined. To reach  $u_{-1}$ , the dimensions  $a$ ,  $c$  and  $ac$  have to be traversed from  $u_0$ , for  $u_{-2}$  the dimensions  $a$ ,  $d$  and  $ad$ . Each  $2^m - 2$  edges are added to  $u_{-1}$  and  $u_{-2}$  and from the vertices in dimensions  $c$  and  $d$  adjacent to  $u_{-1}$  and  $u_{-2}$ , all edges to  $b$  are removed. The added edges are called the plume of these vertices. Parts of the  $P$ -gadget now still look like a normal Hypercube, but a few of the dimensions have been altered (see fig. 14).

This Hypercube part of the  $P$ -gadget, as well as the  $C$ - and  $N$ -gadgets, can be colored nonrepetitively by exactly seven colors each. The reason for this is that they're all based on Hypercubes and that an  $n$ -Hypercube can be colored nonrepetitively with  $n$  colors. The 7-Hypercubes and the  $P$ -gadget's Hypercube portion are of course 7-Hypercubes. The  $C$ - and  $N$ -gadget consist of smaller Hypercubes whose dimensions add up to seven. Coloring them works in the following way: Hypercubes can be imagined as square or diamond like structures stacked upon and connected to each other, with each diamond connected to  $n - 2$  dimensions (each one except the two it's already in). Each of these diamonds will be colored with the same two colors. What's left are the connections between the diamonds and the other dimensions. The edges to different dimensions from each diamond have  $n - 2$  distinct colors, one for every dimension. Every path containing a square path would have to contain an even number of edges from every dimension it passed. The vertices in a Hypercube, however, are connected to exactly one vertex of each dimension, the one that differs in exactly one place from its own labeling. With each change to or from dimension  $i$  meaning a change in the  $i$ -th place

of the vertex labels, passing through dimension  $i$  an even number of times ends with the  $i$ -th place being changed from 0 to 1 or 1 to 0 an even number of times - so the  $i$ -th place in the labeling is the same for the starting and end vertex. As the square portion of a square path must run through every dimension it visited an even number of times, the starting and end vertex must have the same labeling, thus forming a circle.

To color the  $C$ - and  $N$ -gadgets, this technique is used with distinct sets of colors for all their different parts. The consistency gadgets, as they're Cliques, each need a lot more colors,  $2^m - 1$  as shown in [2].

The plumes in the  $P$ -gadget of course pose a problem to this way of coloring a Hypercube. They will be colored with colors that are distinct from the seven used for the rest of the gadget.

### 4.3 The coloring of $H$

The problem with coloring  $H$  is that the number of colors that can be used is fixed.  $7c$  of the colors will be used later, until then  $H$  will be colored with  $2^m + 6$  colors. That isn't exactly a lot seeing that each consistency gadget is a  $2^m$ -Clique. The question now is how the colors can be distributed: can the color set  $Z$  of one of the 7-Hypercubes or gadgets contain colors from the color set  $G^*$  coloring the Clique it's connected to? If not, the consistency gadget has to be colored with the color set  $G^*$  which doesn't contain the seven colors that are left to make up  $Z$  which colors its edge gadgets. This is actually the easiest possible outcome for trying to show later on whether a coloring of  $H$  is nonrepetitive because in every other case the number of possible colorings would be much higher. In fact,  $H$  is constructed in such a way that the consistency gadgets and the rest of the added subgraphs have distinct color sets to make the reduction easier. To show this, every subgraph has to be examined individually. The Hypercubes are a good starting point because they're independent from the rest of the gadgets.

First, recall how Hypercubes are connected to consistency gadgets: the two vertices  $u_0$  and  $u_1$  belong to the consistency gadgets. Let  $u$ ,  $v$  and  $w$  be vertices of a consistency gadget  $Q$ . If an edge in the Hypercube adjacent to  $w$  is colored the same way as the edge from  $u$  to  $v$ , the  $[uv]$  edge in the Hypercube, the edge from  $w$  to  $u$  colored  $uw$ , the edge from  $u$  to  $v$  and the  $[uw]$  edge from  $v$  ( $v$  has this edge as every vertex in  $Q$  has an adjacent edge for every color in  $G^*(Q)$ ) form a square path. So the first two layers of a Hypercube have to be colored with the colors from  $Z(Q)$ . The same applies to all edges connected to either  $u_0$  or  $u_1$ . For edges with a distance of two and three from  $u_i$ , the same argument can be made, the square path simply starts one or two edges further into the Hypercube and continues for one or

two edges more after passing  $v$ . Of course this could be continued, but the more interesting question about edges of distance four from  $u_i$  is whether the coloring around the  $u_i$  matches. The vertices  $u_0$  and  $u_1$  have been selected in such a way, that there is no edge with a distance greater than four from the closer one. Let the vertex  $x$  be separated from  $u_0$  by a path colored  $c_1c_2c_3$  and  $y$  by a path colored  $c_1c_2c_4$ . Then  $e_4$  is an edge connected to  $x$  and  $e_3$  an edge connected to  $y$ . These edges are adjacent.  $e_3$  and  $e_4$  have a distance of four from  $u_0$ , all the edges closer to  $u_0$  are colored according to what was described in 4.2. If the colors of  $e_3$  and  $e_4$  are both in  $G^*$ , so  $C(e_3), C(e_4) \in G^*$ , there is a square path starting in the consistency gadget with two edges colored  $C(e_3)$  and  $C(e_4)$  until it reaches  $u_0$ . It continues via a  $c_1c_2$  colored path to  $e_3$  and from there  $e_4$ . From  $e_4$ ,  $c_1c_2$  can be repeated again, completing the square path. If  $C(e_3) \in G^*$  and  $C(e_4) \in Z$ , the square path would now look like this:  $C(e_4)$  from an edge adjacent to the edge colored  $C(e_3)$  in the consistency gadget,  $C(e_3)$  and then  $[c_1][c_2]e_4e_3[c_1][c_2]$ . For  $C(e_4) \in G^*$  and  $C(e_3) \in Z$ , the path is analogous.<sup>6</sup> Therefore,  $C(e_3)$  and  $C(e_4)$  both have to be in  $Z$ . This can be shown in the same way for  $u_1$ , so the whole Hypercube has to be colored with the seven colors of  $Z$ .

