

Sparsification Lower Bounds for List H -Coloring

Hubie Chen

Birkbeck, University of London, Malet Street, Bloomsbury, UK
hubie@dcs.bbk.ac.uk

Bart M. P. Jansen 

Eindhoven University of Technology, The Netherlands
b.m.p.jansen@tue.nl

Karolina Okrasa 

University of Warsaw, Institute of Informatics, Poland
Warsaw University of Technology, Faculty of Mathematics and Information Science, Poland
k.okrasa@mini.pw.edu.pl

Astrid Pieterse 

Department of Computer Science, Humboldt-Universität zu Berlin, Germany
astrid.pieterse@informatik.hu-berlin.de

Paweł Rzażewski 

Warsaw University of Technology, Faculty of Mathematics and Information Science, Poland
University of Warsaw, Institute of Informatics, Poland
p.rzazewski@mini.pw.edu.pl

Abstract

We investigate the LIST H -COLORING problem, the generalization of graph coloring that asks whether an input graph G admits a homomorphism to the undirected graph H (possibly with loops), such that each vertex $v \in V(G)$ is mapped to a vertex on its list $L(v) \subseteq V(H)$. An important result by Feder, Hell, and Huang [JGT 2003] states that LIST H -COLORING is polynomial-time solvable if H is a so-called *bi-arc graph*, and NP-complete otherwise. We investigate the NP-complete cases of the problem from the perspective of polynomial-time sparsification: can an n -vertex instance be efficiently reduced to an equivalent instance of bitsize $\mathcal{O}(n^{2-\epsilon})$ for some $\epsilon > 0$? We prove that if H is not a bi-arc graph, then LIST H -COLORING does not admit such a sparsification algorithm unless $\text{NP} \subseteq \text{coNP/poly}$. Our proofs combine techniques from kernelization lower bounds with a study of the structure of graphs H which are not bi-arc graphs.

2012 ACM Subject Classification Mathematics of computing \rightarrow Graph coloring; Theory of computation \rightarrow Problems, reductions and completeness; Theory of computation \rightarrow Graph algorithms analysis; Theory of computation \rightarrow Parameterized complexity and exact algorithms

Keywords and phrases List H -Coloring, Sparsification, Constraint Satisfaction Problem

Digital Object Identifier 10.4230/LIPIcs.ISAAC.2020.58

Related Version The full version of the paper can be found at arxiv.org/abs/2009.08353 [4].

Funding *Bart M. P. Jansen*: Supported by NWO Gravitation grant “Networks”.

Karolina Okrasa: Supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme Grant Agreement no. 714704.

Astrid Pieterse: Supported by DFG Emmy Noether-grant (KR 4286/1), part of this research was supported by NWO Gravitation grant “Networks”.

Paweł Rzażewski: Supported by Polish National Science Centre grant no. 2018/31/D/ST6/00062.

Acknowledgements We acknowledge the productive atmosphere at Dagstuhl Seminar 19271 “Graph Colouring: from Structure to Algorithms”, where this work has been initiated.



© Hubie Chen, Bart M. P. Jansen, Karolina Okrasa, Astrid Pieterse, and Paweł Rzażewski;
licensed under Creative Commons License CC-BY

31st International Symposium on Algorithms and Computation (ISAAC 2020).

Editors: Yixin Cao, Siu-Wing Cheng, and Minming Li; Article No. 58; pp. 58:1–58:17

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Background and motivation. The LIST H -COLORING problem is a generalization of the classic graph coloring problem. For a fixed undirected graph H , possibly with self-loops, an input to the problem consists of an undirected graph G together with a list $L(v) \subseteq V(H)$ for each vertex $v \in V(G)$. The question is whether there is a list homomorphism from G to H : a mapping $f: V(G) \rightarrow V(H)$ such that $\{f(u), f(v)\} \in E(H)$ for all $\{u, v\} \in E(G)$, and such that $f(v) \in L(v)$ for all $v \in V(G)$. When H is a q -clique and $L(v) = V(H)$ for each vertex, LIST H -COLORING is equivalent to traditional graph q -colorability.

