Isomorphism Testing for Graphs Excluding Small Topological Subgraphs*

Daniel Neuen†

Abstract

We give an isomorphism test that runs in time $n^{\text{polylog}(h)}$ on all n-vertex graphs excluding some h-vertex graph as a topological subgraph. Previous results state that isomorphism for such graphs can be tested in time $n^{\text{polylog}(n)}$ (Babai, STOC 2016) and $n^{f(h)}$ for some function f (Grohe and Marx, SIAM J. Comp., 2015).

Our result also unifies and extends previous isomorphism tests for graphs of maximum degree d running in time $n^{\text{polylog}(d)}$ (FOCS 2018) and for graphs of Hadwiger number h running in time $n^{\text{polylog}(h)}$ (FOCS 2020).

1 Introduction

Determining the computational complexity of the Graph Isomorphism Problem (GI) is a long-standing open question in theoretical computer science (see, e.g., [21]). The problem is easily seen to be contained in NP, but it is neither known to be in PTIME nor known to be NP-complete. In a breakthrough result, Babai [1] obtained a quasipolynomial-time algorithm for testing isomorphism of graphs (i.e., an algorithm running in time $n^{\text{polylog}(n)}$ where n denotes the number of vertices of the input graphs), achieving the first improvement over the previous best algorithm running in time $n^{\mathcal{O}(\sqrt{n/\log n})}$ [4] in over three decades. The algorithm advances both group-theoretic techniques for GI, dating back to Luks [25], as well as our understanding of combinatorial approaches such as the Weisfeiler-Leman algorithm (see, e.g., [9, 20]). However, it remains wide open whether GI can be solved in polynomial time.

Polynomial-time algorithms for testing isomorphism are known for various restricted classes of graphs (see, e.g., [3, 7, 12, 18, 19, 26, 27]). One of the most general results in this direction due to Grohe and Marx [13] states that isomorphism can be tested in polynomial time for all graph classes that exclude a fixed graph as a topological subgraph. In particular, this includes previous results solving GI in polynomial time on all graphs excluding a fixed graph as a minor [31] and all graphs of bounded maximum degree [25].

A common feature of all these algorithms is that the exponent of the running time depends at least linearly on the parameter in question. In light of Babai's quasipolynomial-time algorithm it seems natural to ask for which graph parameters k the Graph Isomorphism Problem can be solved in time $n^{\text{polylog}(k)}$. The first step towards answering this question was achieved by Grohe, Schweitzer and the present author [15] presenting an isomorphism algorithm running in time $n^{\text{polylog}(d)}$ where n denotes the number of vertices and d the maximum degree of the input graphs. Further extending the group-theoretic advances of Babai's algorithm, subsequent work resulted in algorithms testing isomorphism in time $n^{\text{polylog}(k)}$ for graphs of tree-width k [36] and time $n^{\text{polylog}(g)}$ for graphs of Euler genus g [29]. Both of these results were generalized in [17] obtaining an isomorphism test running in time $n^{\text{polylog}(h)}$ for n-vertex graphs excluding an arbitrary h-vertex graph as a minor. For a recent survey, I also refer to [14]. In this work we further generalize these results to all graphs that exclude an arbitrary h-vertex graph as a topological subgraph.

Recall that a graph H is a topological subgraph of a graph G if H can be obtained from G by deleting vertices and edges as well as dissolving degree two vertices (i.e., deleting a vertex of degree two and connecting its two neighbors). A graph G excludes H as a topological subgraph if G has no topological subgraph that is isomorphic to H. For example, by Kuratowski's Theorem, planar graphs can be characterized by excluding K_5 and $K_{3,3}$ as topological subgraphs. Note that, whenever a graph H is a topological subgraph of G, then H is also a minor of G. Hence, any graph class that excludes some graph H as a minor in particular excludes H as a topological subgraph. As another observation, the maximum degree of a topological subgraph H of G is at most

*Research supported by the European Research Council (ERC) consolidator grant No. 725978 SYSTEMATICGRAPH. The full version of the paper can be accessed at `https://arxiv.org/abs/2011.14730`.

†CISPA Helmholtz Center for Information Security, Saarbrücken, Germany.

Copyright © 2022
Copyright for this paper is retained by authors

the maximum of degree of G. Thus, any graph of maximum degree d excludes the complete graph K_{d+2} on $d+2$ vertices as a topological subgraph.

The main result of this work is a new isomorphism algorithm for graphs that exclude some h-vertex graph as a topological subgraph which significantly improves the previous best algorithm due to Grohe and Marx [13].

THEOREM 1.1. *The Graph Isomorphism Problem for graphs excluding some h-vertex graph as a topological subgraph can be solved in time $n^{\text{polylog}(h)}$.*

By the comments above, this result also unifies and extends previous isomorphism tests for graphs of maximum degree d running in time $n^{\text{polylog}(d)}$ [15] and for graphs of Hadwiger number h (i.e., the maximum h such that K_h is a minor of the input graph) running in time $n^{\text{polylog}(h)}$ [17].

Observe that a graph G excludes some h-vertex graph as a topological subgraph if and only if it excludes K_h, the complete graph on h vertices, as a topological subgraph. Hence, in the following, we restrict our attention to graphs that exclude K_h as a topological subgraph.

On a high level, the algorithm follows the same decomposition strategy that is already used by Grohe et al. in [17] for testing isomorphism of graphs excluding K_h as a minor. The main idea is to decompose an input graph G into parts $D \subseteq V(G)$ such that the interplay between the parts is simple, and G restricted to D is *t-CR-bounded* for some number t that is polynomially bounded in h. Intuitively speaking, a graph G is *t-CR-bounded* (where $t \in \mathbb{N}$) if an initially uniform coloring can be transformed into a discrete coloring (i.e., a coloring where every vertex has its own color) by repeatedly applying the standard Color Refinement algorithm and splitting color classes of size at most t. Building on the group-theoretic isomorphism machinery dating back to Luks [25] as well as recent extensions [29], isomorphism of t-CR-bounded graphs can be decided in time $n^{\text{polylog}(t)}$ (where n denotes the number of vertices of the input graphs).

In order to decompose the input graph G into suitable parts $D \subseteq V(G)$, Grohe et al. [17] introduce the *t-closure* $\text{cl}_t^G(X)$ of a set $X \subseteq V(G)$ as the set of all uniquely colored vertices after artificially individualizing all vertices from X and applying the t-CR procedure. Now, the central idea is to define $D \coloneqq \text{cl}_t^G(X)$ for a suitable set X (we will refer to X as the *initial set*). In order for this approach to work out, the following two statements are crucial where t is some number polynomially bounded in h.

(A) For every $X \subseteq V(G)$ it holds that $|N_G(Z)| < h$ for every connected component of $G - D$ where $D \coloneqq \text{cl}_t^G(X)$.

(B) There is a polynomial-time algorithm that computes an isomorphism-invariant initial set $X \subseteq V(G)$ such that $X \subseteq \text{cl}_t^G(v)$ for every $v \in X$.

Assuming both statements are true, one can build an isomorphism test for graphs G_1 and G_2 as follows. First, we compute sets $X_1 \subseteq V(G_1)$ and $X_2 \subseteq V(G_2)$ using Property (B) and define $D_i \coloneqq \text{cl}_t^{G_i}(X_i)$ for both $i \in \{1,2\}$. Afterwards, we recursively compute isomorphisms between all pairs of connected components of $G_1 - D_1$ and $G_2 - D_2$. By Property (B), G_i is t-CR-bounded on D_i after individualizing a single vertex $v \in X_i$. Hence, isomorphism between $G_1[D_1]$ and $G_2[D_2]$ can be decided in time $n^{\text{polylog}(t)} = n^{\text{polylog}(h)}$. Also, using the techniques from [16, 36] and Property (A), it is possible to incorporate the partial isomorphisms between the connected components of $G_1 - D_1$ and $G_2 - D_2$, overall resulting in an isomorphism test between G_1 and G_2 running in time $n^{\text{polylog}(h)}$. We remark that, for this strategy to work out, it is crucial to define D_i in an isomorphism-invariant manner as to not compare two graphs that are decomposed in structurally different ways. Observe that D_i is indeed defined in an isomorphism-invariant way since the initial set X_i is by Property (B).

While Grohe et al. [17] already prove Property (A) for graphs excluding K_h as a topological subgraph, their proof of Property (B) crucially requires closure under taking minors. More precisely, for Property (B), Grohe et al. provide a complicated algorithm that essentially contracts certain parts of the input graph G, and then builds the set X from a solution X' computed by recursion for the contracted graph G'. Unfortunately, such a strategy is infeasible for graphs excluding K_h as a topological subgraph since the corresponding class of graphs is not closed under taking minors.

The main technical contribution of this work is to provide an alternative algorithm for Property (B) that works for all graphs G excluding K_h as topological subgraph. Indeed, our algorithm is much simpler than the one from [17] and solely relies on the well-known Weisfeiler-Leman algorithm (see, e.g., [9, 20]). The Weisfeiler-Leman algorithm is a standard tool in the context of isomorphism testing and computes an isomorphism-invariant coloring of k-tuples of vertices of a graph G. Our main technical result is that the 3-dimensional Weisfeiler-Leman algorithm is able to provide a suitable initial set X.

Copyright © 2022
Copyright for this paper is retained by authors

THEOREM 1.2. *Let G be a connected graph that excludes K_h as a topological subgraph and let $t = \Omega(h^5)$. Then there is a color $c_0 \in \{\chi^3_{\mathsf{WL}}[G](v,v,v) \mid v \in V(G)\}$ such that, for $\chi(v,w) \coloneqq \chi^3_{\mathsf{WL}}[G](v,w,w)$ for all $v,w \in V(G)$ and $X \coloneqq \{v \in V(G) \mid \chi^3_{\mathsf{WL}}[G](v,v,v) = c_0\}$, it holds that*

$$X \subseteq \mathrm{cl}_t^{(G,\chi)}(v)$$

for all $v \in X$.

Here, $\chi^3_{\mathsf{WL}}[G]\colon (V(G))^3 \to C$ denotes the coloring of 3-tuples computed by the 3-dimensional Weisfeiler-Leman algorithm. We remark that the t-closure needs to be taken over a colored version of G which, however, does not pose any problems for the final algorithm.

Observe that Theorem 1.2 implies Property (B) since the coloring $\chi^3_{\mathsf{WL}}[G]$ can be computed in polynomial time and we can simply try all possible colors $c_0 \in \{\chi^3_{\mathsf{WL}}[G](v,v,v) \mid v \in V(G)\}$ to find a good set X (if more than one color yields a good set X, we simply take the smallest color according to some fixed linear order on the colors).

The proof of Theorem 1.2 builds on a lengthy and technical analysis of the coloring χ. As a main tool for the proof, we introduce the *t-closure graph of* (G,χ) which is a directed graph H on the same vertex set as G and with an edge (v,w) for all $v,w \in V(G)$ such that $w \in \mathrm{cl}_t^{(G,\chi)}(v)$. Building on properties of the 3-dimensional Weisfeiler-Leman algorithm, we show that it is possible to choose $X \coloneqq \{v \in V(G) \mid \chi^3_{\mathsf{WL}}[G](v,v,v) = c_0\}$ in such a way that X only contains vertices appearing in maximal strongly connected components of H (a strongly connected component of H is maximal if it has no outgoing edges). This implies that $\mathrm{cl}_t^{(G,\chi)}(v)$ and $\mathrm{cl}_t^{(G,\chi)}(v')$ are either disjoint or equal for all $v,v' \in X$, since $\mathrm{cl}_t^{(G,\chi)}(v)$ is exactly the strongly connected component of H that contains v. Assuming there are $v,v' \in X$ such that $\mathrm{cl}_t^{(G,\chi)}(v)$ and $\mathrm{cl}_t^{(G,\chi)}(v')$ are disjoint, we proceed by constructing a large number of pairwise internally vertex-disjoint paths between such sets leading to a topological subgraph of G with high edge-density. Eventually, this contradicts the fact that the average degree of every topological subgraph of G is bounded by a polynomial function in h [8, 24].

Besides the isomorphism test for graphs excluding K_h as a topological subgraph, the algorithmic approach taken in this paper also provides some structural insights into the automorphism groups of graphs without a topological subgraph isomorphic to K_h. We show that every such graph G admits a tree decomposition of adhesion width at most $h-1$ (i.e., the intersection between any two bags has size at most $h-1$) such that the automorphism group of G restricted to a single bag is similar to those of graphs of bounded maximum degree. More precisely, for $d \geq 1$, let $\widehat{\Gamma}_d$ denote the class of groups Γ such that every composition factor of Γ is isomorphic to a subgroup of S_d (the symmetric group on d points). It is well-known that the automorphism group of every connected graph G of maximum degree d is in the class $\widehat{\Gamma}_d$ after individualizing a single vertex of G [25]. We obtain the following structural insights on the automorphism group $\mathrm{Aut}(G)$ of a graph G excluding K_h as a topological subgraph.

THEOREM 1.3. *Let G be a graph that excludes K_h as a topological subgraph. Then there is an isomorphism-invariant tree decomposition (T,β) of G such that*

1. *the adhesion-width of (T,β) is at most $h-1$, and*
2. *for every $t \in V(T)$ there is some $v \in \beta(t)$ such that $(\mathrm{Aut}(G))_v[\beta(t)] \in \widehat{\Gamma}_d$ for some $d = \mathcal{O}(h^5)$.*

Here, $(\mathrm{Aut}(G))_v[\beta(t)]$ denotes the restriction of $\mathrm{Aut}(G)$ to the bag $\beta(t)$ after individualizing the vertex v.

The remainder of this work is structured as follows. In the next section we give the necessary preliminaries. Afterwards, we define t-CR-bounded graphs and the corresponding closure operator in Section 3 and provide a more detailed overview on the main algorithm. In Section 4 we state the main technical theorem of this work which provides a suitable initial set X and give a detailed overview on its proof. Finally, we state the main results in Section 5.

2 Preliminaries

2.1 Graphs A *graph* is a pair $G = (V(G), E(G))$ consisting of a *vertex set* $V(G)$ and an *edge set* $E(G)$. All graphs considered in this paper are finite and simple (i.e., they contain no loops or multiple edges). Moreover,

Copyright © 2022
Copyright for this paper is retained by authors

unless explicitly other stated, all graphs are undirected. For an undirected graph G and $v, w \in V(G)$, we write vw as a shorthand for $\{v, w\} \in E(G)$. The *neighborhood* of a vertex $v \in V(G)$ is denoted by $N_G(v)$. The *degree* of v, denoted by $\deg_G(v)$, is the number of edges incident with v, i.e., $\deg_G(v) = |N_G(v)|$. For $X \subseteq V(G)$, we define $N_G[X] := X \cup \bigcup_{v \in X} N_G(v)$ and $N_G(X) := N_G[X] \setminus X$. If the graph G is clear from context, we usually omit the index and simply write $N(v)$, $\deg(v)$, $N[X]$ and $N(X)$.

We write K_n to denote the complete graph on n vertices. A graph is *regular* if every vertex has the same degree. A bipartite graph $G = (V_1, V_2, E)$ is called (d_1, d_2)-*biregular* if all vertices $v_i \in V_i$ have degree d_i for both $i \in \{1, 2\}$. In this case $d_1 \cdot |V_1| = d_2 \cdot |V_2| = |E|$. By a double edge counting argument, for each subset $S \subseteq V_i$, $i \in \{1, 2\}$, it holds that $|S| \cdot d_i \leq |N_G(S)| \cdot d_{3-i}$. A bipartite graph is *biregular*, if there are $d_1, d_2 \in \mathbb{N}$ such that G is (d_1, d_2)-biregular. Each biregular graph satisfies the Hall condition, i.e., for all $S \subseteq V_1$ it holds $|S| \leq |N_G(S)|$ (assuming $|V_1| \leq |V_2|$). Thus, by Hall's Marriage Theorem, each biregular graph contains a matching of size $\min(|V_1|, |V_2|)$.