The  $P$ -gadget's structure is similar to that of a 7-Hypercube, save for the removed edges, so for the most part, the same arguments apply. The removed edges have a distance of at least 3 from  $u_0$  and 4 from  $u_1$ , so below these distances, the coloring of the  $P$ -gadgets has to correspond to the coloring of the Hypercubes. For edges with a distance of 3 from  $u_0$  and 4 from  $u_1$ , the arguments from above can still be used since none of the removed edges share a vertex, so the  $P$ -gadgets have to be colored with  $Z$ . The plumes of the  $u_{-i}$  vertices are an exception, they should be colored by  $G^*$  excluding the color of the edge between  $u_0$  and  $u_1$ . This way, there is no square path with an edge in the plume as beginning or end: it would have to contain exactly one edge in the consistency gadget and the only edge connecting a vertex of the  $P$ -gadget with another vertex of the  $P$ -gadget in the consistency gadget is the one between  $u_0$  and  $u_1$ .

This translates very easily to the  $C$ - and  $N$ -gadgets as they're connected to the consistency gadgets by a very central vertex  $u$  from which the distance of every edge, except the ones connected to  $v_1$ , is 3, so all these edge colors need to be from  $Z$ .

As the edge gadgets aren't only connected by the consistency gadgets but also by their  $v_0$  or  $v_1$  vertices, a few conditions need to be fulfilled. The

---

<sup>6</sup>In [5], this part is almost identical, but Manin claims that the last two square paths start inside the consistency gadget. As  $Z \cap G^* = \emptyset$  and either  $C(e_3)$  or  $C(e_4)$  is in  $Z$  and the other in  $G^*$ , the first edge can't be located inside the consistency gadget but has to be adjacent to it.

Hypercubes as well as the  $C$ -,  $N$ - and  $P$ -gadgets are connected to the gadgets of the consistency gadget representing the edges before and after the edge gadget's edge in  $G$ . Preventing square paths through two neighboring gadgets will be necessary, of course, but this connection also connects two consistency gadgets. The conditions these consistency gadgets need to fulfill specify the coloring of the  $C$ - and  $N$ -gadgets further, so they will be treated first.

Assume a gadget  $P^i$ 's vertex  $v_0$  is also the  $v_1$  vertex of another edge gadget  $E^e$  and the color of the edge  $e$  in  $G$  this Hypercube represents is  $x_i^i$ . This edge gadget can only be a Hypercube as the edges in  $G$  corresponding to universally quantified variables are enclosed by edges with direction determining colors. Therefore  $C$ - and  $N$ -gadgets aren't relevant in this case.

There are four gadgets involved:  $P^i$  and its consistency gadget  $Q^{x_i}$  as well as  $E^e$  and its consistency gadget  $Q^{x_i^i}$ . Excluding a square path only in  $P^i$  and  $E^e$ , a square path would need to have a part in an edge gadget that repeats itself in a consistency gadget. Interesting here is the case when a path has a part in  $P^i$  that repeats itself in  $Q^{x_i^i}$ . If  $Z(P^i) \in G * (Q_{x_i^i})$ , then there is a path of length three from  $v_1$  to  $u_1$  in  $E^e$  (colored  $[k_1][k_2][k_3]$ ) and a few already known paths in  $P^i$  to its  $v_0$  that might be repeated in  $Q_{x_i^i}$ . Those include  $\tau_0 = [a][b][ab]$  from  $u_0$  to  $v_0$ ,  $\tau_1 = [c][d][ac][ad]$  from  $u_1$  to  $v_0$ ,  $\tau_2 = [b][c][ab][ac]$  from  $u_{-1}$  to  $v_0$  and  $\tau_3 = [b][d][ab][ad]$  from  $u_{-2}$  to  $v_0$ .  $[k_1][k_2][k_3]$  can be repeated either partly in  $P^i$  and  $E^e$  or completely in one of these gadgets, so there is a square path looking somewhat like  $[k_1]\tau_i[k_1][k_2][k_3]C(\tau_i)[k_2][k_3]$ ,  $\tau_i$  being in  $P^i$  and  $C(\tau_i)$  in  $Q^{x_i^i}$ . All paths that are a permutation of  $C(\tau_i)$  in  $Q^{x_i^i}$  thus need to form a circle. Two more paths that must be loops can be inferred from them:  $[a][c][ac]$  as this is the result of combining paths  $\tau_0$  and  $\tau_2$  and then removing the circles in that path and  $[a][d][ad]$  by the same logic for  $\tau_0$  and  $\tau_3$ . Removing the circles in the paths means that this new path has the same start and end vertices, so it could also be repeated in  $Q^{x_i^i}$ . If  $P^i$ 's  $v_1$  is  $E^f$ 's  $v_1$ , permutations of the same paths must be loops in  $Q^{x_i^f}$  corresponding to  $E^f$  because the same argument can be made for the paths between  $u_1$  and  $v_1$ ,  $u_{-1}$  and  $v_1$  and  $u_{-2}$  and  $v_1$ . There, of course, could be other paths repeating themselves, but these will be remedied later. Specifically because of these forced loops in  $Q^{x_i^i}$  and  $Q^{x_i^f}$ , better statements about the coloring of the  $C$ - and  $N$ -gadgets can be made.

In  $N^i$ , there would be a square path from the  $u_{-i}$  plume of  $P^i$ , through its  $u_0$  and then  $u$ , similar to the reason the color between  $u_0$  and  $u_1$  in  $P^i$  can't be contained in the plume of  $u_{-i}$ . This path  $p$  would be  $c_1[ac][c][a]c_1[ac][c][a]$  or  $c_1[ad][d][a]c_1[ad][d][a]$ . To avoid this, the  $[ac][c][a]$  and  $[ad][d][a]$  parts repeating the paths from  $u_{-1}$  and  $u_{-2}$  in  $P^i$  respectively can't exist from  $u$  in

$N^i$ . If  $N^i$  is connected to  $E^e$  in the same way as  $P^i$  above, both parts of  $N^i$  have to be colored with colors that form a loop in  $Q^{x_i^i}$  and  $Q^{x_i^f}$ , respectively, for the same reasons as presented above. There are three different paths of length three that form loops in  $Q^{x_i^i}$ , but as just shown, two of them can't occur in the first part of  $N^i$ . As this first part needs to be colored with three colors, though, only one combination of colors remains:  $a, b, ab$ . This leaves  $\{c, d, ac, ad\}$  for the 4-Hypercube part of  $N^i$ . To prevent a square path in this part, all the edges, even the ones closest to  $v_1$  that wouldn't strictly have to be colored in  $Z$  by the induction above, need to be colored from this set. So now all of  $N^i$  has to be colored from  $Z(Q^{X_i})$ .