The classic computational complexity of LIST H -COLORING for other graphs H has been investigated, next to a long line of work for the non-list version of the problem [1, 3, 21, 23, 29, 32, 33, 39]. As the first step towards the dichotomy, Feder and Hell [12] proved that if H is reflexive (i.e., every vertex has a self-loop), then LIST H -COLORING is polynomial-time solvable if H is an interval graph, and NP-complete otherwise. Next, a dichotomy for irreflexive graphs H was proven by Feder, Hell, and Huang [13]: the problem is polynomial-time solvable if H is bipartite and additionally its complement is a circular-arc graph, and in all other cases the problem is NP-complete. It is interesting to mention that the subclass of bipartite graphs consisting of those which are complements of circular-arc graphs, was already studied by Trotter and Moore in the context of classifying some posets [41]. Finally, Feder, Hell, and Huang [14] defined a new class of geometric intersection graphs (potentially with loops), called *bi-arc graphs*, which encapsulates reflexive interval graphs and (irreflexive) bipartite co-circular-arc graphs. We postpone the definition of bi-arc graphs to Section 4.1. Feder, Hell, and Huang proved a powerful dichotomy theorem: LIST H -COLORING is polynomial-time solvable if H is a bi-arc graph, but NP-complete otherwise.

In this work we investigate LIST H -COLORING from the perspective of polynomial-time sparsification (cf. [5, 7, 27]). From this viewpoint, the goal is to develop a polynomial-time algorithm that maps a (potentially dense) n -vertex instance G to a smaller instance G' that can be encoded in $f(n)$ bits for some size function f , yet which has the same YES/NO answer as G . Observe that this is trivial if $f(n) = n^2$; we refer to a sparsification algorithm as *nontrivial* if it achieves a size bound of $f(n) \in \mathcal{O}(n^{2-\varepsilon})$ bits for some $\varepsilon > 0$.

The general quest for sparsification algorithms is motivated by the fact that they allow instances to be stored, manipulated, and solved more efficiently: since sparsification preserves the exact answer to the problem, it suffices to solve the sparsified instance. Our interest in sparsification for LIST H -COLORING has a number of motivations, which we now describe.

There is a growing list of problems for which the existence of nontrivial sparsification algorithms has been ruled out under the established assumption $\text{NP} \not\subseteq \text{coNP/poly}$, which includes VERTEX COVER [7], DOMINATING SET [27], FEEDBACK ARC SET [27], and TREewidth [25]. To the best of our knowledge, to date there is no non-trivial sparsification algorithm for *any* NP-hard problem that is defined on general graphs. Could it be that there is *no* natural NP-hard graph problem that admits a nontrivial sparsification algorithm? The surprising richness of problems that admit a polynomial kernelization, a desirable outcome in a different regime of efficient preprocessing (cf. [16, 19]), may tempt one to believe that for the right problem, something nontrivial can be done. In an attempt to identify a problem that admits nontrivial sparsification, we target the broad class of LIST H -COLORING decision problems.

A second motivation for studying LIST H -COLORING comes from its interpretation as a constraint satisfaction problem: an instance of LIST H -COLORING corresponds to a CSP that has a variable for each vertex of the input graph G , which has to be assigned a value from the set $V(H)$. For each edge $\{u, v\}$ of G there is a constraint that the value assigned to u should

be a neighbor (in graph H) of the value assigned to v , and for each vertex $v \in V(G)$ there is a constraint that the value of v belongs to $L(v)$. Hence any NP-hard LIST H -COLORING problem translates into a CSP with a non-Boolean domain in which constraints have arity at most two. Recent work [5, 30] has led to a number of nontrivial advances in the study of sparsification for CSPs with a Boolean domain. A natural next step in that line of research is to target non-Boolean CSPs, of which the LIST H -COLORING problems form a rich subset.

The last motivation for studying sparsification for LIST H -COLORING is that it forms the logical next step in the study of sparsification for coloring problems. Recent work [26] showed that GRAPH (LIST) q -COLORABILITY does not admit nontrivial polynomial-time sparsification for $q \geq 3$ unless $\text{NP} \subseteq \text{coNP/poly}$, but left the case of LIST H -COLORING open.