A *path of length k* from v to w is a sequence of distinct vertices $v = u_0, u_1, \ldots, u_k = w$ such that $u_{i-1}u_i \in E(G)$ for all $i \in [k] := \{1, \ldots, k\}$. The *distance* between two vertices $v, w \in V(G)$, denoted by $\mathrm{dist}_G(v, w)$, is the length of a shortest path between v and w. As before, we omit the index G it it is clear from context. For two sets $A, B \subseteq V(G)$, we denote by $E_G(A, B) := \{vw \in E(G) \mid v \in A, w \in B\}$. Also, $G[A, B]$ denotes the graph with vertex set $A \cup B$ and edge set $E_G(A, B)$. We write $G[A] := G[A, A]$ to denote the induced subgraph on vertex set A. Also, we denote by $G - A$ the subgraph induced by the complement of A, that is, the graph $G - A := G[V(G) \setminus A]$. A graph H is a *subgraph* of G, denoted by $H \subseteq G$, if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. A set $S \subseteq V(G)$ is a *separator* of G if $G - S$ has more connected components than G. A *k-separator* of G is a separator of G of size k.

An *isomorphism* from G to a graph H is a bijection $\varphi\colon V(G) \to V(H)$ that respects the edge relation, that is, for all $v, w \in V(G)$, it holds that $vw \in E(G)$ if and only if $\varphi(v)\varphi(w) \in E(H)$. Two graphs G and H are *isomorphic*, written $G \cong H$, if there is an isomorphism from G to H. We write $\varphi\colon G \cong H$ to denote that φ is an isomorphism from G to H. Also, $\mathrm{Iso}(G, H)$ denotes the set of all isomorphisms from G to H. The automorphism group of G is $\mathrm{Aut}(G) := \mathrm{Iso}(G, G)$. Observe that, if $\mathrm{Iso}(G, H) \neq \emptyset$, it holds that $\mathrm{Iso}(G, H) = \mathrm{Aut}(G)\varphi := \{\gamma\varphi \mid \gamma \in \mathrm{Aut}(G)\}$ for every isomorphism $\varphi \in \mathrm{Iso}(G, H)$.

A *vertex-colored graph* is a tuple (G, χ_V) where G is a graph and $\chi_V\colon V(G) \to C$ is a mapping into some set C of colors, called *vertex-coloring*. Similarly, an *arc-colored graph* is a tuple (G, χ_E), where G is a graph and $\chi_E\colon \{(u, v) \mid \{u, v\} \in E(G)\} \to C$ is a mapping into some color set C, called *arc-coloring*. Observe that colors are assigned to directed edges, i.e., the directed edge (v, w) may obtain a different color than (w, v). We also consider vertex- and arc-colored graphs (G, χ_V, χ_E) where χ_V is a vertex-coloring and χ_E is an arc-coloring. Also, a *pair-colored graph* is a tuple (G, χ_P), where G is a graph and $\chi_P\colon (V(G))^2 \to C$ is a mapping into some color set C. Typically, C is chosen to be an initial segment $[n]$ of the natural numbers. To be more precise, we generally assume that there is a total order on the set of all potential colors which, for example, allows us to identify a minimal color appearing in a graph in a unique way. Isomorphisms between vertex-, arc- and pair-colored graphs have to respect the colors of the vertices, arcs and pairs.

2.2 Topological Subgraphs A graph H is a *topological subgraph* of G if H can be obtained from G by deleting edges, deleting vertices and dissolving degree 2 vertices (which means deleting the vertex and making its two neighbors adjacent). More formally, we say that H is a topological subgraph of G if a subdivision of H is a subgraph of G (a subdivision of a graph H is obtained by replacing each edge of H by a path of length at least 1). The following theorem states the well-known fact that graphs excluding a topological subgraph have bounded average degree.

THEOREM 2.1. ([8, 24]) *There is an absolute constant $a_{\mathsf{deg}} \geq 1$ such that for every $h \geq 1$ and every graph G that excludes K_h as a topological subgraph, it holds that*

$$\frac{1}{|V(G)|} \sum_{v \in V(G)} \deg_G(v) \leq a_{\mathsf{deg}} h^2.$$

In the remainder of this work, we always use a_{deg} to refer to the constant from the theorem above.

2.3 Weisfeiler-Leman Algorithm The Weisfeiler-Leman algorithm, originally introduced by Weisfeiler and Leman in its 2-dimensional version [35], forms one of the most fundamental subroutines in the context of

Copyright © 2022
Copyright for this paper is retained by authors

isomorphism testing.

Let $\chi_1, \chi_2 \colon V^k \to C$ be colorings of the k-tuples of vertices of G, where C is some finite set of colors. We say χ_1 *refines* χ_2, denoted $\chi_1 \preceq \chi_2$, if $\chi_1(\bar{v}) = \chi_1(\bar{w})$ implies $\chi_2(\bar{v}) = \chi_2(\bar{w})$ for all $\bar{v}, \bar{w} \in V^k$. The colorings χ_1 and χ_2 are *equivalent*, denoted $\chi_1 \equiv \chi_2$, if $\chi_1 \preceq \chi_2$ and $\chi_2 \preceq \chi_1$.

The *Color Refinement algorithm* (i.e., the 1-dimensional Weisfeiler-Leman algorithm) is a procedure that, given a graph G, iteratively computes an isomorphism-invariant coloring of the vertices of G. In this work, we actually require an extension of the Color Refinement algorithm that apart from vertex-colors also takes arc-colors into account. For a vertex- and arc-colored graph (G, χ_V, χ_E) define $\chi_{(0)}[G] \coloneqq \chi_V$ to be the initial coloring for the algorithm. This coloring is iteratively refined by defining $\chi_{(i+1)}[G](v) \coloneqq (\chi_{(i)}[G](v), \mathcal{M}_i(v))$ where

$$\mathcal{M}_i(v) \coloneqq \{\!\!\{ \big(\chi_{(i)}[G](w), \chi_E(v,w), \chi_E(w,v)\big) \mid w \in N_G(v) \}\!\!\}$$

(and $\{\!\!\{ \dots \}\!\!\}$ denotes a multiset). By definition, $\chi_{(i+1)}[G] \preceq \chi_{(i)}[G]$ for all $i \geq 0$. Hence, there is a minimal value i_∞ such that $\chi_{(i_\infty)}[G] \equiv \chi_{(i_\infty+1)}[G]$. We define $\chi_{\mathsf{CR}}[G] \coloneqq \chi_{(i_\infty)}[G]$. The Color Refinement algorithm takes as input a vertex- and arc-colored graph (G, χ_V, χ_E) and returns (a coloring that is equivalent to) $\chi_{\mathsf{CR}}[G]$. The procedure can be implemented in time $\mathcal{O}((m+n)\log n)$ (see, e.g., [6]).

Next, we describe the *k-dimensional Weisfeiler-Leman algorithm* (k-WL) for all $k \geq 2$. For an input graph G let $\chi^k_{(0)}[G] \colon (V(G))^k \to C$ be the coloring where each tuple is colored with the isomorphism type of its underlying ordered subgraph. More precisely, $\chi^k_{(0)}[G](v_1, \dots, v_k) = \chi^k_{(0)}[G](v'_1, \dots, v'_k)$ if and only if, for all $i, j \in [k]$, it holds that $v_i = v_j \Leftrightarrow v'_i = v'_j$ and $v_iv_j \in E(G) \Leftrightarrow v'_iv'_j \in E(G)$. If the graph is equipped with a coloring the initial coloring $\chi^k_{(0)}[G]$ also takes the input coloring into account. More precisely, for a vertex-coloring χ_V, it additionally holds that $\chi_V(v_i) = \chi_V(v'_i)$ for all $i \in [k]$. For an arc-coloring χ_E, it is the case that $\chi_E(v_i, v_j) = \chi_E(v'_i, v'_j)$ for all $i, j \in [k]$ such that $v_iv_j \in E(G)$. Finally, for a pair coloring χ_P, it holds that $\chi_P(v_i, v_j) = \chi_P(v'_i, v'_j)$ is additionally satisfied for all $i, j \in [k]$.

We then recursively define the coloring $\chi^k_{(i)}[G]$ obtained after i rounds of the algorithm. For $\bar{v} = (v_1, \dots, v_k) \in (V(G))^k$ let

$$\chi^k_{(i+1)}[G](\bar{v}) \coloneqq \Big(\chi^k_{(i)}[G](\bar{v}), \mathcal{M}_i(\bar{v})\Big)$$

where

$$\mathcal{M}_i(\bar{v}) \coloneqq \Big\{\!\!\Big\{ \big(\chi^k_{(i)}[G](\bar{v}[w/1]), \dots, \chi^k_{(i)}[G](\bar{v}[w/k])\big) \Bigm| w \in V(G) \Big\}\!\!\Big\}$$

and $\bar{v}[w/i] \coloneqq (v_1, \dots, v_{i-1}, w, v_{i+1}, \dots, v_k)$ is the tuple obtained from $\bar{v}$ by replacing the i-th entry by w. Again, there is a minimal i_∞ such that $\chi^k_{(i_\infty)}[G] \equiv \chi^k_{(i_\infty+1)}[G]$ and for this i_∞ we define $\chi^k_{\mathsf{WL}}[G] \coloneqq \chi^k_{(i_\infty)}[G]$.

The k-dimensional Weisfeiler-Leman algorithm takes as input a (vertex-, arc- or pair-)colored graph G and returns (a coloring that is equivalent to) $\chi^k_{\mathsf{WL}}[G]$. This can be implemented in time $\mathcal{O}(n^{k+1}\log n)$ (see [20]).

Let G be a graph. Let $k \geq 1$ and let $\chi \colon (V(G))^k \to C$ be a coloring of k-tuples. We say that χ is *k-stable on G* if $\chi \preceq \chi^k_{(0)}[G]$ and χ is not refined by applying one round of the k-dimensional Weisfeiler-Leman algorithm (the 1-dimensional Weisfeiler-Leman algorithm is defined as the Color Refinement algorithm). In particular, $\chi^k_{\mathsf{WL}}[G]$ is k-stable on G. The following facts are well-known.

FACT 2.1. *Let G be a graph and let $1 \leq \ell \leq k$. Also, define*

$$\chi(v_1, \dots, v_\ell) \coloneqq \chi^k_{\mathsf{WL}}[G](v_1, \dots, v_\ell, v_\ell, \dots, v_\ell)$$

for all $v_1, \dots, v_\ell \in V(G)$. Then χ is ℓ-stable on G.

FACT 2.2. *Let G be a graph and let $1 \leq \ell < k$. Let $w \in V(G)$ and define*

$$\chi(v_1, \dots, v_\ell) \coloneqq \chi^k_{\mathsf{WL}}[G](w, v_1, \dots, v_\ell, v_\ell, \dots, v_\ell)$$

for all $v_1, \dots, v_\ell \in V(G)$. Then χ is ℓ-stable on (G, χ_w) where χ_w is the vertex-coloring defined via $\chi_w(w) \coloneqq 1$ and $\chi_w(v) \coloneqq 0$ for all $v \in V(G) \setminus \{w\}$.

Copyright © 2022
Copyright for this paper is retained by authors

2.4 Permutation Groups We also require some basic notation from group theory. For further background we refer to [11, 32].

A *permutation group* acting on a set Ω is a subgroup $\Gamma \leq \mathrm{Sym}(\Omega)$ of the symmetric group. The size of the permutation domain Ω is called the *degree* of Γ. For $\gamma \in \Gamma$ and $\alpha \in \Omega$ we denote by α^γ the image of α under the permutation γ. For $\alpha \in \Omega$ the group $\Gamma_\alpha = \{\gamma \in \Gamma \mid \alpha^\gamma = \alpha\} \leq \Gamma$ is the *stabilizer* of α in Γ. A set $A \subseteq \Omega$ is Γ-*invariant* if $A = A^\gamma \coloneqq \{\alpha^\gamma \mid \alpha \in A\}$ for all $\gamma \in \Gamma$.

For $A \subseteq \Omega$ and a bijection $\theta \colon \Omega \to \Omega'$ we denote by $\theta[A]$ the restriction of θ to the domain A. For a Γ-invariant set $A \subseteq \Omega$, we denote by $\Gamma[A] \coloneqq \{\gamma[A] \mid \gamma \in \Gamma\}$ the induced action of Γ on A, i.e., the group obtained from Γ by restricting all permutations to A. More generally, for every set Λ of bijections with domain Ω, we denote by $\Lambda[A] \coloneqq \{\theta[A] \mid \theta \in \Lambda\}$.

In order to perform computational tasks for permutation groups efficiently the groups are typically represented by generating sets of small size (see [34] for more details). Indeed, most algorithms are based on so-called strong generating sets, which can be chosen of size quadratic in the degree of the group and can be computed in polynomial time given an arbitrary generating set (see, e.g., [34]).

In this work we are interested in permutation groups with restricted composition factors. Let Γ be a group. A *subnormal series* is a sequence of subgroups $\Gamma = \Gamma_0 \trianglerighteq \Gamma_1 \trianglerighteq \cdots \trianglerighteq \Gamma_k = \{\mathrm{id}\}$. The length of the series is k and the groups Γ_{i-1}/Γ_i are the factor groups of the series, $i \in [k]$. A *composition series* is a strictly decreasing subnormal series of maximal length. For every finite group Γ all composition series have the same family (considered as a multiset) of factor groups (cf. [32]). A *composition factor* of a finite group Γ is a factor group of a composition series of Γ.

Definition 2.1. *For $d \geq 2$ let $\widehat{\Gamma}_d$ denote the class of all groups Γ for which every composition factor of Γ is isomorphic to a subgroup of S_d.*

3 Allowing Color Refinement to Split Small Color Classes

In the following, we provide a more detailed overview for the main algorithm testing isomorphism of graphs excluding K_h as a topological subgraph. On a high-level, the algorithm builds on a decomposition strategy. Let G_1 and G_2 denote the two input graphs. By testing isomorphisms of connected components separately, we may assume without loss of generality that G_1 and G_2 are connected. The algorithm aims at finding isomorphism-invariant sets D_1 and D_2 and recursively computes isomorphisms between all pairs of components of $G_1 - D_1$ and $G_2 - D_2$. In order to combine the partial isomorphisms between connected components of $G_1 - D_1$ and $G_2 - D_2$ into full isomorphisms between G_1 and G_2 the algorithm builds on the various existing group-theoretic tools from [17, 29, 36]. Here, one of the crucial properties is that $|N_{G_i}(Z)|$ is polynomially bounded in h for every connected component Z of $G_i - D_i$. To test isomorphisms between $G_1[D_1]$ and $G_2[D_2]$, the algorithms builds on the notion of t-CR-bounded graphs originally introduced by Ponomarenko [30].

Intuitively speaking, a (vertex- and arc-colored) graph G is t-CR-bounded, $t \in \mathbb{N}$, if the vertex coloring of G can be turned into the discrete coloring (i.e., each vertex has its own color) by repeatedly

- performing the Color Refinement algorithm (expressed by the letters 'CR'), and
- taking a color class $[v]_\chi \coloneqq \{w \in V(G) \mid \chi(w) = \chi(v)\}$ of size $|[v]_\chi| \leq t$ and assigning each vertex from the class its own color.

The next definition formalizes this intuition.