The argument for  $C^i$  is slightly different, because the overall structure of the  $C$ -gadget is a bit more complicated than the  $N$ -gadget. The goal is similar, though: show that all the edges, even the ones adjacent to  $v_1$ , have to be colored with  $Z$ . Again, there are possible square paths through  $P^i$ 's  $u_{-i}$  and  $u_0$  or  $u_{-i}$  and  $u_1$ . To avoid a repetition of a path from  $u_{-i}$  to  $u_1$ , either  $C^i$  can be colored in  $\{a, b, ab\}$  from  $v_0$  to  $u$  (the color set for this part is called  $A_0$ ), following the same logic as above, or  $n$  can be colored  $b$  and thus leaving no possible path from  $u_1$  to  $u_{-i}$  as none of them contain  $b$ . If the first part of  $C^i$  is colored in  $\{a, b, ab\}$ , then  $n$  has to be colored in either  $c$  or  $d$ , else  $m$  has to be colored  $c$  or  $d$ . The colors of  $m$  and  $n$  can't be repeated in the plumes of vertices with distance two or less, this would yield an easy square path. Also, if a  $p_i$  is colored with a color that is contained in the 2-Hypercube, there could be a square path starting in the plume of a  $w_j$  connected to that edge, and going through  $p_i$  to the plume of a  $z_k$ . To avoid this, the 3-Hypercube can't be colored with the same colors as the 2-Hypercube and thus the color sets coloring  $C^i$  from  $v_0$  to  $u$  and from  $u$  to  $v_1$  have to be distinct. As before, the colors from  $v_0$  to  $u$  have to form a loop in  $Q^{x_i^i}$  and the ones from  $u$  to  $v_1$  have to form a loop in  $Q^{x_i^f}$ . Combining all this information means that the first part has to be colored in  $A_0$  as the other two possible paths that are loops in  $Q^{x_i^i}$  would form repetitions for the path from  $u_{-i}$  to  $u_0$  in  $P^i$ . Therefore the second part must be colored in  $A_1 = Z(Q^{X_i}) \setminus A_0$ , which must be a loop in  $Q^{x_i^f}$ . By the same argument as above, all edges in  $C^i$  are now colored in  $Z$ .

Lastly, the  $7c$  colors that were set aside earlier will be used to guarantee that two adjacent edge gadgets won't form repetitive paths. Each consistency gadget will have 7 assigned colors that are unique in  $H$ . For the consistency gadgets corresponding to an edge in  $G$ , there are up to four Hypercubes associated with it. The first three get a path colored in these unique colors spanning the distance from  $u_0$  to  $v_0$  and  $u_0$  to  $v_1$  each. For the first, these paths are  $\{a_i, b_i, ab_i\}$  and  $\{c_i, d_i, ac_i, ad_i\}$ , for the second  $\{a_i, c_i, ac_i\}$



and  $\{b_i, d_i, ab_i, ad_i\}$  and the for the third  $\{a_i, d_i, ad_i\}$  and  $\{b_i, c_i, ab_i, ac_i\}$ . If there's a fourth Hypercube, it's colored in one of those three ways but not the way the Hypercube corresponding to the edge in the snout is colored. Now the paths between those key vertices in the Hypercubes are unique except for maybe two Hypercubes in the same consistency gadget. This excludes any square paths in neighboring edge gadgets.

In consistency gadgets added for universally quantified variables, the colors in the  $N$ - and  $P$ -gadgets are all made unique for this consistency gadget. The  $C$ -gadget's coloring depends on the value of the variable. If its value is 1, the  $C$ -gadget is colored so that it contains a path through the corresponding  $N$ -gadget. The first part of  $N^i$  is now colored in  $\{a_i, b_i, ab_i\}$  and the second one in  $\{c_i, d_i, ac_i, ad_i\}$ , so the path in  $C^i$  has to look like  $(a_i, b_i, ab_i, c_i, d_i, ac_i, ad_i)$ . If  $x_i$  is 0,  $C^i$  won't contain this path and thus there is no path through both  $C^i$  and  $N^i$ . Now, however there is a path through  $P^i$ . If  $C^i$  contains a path through  $N^i$ ,  $\{a_i, b_i, ab_i\}$  make up the first three edges of the path. In  $P^i$ , all edges have been removed from  $v_1$  except  $[a]$ ,  $[b]$  and  $[ab]$ . If those colors are only in the beginning of the path, it won't be able to reach  $v_1$  in  $P^i$ . As soon as the path in  $C^i$  is the same as for  $N^i$  anymore, though, it can pass  $P^i$  with a path  $(a_i, c_i, ac_i, b_i, d_i, ad_i, ab_i)$ . The conditions for a coloring of  $C^i$  as seen above are fulfilled and this path passes through  $P^i$ , but not  $N^i$ .

The last seven colors are called  $\{\beta_1, \dots, \beta_7\}$ . Those are used for the  $u$  vertices of the edge gadgets. Each consistency gadget contains at most eight such vertices. The edges between these vertices are now colored  $\beta_i$ , with  $\beta_1$  being the color between the  $u_0$  and  $u_1$  vertex of every Hypercube or  $P$ -gadget. Coloring this smaller  $n$ -clique (with  $n \leq 8$ ) inside the consistency gadget is possible with 7 colors. In the consistency gadgets corresponding to universally quantified variables, it's important that all the edges connecting  $u$  vertices are colored differently. This way, square paths containing such connections inside a consistency gadget have to repeat those in other consistency gadgets in the same way.

The remaining edges in the consistency gadgets are then also colored with the new unique colors. To ensure that the consistency gadgets' colorings differ from each other, the group coloring method described in [2] for cliques is used in a different variation. The  $u$  vertices have already been colored, so their colors remain the same. The rest of the consistency gadget vertices are randomly assigned an element of a subgroup of  $\mathbb{Z}_2^m = \underbrace{\mathbb{Z}_2 \times \mathbb{Z}_2 \times \dots \times \mathbb{Z}_2}_{m \text{ times}}$

isomorphic to  $\mathbb{Z}_2^4$ . For each nonzero element of  $\mathbb{Z}_2^{\frac{m}{4}}$ , there is such a subgroup of  $\mathbb{Z}_2^m$ , so the number of subgroups is large enough to assign each consistency gadget a distinct one. The four generators  $g^j, j \in \{0, 3\}$  of each subgroup look

the following way: the  $i$ -th digit of  $g^j$  is  $\frac{(j-i-3)}{4}$  if 4 is divisible by  $(j-i-3)$  and else zero.