Our results. We prove that for all undirected, possibly non-simple, graphs H for which LIST H -COLORING is NP-complete, the problem does not admit nontrivial sparsification unless an unlikely complexity-theoretic collapse occurs. Our proofs combine techniques from kernelization lower bounds with a careful analysis of the common structures of hard graphs H . To state our sparsification lower bounds in full generality, we use the notion of generalized kernelization (see Definition 2), where the number of vertices n of the instance plays the role of the complexity parameter k . A generalized kernelization for LIST H -COLORING of size $f(n)$ is therefore a polynomial-time algorithm that maps any n -vertex input G , to an equivalent instance (of a potentially different but fixed decision problem) of bitsize $f(n)$. Since a polynomial-time sparsification algorithm mapping to instances of bitsize $f(n)$ yields a generalized kernelization of size $f(n)$, lower bounds on the latter also apply to the former.

► **Theorem 1.** *If H is an undirected graph that is not a bi-arc graph, possibly with loops, then LIST H -COLORING parameterized by the number of vertices n admits no generalized kernel of size $\mathcal{O}(n^{2-\varepsilon})$ for any $\varepsilon > 0$, unless $\text{NP} \subseteq \text{coNP/poly}$.*

The techniques employed in the proof of Theorem 1 are rather different from those in the NP-completeness proof for the hard cases of LIST H -COLORING. Feder, Hell, and Huang [14] establish the NP-completeness of LIST H -COLORING when H is not a bi-arc graph, by reducing from 3-COLORING. They build gadgets in LIST H -COLORING instances to mimic the effect of a normal edge in 3-COLORING, and then replace each edge with such a gadget. Although 3-COLORING is known not to admit any nontrivial sparsification unless $\text{NP} \subseteq \text{coNP/poly}$ [28], the mentioned NP-completeness reduction does not transfer this lower bound from 3-COLORING to LIST H -COLORING: as the reduction introduces a gadget (with new vertices) for every edge of the 3-COLORING instance, it blows up the number of variables.

Our sparsification lower bound therefore follows a different route. We introduce a technical annotated version of the LIST P_4 -COLORING problem. For this annotated problem, we prove a sparsification lower bound via cross-composition [2], a technique from kernelization lower bounds. We give a polynomial-time algorithm that embeds a sequence of t^2 instances of the CLIQUE problem, on n vertices each, into a single instance (G', L') of ANNOTATED LIST P_4 -COLORING, on $\mathcal{O}(t \cdot n^{\mathcal{O}(1)})$ vertices, which acts as the logical OR of the CLIQUE inputs: there is a list coloring if and only if at least one CLIQUE instance has a solution. The fact that the information from t^2 distinct inputs is packed into a single instance of $\mathcal{O}(t \cdot n^{\mathcal{O}(1)})$ vertices, means that the embedding is very efficient: the t^2 n -vertex instances of CLIQUE carry $t^2 \cdot n^2$ bits of information (for each instance, which edges are present?), while G' has $t^2 \cdot n^{\mathcal{O}(1)}$ potential edges, and therefore carries $t^2 \cdot n^{\mathcal{O}(1)}$ bits of information. Applying this reduction for t a polynomial in n whose degree depends on the constant in $n^{\mathcal{O}(1)}$, this intuitively

implies that G' cannot be sparsified without losing information. Via the framework of cross-composition [2] we get the formal result that ANNOTATED LIST P_4 -COLORING parameterized by the number of vertices n does not admit a generalized kernelization of size $\mathcal{O}(n^{2-\varepsilon})$ for any $\varepsilon > 0$ unless $\text{NP} \subseteq \text{coNP}/\text{poly}$.

To transfer the lower bound for ANNOTATED LIST P_4 -COLORING to LIST H -COLORING for all graphs H which are not bi-arc, we first use a reduction inspired by Feder, Hell, and Huang [14], to reduce to the case of bipartite graphs H . Then we investigate the common structure of simple bipartite non-bi-arc graphs H , which are known to be the simple bipartite graphs H whose complement is not a circular-arc graph [14]. We uncover a common structure of such graphs which can be used to prove the incompressibility of the related LIST H -COLORING problems: we prove all such graphs H contain five vertices (a, b, c, d, e) such that $H[\{a, b, c, d\}]$ is an induced P_4 , the open neighborhoods $N_H(a), N_H(c)$, and $N_H(e)$ are incomparable (i.e., none of them is contained in another), and such that also the open neighborhoods $N_H(b), N_H(d)$ are incomparable. This 5-tuple in a bipartite graph H is sufficient to prove hardness of sparsification, which we consider one of the main contributions of the paper: We prove that the 5-tuple can be used to implement certain gadgets to enforce pairs of vertices to receive different colors in LIST H -COLORING. By applying these gadgets sparingly – and not for all edges – we reduce ANNOTATED LIST P_4 -COLORING to LIST H -COLORING without blowing up the number of vertices, and obtain Theorem 1.