Definition 3.1. *A vertex- and arc-colored graph $G = (V, E, \chi_V, \chi_E)$ is* t-CR-bounded *if the sequence $(\chi_i)_{i\geq 0}$ reaches a discrete coloring where $\chi_0 \coloneqq \chi_V$,*

$$\chi_{2i+1} \coloneqq \chi^1_{\mathsf{WL}}[V, E, \chi_{2i}, \chi_E]$$

and

$$\chi_{2i+2}(v) \coloneqq \begin{cases} (v, 1) & \text{if } |[v]_{\chi_{2i+1}}| \leq t \\ (\chi_{2i+1}(v), 0) & \text{otherwise} \end{cases}$$

for all $i \geq 0$.

Copyright © 2022
Copyright for this paper is retained by authors

Also, for the minimal $i_\infty \geq 0$ such that $\chi_{i_\infty} \equiv \chi_{i_\infty+1}$, we refer to χ_{i_∞} as the t-CR-stable *coloring of G and denote it by $\chi_{t\text{-}\mathsf{CR}}[G]$.*

Assuming $G_1[D_1]$ and $G_2[D_2]$ are t-CR-bounded, isomorphisms between the subgraphs can be found in time $n^{\mathrm{polylog}(t)}$ building on the group-theoretic graph isomorphism machinery [29]. Also, using the tools from [36], one can incorporate the partial isomorphisms between connected components of $G_1 - D_1$ and $G_2 - D_2$, assuming $|N_{G_i}(Z)|$ is polynomially bounded in h for every connected component Z of $G_i - D_i$. Hence, the main task is to find suitable sets D_1 and D_2. Here, we follow the same strategy as in [17] and rely on a closure operator associated with t-CR-bounded graphs. Let G be a graph and let $X \subseteq V(G)$ be a set of vertices. Let $\chi_V^*\colon V(G) \to C$ be the vertex-coloring obtained from individualizing all vertices in the set X, i.e., $\chi_V^*(v) := (v, 1)$ for $v \in X$ and $\chi_V^*(v) := (0, 0)$ for $v \in V(G) \setminus X$. Let $\chi := \chi_{t\text{-}\mathsf{CR}}[G, \chi_V^*]$ denote the t-CR-stable coloring with respect to the input graph (G, χ_V^*). We define the *t-closure* of the set X (with respect to G) to be the set

$$\mathrm{cl}_t^G(X) := \{v \in V(G) \mid |[v]_\chi| = 1\}\,.$$

Observe that $X \subseteq \mathrm{cl}_t^G(X)$. For $v_1, \dots, v_\ell \in V(G)$ we also use $\mathrm{cl}_t^G(v_1, \dots, v_\ell)$ as a shorthand for $\mathrm{cl}_t^G(\{v_1, \dots, v_\ell\})$. If the input graph is equipped with a vertex- or arc-coloring, all definitions are extended in the natural way.

We also build the t-closure for vertex sets over pair-colored graphs (G, χ). Let n denote the number of vertices of G. We define a vertex- and arc-coloring $\widetilde{\chi}_V$ and $\widetilde{\chi}_E$ of the complete graph K_n by $\widetilde{\chi}_V(v) := \chi(v, v)$ and $\widetilde{\chi}_E(v, w) := (1, \chi(v, w))$ for all $vw \in E(G)$ and $\widetilde{\chi}_E(v, w) := (0, \chi(v, w))$ for all $v, w \in V(G)$ such that $v \neq w$ and $vw \notin E(G)$. Now, for $X \subseteq V(G)$ let

$$\mathrm{cl}_t^{(G,\chi)}(X) := \mathrm{cl}_t^{(K_n, \widetilde{\chi}_V, \widetilde{\chi}_E)}(X).$$

As before, for $v_1, \dots, v_\ell \in V(G)$, we use $\mathrm{cl}_t^{(G,\chi)}(v_1, \dots, v_\ell)$ as a shorthand for $\mathrm{cl}_t^{(G,\chi)}(\{v_1, \dots, v_\ell\})$.

Now, following the general strategy outlined above, the goal is to compute suitable isomorphism-invariant sets $X_i \subseteq V(G_i)$ and define $D_i := \mathrm{cl}_t^{G_i}(X_i)$ for both $i \in \{1, 2\}$ (for some suitable choice of t). We refer to the sets X_1 and X_2 as the *initial sets*. More precisely, the algorithms aims at finding isomorphism-invariant initial sets $X_i \subseteq V(G_i)$ and pair-colorings $\chi_i\colon (V(G_i))^2 \to C$ such that the following properties hold for the some $t \in \mathbb{N}$ that is polynomially bounded in h.

(A) $|N_{G_i}(Z)| < h$ for every connected component of $G_i - D_i$ where $D_i := \mathrm{cl}_t^{(G_i,\chi_i)}(X_i)$, and

(B) $X_i \subseteq \mathrm{cl}_t^{(G_i,\chi_i)}(v)$ for every $v \in X_i$.

The second property implies that $D_i = \mathrm{cl}_t^{(G_i,\chi_i)}(v)$ for every $v \in X_i$. This allows us to test isomorphisms between $G_1[D_1]$ and $G_2[D_2]$ using the group-theoretic graph isomorphism machinery after individualizing a single vertex. Also, as already discussed above, the first property allows to incorporate the partial isomorphisms between connected components of $G_1 - D_1$ and $G_2 - D_2$ using tools from [16, 36].

Now, Property (A) has already been proved in [17] for graphs excluding K_h as a topological subgraph.

Theorem 3.1. ([17, Theorem IV.1]) *Let G be a graph that excludes K_h as a topological subgraph and let $X \subseteq V(G)$. Let $t \geq 3h^3$ and define $D := \mathrm{cl}_t^G(X)$. Let Z be the vertex set of a connected component of $G - D$. Then $|N_G(Z)| < h$.*

Hence, the remaining task is to find suitable isomorphism-invariant initial sets X_1 and X_2 that satisfy Property (B). This is the main technical contribution of this paper. We show that, after applying the 3-dimensional Weisfeiler-Leman algorithm, there is some color c_0 such that, setting $\chi_i(v, w) := \chi_{\mathsf{WL}}^3[G_i](v, w, w)$ for all $v, w \in V(G_i)$ and $X_i := \{v \in V(G_i) \mid \chi_{\mathsf{WL}}^3[G_i](v, v, v) = c_0\}$, Property (B) is satisfied. Observe that X_i and χ_i are clearly defined in an isomorphism-invariant manner since the coloring computed the 3-dimensional Weisfeiler-Leman algorithm is preserved by isomorphisms. This completes the high-level description of the algorithm.

4 Finding the Initial Set

In this section, we argue how to compute the initial sets X_1 and X_2 with the desired properties as discussed in the last section. Recall the definition of the constant a_{deg} from Theorem 2.1. Without loss of generality assume $a_{\mathsf{deg}} \geq 2$. We define

$$t(h) := \max\{3h^3, ah^2, a_{\mathsf{deg}}^2 h^4, 12a_{\mathsf{deg}} h^4, 36a_{\mathsf{deg}}^2 h^5, 144a_{\mathsf{deg}}^2 h^5\} = 144a_{\mathsf{deg}}^2 h^5. \tag{4.1}$$

Copyright © 2022
Copyright for this paper is retained by authors

Here, the term $t(h)$ provides a lower bound on the parameter t for the t-closure of a set that is required to achieve the desired properties. While it is clear that the last term achieves the maximum, the other terms are also stated for later reference. Indeed, each bound on t will allow us to derive a specific property when building the t-closure of a set. For example, $t \geq 3h^3$ allows the application of Theorem 3.1. Combining all these properties eventually allows to prove that the computed initial sets X_1 and X_2 have the desired properties. The next theorem forms the key technical contribution of this paper.

THEOREM 4.1. *Let G be a connected graph that excludes K_h as a topological subgraph and let $t \geq t(h)$. Then there is a color $c_0 \in \{\chi_{\mathsf{WL}}^3[G](v,v,v) \mid v \in V(G)\}$ such that, for $\chi(v,w) := \chi_{\mathsf{WL}}^3[G](v,w,w)$ for all $v,w \in V(G)$ and $X := \{v \in V(G) \mid \chi_{\mathsf{WL}}^3[G](v,v,v) = c_0\}$, it holds that*

$$X \subseteq \mathrm{cl}_t^{(G,\chi)}(v)$$

for all $v \in X$.

Before diving into the proof of Theorem 4.1, let us state the main corollary that is used for our isomorphism test for graphs excluding K_h as a topological subgraph.

COROLLARY 4.1. *There is a polynomial-time algorithm that, given a connected vertex- and arc-colored graph G and a number $t \geq t(h)$, either correctly concludes that G has a topological subgraph isomorphic to K_h or computes a pair-coloring $\chi\colon (V(G))^2 \to C$ and a non-empty set $X \subseteq V(G)$ such that*

1. *$X = \{v \in V(G) \mid \chi(v,v) = c\}$ for some color $c \in C$, and*
2. *$X \subseteq \mathrm{cl}_t^{(G,\chi)}(v)$ for every $v \in X$.*

Moreover, the output of the algorithm is isomorphism-invariant with respect to G and t.

Here, the output is isomorphism-invariant if it depends only on the isomorphism type of G and the number t. More formally, let (G_1,t_1) and (G_2,t_2) be two input pairs such that $G_1 \cong G_2$ and $t_1 = t_2$. Then the algorithm either concludes in both cases that G_i contains a topological subgraph isomorphic to K_h, or it computes colorings $\chi_i\colon (V(G_i))^2 \to C$ and sets $X_i \subseteq V(G_i)$, $i \in \{1,2\}$, such that $\chi_1(v,w) = \chi_2(\varphi(v),\varphi(w))$ and $X_1^\varphi = X_2^\varphi$ for all $v,w \in V(G_1)$ and all $\varphi \in \mathrm{Iso}(G_1,G_2)$.

Proof. The algorithms sets $\chi(v,w) := \chi_{\mathsf{WL}}^3[G](v,w,w)$ for all $v,w \in V(G)$. Also, it computes the set C_0 of all colors $c_0 \in \{\chi_{\mathsf{WL}}^3[G](v,v,v) \mid v \in V(G)\}$ such that, for $X_{c_0} := \{v \in V(G) \mid \chi_{\mathsf{WL}}^3[G](v,v,v) = c_0\}$, it holds that

$$X_{c_0} \subseteq \mathrm{cl}_t^{(G,\chi)}(v)$$

for all $v \in X_{c_0}$. If $C_0 = \emptyset$ the algorithm outputs that G contains a topological subgraph isomorphic to K_h. This is correct by Theorem 4.1.

Otherwise, let c_0 be the minimal color contained in C_0 (recall that we always assume colors to be linearly ordered). The algorithm outputs χ and X where $X := X_{c_0}$. Clearly, both requirements are satisfied by definition.

Also, the algorithm runs in polynomial time since $\chi_{\mathsf{WL}}^3[G]$ and $\mathrm{cl}_t^{(G,\chi)}(v)$ for every $v \in V(G)$ can be computed in polynomial time. □

Now, let us return to Theorem 4.1. Its lengthy and technical proof is discussed in the remainder of this section[1]. For better readability it is split into several steps. Let us start by giving a brief outline. Further intuition is provided throughout the proof when the single steps can be formulated more clearly.

The central tool for the analysis of the closure sets is the *t-closure graph of* (G,χ) which is defined as the directed graph H with vertex set $V(H) := V(G)$ and edge set

$$E(H) := \{(v,w) \mid w \in \mathrm{cl}_t^{(G,\chi)}(v)\}.$$

[1]Some details are omitted from this extended abstract. The complete proof can be found in the full version of the paper.

Copyright © 2022
Copyright for this paper is retained by authors

A key property of the 3-dimensional Weisfeiler-Leman algorithm is that it detects the edge relation of H, i.e., there is some set of colors $C_H \subseteq \{\chi(v,w) \mid v \neq w \in V(H)\}$ such that $(v,w) \in E(H)$ if and only if $\chi(v,w) \in C_H$ for all $v, w \in V(G)$. Actually, this is the only part of the proof that requires us to use the 3-dimensional Weisfeiler-Leman algorithm. For all remaining parts, it turns out to be sufficient to use the 2-dimensional Weisfeiler-Leman algorithm.

We define $c_0 := \chi^3_{\mathsf{WL}}[G](v_0, v_0, v_0)$ for some $v_0 \in V(G)$ that appears in a maximal strongly connected component of H (a strongly connected component of H is *maximal* if it has no outgoing edges). Let $D(v) := \mathrm{cl}_t^{(G,\chi)}(v)$ for all $v \in X$. In order to show that $X \subseteq \mathrm{cl}_t^{(G,\chi)}(v)$ it suffices to show that $D(v) = D(w)$ for all $v, w \in X$. Building on the fact that the 3-dimensional Weisfeiler-Leman algorithm detects the edge relation of H, we first show that $D(v) = D(w)$ or $D(v) \cap D(w) = \emptyset$ for all $v, w \in X$. This allows us to partition the set $D := \bigcup_{v \in X} D(v)$ and into classes $D_1, \dots, D_k$ where $\{D_1, \dots, D_k\} = \{D(v) \mid v \in X\}$, i.e., $D_1, \dots, D_k$ is an arbitrary enumeration of all distinct sets $D(v)$, $v \in X$. Assuming $k \geq 2$, the basic strategy is to construct a large number of internally vertex-disjoint paths connecting vertices of different partition classes (see Lemma 4.14). The construction of these paths (see Section 4.4) turns out to the most technical and complicated part of the proof. Given a large number of such paths then results in a topological subgraph of G that has high edge density which eventually contradicts Theorem 2.1.

The remainder of this section is structured as follows. In Section 4.1 we provide additional notation and basic tools for the proof of Theorem 4.1. The closure graph and its basic properties are covered in Section 4.2. In Section 4.3, we investigate the interaction between the closure sets $D(v) := \mathrm{cl}_t^{(G,\chi)}(v)$, $v \in X$, in the graph G. In particular, we reduce the task of proving Theorem 4.1 to proving the existence of certain vertex-disjoint paths, as formulated in Lemma 4.14. Finally, the proof of Lemma 4.14 is covered in Section 4.4.

4.1 Basic Tools Among other things, the proof builds on various properties of the 2-dimensional Weisfeiler-Leman algorithm. Towards this end, we first introduce additional notation as well as some basic tools.

Let G be a graph and let χ be a pair coloring that is 2-stable on G. We refer to $C_V := C_V(G,\chi) := \{\chi(v,v) \mid v \in V(G)\}$ as the set of *vertex colors* and $C_E := C_E(G,\chi) := \{\chi(v,w) \mid vw \in E(G)\}$ as the set of *edge colors*. For a vertex color $c \in C_V(G,\chi)$, we define $V_c := V_c(G,\chi) := \{v \in V(G) \mid \chi(v,v) = c\}$ as the set of all vertices with color c. Similar, for an edge color $c \in C_E(G,\chi)$ we define $E_c := E_c(G,\chi) := \{v_1v_2 \in E(G) \mid \chi(v_1,v_2) = c\}$. We say a set $U \subseteq V(G)$ is *χ-invariant* if there is a set of vertex colors $C_U \subseteq C_V$ such that $U = \bigcup_{c \in C_U} V_c$.