To color the consistency gadgets, the generators of the subgroup of a consistency gadget  $Q^i$  are assigned colors instead of the preexisting integers, e.g.  $g^0 = a_i, g^1 = b_i, g^2 = c_i, g^3 = d_i$ . The remaining colors are defined as the sum of the other colors,  $ab_i$  as  $a_i + b_i$ ,  $ac_i$  as  $a_i + c_i$  and  $ad_i$  as  $a_i + d_i$ . The subgroups now still have the same distinct structure as before, but every vertex is labeled with a color instead of a number. the edges are colored as sums of their numbers, with an edge between a vertex labeled  $a_i$  and  $b_i$  colored  $ab_i$ . The color combinations in the consistency gadgets thus are distinct even beyond the unique colors.

Now neighboring edge gadgets won't be a problem and there is always a path in  $C^i$  that can be repeated in either  $N^i$  or  $P^i$ . With this, the coloring of  $H$  is complete and what's left is to prove that it's indeed nonrepetitive if  $f$  is not satisfiable.

#### 4.4 The final part of the reduction

As the colors of the original graph  $G$  stayed the same and there is a path through each  $C^i$  that can be matched by either  $N^i$  or  $P^i$ , if  $f$  is satisfiable, the square path from the tip of the snout to  $d$  exists just like it did in  $G$ . The more difficult question here is whether it's the only square path.

**Claim 3** *There is no square path within a consistency gadget and its edge gadgets.*

As the section above showed, there are no square paths within a consistency gadget and one of its neighboring gadgets. So to form a square path, one part of the path would have to repeat itself in another edge gadget belonging to the same consistency gadget. That means the path  $p$  would have the form  $p = zcz'c'$  with  $z, z' \in Z$ ,  $C(z) = C(z')$  and  $c, c' \in G^*$ ,  $C(c) = C(c')$ . So  $z'$  has to travel through the edge gadget back to the consistency gadget, from either  $u_0$  to  $u_1$  or the other way around. Replacing  $z$  and  $z'$  with the edge between  $u_0$  and  $u_1$  therefore results in a shorter square path with the same structure as just  $z$  and  $z'$  that aren't even part of the consistency gadget have been changed. This, however, also means that  $p$  is contained entirely in the consistency gadget and is therefore a loop. The same vertices visited in the smaller path are also visited in the longer one, so the loop still exists in any path  $p$ .

For  $C$ -,  $N$ - and  $P$ -gadgets, such a square path would have to be more complex. The  $C$ - and  $N$ -gadgets' coloring is constructed so that paths from inside

the  $P$ -gadget's  $u_{-i}$  vertices can't repeat themselves there. The plumes of  $u_{-i}$  also lack the color  $\beta_1$  to avoid a simple repetition from  $u_0$  to  $u_1$ , so that any square path with a part in  $P^i$  now has to be longer than just one edge. The plumes of  $u_{-i}$ , however, are only one edge long, so there's no path repeating the colors in the plumes of  $u_{-i}$  in the consistency gadget in just one edge and then repeating the rest of the path in another edge gadget. Also, as the colors spanning from either  $u_0$  or  $u_1$  to a  $u_{-i}$  and the other pair of  $u$  vertices are distinct, there is no square path through both at the same time with just one section in  $Q^i$ . So a square path in a consistency gadget corresponding to a universally quantified variable needs at least two sections inside the consistency gadget. As the  $u$  vertices in  $Q^i$  are colored uniquely, they can't be contained in such a path. Inside the  $C$ - and  $N$ -gadgets, however, each vertex has a distance of at most four from  $u$ . A square path through  $C^i$  or  $N^i$ , thus, could again not use the plumes in  $P^i$ , as it would have to use both but a distance of four is not enough to repeat all seven colors in the path between the plumes. So this path isn't possible either, consequently no square path can repeat the colors from the plumes in  $P^i$  and as square paths inside the edge gadgets can't occur by construction, the whole consistency gadget with edge gadgets is colored nonrepetitively.

**Claim 4** *In a square path, colors in edge gadgets can't be duplicated in consistency gadgets, except in one plume.*

There are a few different possibilities how a square path  $p$  contradicting this claim might look. The edge gadget in question will be referred to as  $E_0$  and the consistency gadget as  $Q_1$ .

The first possibility would be that  $E_0$  doesn't belong to  $Q_1$  and that  $p$  doesn't start in  $E_0$  but in another gadget. As the connections from and to other vertices in any gadget form the set  $\{u_0, u_1, u_{-1}, u_{-2}, v_0, v_1\}$ , the part of  $p$  inside  $E_0$  has to end at one of these vertices. The coloring of  $Q_1$  is constructed so that such a path is a loop, though, so either  $E_0$  does belong to  $Q_1$  or  $p$  starts in  $E_0$ .

If  $p$  starts in  $E_0$ , there are again two cases of what can follow the edges of  $E_0$ . It could either be the consistency gadget  $E_0$  belongs to,  $Q_0$ , or another edge gadget  $E^*$ .

If  $p$  starts in  $E_0$  and then goes on into  $Q_0$ , it must also leave  $Q_0$  in some way. To do that, it has to traverse another edge gadget to get to  $Q_1$ . It has to leave, however, after traversing enough edges in  $Q_0$  as to not repeat a neighboring edge gadget in  $Q_1$ , as this would mean a loop like before, and also before the path in  $Q_0$  becomes large enough to be repeated in an adjacent edge gadget, again a case that would form a loop. Even if  $p$  avoids both of

that, the color sets  $Z$  of the consistency gadgets are different, so if  $p$  leaves  $Q_0$  to an edge gadget and  $Q_1$  eventually does the same, the colors in  $p$  after those two points are completely distinct and thus can't be a repetition of one another.