Related work. More background on homomorphisms and H -COLORING can be found in the textbook by Hell and Nešetřil [22], or the survey by Hahn and Tardif [20]. The classical complexity of H -COLORING has also been investigated when restricted to planar [31], minor-closed [11], and bounded-degree [17, 40] input graphs G . The complexity of LIST H -COLORING was investigated for bounded-degree graphs [15]. There is also an interesting line of research concerning the descriptive and space complexity [6, 8, 9]. Finally, the fine-grained complexity of both variants was also investigated [10, 18, 34, 36, 37].

Organization. Section 2 contains preliminaries on kernelization and graphs. In Section 3 we present a sparsification lower bound for an annotated version of LIST P_4 -COLORING, which forms the keystone of our hardness results. In Section 4 we analyze the structure of hard graphs H , and use that structure to build certain gadgets. These allow us to reduce the annotated problem to standard LIST H -COLORING problems and prove Theorem 1.

2 Preliminaries

To denote the set of numbers 1 to n , we use the following notation: $[n] := \{1, \dots, n\}$. For a set S we use the notation $\binom{S}{k} := \{S' \subseteq S \mid |S'| = k\}$ to denote the set of all size- k subsets of S , and we define $2^S := \bigcup_{k=0}^{|S|} \binom{S}{k}$. We use the notation $S^k := \{(s_1, \dots, s_k) \mid s_1, \dots, s_k \in S\}$ to denote the set of all k -tuples with elements from S . In particular, $[n]^2$ denotes all 2-tuples of elements from $[n]$.

Graphs. All graphs considered in this paper are finite and undirected, and do not have parallel edges. We allow self-loops, unless explicitly stated otherwise. The vertex set and the edge set of G are denoted by $V(G)$ and $E(G)$, respectively. An edge $\{u, v\} \in E(G)$ is denoted shortly by uv , and by vv we denote the loop on the vertex v . For $v \in V(G)$, by $N_G(v)$ we denote the *open neighborhood* of v , i.e., the set $\{u \mid uv \in E(G)\}$. The closed neighborhood of v is $N_G[v] := N_G(v) \cup \{v\}$. For $S \subseteq V(G)$, by $G[S]$ we denote the subgraph of G induced

by S . A *proper q -coloring* of G is a function $f: V(G) \rightarrow [q]$ such that $f(u) \neq f(v)$ for all $uv \in E(G)$. Let G and H be graphs. We say that G is *H -colorable* if there exists a function $f: V(G) \rightarrow V(H)$ such that for all $uv \in E(G)$ it holds that $f(u)f(v) \in E(H)$. Such a function is also called a *homomorphism* from G to H . Note that a graph G has a homomorphism to the complete graph K_q if and only if G is (properly) q -colorable. If f is a homomorphism from G to H , then we denote it by $f: G \rightarrow H$. We write $G \rightarrow H$ to indicate that some homomorphism from G to H exists. For a graph G and *lists* $L: V(G) \rightarrow 2^{V(H)}$, a *list homomorphism* from (G, L) to H is a homomorphism $f: G \rightarrow H$, such that for every $v \in V(G)$ it holds that $f(v) \in L(v)$. We write $f: (G, L) \rightarrow H$ if f is a list homomorphism from (G, L) to H , and $(G, L) \rightarrow H$ if some $f: (G, L) \rightarrow H$ exists.

Parameterized complexity. A *parameterized problem* \mathcal{Q} is a subset of $\Sigma^* \times \mathbb{N}$, where Σ is a finite alphabet.

► **Definition 2** (Generalized kernel [2]). *Let $\mathcal{Q}, \mathcal{Q}' \subseteq \Sigma^* \times \mathbb{N}$ be parameterized problems and let $h: \mathbb{N} \rightarrow \mathbb{N}$ be a computable function. A generalized kernel for \mathcal{Q} into \mathcal{Q}' of size $h(k)$ is an algorithm that, on input $(x, k) \in \Sigma^* \times \mathbb{N}$, takes time polynomial in $|x| + k$ and outputs an instance (x', k') such that: (i) $|x'|$ and k' are bounded by $h(k)$, and (ii) $(x', k') \in \mathcal{Q}'$ if and only if $(x, k) \in \mathcal{Q}$. A generalized kernel is a kernel for \mathcal{Q} if $\mathcal{Q} = \mathcal{Q}'$.*

In our applications, the complexity parameter k will be the number of vertices n . We will use the framework of cross-composition, introduced by Bodlaender, Jansen, and Kratsch [2], to establish kernelization lower bounds.