Let $C \subseteq \{\chi(v,w) \mid v, w \in V(G), v \neq w\}$ be a set of colors. We define the graph $G[C]$ with vertex set

$$V(G[C]) := \{v, w \mid \chi(v,w) \in C\}$$

and edge set

$$E(G[C]) := \{vw \mid \chi(v,w) \in C\}.$$

Let $A_1, \dots, A_\ell$ be the vertex sets of the connected components of $G[C]$. We also define the graph G/C as the graph obtained from contracting every set A_i to a single vertex. Formally,

$$V(G/C) := \{\{v\} \mid v \in V(G) \setminus V(G[C])\} \cup \{A_1, \dots, A_\ell\}$$

and edge set

$$E(G/C) := \{X_1X_2 \mid \exists v_1 \in X_1, v_2 \in X_2 \colon v_1v_2 \in E(G)\}.$$

Lemma 4.1. (see [10, Theorem 3.1.11]) *Let G be a graph and $C \subseteq \{\chi(v,w) \mid v, w \in V(G), v \neq w\}$ be a set of colors. Define*

$$(\chi/C)(X_1, X_2) := \{\!\{\chi(v_1,v_2) \mid v_1 \in X_1, v_2 \in X_2\}\!\}$$

for all $X_1, X_2 \in V(G/C)$. Then χ/C is 2-stable on G/C.

Moreover, for all $X_1, X_2, X_1', X_2' \in V(G/C)$, either it holds $(\chi/C)(X_1,X_2) = (\chi/C)(X_1',X_2')$ or $(\chi/C)(X_1,X_2) \cap (\chi/C)(X_1',X_2') = \emptyset$.

For every edge color c, the endvertices of all c-colored edges have the same vertex colors, that is, for all edges $vw, v'w' \in E(G)$ with $\chi(v,w) = \chi(v',w') = c$ we have $\chi(v,v) = \chi(v',v')$ and $\chi(w,w) = \chi(w',w')$. This implies $1 \leq |C_V(G[\{c\}],\chi)| \leq 2$. We say that $G[c] := G[\{c\}]$ is *unicolored* if $|C_V(G[c],\chi)| = 1$. Otherwise $G[c]$ is called *bicolored*. The next two lemmas investigate properties of connected components of bicolored graphs $G[c]$ for an edge color c. Again, recall the definition of the constant a_{deg} from Theorem 2.1.

Copyright © 2022
Copyright for this paper is retained by authors

LEMMA 4.2. *Let $G = (V_1, V_2, E)$ be a connected, bipartite graph that excludes K_h as a topological subgraph and let χ be a pair-coloring that is 2-stable on G. Suppose that $\chi(v_1, v_2) = \chi(v_1', v_2')$ for all $(v_1, v_2), (v_1', v_2') \in V_1 \times V_2$ with $v_1v_2, v_1'v_2' \in E$. Also assume that $|V_2| > a_{\mathsf{deg}}h^2|V_1|$. Let*

$$E^* \coloneqq \left\{ v_1v_2 \in \binom{V_1}{2} \mid \exists w \in V_2 \colon v_1w, v_2w \in E(G) \right\}.$$

Then there are colors $c_1, \ldots, c_r \in \chi(V_1^2)$ such that

1. *$E^* = \bigcup_{i\in[r]} E_{c_i}$ where $E_{c_i} \coloneqq \{v_1v_2 \in V(G)^2 \mid \chi(v_1, v_2) = c_i\}$,*

2. *$H \coloneqq (V_1, E^*)$ is connected, and*

3. *H_i is a topological subgraph of G for all $i \in [r]$ where $H_i = (V_1, E_{c_i})$.*

Proof. Clearly, H is connected (since G is connected) and there are colors $c_1, \ldots, c_r \in \chi(V_1^2)$ such that $E^* = \bigcup_{i\in[r]} E_{c_i}$.

So let $i \in [r]$ and consider the bipartite graph $B = (V_2, E_{c_i}, E(B))$ where $E(B) \coloneqq \{(u, v_1v_2) \mid u \in N_G(v_1) \cap N_G(v_2)\}$. By the properties of the 2-dimensional Weisfeiler-Leman algorithm the graph B is biregular. It follows from Hall's Marriage Theorem that B contains a matching M of size $\min(|V_2|, |E_{c_i}|)$ as explained in the preliminaries. If $|V_2| \geq |E_{c_i}|$ then H_i is a topological subgraph of G.

So suppose that $|V_2| < |E_{c_i}|$. Let $F_i \subseteq E_{c_i}$ be those vertices that are matched by the matching M in the graph B. Then $H_i' \coloneqq (V_1, F_i)$ is a topological subgraph of G, and thus it excludes K_h as a topological subgraph. However, $|F_i| = |V_2| > a_{\mathsf{deg}}h^2|V_1|$ which contradicts Theorem 2.1. ∎

LEMMA 4.3. *Let $t \geq a_{\mathsf{deg}}^2 h^4$. Let $G = (V_1, V_2, E)$ be a connected bipartite graph that excludes K_h as a topological subgraph and let χ be a pair-coloring that is 2-stable on G. Suppose that $\chi(v_1, v_2) = \chi(v_1', v_2')$ for all $(v_1, v_2), (v_1', v_2') \in V_1 \times V_2$ with $v_1v_2, v_1'v_2' \in E$. Also assume that $|V_1| \leq |V_2|$. Then $V_1 \subseteq \mathrm{cl}_t^{(G,\chi)}(v)$ for all $v \in V_1 \cup V_2$.*

Proof. The graph G is biregular and it holds that $\deg(v_1) \cdot |V_1| = \deg(v_2) \cdot |V_2|$ for all $v_1 \in V_1$ and $v_2 \in V_2$. Hence, $\deg(v_2) \leq a_{\mathsf{deg}}h^2$ for all $v_2 \in V_2$ by Theorem 2.1. This means $\mathrm{cl}_t^{(G,\chi)}(v) \cap V_1 \neq \emptyset$, because either $v \in \mathrm{cl}_t^{(G,\chi)}(v) \cap V_1$ or $v \in V_2$ and $N_G(v) \subseteq \mathrm{cl}_t^{(G,\chi)}(v) \cap V_1$.

First suppose that $|V_2| \leq a_{\mathsf{deg}}h^2|V_1|$. Then $\deg(v_1) = \deg(v_2)\frac{|V_2|}{|V_1|} \leq t$ and $\deg(v_2) \leq t$ for all $v_1 \in V_1, v_2 \in V_2$. It follows that $\mathrm{cl}_t^{(G,\chi)}(v) = V(G)$.

So assume that $|V_2| > a_{\mathsf{deg}}h^2|V_1|$. By Lemma 4.2, there are colors $c_1, \ldots, c_r \in \chi(V_1^2)$ such that

1. H_i excludes K_h as a topological subgraph for all $i \in [r]$ where $H_i = (V_1, E_{c_i})$ and $E_{c_i} \coloneqq \{v_1v_2 \in V(G)^2 \mid \chi(v_1, v_2) = c_i\}$, and

2. $H = (V_1, \bigcup_{i\in[r]} E_{c_i})$ is connected.

For all $i \in [r]$, the graph H_i is d-regular for some d, and by Theorem 2.1 it holds that $d \leq a_{\mathsf{deg}}h^2 \leq t$. This implies that $V_1 \subseteq \mathrm{cl}_t^{(G,\chi)}(v_1)$ for all $v_1 \in V_1$, and since $V_1 \cap \mathrm{cl}_t^{(G,\chi)}(v) \neq \emptyset$ for all $v \in V_1 \cup V_2$, it follows that $V_1 \subseteq \mathrm{cl}_t^{(G,\chi)}(v)$. ∎

4.2 The Closure Graph We now turn to the proof of Theorem 4.1. For the remainder of this section, let us fix a connected graph G and a number $t \geq t(h)$ as the input for Theorem 4.1. Also, fix the pair-coloring χ defined via $\chi(v, w) \coloneqq \chi_{\mathsf{WL}}^3[G](v, w, w)$ for all $v, w \in V(G)$. Observe that χ is 2-stable on G by Fact 2.1.

We define the *t-closure graph of* (G, χ) to be the directed graph H defined via $V(H) \coloneqq V(G)$ and

$$E(H) \coloneqq \{(v, w) \mid w \in \mathrm{cl}_t^{(G,\chi)}(v), v \neq w\}.$$

LEMMA 4.4. *The coloring χ is 2-stable on H.*

Copyright © 2022
Copyright for this paper is retained by authors

Proof. Let λ be a pair-coloring that is 2-stable on G such that $\lambda \preceq \chi$. We say a partition $\mathcal{P}$ of the set of vertices of G is λ-*definable* if there is a set $C_{\mathcal{P}} \subseteq \{\lambda(v,w) \mid v \neq w \in V(G)\}$ such that $\mathcal{P}$ is the partition into connected components of $G[C_{\mathcal{P}}]$. To prove the lemma we argue that all partitions into color classes of colorings computed by the t-CR algorithm are λ-definable.

Claim 4.1. *Let $\mathcal{P}$ be a λ-definable partition of the vertex set of G. Also define*

$$\mathcal{P}' \coloneqq \{P \in \mathcal{P} \mid |P| > t\} \cup \{\{v\} \mid v \in P \text{ for some } P \in \mathcal{P} \text{ with } |P| \leq t\}.$$

Then $\mathcal{P}'$ is λ-definable.

Proof. Let $P_1, P_2 \in \mathcal{P}$ such that $|P_1| \neq |P_2|$. Also let $v_1 \in P_1$ and $v_2 \in P_2$. Since the 2-dimensional Weisfeiler-Leman algorithm detects which vertices are reachable from v_1 and v_2 in the graph $G[C_{\mathcal{P}}]$ (see, e.g., [10, Theorem 2.6.7]), it follows that $\lambda(v_1,v_1) \neq \lambda(v_2,v_2)$. Hence, defining

$$C_{\mathcal{P}'} \coloneqq C_{\mathcal{P}} \setminus \{\lambda(v,w) \mid \exists P \in \mathcal{P} \colon |P| \leq t \wedge v,w \in P\},$$

it follows that $\mathcal{P}'$ is λ-definable. ◻

Claim 4.2. *Let $\mathcal{P}$ be a λ-definable partition of the vertex set of G. Also define $\mathcal{P}' \preceq \mathcal{P}$ to be the coarsest partition which is stable with respect to the Color Refinement algorithm. Then $\mathcal{P}'$ is λ-definable.*

Proof. For $v \in V(G)$, define P_v to be the unique set $P \in \mathcal{P}$ such that $v \in P$. Define $v \sim_{\mathcal{P}} w$ if

$$\left\{\!\!\left\{ \big(P_u, \chi(v,u), \chi(u,v)\big) \;\middle|\; u \in V(G) \right\}\!\!\right\} = \left\{\!\!\left\{ \big(P_u, \chi(w,u), \chi(u,w)\big) \;\middle|\; u \in V(G) \right\}\!\!\right\}.$$

Let $\mathcal{P}''$ be the partition into equivalence classes of $\sim_{\mathcal{P}}$. Then $\mathcal{P}''$ is λ-definable using Lemma 4.1. So in other words, applying a single round of the Color Refinement algorithm does not effect λ-definability. Hence, the claim follows. ◻

Now, let $u \in V(G)$ and define $\lambda_u(v,w) \coloneqq \chi^3_{\mathsf{WL}}[G](u,v,w)$ for all $v,w \in V(G)$. Also, let $\chi_u(u,u) \coloneqq (1,1)$ and $\chi_u(v,w) \coloneqq (0,\chi(v,w))$ for all $(u,u) \neq (v,w) \in (V(G))^2$. Then λ_u is 2-stable on (G,χ_u) (i.e., the graph obtained from (G,χ) by individualizing u) by Fact 2.2. Now, the previous two claims imply that the partition $\mathcal{P}$ into color classes of $\chi_{t\text{-}\mathsf{CR}}[G,\chi_u]$ is λ_u-definable. Moreover, the choice of $C_{\mathcal{P}}$ does not depend on u. It follows that there is some set of colors $C^* \subseteq \{\chi^3_{\mathsf{WL}}[G](u,v,w) \mid u,v,w \in V(G)\}$ such that

$$\chi^3_{\mathsf{WL}}[G](u,v,w) \in C^* \quad\Leftrightarrow\quad \chi_{t\text{-}\mathsf{CR}}[G,\chi_u](v) = \chi_{t\text{-}\mathsf{CR}}[G,\chi_u](w).$$

This implies the statement of the lemma since $(u,v) \in E(H)$ if and only if there is no $w \neq v$ such that $\chi^3_{\mathsf{WL}}[G](u,v,w) \in C^*$. ◻

Observation 4.1. *Let $(u,v),(v,w) \in E(H)$ such that $u \neq w$. Then $(u,w) \in E(H)$.*

Proof. By definition of the closure graph, $v \in \mathrm{cl}_t^{(G,\chi)}(u)$ and $w \in \mathrm{cl}_t^{(G,\chi)}(v)$. So $w \in \mathrm{cl}_t^{(G,\chi)}(u)$ which implies $(u,w) \in E(H)$. ◻

Now let A be a strongly connected component of H. Then $(v,w) \in E(H)$ for all distinct $v,w \in A$ by the observation above. We say that a vertex $v \in V(G)$ is *maximal* if it appears in a maximal strongly connected component of H, i.e., $(u,v) \in E(H)$ for all $u \in V(G)$ such that $(v,u) \in E(H)$.

Corollary 4.2. *The set of maximal vertices is χ-invariant.*

Proof. This follows from the fact that χ is 2-stable on H. ◻

Copyright © 2022
Copyright for this paper is retained by authors

We say that a vertex color $c \in C_V \coloneqq C_V(G,\chi)$ is *maximal* if there is some vertex $v \in V_c$ which is maximal. By the corollary, for a maximal color $c \in C_V$, every vertex $v \in V_c$ is maximal. We fix $c_0 \in C_V$ to be a maximal color and define

$$X \coloneqq V_{c_0}. \tag{4.2}$$

Also, for $v \in X$, let

$$D(v) \coloneqq \mathrm{cl}_t^{(G,\chi)}(v) \tag{4.3}$$

denote the t-closure of v.

OBSERVATION 4.2. *Let $v, w \in X$. Then $D(v) = D(w)$ or $D(v) \cap D(w) = \emptyset$.*

Proof. Since v is maximal it holds that $D(v)$ is precisely the strongly connected component of H that contains v. Two such components are either equal or disjoint. ∎

Then next lemma forms the main step in proving Theorem 4.1.

LEMMA 4.5. *Suppose there are $v, w \in X$ such that $D(v) \cap D(w) = \emptyset$. Then G has a topological subgraph isomorphic to K_h.*

REMARK 4.1. *We remark that the proof of Lemma 4.5 only exploits that the coloring χ is 2-stable on the graph G and on the t-closure graph of (G,χ). In other words, Lemma 4.4 is the only part of the proof that actually requires the 3-dimensional Weisfeiler-Leman algorithm.*

Before diving into the proof of the lemma, let us first provide a proof for Theorem 4.1 assuming the lemma holds true.

Proof of Theorem 4.1. Let H denote the t-closure graph of (G,χ) and pick $c_0 \in C_V$ to be a maximal color. Also, let $X \coloneqq V_{c_0}$ and define $D(v) \coloneqq \mathrm{cl}_t^{(G,\chi)}(v)$ for all $v \in X$.

If there are $v, w \in X$ such that $D(v) \neq D(w)$ then G has a topological subgraph isomorphic to K_h by Observation 4.2 and Lemma 4.5. So $D(v) = D(w)$ for all $v, w \in X$. Since $v \in D(v)$ for all $v \in X$ this implies that $X \subseteq D(v)$ for all $v \in X$. ∎

Now, let us turn to the proof of Lemma 4.5 which covers the rest of this section. Assume there are $v, w \in X$ such that $D(v) \cap D(w) = \emptyset$. We define $D \coloneqq \bigcup_{v \in X} D(v)$ and let $k \coloneqq |\{D(v) \mid v \in X\}|$. Also let $\{D_1, \dots, D_k\} = \{D(v) \mid v \in X\}$, i.e., $D_1, \dots, D_k$ is an arbitrary enumeration of all distinct sets $D(v)$, $v \in X$.