So if  $E_0$  doesn't belong to  $Q_1$ ,  $p$  has to start in  $E_0$  but can't continue to  $Q_0$ ; it has to pass through another edge gadget. The path through  $E^*$  also needs to be duplicated somewhere. It can't be duplicated in  $Q_0$ , if it belongs to  $Q_0$  or not, because in the first case,  $Z(Q_0)$  and  $G * (Q_0)$  are distinct and in the second case, the situation mirrors possibility one, thus forming a loop. So  $E^*$  has to belong to  $Q_1$ , connecting  $E_0$  and  $Q_1$  without passing through  $Q_0$ . Thus,  $E_0$  and  $E^*$  are connected at a  $v_i$  vertex and the path from the connecting vertex in  $E^*$  to a  $u$  vertex of the consistency gadget is at least three edges long. This path of course can't be repeated in  $Q_1$ , so the part of  $p$  through  $E_0$  is repeated in  $Q_1$  and the part through  $E^*$  is repeated in another edge gadget  $E_1$  starting from one of its  $u$  vertices. If the path through  $E^*$  ends at its second  $v$  vertex, the issue of possibility one surfaces again, as  $Q_1$  is colored to explicitly prevent such square paths. The same goes if the path through  $E^*$  ends at a  $u$  vertex, then the path in  $E_1$  ends at one of its  $v$  vertices. Either this forms a loop in  $Q_1$  or  $p$  ends here and creates a contradiction in the construction as the  $u$  vertices in  $Q_1$  are connected by  $\beta_i$  colored edges and the coloring is constructed in such a way that those colors can't exist anywhere but between those vertices.

Thus  $E_0$  has to belong to  $Q_1$ . Then, however, the path through  $E_0$  can only be repeated in a plume of  $Q_1$  so the colors match. That also means that  $p$  ends at the plume, so colors of edge gadgets can't be repeated in more than one plume of a consistency gadget. This implies that colors from edge gadgets in square paths need to be duplicated in other edge gadgets and not in consistency gadgets.

**Claim 5** *No square path in this coloring goes through a consistency gadget.*

As edge gadgets need to repeat colors from edge gadgets and the coloring of a consistency gadget and the edge gadgets belonging to it is nonrepetitive, the only possibility that hasn't been excluded by the construction are bigger square paths through multiple edge gadgets. So suppose the path  $p$  is a square path going through the consistency gadget  $Q$ . Then, as before, there are several possibilities of how  $p$  looks. Certainly, however,  $p$  has to pass through some edge gadgets  $E_1$  and  $E_2$  before and after going through  $Q$  respectively.

As before, there are a few different ways  $p$  might look. One is, that the path through  $E_1$  in  $p$  is from its  $u_1$  to  $u_0$  vertex or the other way around. To

replicate this,  $p$  has to contain the same path in a different edge gadget of  $Q$ , as shown above. A square of that form has earlier been shown to form a square path in  $Q$  (by replacing the subpaths with single edges), so  $p$  leads to a  $v$  vertex in  $E_1$  and the edge gadget that repeats the path. By claim 5, a path through  $E_1$  and  $Q$  can't be repeated in another edge gadget of  $Q$ , so  $p$  has to be bigger. Thus, it has to go through edge gadgets of  $Q$  twice and two times through  $Q$  itself. At least three of the four times traversing an edge gadget, the path has to be between a  $u_i$  and a  $v$  vertex, as the paths from one  $v_i$  to the other  $v$  contain all seven colors of an edge gadget but edge gadgets from other consistency gadgets don't share all of these. The fourth time, this path can be split so that one part of it is at the beginning and the other at the end of  $p$ .

If  $Q$  corresponds to an existentially quantified variable,  $p$  can have all of its four parts in edge gadgets of  $Q$  in different edge gadgets or two parts in the same edge gadget. In the second case, this would mean that the edge gadget  $E_1$  that contains two parts of  $p$  is connected to two edge gadgets with the same colors. The colors in the edge gadgets neighboring  $E_1$  don't match, however, as the preceding and following color of every edge in  $G$  are distinct, and so  $p$  has to traverse four different edge gadgets. Only up to two of  $Q$ 's edge gadgets are colored the same and none of them are the edge gadget corresponding to the edge in the snout of  $G$ . That means a square path would have to either replicate the coloring of one edge gadget of  $Q$  outside of  $Q$  (which doesn't work due to the unique coloring of the consistency gadgets) or traverse those gadgets in  $Q$  that are colored the same and also two other edge gadgets from the same consistency gadget. As no two edge gadgets of  $Q$  that could be used for this have same edge colors in both preceding and following edges, however, these two gadgets can't be connected by two subpaths through other edge gadgets that are colored identically and thus there is no square path through  $Q$ .

If  $Q$  corresponds to a universally quantified variable  $x_j$ , it's colored so that the colors between  $u$  vertices are never the same, so  $p$  has to end in  $Q$  and therefore begin in an edge gadget. Again, the paths between the  $u$  and  $v$  vertices of the edge gadgets are interesting here. The consistency gadgets corresponding to universally quantified variables have a coloring dependent on the assignment of the variable, so if  $x_j = 0$ , the paths between  $v_0$  and  $u$  in  $N^{x_j}$  and between  $u_0$  and  $v_0$  of  $P^{x_j}$  are colored identically, while if  $x_j = 1$ , the path between  $v_0$  and  $u$  in  $C^{x_j}$  can be matched by  $N^{x_j}$  (between  $v_0$  and  $u$ ) and  $P^{x_j}$  (between  $V - 0$  and  $u_0$ ) and the path from  $v_1$  to  $u$  in  $C^{x_j}$  can also be repeated between  $v_1$  and  $u$  in  $N^{x_j}$ . Just because the paths are colored the same, however, doesn't mean they can be part of the same square path:  $N^{x_j}$  and  $P^{x_j}$  can't repeat a path through the other gadget, as their direction

determining colors specify the same direction. So a second repetition has to contain a part in  $C^{x_j}$ . This part can't go through  $v_0$ , as  $v_0$  lies in the direction of the tip of the snout and after visiting it, no edges farther away from the snout can repeat the direction determining colors of this part. So the only possibility for repetition that's left is the path between  $v_1$  and  $u$  in  $C^{x_j}$ . As the beginning and also end of  $p$  is in edge gadgets of  $Q$ , no other consistency gadgets but  $Q$  can be traversed. That means  $p$  would have to move from  $v_1$  of  $C^{x_j}$  to another edge gadget of  $Q$  only through edge gadgets. From  $C^{x_j}$ , the path would have to traverse an edge gadget belonging to a consistency gadget of the direction determining color  $a_i^1$ ,  $Q^{a_i^1}$ , to get to  $N^{x_j}$ , but this would have to be repeated at the beginning of the snout. As  $p$  has to start in an edge gadget of  $Q$  and can't reach another edge gadget colored the same as the one from  $Q^{a_i^1}$ , there is no possible square path through  $Q$  in this possibility either.