► **Definition 3** (Polynomial equivalence relation, [2, Def. 3.1]). *An equivalence relation \mathcal{R} on Σ^* is called a polynomial equivalence relation if the following conditions hold.*

- *There is an algorithm that, given two strings $x, y \in \Sigma^*$, decides whether x and y belong to the same equivalence class in time polynomial in $|x| + |y|$.*
- *For any finite set $S \subseteq \Sigma^*$ the equivalence relation \mathcal{R} partitions the elements of S into a number of classes that is polynomially bounded in the size of the largest element of S .*

► **Definition 4** (Cross-composition, [2, Def. 3.7]). *Let $L \subseteq \Sigma^*$ be a language, let \mathcal{R} be a polynomial equivalence relation on Σ^* , let $\mathcal{Q} \subseteq \Sigma^* \times \mathbb{N}$ be a parameterized problem, and let $f: \mathbb{N} \rightarrow \mathbb{N}$ be a function. An OR-cross-composition of L into \mathcal{Q} (with respect to \mathcal{R}) of cost $f(t)$ is an algorithm that, given t instances $x_1, x_2, \dots, x_t \in \Sigma^*$ of L belonging to the same equivalence class of \mathcal{R} , takes time polynomial in $\sum_{i=1}^t |x_i|$ and outputs an instance $(y, k) \in \Sigma^* \times \mathbb{N}$ such that:*

- *The parameter k is bounded by $\mathcal{O}(f(t) \cdot (\max_i |x_i|)^c)$, where c is some constant independent of t , and*
- *instance $(y, k) \in \mathcal{Q}$ if and only if there is an $i \in [t]$ such that $x_i \in L$.*

► **Theorem 5** ([2, Theorem 3.8]). *Let $L \subseteq \Sigma^*$ be a language, let $\mathcal{Q} \subseteq \Sigma^* \times \mathbb{N}$ be a parameterized problem, and let d, ε be positive reals. If L is NP-hard under Karp reductions, has an OR-cross-composition into \mathcal{Q} with cost $f(t) = t^{1/d+o(1)}$, where t denotes the number of instances, and \mathcal{Q} has a polynomial (generalized) kernelization with size bound $\mathcal{O}(k^{d-\varepsilon})$, then $\text{NP} \subseteq \text{coNP/poly}$.*

We will refer to an OR-cross-composition of cost $f(t) = \sqrt{t} \log(t)$ as a *degree-2 cross-composition*. By Theorem 5, a degree-2 cross-composition can be used to rule out generalized kernels of size $\mathcal{O}(k^{2-\varepsilon})$ and thus provides a way to obtain sparsification lower bounds. Generalized kernelization lower bounds can be transferred using the notion of linear-parameter transformations.

► **Definition 6** (Linear-parameter transformation). *Let $\mathcal{P}, \mathcal{Q} \subseteq \Sigma^* \times \mathbb{N}$ be two parameterized problems. A linear-parameter transformation from \mathcal{P} to \mathcal{Q} is a polynomial-time algorithm that, given an instance $(x, k) \in \Sigma^* \times \mathbb{N}$ of \mathcal{P} , outputs an instance $(x', k') \in \Sigma^* \times \mathbb{N}$ of \mathcal{Q} such that the following holds: (i) $(x, k) \in \mathcal{P}$ if and only if $(x', k') \in \mathcal{Q}$, and (ii) $k' \in \mathcal{O}(k)$.*

It is well-known [2] that the existence of a linear-parameter transformation from problem \mathcal{P} to \mathcal{Q} implies that any generalized kernelization lower bound for \mathcal{P} , also holds for \mathcal{Q} .

3 Lower bound for Annotated List P_4 -Coloring

We prove a sparsification lower bound for the following problem, where we take P_4 to be the graph on vertices $\{a, b, c, d\}$ with edges ab, bc, cd .