COROLLARY 4.3. *The set D is χ-invariant. Moreover, there is a set of colors*

$$C_\sim \subseteq \{\chi(v,w) \mid v, w \in V(G), v \neq w\} \tag{4.4}$$

such that $D_1, \dots, D_k$ are precisely the connected components of $G[C_\sim]$. Also,

$$(\chi/C_\sim)(D_i, D_i) = (\chi/C_\sim)(D_j, D_j) \tag{4.5}$$

for all $i, j \in [k]$.

Proof. By definition, the set X is χ-invariant. Recall that H denotes the t-closure graph of (G,χ). By Lemma 4.4 χ is 2-stable on H. By definition of the closure graph $D = \{v \in V(H) \mid \exists w \in X \colon (w,v) \in E(H)\}$. It follows that D is χ-invariant.

Moreover, the sets D_i, $i \in [k]$, are precisely the strongly connected components of $G[D]$. Hence, there is a set $C_\sim \subseteq \{\chi(v,w) \mid v, w \in V(G), v \neq w\}$ such that $D_1, \dots, D_k$ are precisely the connected components of $G[C_\sim]$.

Finally, by Lemma 4.1, it holds that $(\chi/C_\sim)(D_i, D_i) = (\chi/C_\sim)(D_j, D_j)$ or $(\chi/C_\sim)(D_i, D_i) \cap (\chi/C_\sim)(D_j, D_j) = \emptyset$ for all $i, j \in [k]$. Since $X \cap D_i \neq \emptyset$ for all $i \in [k]$ and $X = V_{c_0}$ by definition, it follows that $c_0 \in (\chi/C_\sim)(D_i, D_i)$ for all $i \in [k]$. So $(\chi/C_\sim)(D_i, D_i) = (\chi/C_\sim)(D_j, D_j)$ for all $i, j \in [k]$. ∎

Copyright © 2022
Copyright for this paper is retained by authors

On a high level, the main target for the proof of Lemma 4.5 is to construct a topological subgraph of G which violates the bound on the average degree from Theorem 2.1. Towards this end, we shall consider paths of minimum length connecting sets D_i and D_j for distinct $i, j \in [k]$. We start by covering some simple cases using the tools from Section 4.1. This simplifies the analysis later on since we can exclude certain corner cases.

LEMMA 4.6. *Let $v, w \in X$ such that $D(v) \cap D(w) = \emptyset$ and $E_G(D(v), D(w)) \neq \emptyset$. Then G has a topological subgraph isomorphic to K_h.*

Proof. Suppose that G has no topological subgraph isomorphic to K_h and there are $v' \in D(v)$ and $w' \in D(w)$ such that $v'w' \in E(G)$. We argue that $D(v) \cap D(w) \neq \emptyset$.

Let $c_E := \chi(v', w')$ and consider graph $F := G[c_E]$. Let A be the vertex set of the connected component of F such that $v', w' \in A$. If F is unicolored then $G[A]$ is d-regular for some $d \leq a_{\mathsf{deg}} h^2 \leq t$ by Theorem 2.1. Hence, $A \subseteq D(v)$ and $D(v) \cap D(w) \neq \emptyset$.

Otherwise F is bicolored with color classes V_1 and V_2. Without loss of generality suppose that $|V_1| \leq |V_2|$ and $w' \in V_1$. Hence, $v' \in V_2$ which implies that $w' \in D(v)$ using Lemma 4.3. ∎

The next lemma can be proved using similar, but slightly more involved arguments.

LEMMA 4.7. *Let $v, w \in X$ such that $D(v) \cap D(w) = \emptyset$ and $N_G(D(v)) \cap N_G(D(w)) \neq \emptyset$. Then G has a topological subgraph isomorphic to K_h.*

Hence, for the remainder of this section, we assume that the distance between distinct sets D_i and D_j is least three. In order to cover the case of larger distances, we need to understand in more detail how the *closure sets* D_i, $i \in [k]$, interact with one another.

4.3 Interaction between Closure Sets In order to analyze the interaction between closure sets, it turns out to be more convenient to assume that G has no topological subgraph isomorphic to K_h. Hence, for the remainder of this section, we make the following assumptions and eventually derive a contradiction:

(A.1) $k \geq 2$, i.e., there are $v, w \in X$ such that $D(v) \cap D(w) = \emptyset$, and

(A.2) G has no topological subgraph isomorphic to K_h.

The first goal is to argue that each set D_i interacts with the other sets D_j only via a small set of vertices. To be more precise, we argue that there are sets $S_i \subseteq D_i$ of size $|S_i| < h$ such that every shortest path between D_i and D_j for distinct $i, j \in [k]$ starts in S_i and ends in S_j. The following auxiliary lemma turns out to be useful for this task.

LEMMA 4.8. *Let G be a graph and let $X_1, \dots, X_\ell \subseteq V(G)$ be disjoint sets such that $G[X_i, X_{i+1}]$ is a non-empty, biregular graph for all $i \in [\ell - 1]$. Let $k = \min_{i \in [\ell]} |X_i|$.*

Then there exist k vertex-disjoint paths from X_1 to X_ℓ, i.e., there are distinct vertices $v_{i,j} \in X_i$ for all $i \in [\ell]$ and $j \in [k]$ such that $v_{i,j}v_{i+1,j} \in E(G)$ for all $i \in [\ell - 1]$ and $j \in [k]$.

LEMMA 4.9. *Let $v, w \in X$ such that $D(v) \cap D(w) = \emptyset$. Then there is a connected component Z of $G - D(v)$ such that $D(w) \subseteq Z$.*

Proof. Suppose towards a contraction that there are distinct components of Z_1, Z_2 of the graph $G - D(v)$ such that $D(w) \cap Z_i \neq \emptyset$ for both $i \in \{1, 2\}$. Pick vertices $w_i \in D(w) \cap Z_i$ for both $i \in \{1, 2\}$. Since G is connected there is a shortest path $w_1 = u_0, u_1, \dots, u_m, u_{m+1} = w_2$ from w_1 to w_2. Without loss of generality assume that $u_1, \dots, u_m \notin D(w)$.

Let $\lambda := \chi_{t\text{-}\mathsf{CR}}[G, \chi, w]$ be the t-CR-stable coloring after individualizing w and let $X_i := [u_i]_\lambda$ be the color class of u_i with respect to the coloring λ. Note that $X_i \cap X_j = \emptyset$ for all distinct $i, j \in [m]$. Also $|X_i| \geq t$ for all $i \in [m]$. So there are t internally vertex-disjoint paths from w_1 to w_2 by Lemma 4.8.

On the other hand, $|N_G(Z_1)| < h \leq t$ by Theorem 3.1. Since $w_1 \in Z_1$ and $w_2 \notin Z_1$ this is a contradiction. ∎

Copyright © 2022
Copyright for this paper is retained by authors

The lemma builds a main step for understanding the interaction between closure sets. Recall that we currently aim to prove that there are sets $S_i \subseteq D_i$ of size $|S_i| < h$ such that every shortest path between D_i and D_j for distinct $i, j \in [k]$ starts in S_i and ends in S_j. If $G - D_i$ is connected for all $i \in [k]$ this statement follows from directly from Theorem 3.1 setting $S_i := N_G(Z_i)$ where Z_i is the unique connected component of $G - D_i$. So suppose there is some $i \in [k]$ such that $G - D_i$ is not connected, i.e., the set D_i forms a separator. Now, Lemma 4.9 implies that these separators do not "overlap". In particular, we may contract all sets D_i to a single vertex without effectively changing the connected components of $G - D_i$. This way, we can strengthen the last lemma and prove that, indeed, all sets D_j, $i \neq j \in [k]$, appear in the same connected component of $G - D_i$. Here, we exploit the fact that the 2-dimensional Weisfeiler-Leman algorithm detects cut vertices (i.e., 1-separators) as well as the structure of the block-cut tree (see [22, 23]).

Lemma 4.10. ([23, Corollary 7]) *Let G_1 be a graph and suppose $v_1 \in V(G_1)$ is a cut vertex of G_1. Also let G_2 be a second graph and let $v_2 \in V(G_2)$ such that $\chi^2_{\mathsf{WL}}[G_1](v_1, v_1) = \chi^2_{\mathsf{WL}}[G_2](v_2, v_2)$. Then v_2 is a cut vertex of G_2.*

Lemma 4.11. *For each $v \in X$ there is a connected component $Z(v)$ of the graph $G - D(v)$ such that*

1. *$D(w) \subseteq Z(v)$, or*
2. *$D(w) = D(v)$*

for all $w \in X$. Moreover, the set

$$R := \bigcap_{v \in X} Z(v). \tag{4.6}$$

is χ-invariant.

Proof. We first argue that there is some $i \in [k]$ and a connected component Z_i of $G - D_i$ such that $D_j \subseteq Z_i$ for all $i \neq j \in [k]$. For $i \in [k]$ and Z a connected component of $G - D_i$ define $s(i, Z) := |\{j \in [k] \mid D_j \subseteq Z\}|$. Pick $i \in [k]$ and Z_i a connected component of $G - D_i$ such that $s(i, Z_i)$ is maximal. Suppose towards a contradiction that $s(i, Z_i) < k - 1$. Then, using Lemma 4.9, there is a second component Z_i' of the graph $G - D_i$ such that $s(i, Z_i') \geq 1$. Suppose that $D_j \subseteq Z_i'$ and let Z_j be the connected component of $G - D_j$ such that $D_i \subseteq Z_j$. Then $Z_i \subseteq Z_j$ and hence, $s(j, Z_j) \geq s(i, Z) + 1$. This contradicts the maximality of $s(i, Z_i)$. Hence, $s(i, Z_i) = k - 1$ which means that $D_j \subseteq Z_i$ for all $i \neq j \in [k]$.

For two vertices $v \in D$ and $u \in V(G) \setminus D$ we say that v *is directly reachable from* u if there is a path $u = u_1, \dots, u_m = v$ from u to v such that $u_\mu \notin D$ for all $\mu \in [m-1]$. For $j \in [k]$ we say that D_j *is directly reachable from* u if there is some $v \in D_j$ such that v is directly reachable from u. Finally, define

$$d(u) := |\{j \in [k] \mid D_j \text{ is directly reachable from } u\}|.$$

Observe that $d(u) \geq 1$ for all $u \in V(G) \setminus D$ because G is connected. Let $U := \{u \in V(G) \setminus D \mid d(u) = 1\}$.

Claim 4.3. *U is χ-invariant.*

Proof. Let $C_\sim \subseteq \{\chi(v, w) \mid v, w \in V(G), v \neq w\}$ be the set of colors defined in Corollary 4.3 such that $D_1, \dots, D_k$ are precisely the connected components of $G[C_\sim]$. Consider the graph $G/C_\sim$. By Lemma 4.1 the coloring $\chi/C_\sim$ is 2-stable on $G/C_\sim$. Moreover, $\mathcal{D} := \{D_1, \dots, D_k\}$ is $(\chi/C_\sim)$-invariant by Corollary 4.3. Now, U contains all vertices that can reach only one vertex of $\mathcal{D}$ without visiting another vertex of $\mathcal{D}$. This property is detected by the 2-dimensional Weisfeiler-Leman algorithm which implies that $\{\{u\} \mid u \in U\}$ is $(\chi/C_\sim)$-invariant. So U is χ-invariant. □

Now let $\widetilde{G}$ be the graph obtained from G by turning each set D_i, $i \in [k]$, into a clique. Formally, $V(\widetilde{G}) := V(G)$ and

$$E(\widetilde{G}) := E(G) \cup \{vw \mid v \neq w \wedge \exists i \in [k] \colon v, w \in D_i\}.$$

By Lemma 4.9 the connected components of $G - D_j$ are the same as the connected components of the graph $\widetilde{G} - D_j$ for all $j \in [k]$. Also, Corollary 4.3 and Claim 4.3 imply that $\chi|_{(V(G) \setminus U)^2}$ is 2-stable on the graph $\widetilde{G} - U$.

Copyright © 2022
Copyright for this paper is retained by authors

Now $(\widetilde{G} - U) - D_i$ is connected. We claim that $(\widetilde{G} - U) - D_j$ is connected for all $j \in [k]$. Consider the graph $G^* := (G - U)/C_\sim$ where $C_\sim$ denotes the set of colors from Corollary 4.3. Observe that G^* is the graph obtained from $\widetilde{G} - U$ by contracting each clique D_i, $i \in [k]$, to a single vertex. Since $(\widetilde{G} - U) - D_i$ is connected, we conclude that the vertex $D_i \in V(G^*)$ is not a cut vertex of G^*. Hence, by Corollary 4.3, Lemma 4.1 and 4.10 $D_j \in V(G^*)$ is not a cut vertex of G^* for all $j \in [k]$. In other words, $(\widetilde{G} - U) - D_j$ is connected for all $j \in [k]$. This implies the first part of the lemma.

Also, $R = V(G) \setminus (D \cup U)$. Hence, R is χ-invariant by Corollary 4.3 and Claim 4.3. ◻

By Lemma 4.6 it holds that $E_G(D_i, D_j) = \emptyset$ for all distinct $i, j \in [k]$. It follows that $N_G(D(v)) \cap Z(v) \subseteq R$ for all $v \in X$. In particular, $N_G(Z(v)) = N_G(R) \cap D(v)$ for all $v \in X$. Hence, $|N_G(R) \cap D(v)| < h$ by Theorem 3.1. For $i \in [k]$ define

$$S_i := N_G(R) \cap D_i \tag{4.7}$$

and let $S := \bigcup_{i \in [k]} S_i$.

Observation 4.3. *For all $i \in [k]$ it holds that $|S_i| < h$. Also, S is χ-invariant.*

Proof. As already argued above, the first part follows from Theorem 3.1. Moreover, S is χ-invariant because D and R are χ-invariant by Corollary 4.3 and Lemma 4.11. ◻

Now let

$$p := \min_{i \neq j \in [k]} \min_{v \in D_i, w \in D_j} \operatorname{dist}_G(v, w).$$

Observe that p is indeed a natural number since G is connected. Also note that $p \geq 3$ by Lemma 4.6 and 4.7.

Fix some $i \neq j \in [k]$ and $v \in D_i$, $w \in D_j$ such that $\operatorname{dist}_G(v, w) = p$. Let $v = u_0, \dots, u_p = w$ be a path from v to w to length p. Observe that $u_\mu \in R$ for all $\mu \in [p-1]$. Moreover, let

$$\bar{c} := (\chi(u_0, u_0), \chi(u_0, u_1), \chi(u_1, u_1), \chi(u_1, u_2), \dots, \chi(u_{p-1}, u_{p-1}), \chi(u_{p-1}, u_p), \chi(u_p, u_p))$$

be the sequence of vertex- and arc-colors appearing along the path. A path $w_0, \dots, w_\ell$ is a *$\bar{c}$-path* if

$$\bar{c} = (\chi(w_0, w_0), \chi(w_0, w_1), \chi(w_1, w_1), \chi(w_1, w_2), \dots, \chi(w_{\ell-1}, w_{\ell-1}), \chi(w_{\ell-1}, w_\ell), \chi(w_\ell, w_\ell)).$$

Note that every $\bar{c}$-path has length exactly p. We define the graph F with vertex set $V(F) := \{D_{i'} \mid i' \in [k]\}$ and edge set

$$E(F) := \{D_{i'} D_{j'} \mid (\chi/C_\sim)(D_{i'}, D_{j'}) = (\chi/C_\sim)(D_i, D_j)\}.$$

Observe that the graph F contains at least one edge. We collect some basic properties of the graph F.