This way, there is no square path in  $H$  that didn't exist in  $G$  - and thus none at all if  $f$  isn't satisfiable. The steps to construct  $H$  from  $G$  are of course polynomial, so  $\text{co-}\forall\exists\text{SAT}$  can be reduced to Thue Number, making the Thue Number problem  $\Sigma_2^P$  - *hard*.

**Claim 6** *Thue Number is in  $\Sigma_2^P$ .*

Using an oracle for the nonrepetitive coloring problem for edge colorings, a Turing machine can non-deterministically guess a coloring of a graph and then get the decision whether it's repetitive or not from the oracle in  $O(|E|)$  steps. So the Thue Number problem is in  $\Sigma_2^P$  and thus  $\Sigma_2^P$  - *complete*.

## 5 The Thue chromatic number

As there is no general proof of complexity but for the Thue chromatic Number but rather many smaller results regarding bounds on colors or specific graphs, its section in this thesis is to be thought of as a snippet of these results. Most of them won't be explained very thoroughly because of this, the last part about trees, however, will be more detailed, featuring two proofs for the prominent results. One of them has been taken directly from [13], the other one has been altered as the original proofed more than just the nonrepetitiveness of the coloring.

### 5.1 Algorithms for vertex-colorings

Even though the complexity of the Thue chromatic number is still an open problem, some restricted versions have been found to have an upper bound. A relatively simple restriction is to limit the length of a possible square path.

**Theorem 4** *For any graph  $G(V, E)$  and a given path length  $2k$ , checking  $G$  for a square path of length  $2k$  takes  $k^{O(k)} \cdot |V|^5 \log |V|$  [3].*

This number is based on a randomized algorithm developed to find such paths. The first part,  $k^{O(k)}$ , represents the number of times the randomized part of the algorithm has to "guess" repetitive paths, that is to assign a label from  $\{1, \dots, 2k\}$  to each vertex of  $G$ , until it succeeds. The probability for the assignment to be correct is  $\frac{1}{2k^{2k}}$ , so after trying  $2k^{2k}$  assignments, every square path will have been found, leading to a duration of  $k^{O(k)}$  for this part. The actual checking of the assignment happens in a directed graph  $D(U, A)$  constructed from  $G$ , that has a directed path with a length of  $k$  vertices only if  $G$  has a repetitive path of  $2k$  vertices. The vertices of  $D$  are a subset of  $V \times V$ , so  $D$  has at most  $|V|^2$  vertices.  $D$  is constructed to be acyclic, so its longest directed path can be found in  $O(|A|)$ . As  $D$  has at most  $|V|^2$  vertices, correspondingly it has at most  $|V|^4$  edges, so finding the longest path can be done in  $|V|^4$ . As this has to be done for every possible ending point of the path, so every vertex in  $V$ , the resulting time limit is  $|V|^5$ . The last part,  $\log |V|$ , is the running time of the derandomization of the algorithm.

A special case of bounded-length paths are so called *star-free* colorings. A coloring is star-free if there are no repetitive paths of length four or less. This problem has been proven to be *NP-complete* [14].

An algorithm for constructing nonrepetitive colorings rather than finding repetitive ones was described in [15]. It's an application of the asymmetrical Lovász local lemma which can be used to prove the existence of combinatorial structures.

**Theorem 5** *There is a constant  $c$  so that for every value of  $\epsilon > 0$  there is a Monte Carlo algorithm constructing a nonrepetitive coloring for any given graph  $H$  of maximum degree  $\Delta$  using  $c\Delta^{2+\epsilon}$  colors.*

More specifically, it says that for some set of mutually independent variables  $P = P_1, P_2, \dots, P_n$ , some set of undesirable or "bad" events  $A = A_1, A_2, \dots, A_n$  and any event  $B$ , with both  $A$  and  $B$  determined by a subset of  $P$ , there exists an assignment of probabilities to  $P$  so that the probability for a desirable event is positive if for every element of  $A$ , its probability is smaller or equal to the product of its probability with the counter probabilities of all events in  $B$  that are disjoint from it.

Applied to nonrepetitive graph colorings, bad events would mean the existence of repetitive paths, so the probability for such a bad event  $A$  for a path  $P$  of length  $2i$  with  $i \geq 1$  to be repetitively colored is  $\frac{1}{C^i}$ ,  $C$  being the amount of colors that can be used. As a  $2i$  path can intersect a lot of other paths ( $4ij\Delta^{2j}$  paths of length  $2j$ ),  $A$  shares variables with just as many other events. Using the correct assignment of probabilities and  $(2e^{16} + 1)\Delta^2$ , the condition is fulfilled. That means there is a nonrepetitive coloring or, to put it differently, the probability of no bad events occurring is greater than zero. If the condition also holds for a fixed "slack"  $\epsilon$ , so that  $PR[A_i] \leq x(A_i) \prod_{B \in \Gamma(A_i)} (1 - x(B))$ ,

a Monte Carlo algorithm, a kind of randomized algorithm that may return an incorrect result with some probability, can be used to find one of these good assignments. As the condition depends on the number of colors used, increasing it is the simple solution.

The algorithm works similarly to the one above, sampling a coloring and then verifying if any events in  $A$  occur. The number of resamplings is in  $O(\frac{n}{\epsilon^2} \log \frac{n}{\epsilon^2})$ ,  $\epsilon$  being the constant from before and  $n$  the size of the input, so the number of vertices in the graph in question, in this case. Again similarly to above, breadth-first-search though the graph is used to check whether any paths are colored nonrepetitively. The runtime to find a nonrepetitive vertex-coloring for a graph  $G$  with maximum degree  $\Delta$  with  $c\Delta^{2+\epsilon}$  colors is therefore in  $O(|V| + |E|)$  for the search plus  $O(\frac{|V|}{\epsilon^2} \log \frac{|V|}{\epsilon^2})$  because of the number of samplings that might be needed. The failure probability of the algorithm is also important, of course: it's an inverse polynomial of the size of  $G$  that is arbitrarily small.



## 5.2 Bounds on Planar graphs

A different way to make the problem more manageable, is to look at specific graphs or subgraphs. One type of graph especially interesting is the planar graph.

**Definition 25** *A graph  $G$  is planar if its realization is homeomorphic to  $R^2$ .*

Realization means all vertices of  $G$  are represented by distinct points in  $R^3$  and connected to each other by straight lines corresponding to the edges of  $G$ . If this realization lies in a single plane, the graph is planar [6, p. 21]. For such graphs, a lower bound of  $\sigma(G) \leq 11$  and an upper bound of  $\sigma(G) \geq O(\log(n))$  are known [16].