ANNOTATED LIST P_4 -COLORING

Input: A tuple (G, L, \mathcal{S}, F) , such that G is a simple undirected bipartite graph with bipartition $V(G) = V_1 \cup V_2$, $L: V(G) \rightarrow 2^{\{a, b, c, d\}}$ with $L(v) \subseteq \{a, c\}$ for all $v \in V_1$ and $L(v) \subseteq \{b, d\}$ for all $v \in V_2$, $\mathcal{S} = S_1, \dots, S_m$ is a sequence such that $S_i \subseteq V_1$ for each $i \in [m]$ and \mathcal{S} satisfies $\sum_{i=1}^m |S_i| \leq 3|V(G)|$, and $F \subseteq \binom{V_1}{2} \cup \binom{V_2}{2}$ is a set with $|F| \leq |V(G)|$.

Question: Does G admit a homomorphism $f: V(G) \rightarrow \{a, b, c, d\}$ to the graph P_4 with $f(v) \in L(v)$ for all $v \in V(G)$, such that for all $i \in [m]$ there is a vertex $v \in S_i$ with $f(v) \neq c$, and such that for all $\{u, v\} \in F$ we have $f(u) \neq f(v)$?

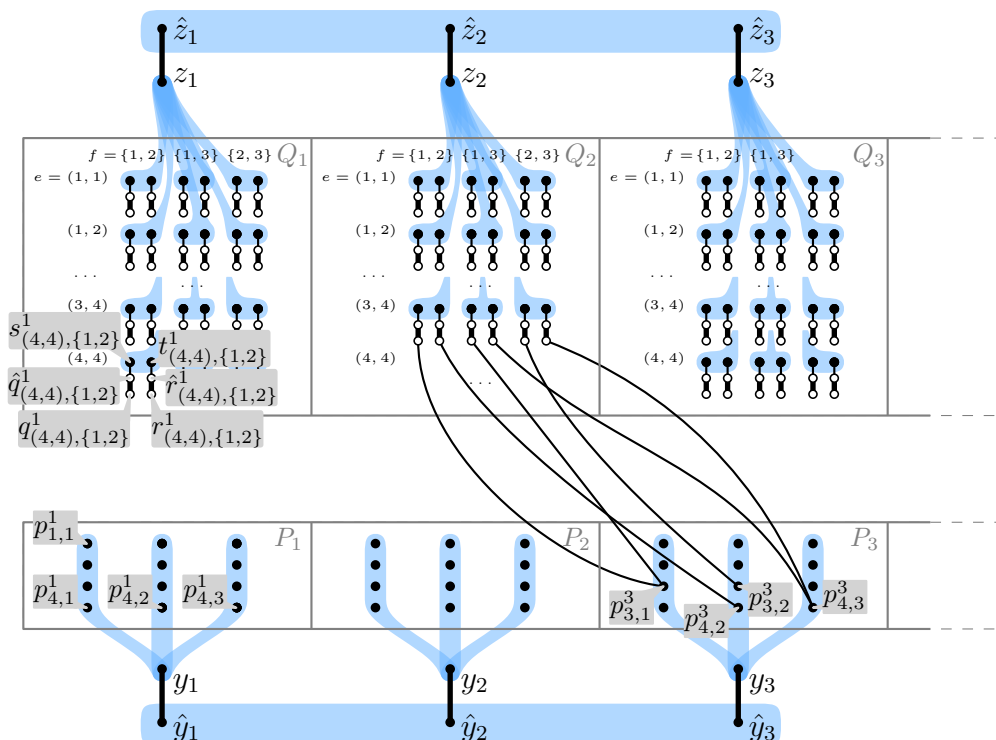
Intuitively, the annotations allow one to express two types of additional constraints on the coloring f . Using a set S_i , one can enforce that at least one vertex is not colored c . Using a pair $\{u, v\} \in F$, one can enforce that u and v do not receive the same color. While the latter can easily be expressed by simply inserting an edge between u and v in a K_q -COLORING instance, this needs a nontrivial gadget for general graphs H .