Observation 4.4. *F is regular, i.e., $\deg_F(D_i) = \deg_F(D_j)$ for all $i, j \in [k]$.*

Proof. We have $(\chi/C_\sim)(D_i, D_i) = (\chi/C_\sim)(D_j, D_j)$ by Corollary 4.3. So F is regular by Lemma 4.1. ◻

Lemma 4.12. *For all $i \in [k]$ it holds that $\deg_F(D_i) \geq 12 a_{\mathsf{deg}} h^3$.*

Proof. Suppose towards a contradiction that there is some $i \in [k]$ such that $\deg_F(D_i) \leq 12 a_{\mathsf{deg}} h^3$. Let $J := \{j \in [k] \mid D_j \in N_F(D_i)\}$. Now pick arbitrary elements $j \in J$, $v \in S_i$, and a color $c \in \{\chi(v, w) \mid w \in S_j\}$. Let $W := \{w \in V(G) \mid \chi(v, w) = c\}$. Then $W \cap S_j \neq \emptyset$. Also,

$$W \subseteq \bigcup_{j \in J} S_j$$

by Observation 4.3 and Lemma 4.1. So $|W| \leq |J| \cdot h \leq 12 a_{\mathsf{deg}} h^4$ by Observation4.3. Now let $v' \in X \cap D_i$. Then $v \in \operatorname{cl}_t^{(G,\chi)}(v')$ and hence, $W \subseteq \operatorname{cl}_t^{(G,\chi)}(v')$ since $t \geq |W|$. But $\operatorname{cl}_t^{(G,\chi)}(v') = D_i$ by definition and $W \not\subseteq D_i$. A contraction. ◻

Copyright © 2022
Copyright for this paper is retained by authors

LEMMA 4.13. *For every $D_iD_j \in E(F)$ there are $v \in D_i$ and $w \in D_j$ such that there is a $\bar{c}$-path from v to w or there is a $\bar{c}$-path from w to v.*

Proof. By definition of the graph F there exists $D_iD_j \in E(F)$ and $v \in D_i$ and $w \in D_j$ such that there is a $\bar{c}$-path from v to w. Now let $D_{i'}D_{j'} \in E(F)$ be another edge. By the definition of F, Corollary 4.3 and Lemma 4.1 there are $v' \in D_{i'}$ and $w \in D_{j'}$ such that $\chi(v, w) = \chi(v', w')$ or $\chi(v, w) = \chi(w', v')$. Without loss of generality assume the former holds. By the properties of the 2-dimensional Weisfeiler-Leman algorithm there is a $\bar{c}$-*walk* from v' to w', i.e., a sequence of vertices $v' = w'_0, \ldots, w'_\ell = w'$ such that

$$\bar{c} = (\chi(w_0, w_0), \chi(w_0, w_1), \chi(w_1, w_1), \chi(w_1, w_2), \ldots, \chi(w_{\ell-1}, w_{\ell-1}), \chi(w_{\ell-1}, w_\ell), \chi(w_\ell, w_\ell)).$$

(In comparison to a $\bar{c}$-path, a vertex may occur multiple times on the walk.) Since p is the minimal distance between distinct sets D_i, D_j, $i, j \in [k]$, it follows that $w'_0, \ldots, w'_\ell$ is a path. ◻

The following lemma is the crucial step towards the proof of Lemma 4.5.

LEMMA 4.14. *Let $d \coloneqq 4a_{\mathsf{deg}}h^3$. There is a subgraph $\widehat{F} \subseteq F$ with $V(\widehat{F}) = V(F)$ such that*

(A) $\deg_{\widehat{F}}(D_i) \leq d$ *for all* $i \in [k]$,

(B) $\sum_{i\in[k]} \deg_{\widehat{F}}(D_i) = 2 \cdot |E(\widehat{F})| \geq a_{\mathsf{deg}}h^3k$, *and*

(C) for every $e = D_iD_j \in E(\widehat{F})$ there is a path P_e from D_i to D_j of length p such that all paths P_e, $e \in E(\widehat{F})$, are internally vertex-disjoint.

Before proceeding to the proof of the lemma let us first prove Lemma 4.5 based on the lemma above.

Proof of Lemma 4.5. Let $\widehat{F} \subseteq F$ be the subgraph from Lemma 4.14 and fix a set of paths P_e, $e \in E(\widehat{F})$, satisfying Property (C). By the length constraint, all internal vertices of a path P_e, $e \in E(\widehat{F})$, are contained in R and both endvertices are contained in S. Now consider the graph $\widetilde{F}$ with vertex set $V(\widetilde{F}) \coloneqq S$ and $vw \in E(\widetilde{F})$ whenever there is a path P_e, $e \in E(\widehat{F})$, from v to w. Clearly, $\widetilde{F}$ is a topological subgraph of G. Then $\sum_{v\in S} \deg_{\widetilde{F}}(v) = 2|E(\widetilde{F})| \geq 2|E(\widehat{F})| \geq a_{\mathsf{deg}}h^3k$. On the other hand, $|S| < hk$ by Observation 4.3. But this contradicts Theorem 2.1. Hence, one of the Assumptions (A.1) and (A.2) is false. ◻

Hence, it remains to prove Lemma 4.14. This is achieved in the next subsection.

4.4 Constructing Disjoint Paths Consider some $i \neq j \in [k]$ such that $D_iD_j \in E(F)$ as well as some $v \in D_i$ and $w \in D_j$ such that $\mathrm{dist}_G(v, w) = p$. Let $v = u_0, \ldots, u_p = w$ be a path from v to w of length p. Then $v \in S_i$, $w \in S_j$, and $u_\mu \in R$ for all $\mu \in [p - 1]$. Recall that $p \geq 3$ by Lemma 4.6 and 4.7.

We define $r \coloneqq \lceil \frac{p-1}{2} \rceil$ and

$$L_i^{\leq r} \coloneqq \{u \in R \mid \min_{v'\in D_i} \mathrm{dist}(v', u) \leq r\}.$$

By the definition of the parameter r we have that $L_i^{\leq r} \cap L_j^{\leq r} = \emptyset$ for all distinct $i, j \in [k]$, but there exist $i, j \in [k]$ such that $N_G[L_i^{\leq r}] \cap N_G[L_j^{\leq r}] \neq \emptyset$. Furthermore, for $1 \leq \mu \leq r$ let

$$L_i^{=\mu} \coloneqq \{u \in R \mid \min_{v'\in D_i} \mathrm{dist}(v', u) = \mu\}.$$

A visualization can be found in Figure 1.

OBSERVATION 4.5. *For every $\mu \in [r]$ the set $L^{=\mu} \coloneqq \bigcup_{i\in[k]} L_i^{=\mu}$ is χ-invariant. Moreover, there is a set of colors*

$$C^{\leq r} \subseteq \{\chi(v, w) \mid v, w \in V(G), v \neq w\} \tag{4.8}$$

such that $L_1^{\leq r}, \ldots, L_k^{\leq r}$ are precisely the connected components of $G[C^{\leq r}]$. Also,

$$(\chi/C^{\leq r})(L_i^{\leq r}, L_i^{\leq r}) = (\chi/C^{\leq r})(L_j^{\leq r}, L_j^{\leq r}) \tag{4.9}$$

for all $i, j \in [k]$.

Copyright © 2022
Copyright for this paper is retained by authors

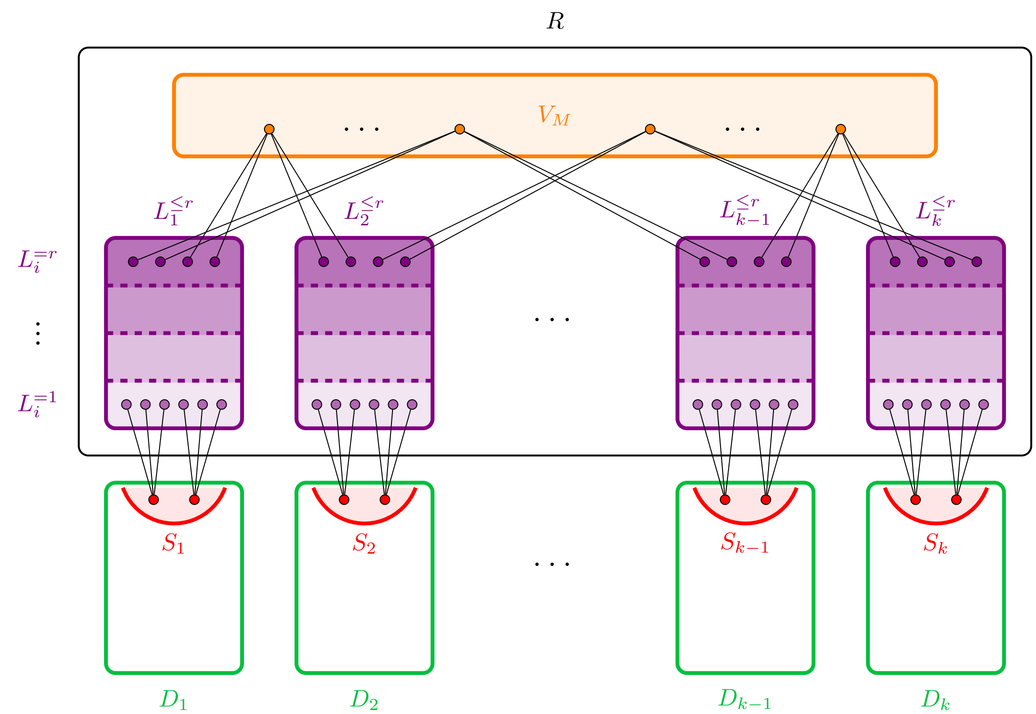

Figure 1: Visualization of the interaction between the closure sets (for p even).

Proof. The sets D and R are χ-invariant by Corollary 4.3 and Lemma 4.11. Define $L^{=0} := D$. Then

$$L^{=\mu} = \left(N_G(L^{=\mu-1}) \cap R\right) \setminus \bigcup_{\mu' < \mu} L^{=\mu'}.$$

Hence, $L^{=\mu}$ is χ-invariant for all $\mu \in [r]$ by induction. For the second part observe that $L_1^{\leq r} \cup D_1, \dots, L_k^{\leq r} \cup D_k$ are the connected components of the graph $G[C]$ where

$$C := \{\chi(v,w) \mid vw \in E(G), v, w \in \bigcup_{i \in [k]} L_i^{\leq r} \cup D_i\} \cup C_\sim$$

where $C_\sim$ is the set of colors from Corollary 4.3. In combination with Corollary 4.3 and Lemma 4.1 this also implies the third statement. $\square$

We split the problem of constructing the desired paths into two parts: constructing paths within the sets $L_i^{\leq r}$ and building the middle part of constant size. We start by considering the paths within the sets $L_i^{\leq r}$. For these parts, we are interested in which vertices from the set $L_i^{=r}$ can be extended to a path that is disjoint from a set of existing paths on $L_i^{\leq r}$. Actually, for the overall argument to work out, we may even modify the set of existing paths on $L_i^{\leq r}$ as long as we do not change the "interface" in the set $L_i^{=r}$, i.e., the endpoints of the paths have to remain the same.

Fix $i \in [k]$. Let $P_1, \dots, P_\ell$ be a set of ℓ vertex-disjoint paths of length $r-1$ from $L_i^{=1}$ to $L_i^{=r}$. We define $L_i^\mu(P_1, \dots, P_\ell)$ to be the set of vertices lying on the paths $P_1, \dots, P_\ell$ that are contained in $L_i^{=\mu}$. Observe that $|L_i^\mu(P_1, \dots, P_\ell)| = \ell$ for all $\mu \in [r]$. Now we define the *expansion set of* $P_1, \dots, P_\ell$ to be the set

$$\mathrm{Exp}_i(P_1, \dots, P_\ell) := \Big\{ v \in L_i^{=r} \;\Big|\; \text{there exist } \ell+1 \text{ vertex-disjoint paths } P'_1, \dots, P'_{\ell+1} \text{ of length } r-1 \text{ from } L_i^{=1} \text{ to } L_i^{=r} \text{ such that } L_i^r(P'_1, \dots, P'_{\ell+1}) = L_i^r(P_1, \dots, P_\ell) \cup \{v\} \Big\}$$

Copyright © 2022
Copyright for this paper is retained by authors

LEMMA 4.15. *Let $P_1, \dots, P_\ell$ be a set of $\ell < 12a_{\mathsf{deg}}^2 h^5$ vertex-disjoint paths of length $r-1$ from $L_i^{=1}$ to $L_i^{=r}$. Also, let $c \in C_V(G,\chi)$. Then*

$$|\operatorname{Exp}_i(P_1, \dots, P_\ell) \cap V_c| \geq \left(1 - \frac{24a_{\mathsf{deg}}^2 h^5}{t}\right) |L_i^{=r} \cap V_c|.$$

Proof. Let $x \in X$ such that $D_i = D(x)$ and let $\lambda := \chi_{t\text{-}\mathsf{CR}}[G,\chi,x]$ be the t-CR-stable coloring after individualizing x. Observe that $D_i = \{w \in V(G) \mid |[w]_\lambda| = 1\}$ and the sets $F_i^{=\mu}$ are λ-invariant for all $\mu \in r$.

Let $u_r \in L_i^{=r} \cap V_c$ and let $u_1, \dots, u_r$ be a path from $L_i^{=1}$ to $L_i^{=r} \cap V_c$. Also let $c_\mu := \lambda(u_\mu)$ for all $\mu \in [r]$. We argue that

$$|\operatorname{Exp}_i(P_1, \dots, P_\ell) \cap \lambda^{-1}(c_r)| \geq \left(1 - \frac{24a_{\mathsf{deg}}^2 h^5}{t}\right) |\lambda^{-1}(c_r)|. \tag{4.10}$$

Clearly, this implies the lemma since

$$\begin{aligned} |\operatorname{Exp}_i(P_1, \dots, P_\ell) \cap V_c| &= \sum_{c' \in \lambda^{-1}(F_i^{=r} \cap V_c)} |\operatorname{Exp}_i(P_1, \dots, P_\ell) \cap \lambda^{-1}(c')| \\ &\geq \sum_{c' \in \lambda^{-1}(F_i^{=r} \cap V_c)} \left(1 - \frac{24a_{\mathsf{deg}}^2 h^5}{t}\right) |\lambda^{-1}(c')| \\ &= \left(1 - \frac{24a_{\mathsf{deg}}^2 h^5}{t}\right) \sum_{c \in \lambda^{-1}(F_i^{=r} \cap V_c)} |\lambda^{-1}(c')| \\ &= \left(1 - \frac{24a_{\mathsf{deg}}^2 h^5}{t}\right) |L_i^{=r} \cap V_c|. \end{aligned}$$

To prove Equation (4.10) we use an alternating-paths argument and define the following directed graph H with vertex set $V(H) := F_i^{\leq r}$ and edge set $E(H) := E_{\mathrm{fw}} \cup E_{\mathrm{bw}}$. The *forward edges* are defined as

$$E_{\mathrm{fw}} := \{(v,w) \mid vw \in E(G) \setminus \bigcup_{j \in [\ell]} E(P_j), v \in L_i^{=\mu}, w \in L_i^{=\mu+1} \text{ for some } \mu \in [r-1]\}.$$

The *backward edges* are defined as

$$E_{\mathrm{bw}} := \{(v,w) \mid vw \in \bigcup_{j \in [\ell]} E(P_j), v \in L_i^{=\mu+1}, w \in L_i^{=\mu} \text{ for some } \mu \in [r-1]\}.$$

We consider directed paths that start in $L_i^{=1} \setminus L_i^1(P_1, \dots, P_\ell)$. A directed path $v_1, \dots, v_q$ in H is *admissible* if

1. $v_1 \in L_i^{=1} \setminus L_i^1(P_1, \dots, P_\ell)$,
2. $(v_\eta, v_{\eta+1}) \in E(H)$, and
3. if $(v_\eta, v_{\eta+1}) \in E_{\mathrm{fw}}$ and $v_{\eta+1} \in \bigcup_{j \in [\ell]} V(P_j)$ then $\eta \leq q-2$ and $(v_{\eta+1}, v_{\eta+2}) \in E_{\mathrm{bw}}$

for all $\eta \in [q-1]$. Let

$$A := \{v \in V(G) \mid \text{there is an admissible path } v_1, \dots, v_q \text{ such that } v_q = v\}.$$

Also let $A_\mu := A \cap F_i^{=\mu}$ for all $\mu \in [r]$.