**Definition 26** *If  $G$  can be drawn so that all vertices are on the boundary of a face, one of the regions  $G$  separates the plane into, then  $G$  is an outerplanar graph [6, p. 36].*

If  $G$  is an outerplanar graph,  $\sigma(G)$  is known to be at most 12 [17]. Even if  $G$  is not an outerplanar graph, there is an upper bound for so called facial nonrepetitive colorings, meaning that every path on the boundary of a face is colored nonrepetitively. For planar graphs, the upper bound is 24 colors [18] and 16 for hamiltonian planar graphs [19].

## 5.3 Colorings of Trees

Trees are another class of graphs for which good bounds on the Thue chromatic number exist. For small trees, only three colors suffice for a nonrepetitive coloring.

**Theorem 6** *Smaller trees with a degree of four or less can be colored non-repetitively with three colors. [13]*

The degree  $rad(T)$  of a tree  $T$  is the maximum distance a vertex  $u$  has to any other vertex. The condition of the maximum distance being four also limits the maximum number of levels to five. Coloring the vertices according to their distance from the root with  $c \in \{0, 1, 2\}$  and  $c_0 = c_3$  as well as  $c_1 = c_4$  will result in a vertex-coloring that is nonrepetitive [13]. As the maximum distance from any vertex is at most four, no open path can be longer than five vertices. A square path would have to either visit two vertices of the same distance from the root directly following one another, visit a vertex of distance zero or one and then one of distance three or four, respectively, or have length four. The two former possibilities can be excluded as they

contradict the structure of trees. So are square paths of length four possible? Such a path  $p$  would have the structure  $p = c_i c_j c_i c_j$ , as two vertices of the same color can't follow each other. This can't work either, though, as  $p$  starts at level  $i$  (or  $i + 2$ ), then proceeds to level  $j$  (or  $j + 2$ , respectively) and now needs to connect to another vertex of either level  $i$  or  $i + 2$ . It's only possible for  $p$  to visit the level it hasn't started on to avoid visiting a vertex twice, so it has now moved through three levels. That means it contains  $c_2$  either as first color (in this case it would have visited vertex one again for vertex three) or the second color, which would mean that now the path would have to go back to the second vertex as  $c_2$  is only present in level two. Therefore a square path of length four can't exist and the coloring is nonrepetitive.

Even if  $T$  has a degree greater than four, it at least has a subgraph which can be colored nonrepetitively with three colors [13]. The whole tree, in this case, is colorable by four colors.

**Theorem 7** *Every tree is nonrepetitively colorable by 4 colors [13].*

To construct a nonrepetitive 4-coloring of a tree, a nonrepetitive word with four characters is needed. To construct it, the fourth character can be inserted after each two characters of a three character word. This will be the color sequence used from the root to the leaves. As Thue constructed an infinite word with three characters, there is also one with four and thus a color sequence for arbitrarily big (but finite) trees. The root is colored by the color represented by the first character, the vertices with distance one from the root by the second and so forth. Using four characters mitigates the problem in the trees above and using a nonrepetitive sequence guarantees it for the whole tree.

To show that this coloring is in fact nonrepetitive, assume there is a square path  $p$  in  $T$ . This path can't only contain vertices with their degree in strictly ascending or descending order, as by construction this would be a nonrepetitive sequence. So  $p$  has to have one part ascending and one descending in degree. Multiple parts of the sort aren't possible as once a path starts from the root towards the leaves, it can't go into the opposite direction without revisiting vertices. One of those parts of the path has to be shorter than the other, as they both share a vertex  $v$  which has the highest degree in  $p$ . So  $p$  has the form  $p = p_1 v p_2$ . Both  $p_1$  and  $p_2$  have to contain more than one vertex, as in the sequence used for the coloring between a character and its repetition, there are at least two other characters. If they do contain more than one color, the coloring in  $p_2$  will not be a repetition but a reversal of the coloring of  $p_1$ . Thus, there is no possible square path in  $T$ .

Trees, therefore, can't only be colored in very few colors, but also efficiently. On the basis of Thue's or any other nonrepetitive sequence, constructing a word with one character more is straightforward. Then, all that's left is to color the vertices and thus this forms an Algorithm that colors any tree nonrepetitively in  $O(|V|)$  steps.

## 6 Other Thue type problems

As mentioned before, graph colorings are not the only area in which the concept of nonrepetitiveness can be applied. An example of this is the real line: a coloring of the real line is nonrepetitive if there are no adjacent intervals that are colored the same. For this, only two colors are needed. One possibility to color the real line nonrepetitively is to color all points  $x$  for which  $\ln|x|$  is rational with one color and every other point with the other color [12]. In number theory, Thue numbers are not directly relevant, but nonrepetitive sequences can be used none the less. Erdős and Selfridges' proof that perfect powers can't be formed products of consecutive integers, e.g., is based upon the examination of square free parts of those sequences.

Of course there are also variations of the definition of nonrepetitiveness. One possibility is to define a stronger form of nonrepetitiveness through anagrams: if the first part of a sequence is a permutation, and therefore an anagram, of the second, it would not be repetitive. It's not strongly-nonrepetitive, though. Trying to solve the problems mentioned before for strong nonrepetitiveness, they become a lot harder. Thue's infinite nonrepetitive number, e.g., isn't strongly-nonrepetitive for only three symbols [12]. Surprisingly, four symbols suffice, although tests on computers had to be used to confirm this because the number of possibilities to take into account grows very large [20].

A Thue type problem on graphs that hasn't been mentioned yet is the Thue choice Number. The Thue choice number is derived from the Thue chromatic Number. Every vertex  $v$  in a graph  $G$  isn't assigned a single but a list of colors. The Thue choice Number  $\sigma_l(G)$  is the smallest number of colors the lists have to contain so that no repetitive paths exist in  $G$ . In a repetitive path on a list coloring, no color in the list of a vertex can be repeated in the next. In some cases, this number can be put in direct relation to  $\sigma(G)$ , for complete graphs  $\sigma_l(K_n) = \sigma(K_n)$ . The lower bounds for planar and outerplanar graphs are the same, too, for Thue choice Number and Thue chromatic Number. In other cases, the two are quite different. Trees for example, have a Thue chromatic Number of at most four while there is no upper bound for the Thue choice Number. For every value in  $\mathbb{R}$ , there is a tree whose Thue choice number is at least as large [21].