► **Lemma 7.** *ANNOTATED LIST P_4 -COLORING parameterized by the number of vertices n admits no generalized kernel of size $\mathcal{O}(n^{2-\varepsilon})$ for any $\varepsilon > 0$, unless $\text{NP} \subseteq \text{coNP/poly}$.*

Proof. We will prove this lower bound by giving a degree-2 cross-composition from CLIQUE to ANNOTATED LIST P_4 -COLORING. We define a polynomial equivalence relation \mathcal{R} on instances of CLIQUE. Let any two instances that ask for a clique that is larger than their respective number of vertices be equivalent; these are always no-instances. Let two instances of CLIQUE be equivalent under \mathcal{R} , when the input graphs have same number of vertices and the problems ask for a clique of the same size. It is easy to verify that \mathcal{R} is indeed a polynomial equivalence relation.

By duplicating one of the inputs multiple times as needed, we can assume the number of inputs to the cross-composition is a square. Therefore, assume we are given t instances of CLIQUE, such that $t' := \sqrt{t}$ is integer and such that each instance has n vertices and asks for a size- k clique. Enumerate the given input instances as $X_{i,j}$ for $i, j \in [t']$ and let $G_{i,j}$ denote the corresponding graph. Label the vertices in each instance arbitrarily as x_1, \dots, x_n . We show how to create an instance (G, L, \mathcal{S}, F) that is a yes-instance for ANNOTATED LIST P_4 -COLORING if and only if at least one of the given instances for CLIQUE is a yes-instance. Refer to Figure 1 for a sketch.

1. For each $j \in [t']$, $\ell \in [n]$, and $m \in [k]$ create a vertex $p_{\ell, m}^j$. Let $L(p_{\ell, m}^j) := \{a, c\}$. Let P_j contain all created vertices $p_{\ell, m}^j$ for $\ell \in [n]$, $m \in [k]$. Let $P := \bigcup_{j \in [t']} P_j$.



■ **Figure 1** A sketch of the created graph G , for $n = 4$, and $k = 3$ where $x_3x_4 \notin E(G_{2,3})$. Edges between P and Q are omitted, except for the edges that result from the fact that $x_3x_4 \notin E(G_{2,3})$. A fat edge between u and v indicates that $\{u, v\} \in F$. Vertex sets contained in \mathcal{S} are marked in blue. White vertices have lists $\{b, d\}$ while black vertices have list $\{a, c\}$. Note that the constructed graph is bipartite with the white and black vertices as partite sets.

2. For each $f \in \binom{[k]}{2}$, each $e = (e_1, e_2) \in [n]^2$, and each $i \in [t']$, create vertices $q_{e,f}^i, r_{e,f}^i, \hat{q}_{e,f}^i, \hat{r}_{e,f}^i, s_{e,f}^i$, and $t_{e,f}^i$. Let $Q_i := \{q_{e,f}^i, r_{e,f}^i, \hat{q}_{e,f}^i, \hat{r}_{e,f}^i, s_{e,f}^i, t_{e,f}^i \mid f \in \binom{[k]}{2}, e \in [n]^2\}$. Note that Q_i contains $6\binom{k}{2}$ vertices for each ordered pair of vertices in an n -vertex graph; these pairs model edges and self-loops. Let $Q := \bigcup_{i \in [t']} Q_i$. Now let $L(q_{e,f}^i) := L(r_{e,f}^i) := L(\hat{q}_{e,f}^i) := L(\hat{r}_{e,f}^i) := \{b, d\}$ and $L(s_{e,f}^i) := L(t_{e,f}^i) := \{a, c\}$.
3. For each $f \in \binom{[k]}{2}$, each $e = (e_1, e_2) \in [n]^2$, and each $i \in [t']$, do the following. Connect vertex $\hat{q}_{e,f}^i$ to vertex $s_{e,f}^i$, and connect vertex $\hat{r}_{e,f}^i$ to vertex $t_{e,f}^i$. This ensures that when $\hat{q}_{e,f}^i$ (respectively, $\hat{r}_{e,f}^i$) gets color d , then $s_{e,f}^i$ (respectively $t_{e,f}^i$) always gets color c , since vertex c is the unique neighbor of vertex d in P_4 . If however $\hat{q}_{e,f}^i$ gets color b , then $s_{e,f}^i$ can receive color a or c . Add the pairs $\{q_{e,f}^i, \hat{q}_{e,f}^i\}$ and $\{\hat{r}_{e,f}^i, r_{e,f}^i\}$ to F . Verify that when both $q_{e,f}^i$ and $r_{e,f}^i$ get color b , then $s_{e,f}^i$ and $t_{e,f}^i$ must get color c .