CLAIM 4.4. *$A_r \subseteq \operatorname{Exp}_i(P_1, \dots, P_\ell)$.*

Copyright © 2022
Copyright for this paper is retained by authors

Proof. Let $v \in A_r$ and let $v_1, \ldots, v_q$ be an admissible path of minimal length q such that $v = v_q$. Let $P_{\ell+1}$ be the corresponding path graph with $V(P_{\ell+1}) \coloneqq \{v_1, \ldots, v_q\}$ and $E(P_{\ell+1}) \coloneqq \{v_i v_{i+1} \mid i \in [q-1]\}$. Consider the graph P with vertex set $V(P) \coloneqq \bigcup_{j\in[\ell+1]} V(P_j)$ and edge set

$$E(P) \coloneqq \left(\bigcup_{j\in[\ell]} E(P_j) \setminus E(P_{\ell+1})\right) \cup \left(E(P_{\ell+1}) \setminus \bigcup_{j\in[\ell]} E(P_j)\right).$$

Then P is the disjoint union of $(\ell+1)$ many paths $P_1', \ldots, P_{\ell+1}'$ (and possibly isolated vertices) from $L_i^{=1}$ to $L_i^{=r}$ such that $L_i^r(P_1', \ldots, P_{\ell+1}') = L_i^r(P_1, \ldots, P_\ell) \cup \{v\}$. To see this, observe that if there is a vertex $w \in V(P_{\ell+1}) \cap V(P_j)$ for some $j \in [\ell]$, then there is also a common adjacent edge $ww' \in E(P_{\ell+1}) \cap E(P_j)$ by the definition of an admissible path. This implies that $v \in \mathrm{Exp}_i(P_1, \ldots, P_\ell)$. ◻

By the claim, it suffices to provide a lower bound on the size of the set $A_r \cap \lambda^{-1}(c_r)$. Towards this end, we analyze the structure of the set A. For $j \in [\ell]$ let $u_{\mu,j}$ be the unique vertex in the set $V(P_j) \cap F_i^{=\mu}$, $\mu \in [r]$. First observe that, if $u_{\mu,j} \in A$, then also $u_{\mu',j} \in A$ for all $\mu' < \mu$ since all vertices $u_{\mu',j}$ are reachable with backward edges in E_{bw}.

We call a vertex $b \in \bigcup_{j\in[\ell]} V(P_j) \setminus A$ a *blocking vertex* if there is a vertex $v \in \bigcup_{j\in[\ell]} V(P_j) \cap A$ such that $(b, v) \in E_{\mathrm{bw}}$. In other words, the vertex $u_{\mu,j}$ is a blocking vertex if $u_{\mu,j} \notin A$ and $u_{\mu-1,j} \in A$ (and therefore $u_{\mu',j} \in A$ for all $\mu' < \mu$). Let B be the set of blocking vertices. By the above observation, $|B \cap V(P_j)| \leq 1$ for all $j \in [\ell]$. Hence, $|B| \leq \ell$. Let $\ell_\mu \coloneqq |B \cap F_i^{=\mu}|$ be the number of blocking vertices on level μ, $\mu \in [r]$.

Claim 4.5. $|A_\mu \cap \lambda^{-1}(c_\mu)| \geq \left(\dfrac{|A_1 \cap \lambda^{-1}(c_1)|}{|\lambda^{-1}(c_1)|} - \dfrac{\ell_1 + \ldots + \ell_\mu}{t}\right) |\lambda^{-1}(c_\mu)|$ *for all* $\mu \in [r]$.

Proof. The claim is proved by induction on $\mu \in [r]$. The base case $\mu = 1$ is immediately clear.

For the inductive step assume that $\mu \geq 1$. Since $G[\lambda^{-1}(c_\mu), \lambda^{-1}(c_{\mu+1})]$ is a non-empty biregular graph, for each subset $S \subseteq \lambda^{-1}(c_\mu)$ it holds that

$$\frac{|N_G(S) \cap \lambda^{-1}(c_{\mu+1})|}{|\lambda^{-1}(c_{\mu+1})|} \geq \frac{|S|}{|\lambda^{-1}(c_\mu)|}$$

as argued in the preliminaries. We first argue that

$$N\big(A_\mu \cap \lambda^{-1}(c_\mu)\big) \;\cap\; \lambda^{-1}(c_{\mu+1}) \quad\subseteq\quad A_{\mu+1} \cup B.$$

Let $v \in A_\mu \cap \lambda^{-1}(c_\mu)$ and $w \in N(v) \cap \lambda^{-1}(c_{\mu+1})$. If $w \in \bigcup_{j\in[\ell]} V(P_j)$ then $w \in B$ or $w \in A$. Otherwise $w \in V(G) \setminus \bigcup_{j\in\ell]} V(P_j)$ and $(v, w) \in E_{\mathrm{fw}}$ which means $w \in A$. This shows the inclusion and therefore

$$\frac{|A_{\mu+1} \cap \lambda^{-1}(c_{\mu+1})|}{|\lambda^{-1}(c_{\mu+1})|} \geq \frac{|N(A_\mu \cap \lambda^{-1}(c_\mu)) \cap \lambda^{-1}(c_{\mu+1})| - \ell_{\mu+1}}{|\lambda^{-1}(\mu_{i+1})|} \geq \frac{|A_\mu \cap \lambda^{-1}(c_\mu)|}{|\lambda^{-1}(c_\mu)|} - \frac{\ell_{\mu+1}}{t}.$$

By the induction hypothesis,

$$\frac{|A_\mu \cap \lambda^{-1}(c_\mu)|}{|\lambda^{-1}(c_\mu)|} \geq \frac{|A_1 \cap \lambda^{-1}(c_1)|}{|\lambda^{-1}(c_1)|} - \frac{\ell_1 + \cdots + \ell_\mu}{t}.$$

In combination this means

$$\frac{|A_{\mu+1} \cap \lambda^{-1}(c_{\mu+1})|}{|\lambda^{-1}(c_{\mu+1})|} \geq \frac{|A_1 \cap \lambda^{-1}(c_1)|}{|\lambda^{-1}(c_1)|} - \frac{\ell_1 + \cdots + \ell_\mu + \ell_{\mu+1}}{t}.$$

◻

Now, we can prove Equation (4.10). We already observed in Claim 4.4 that

$$A_r \subseteq \mathrm{Exp}_i(P_1, \ldots, P_\ell).$$

Copyright © 2022
Copyright for this paper is retained by authors

Moreover, it holds that $|A_1 \cap \lambda^{-1}(c_1)| = |\lambda^{-1}(c_1)| - \ell$. Combining this with Claim 4.5, we obtain

$$\begin{aligned}
|\operatorname{Exp}_i(P_1, \dots, P_\ell) \cap \lambda^{-1}(c_r)| &\geq |A_r \cap \lambda^{-1}(c_r)| \\
&\geq \left(\frac{|A_1 \cap \lambda^{-1}(c_1)|}{|\lambda^{-1}(c_1)|} - \frac{\ell}{t} \right) |\lambda^{-1}(c_r)| \\
&\geq \left(\frac{|\lambda^{-1}(c_1)| - 12a_{\mathsf{deg}}^2 h^5}{|\lambda^{-1}(c_1)|} - \frac{12a_{\mathsf{deg}}^2 h^5}{t} \right) |\lambda^{-1}(c_r)| \\
&\geq \left(\frac{t - 12a_{\mathsf{deg}}^2 h^5}{t} - \frac{12a_{\mathsf{deg}}^2 h^5}{t} \right) |\lambda^{-1}(c_r)| \\
&= \left(1 - \frac{24a_{\mathsf{deg}}^2 h^5}{t} \right) |\lambda^{-1}(c_r)|.
\end{aligned}$$

□

Building on the previous lemma, the critical step becomes the construction of the middle part of the paths P_e that need to be constructed in order to prove Lemma 4.14. Here, we distinguish between two cases depending on the parity of the path length p. In fact, in this extended abstract, we shall only discuss the case p odd. The arguments for p even are similar, but they are much more involved on a technical level.

4.4.1 Paths of Odd Length We give a proof for Lemma 4.14 in the simpler case in which p is odd. The basic idea is to construct the paths one-by-one, i.e., initially we define $E(\widehat{F})$ to be empty. In each iteration, the set of edges (as well as the corresponding set of paths) is extended by one until Property (B) is satisfied while always maintaining Properties (A) and (C). Observe that Properties (A) and (C) are satisfied initially.

Hence, let us fix a subgraph $\widehat{F} \subseteq F$ which satisfies Properties (A) and (C), but violates Property (B). Let P_e, $e \in E(\widehat{F})$, be the corresponding set of paths. We argue how to extend $\widehat{F}$ by a single edge while maintaining Properties (A) and (C).

Consider the color $c_M \coloneqq \bar{c}_{2r+2}$ of the *middle edge* of a $\bar{c}$-path as well as the color of its two incident vertices $c_L \coloneqq \bar{c}_{2r+1}$ and $c_R \coloneqq \bar{c}_{2r+3}$. Let $E_M \coloneqq \{vw \in E(G) \mid \chi(v,w) = c_M\}$.

For each $D_iD_j \in E(F)$ we define the *witness set* $W(D_iD_j) \coloneqq \{vw \in E_M \mid v \in L_i^{\leq r}, w \in L_j^{\leq r}\}$. I remark that each $e \in W(D_iD_j)$ appears on a $\bar{c}$-path from some $v \in D_i$ to some $w \in D_j$, or on a $\bar{c}$-path from some $w \in D_j$ to some $v \in D_i$.

A partition $\mathcal{P}$ of a set A is an *equipartition* if $|P| = |P'|$ for all $P, P' \in \mathcal{P}$. Observe that, for an equipartition $\mathcal{P}$, we have that $|A| = |\mathcal{P}| \cdot |P|$ for all $P \in \mathcal{P}$.

Lemma 4.16. *The sets $W(D_iD_j)$, $D_iD_j \in E(F)$, form an equipartition of the set E_M.*

Proof. Consider the set $C^{\leq r} \subseteq \{\chi(v,w) \mid v,w \in V(G), v \neq w\}$ defined in Observation 4.5 such that $L_1^{\leq r}, \dots, L_k^{\leq r}$ are precisely the connected components of $G[C^{\leq r}]$. By Observation 4.5,

$$\{\{\chi(v,v) \mid v \in L_i^{\leq r}\}\} = \{\{\chi(v,v) \mid v \in L_j^{\leq r}\}\}$$

for all $i,j \in [k]$. Also,

$$c_M \in \{\{\chi(v,w) \mid v \in L_i^{\leq r}, w \in L_j^{\leq r}\}\}$$

for some $D_iD_j \in E(F)$ by definition of the graph F. Now, Lemma 4.1 implies that

$$\{\{\chi(v,w) \mid v \in L_i^{\leq r}, w \in L_j^{\leq r}\}\} = \{\{\chi(v,w) \mid v \in L_{i'}^{\leq r}, w \in L_{j'}^{\leq r}\}\}$$

for all $D_iD_j, D_{i'}D_{j'} \in E(G)$ and c_M only appears in sets $\{\{\chi(v,w) \mid v \in L_i^{\leq r}, w \in L_j^{\leq r}\}\}$ for $D_iD_j \in E(G)$. □

Now let $i \in [k]$. Let $P_1^i, \dots, P_{d_i}^i$ denote the paths which are obtained from intersecting the paths P_e, $e \in E(\widehat{F})$, with the set $L_i^{\leq r}$. Observe that $d_i = \deg_{\widehat{F}}(D_i)$.

Copyright © 2022
Copyright for this paper is retained by authors

The basic idea to extend the graph $\widehat{F}$ by a single edge is to apply a counting argument. Consider the set E_M. We argue that there is some $e \in E_M$ which can be extended to a $\bar{c}$-path which is internally-vertex disjoint from all paths P_e, $e \in E(\widehat{F})$, and provides a witness for extending $\widehat{F}$ by a single edge. More precisely, in light of Lemma 4.15, it suffices to prove the following lemma. Recall the definition of the parameter d from Lemma 4.14.

LEMMA 4.17. *There is an edge $vw \in E_M$ such that $vw \in W(D_iD_j)$ and*

(a) $\deg_{\widehat{F}}(D_i) < d$ *and* $\deg_{\widehat{F}}(D_j) < d$,

(b) $D_iD_j \notin E(\widehat{F})$, *and*

(c) $v \in \operatorname{Exp}_i(P_1^i, \dots, P_{d_i}^i)$ *and* $w \in \operatorname{Exp}_j(P_1^j, \dots, P_{d_j}^j)$.

Proof. We first provide upper bounds on the number edges in E_M violating one of the first two properties. Let

$$\widehat{E}_M^1 \coloneqq \bigcup_{D_i\colon \deg_{\widehat{F}}(D_i)=d} \bigcup_{D_j \in N_F(D_i)} W(D_iD_j)$$

denote the set of edges violating Property (a). Let $U \coloneqq \{i \in [k] \mid \deg_{\widehat{F}}(D_i) = d\}$. Then $|E(\widehat{F})| \geq \frac{d}{2}|U| = 2a_{\mathsf{deg}}h^3|U|$. Since $|E(\widehat{F})| < \frac{1}{2}a_{\mathsf{deg}}h^3k$ it follows that $|U| < \frac{k}{4}$. By Observation 4.4 there is a number d_F such that $\deg_F(D_i) = d_F$ for all $i \in [k]$ Hence,

$$|\{D_iD_j \in E(F) \mid \deg_{\widehat{F}}(D_i) = d \vee \deg_{\widehat{F}}(D_j) = d\}| \leq d_F \cdot |U| < \frac{d_F k}{4} = \frac{|E(F)|}{2}.$$

So $|\widehat{E}_M^1| \leq \frac{1}{2}|E_M|$ by Lemma 4.16. Next, let

$$\widehat{E}_M^2 \coloneqq \bigcup_{D_iD_j \in E(\widehat{F})} W(D_iD_j)$$

denote the set of edges violating Property (b). We have that $\frac{|E(\widehat{F})|}{|E(F)|} < \frac{1}{12}$ by Lemma 4.12 and the fact that $\widehat{F}$ violates Property (B). So $|\widehat{E}_M^2| \leq \frac{1}{12}|E_M|$ by Lemma 4.16.