## 7 Conclusion

Nonrepetitive graph colorings are arguably the most interesting field the concept of nonrepetitiveness has been applied to. This thesis should have shown some of the challenges as well as the variety of different approaches that were taken to overcome them.

In terms of complexity, the proof of Thue Number being  $\Sigma_2^P$ -complete is the main result in this field. The result had been recognized as true before even though the proof has been lacking, now with a few corrections and amendments it should be clear that the reduction does in fact work and Manin was right. The results in vertex-coloring refer for the most part to bounds on the Thue chromatic Number and not its complexity, but the algorithms presented in 5.1 show that nonrepetitive vertex-colorings can be found and constructed in decent time. Furthermore, the results for planar graphs and trees try to examine the coloring themselves to give upper and lower bounds on the amount of colors needed as well as methods for constructing a non-repetitive vertex-coloring.

Even though remarkable results in both nonrepetitive edge- and vertex-colorings have been achieved, some open questions remain.

In [22], Grytczuk calls the question whether the Thue chromatic Number is bounded for planar graphs the most intriguing question concerning nonrepetitive vertex-colorings. Of course, an upper bound was shown in [16]. That isn't a fixed value, however, which would be far more interesting, similar to the bounds known for outerplanar graphs and trees.

Furthermore, the actual complexity of the Thue chromatic Number is still open. While it might be speculated that it's similar to the Thue Number, the proof of complexity for the Nonrepetitive Coloring problem can't be extended for vertex-colorings the same way Manin has done for edge-colorings. Most researchers, however, seem to focus on probabilistic methods and randomized algorithms that have their own complexity classes.

Lastly, an interesting parallel could be drawn between Thue's original findings and subgraphs: every graph has been shown to have a subgraph that has a nonrepetitive vertex-coloring using only four colors. If this bound could be shown to be three, every graph would have a portion that could be colored in the same way Thue's nonrepetitive word is constructed [23].

## References

- [1] A. Thue, “Über unendliche Zeichenreihen”, *Norske Videnskabers Selskabs Skrifter Mat.-Nat. Kl.*, vol. 7, pp. 1–22, 1906.
- [2] N. Alon, J. Grytczuk, M. Haluszczak, and O. Riordan, “Nonrepetitive colorings of graphs”, *Random Structures and Algorithms*, vol. 21, no. 3–4, pp. 336–346, 2002.
- [3] D. Marx and M. Schaefer, “The complexity of nonrepetitive coloring”, *Discrete Applied Mathematics*, vol. 157, no. 1, pp. 13–18, 2009.
- [4] M. Schaefer, “Guest column: Completeness in the polynomial-time hierarchy: A compendium”, *ACM SIGACT News*, pp. 22–36, 2002.
- [5] F. Manin, “The complexity of nonrepetitive edge coloring of graphs”, 2008. arXiv preprint: 0709.4497v2.
- [6] B. Bollobas, *Modern Graph Theory*. Springer Science & Business Media, 1998, ISBN: 0387984887.
- [7] J. L. Gross, J. Yellen, and P. Zhang, *Handbook of Graph Theory*, 2nd. Chapman and Hall/CRC, 2013, ISBN: 978-1439880180.
- [8] M. D. Davis and E. J. Weyuker, *Computability, complexity and languages*. Academic Press, 1983, ISBN: 0-12-206380-5.
- [9] O. Goldreich, *P, NP, and NP-completeness: The Basics of Computational Complexity*. Cambridge University Press, 2010, ISBN: 9780521122542.
- [10] A. R. Meyer and L. J. Stockmeyer, *The equivalence problem for regular expressions with squaring requires exponential space*, 1972.
- [11] L. J. Stockmeyer, “The polynomial-time hierarchy”, *Theoretical Computer Science*, vol. 3, no. 1, pp. 1–22, 1976.
- [12] J. Grytczuk, “Thue type problems for graphs, points, and numbers”, *Discrete Mathematics*, vol. 308, no. 19, pp. 4419–4429, 2008.
- [13] B. Brešar, J. Grytczuk, S. Klavžar, S. Niwczyk, and I. Peterin, “Non-repetitive colorings of trees”, *Discrete Mathematics*, vol. 307, no. 2, pp. 163–172, 2007.
- [14] T. F. Coleman and J. J. Moré, “Estimation of sparse Hessian matrices and graph coloring problems”, *Math. Program.*, vol. 28, no. 3, pp. 243–270, 1984.
- [15] B. Haeupler, B. Saha, and A. Srinivasan, “New constructive aspects of the Lovasz Local Lemma”, *Proceedings - Annual IEEE Symposium on Foundations of Computer Science, FOCS*, pp. 397–406, 2010.

- [16] V. Dujmović, F. Frati, G. Joret, and D. R. Wood, “Nonrepetitive Colourings of Planar Graphs with  $O(\log n)$  Colours”, pp. 1–11, 2012. arXiv: 1202.1569.
- [17] J. Barát and P. P. Varjú, “On square-free vertex colorings of graphs”, *Studia Scientiarum Mathematicarum Hungarica. A Quarterly of the Hungarian Academy of Sciences*, vol. 44, no. 3, pp. 411–422, 2007.
- [18] J. Barát and J. Czap, “Facial Nonrepetitive Vertex Coloring of Plane Graphs”, *Graph Theory*, vol. 74, no. 1, pp. 115–121, 2013.
- [19] J. Harant and S. Jendrol, “Facial non-repetitive vertex colouring of some families of 2-connected plane graphs”, *IfM Preprint M*, vol. 04, no. 11, 2011.
- [20] V. Keränen, “Abelian Squares are Avoidable on 4 Letters”, *Icalp*, vol. 623, no. 8, pp. 41–52, 1992.
- [21] E. Škrabuľáková, “The Thue choice number versus the Thue chromatic number of graphs”, 2015. arXiv preprint: 1508.02559.
- [22] J. Grytczuk, “Nonrepetitive colorings of graphs - A survey”, *International Journal of Mathematics and Mathematical Sciences*, vol. 2007, pp. 1–11, 2007.
- [23] J. Barát and D. R. Woody, “Notes on nonrepetitive graph colouring”, *Electronic Journal of Combinatorics*, vol. 15, no. 1, pp. 1–13, 2008.

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit im Studiengang Informatik selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Arbeit vorher nicht in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

---