Now let us analyze Property (c). By Lemma 4.15 it holds that

$$|\operatorname{Exp}_i(P_1^i, \dots, P_{d_i}^i) \cap V_{c_L}| \geq \left(1 - \frac{1}{6}\right) |L_i^{=r} \cap V_{c_L}|$$

and

$$|\operatorname{Exp}_i(P_1^i, \dots, P_{d_i}^i) \cap V_{c_R}| \geq \left(1 - \frac{1}{6}\right) |L_i^{=r} \cap V_{c_R}|$$

for all $i \in [k] \setminus U$. Hence,

$$\begin{aligned} &|\{vw \in E_M \setminus \widehat{E}_M^1 \mid vw \in W(D_iD_j), v \in \operatorname{Exp}_i(P_1^i, \dots, P_{d_i}^i), w \in \operatorname{Exp}_j(P_1^j, \dots, P_{d_j}^j)\}| \\ \geq\ &\frac{2}{3}(|E_M| - |\widehat{E}_M^1|). \end{aligned}$$

So, the number of edges satisfying Property (a), (b), and (c) it at least

$$\frac{2}{3}\left(|E_M| - |\widehat{E}_M^1|\right) - |\widehat{E}_M^2| \geq \frac{2}{3}\left(|E_M| - \frac{1}{2}|E_M|\right) - \frac{1}{12}|E_M| = \frac{1}{4}|E_M| > 0.$$

□

Copyright © 2022
Copyright for this paper is retained by authors

Proof of Lemma 4.14 for p odd. Let $\widehat{F}$ be a maximal subgraph of F that satisfies Property (A) and (C). Suppose towards a contraction that Property (B) is violated, i.e., $2 \cdot |E(\widehat{F})| < a_{\mathsf{deg}} h^3 k$. Let P_e, $e \in E(\widehat{F})$, be the corresponding set of paths guaranteed by Property (C). For $i \in [k]$ let $P_1^i, \dots, P_{d_i}^i$ denote the paths which are obtained from intersecting the paths P_e, $e \in E(\widehat{F})$, with the set $L_i^{\leq r}$. Observe that $d_i = \deg_{\widehat{F}}(D_i)$ and the path P_j^i has length $r-1$ for all $i \in [k]$ and $j \in [d_i]$.

By Lemma 4.17 there is an edge $vw \in E_M$ such that $vw \in W(D_iD_j)$ satisfying Property (a), (b), and (c). Let $\widehat{F} + D_iD_j$ denote the graph obtained from $\widehat{F}$ by adding the edge D_iD_j. By Property (a) the graph $\widehat{F} + D_iD_j$ satisfies Property (A). By Property (c) and the definition of an extension set there are vertex-disjoint paths $Q_1^i, \dots, Q_{d_i+1}^i$ from $L_i^{=1}$ to $L_i^{=r}$ of length $r-1$ such that

$$L_i^{=r}(Q_1^i, \dots, Q_{d_i+1}^i) = L_i^{=r}(P_1^i, \dots, P_{d_i}^i) \cup \{v\}$$

and vertex-disjoint paths $Q_1^j, \dots, Q_{d_j+1}^j$ from $L_j^{=1}$ to $L_j^{=r}$ of length $r-1$ such that

$$L_j^{=r}(Q_1^j, \dots, Q_{d_j+1}^j) = L_j^{=r}(P_1^j, \dots, P_{d_j}^j) \cup \{w\}.$$

This gives a set of paths Q_e, $e \in E(\widehat{F} + D_iD_j)$, witnessing Property (C) for the graph $\widehat{F} + D_iD_j$. Hence, by the maximality of $\widehat{F}$, it holds that $D_iD_j \in E(\widehat{F})$ contracting Property (b). $\square$

As already pointed out above, the analysis for p even is similar, but turns out to be much more involved on a technical level. We refer the interested reader to the full version of the paper. Overall, this completes the proof of Lemma 4.14 and thus, we have also shown Theorem 4.1.

5 Isomorphism Test for Graphs Excluding a Topological Subgraph

Having established the necessary combinatorial tools, one can now assemble the main algorithm following the high-level description given above. Towards this end, the combinatorial tools introduced in the previous sections are combined with group-theoretic tools from [29, 36]. Since this part is identical to the isomorphism algorithm from [17] for graphs excluding K_h as a minor, we omit the details here and only state the main result. The complete proof can be found in the full version.

Theorem 5.1. *There is an algorithm that, given a number $h \in \mathbb{N}$ and two connected vertex-colored graphs G_1 and G_2 with n vertices, either correctly concludes that G_1 has a topological subgraph isomorphic to K_h, or decides whether G_1 is isomorphic to G_2 in time $n^{\mathcal{O}((\log h)^c)}$ for some absolute constant c.*

We remark that, by standard reduction techniques, there is also an algorithm computing a representation for the set $\mathrm{Iso}(G_1, G_2)$ in time $n^{\mathcal{O}((\log h)^c)}$ assuming G_1 excludes K_h as a topological subgraph.

Moreover, the arguments also reveal some insight into the structure of the automorphism group of a graph that excludes K_h as a topological subgraph.

Let G be a graph. A *tree decomposition* for G is a pair (T, β) where T is a rooted tree and $\beta\colon V(T) \to 2^{V(G)}$ such that

(T.1) for every $e \in E(G)$ there is some $t \in V(T)$ such that $e \subseteq \beta(t)$, and

(T.2) for every $v \in V(G)$ the graph $T[\{t \in V(T) \mid v \in \beta(t)\}]$ is non-empty and connected.

The *adhesion-width* of (T, β) is $\max_{t_1t_2 \in E(T)} |\beta(t_1) \cap \beta(t_2)|$.

Let $v \in V(G)$. Also, recall that $(\mathrm{Aut}(G))_v = \{\varphi \in \mathrm{Aut}(G) \mid v^\varphi = v\}$ denotes the subgroup of the automorphism group of G that stabilizes the vertex v.

Theorem 5.2. *Let G be a graph that excludes K_h as a topological subgraph. Then there is an isomorphism-invariant tree decomposition (T, β) of G such that*

1. *the adhesion-width of (T, β) is at most $h-1$, and*
2. *for every $t \in V(T)$ there is some $v \in \beta(t)$ such that $(\mathrm{Aut}(G))_v[\beta(t)] \in \widehat{\Gamma}_d$ for $d := 144 a_{\mathsf{deg}}^2 h^5$.*

Copyright © 2022
Copyright for this paper is retained by authors

6 Conclusion

We presented an isomorphism test for all graphs excluding K_h as a topological subgraph running in time $n^{\mathrm{polylog}(h)}$. On the technical side, the main contribution towards this algorithm is a combinatorial statement which provides a suitable isomorphism-invariant initial set to apply the t-CR algorithm. As a consequence, we also obtain restrictions on the structure of the automorphism groups of graphs excluding K_h as a topological subgraph.

Overall, the presented result unifies and extends existing isomorphism tests with polylogarithmic parameter dependence in the exponent of the running time, and essentially completes the picture for such algorithms on sparse graph classes. It is an interesting open question whether the techniques can be extended to graph parameters that include dense graphs. As a specific example, can isomorphism of graphs of rank-width k be tested in time $n^{\mathrm{polylog}(k)}$?

Of course, it also remains an important question for which graph parameters the isomorphism problem is fixed-parameter tractable. However, here it is already open whether isomorphism testing parameterized by the maximum degree or the Hadwiger number (i.e., the maximum h for which K_h is a minor of the input graph) is fixed-parameter tractable.

References

[1] László Babai. Graph isomorphism in quasipolynomial time [extended abstract]. In Daniel Wichs and Yishay Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 684–697. ACM, 2016.

[2] László Babai, Peter J. Cameron, and Péter P. Pálfy. On the orders of primitive groups with restricted nonabelian composition factors. *J. Algebra*, 79(1):161–168, 1982.

[3] László Babai, D. Yu. Grigoryev, and David M. Mount. Isomorphism of graphs with bounded eigenvalue multiplicity. In Harry R. Lewis, Barbara B. Simons, Walter A. Burkhard, and Lawrence H. Landweber, editors, *Proceedings of the 14th Annual ACM Symposium on Theory of Computing, May 5-7, 1982, San Francisco, California, USA*, pages 310–324. ACM, 1982.

[4] László Babai, William M. Kantor, and Eugene M. Luks. Computational complexity and the classification of finite simple groups. In *24th Annual Symposium on Foundations of Computer Science, Tucson, Arizona, USA, 7-9 November 1983*, pages 162–171. IEEE Computer Society, 1983.

[5] László Babai and Eugene M. Luks. Canonical labeling of graphs. In David S. Johnson, Ronald Fagin, Michael L. Fredman, David Harel, Richard M. Karp, Nancy A. Lynch, Christos H. Papadimitriou, Ronald L. Rivest, Walter L. Ruzzo, and Joel I. Seiferas, editors, *Proceedings of the 15th Annual ACM Symposium on Theory of Computing, 25-27 April, 1983, Boston, Massachusetts, USA*, pages 171–183. ACM, 1983.

[6] Christoph Berkholz, Paul S. Bonsma, and Martin Grohe. Tight lower and upper bounds for the complexity of canonical colour refinement. *Theory Comput. Syst.*, 60(4):581–614, 2017.

[7] Hans L. Bodlaender. Polynomial algorithms for graph isomorphism and chromatic index on partial k-trees. *J. Algorithms*, 11(4):631–643, 1990.

[8] Béla Bollobás and Andrew Thomason. Proof of a conjecture of Mader, Erdős and Hajnal on topological complete subgraphs. *Eur. J. Comb.*, 19(8):883–887, 1998.

[9] Jin-yi Cai, Martin Fürer, and Neil Immerman. An optimal lower bound on the number of variables for graph identification. *Comb.*, 12(4):389–410, 1992.

[10] Gang Chen and Ilia N. Ponomarenko. Lectures on coherent configurations. `http://www.pdmi.ras.ru/~inp/ccNOTES.pdf`, 2019.

[11] John D. Dixon and Brian Mortimer. *Permutation Groups*, volume 163 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1996.

[12] I. S. Filotti and Jack N. Mayer. A polynomial-time algorithm for determining the isomorphism of graphs of fixed genus (working paper). In Raymond E. Miller, Seymour Ginsburg, Walter A. Burkhard, and Richard J. Lipton, editors, *Proceedings of the 12th Annual ACM Symposium on Theory of Computing, April 28-30, 1980, Los Angeles, California, USA*, pages 236–243. ACM, 1980.

[13] Martin Grohe and Dániel Marx. Structure theorem and isomorphism test for graphs with excluded topological subgraphs. *SIAM J. Comput.*, 44(1):114–159, 2015.

[14] Martin Grohe and Daniel Neuen. Recent advances on the graph isomorphism problem. In Konrad K. Dabrowski, Maximilien Gadouleau, Nicholas Georgiou, Matthew Johnson, George B. Mertzios, and Daniël Paulusma, editors, *Surveys in Combinatorics 2021*, London Mathematical Society Lecture Note Series, page 187–234. Cambridge University Press, 2021.

Copyright © 2022
Copyright for this paper is retained by authors

[15] Martin Grohe, Daniel Neuen, and Pascal Schweitzer. A faster isomorphism test for graphs of small degree. In Mikkel Thorup, editor, *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 89–100. IEEE Computer Society, 2018.

[16] Martin Grohe, Daniel Neuen, Pascal Schweitzer, and Daniel Wiebking. An improved isomorphism test for bounded-tree-width graphs. *ACM Trans. Algorithms*, 16(3):34:1–34:31, 2020.

[17] Martin Grohe, Daniel Neuen, and Daniel Wiebking. Isomorphism testing for graphs excluding small minors. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, November 16-19, 2020 (Virtual Conference)*, pages 625–637. IEEE Computer Society, 2020.

[18] Martin Grohe and Pascal Schweitzer. Isomorphism testing for graphs of bounded rank width. In Venkatesan Guruswami, editor, *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 1010–1029. IEEE Computer Society, 2015.

[19] John E. Hopcroft and Robert Endre Tarjan. A v^2 algorithm for determining isomorphism of planar graphs. *Inf. Process. Lett.*, 1(1):32–34, 1971.

[20] Neil Immerman and Eric Lander. Describing graphs: A first-order approach to graph canonization. In Alan L. Selman, editor, *Complexity Theory Retrospective: In Honor of Juris Hartmanis on the Occasion of His Sixtieth Birthday, July 5, 1988*, pages 59–81. Springer New York, New York, NY, 1990.

[21] Richard M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller and James W. Thatcher, editors, *Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, USA*, The IBM Research Symposia Series, pages 85–103. Plenum Press, New York, 1972.

[22] Sandra Kiefer and Daniel Neuen. The power of the Weisfeiler-Leman algorithm to decompose graphs. In Peter Rossmanith, Pinar Heggernes, and Joost-Pieter Katoen, editors, *44th International Symposium on Mathematical Foundations of Computer Science, MFCS 2019, August 26-30, 2019, Aachen, Germany*, volume 138 of *LIPIcs*, pages 45:1–45:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

[23] Sandra Kiefer, Ilia Ponomarenko, and Pascal Schweitzer. The Weisfeiler-Leman dimension of planar graphs is at most 3. *J. ACM*, 66(6):44:1–44:31, 2019.

[24] János Komlós and Endre Szemerédi. Topological cliques in graphs. II. *Combin. Probab. Comput.*, 5(1):79–90, 1996.

[25] Eugene M. Luks. Isomorphism of graphs of bounded valence can be tested in polynomial time. *J. Comput. Syst. Sci.*, 25(1):42–65, 1982.

[26] Gary L. Miller. Isomorphism testing for graphs of bounded genus. In Raymond E. Miller, Seymour Ginsburg, Walter A. Burkhard, and Richard J. Lipton, editors, *Proceedings of the 12th Annual ACM Symposium on Theory of Computing, April 28-30, 1980, Los Angeles, California, USA*, pages 225–235. ACM, 1980.

[27] Daniel Neuen. Graph isomorphism for unit square graphs. In Piotr Sankowski and Christos D. Zaroliagis, editors, *24th Annual European Symposium on Algorithms, ESA 2016, August 22-24, 2016, Aarhus, Denmark*, volume 57 of *LIPIcs*, pages 70:1–70:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016.

[28] Daniel Neuen. *The Power of Algorithmic Approaches to the Graph Isomorphism Problem.* PhD thesis, RWTH Aachen University, Aachen, Germany, 2019.

[29] Daniel Neuen. Hypergraph isomorphism for groups with restricted composition factors. In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference)*, volume 168 of *LIPIcs*, pages 88:1–88:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

[30] Ilia N. Ponomarenko. The isomorphism problem for classes of graphs. *Dokl. Akad. Nauk SSSR*, 304(3):552–556, 1989.

[31] Ilia N. Ponomarenko. The isomorphism problem for classes of graphs closed under contraction. *Journal of Soviet Mathematics*, 55(2):1621–1643, Jun 1991.

[32] Joseph J. Rotman. *An Introduction to the Theory of Groups*, volume 148 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, fourth edition, 1995.

[33] Pascal Schweitzer and Daniel Wiebking. A unifying method for the design of algorithms canonizing combinatorial objects. In Moses Charikar and Edith Cohen, editors, *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 1247–1258. ACM, 2019.

[34] Ákos Seress. *Permutation Group Algorithms*, volume 152 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2003.

[35] Boris Weisfeiler and Andrei Leman. The reduction of a graph to canonical form and the algebra which appears therein. *NTI, Series 2*, 1968. English translation by Grigory Ryabov available at `https://www.iti.zcu.cz/wl2018/pdf/wl_paper_translation.pdf`.

[36] Daniel Wiebking. Graph isomorphism in quasipolynomial time parameterized by treewidth. In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference)*, volume 168 of *LIPIcs*, pages 103:1–103:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

Copyright © 2022
Copyright for this paper is retained by authors