Recall that the goal of the construction is to ensure that the ANNOTATED LIST P_4 -COLORING instance (G, L, \mathcal{S}, F) acts as the logical OR of the CLIQUE instances $X_{i,j}$, so that G has a coloring respecting the lists and annotations if and only if some input graph $G_{i,j}$ has a clique of size k . The part of G constructed so far allows colorings of G to encode the vertex set of a k -clique through its behavior on P . Finding a proper list coloring of G entails highlighting vertices from one set P_j that correspond to a clique in instance $X_{i,j}$ for some $i \in [t']$. The highlighting property will be enforced by ensuring at least one vertex in each set $\{p_{\ell,m}^j \mid \ell \in [n]\}$ for $m \in [k]$ receives color a . The index of the vertex that is colored a encodes the m -th vertex in the clique to which the coloring corresponds. The vertices in Q_i are then used to verify that the selected vertices form a clique in $G_{i,j}$. The next steps add additional vertices and edges, in order to achieve these properties.

4. For each $i, j \in [t']$, consider instance $X_{i,j}$. For all $f \in \binom{[k]}{2}$ and $e = (e_1, e_2) \in [n]^2$, connect vertex p_{e_1, f_1}^j to $q_{e, f}^i$ and connect p_{e_2, f_2}^j to $r_{e, f}^i$ whenever $x_{e_1} x_{e_2} \notin E(G_{i,j})$. Here $f_1 < f_2$ are such that $f = \{f_1, f_2\}$. Observe that in particular (since $G_{i,j}$ is a simple graph), we have that $x_{e_1} x_{e_1} \notin E(G_{i,j})$ for all $e_1 \in [n]$. Observe also that each vertex $q_{e, f}^i, r_{e, f}^i$ has a *unique* neighbor in P_j for each $j \in [t']$.

The above step will allow using the coloring of vertices $s_{e, f}^i$ and $t_{e, f}^i$ to verify that the vertices selected in P_j correspond to a clique: when $x_{e_1} x_{e_2}$ is not an edge, they will ensure that we cannot select both.

5. Add vertices y_j and \hat{y}_j for all $j \in [t']$ and let $Y := \{y_j \mid j \in [t']\}$, $\hat{Y} := \{\hat{y}_j \mid j \in [t']\}$. Let $L(y_j) := L(\hat{y}_j) := \{a, c\}$ for all $j \in [t']$.
6. Similarly, add vertices z_i and \hat{z}_i for all $i \in [t']$ and let $Z := \{z_i \mid i \in [t']\}$, $\hat{Z} := \{\hat{z}_i \mid i \in [t']\}$. Let $L(z_i) := L(\hat{z}_i) := \{a, c\}$.
7. Add the sets \hat{Y} and \hat{Z} to \mathcal{S} . Furthermore, for all $i \in [t']$, add $\{y_i, \hat{y}_i\}$ and $\{z_i, \hat{z}_i\}$ to F .

The steps above ensure that at least one vertex $y_j \in Y$ receives color c and at least one vertex in $z_i \in Z$ receives color c . This will indicate that instance $X_{i,j}$ is selected. We will now put further constraints on the coloring of P_j and Q_i when they correspond to a selected instance.

8. For all $j \in [t']$, $m \in [k]$, we add the set $\{y_j\} \cup \{p_{\ell, m}^j \mid \ell \in [n]\}$ to \mathcal{S} .
9. For all $i \in [t']$, for all $f \in \binom{[k]}{2}$ and $e \in [n]^2$, add the set $\{s_{e, f}^i, t_{e, f}^i, z_i\}$ to \mathcal{S} .

This concludes the construction of G , L , \mathcal{S} and F . Let us start by counting the number of vertices in G :

$$|V(G)| = \underbrace{t' \cdot n \cdot k}_{|P|} + \underbrace{t' \cdot (n^2 \cdot \binom{k}{2}) \cdot 6}_{|Q|} + \underbrace{t' + t' + t' + t'}_{|Y| + |\hat{Y}| + |Z| + |\hat{Z}|} = \mathcal{O}(\sqrt{t} \cdot n^2 \cdot k^2).$$

Observe that hereby $|V(G)|$ is properly bounded for a degree-2 cross composition.

We continue by showing that G is a valid instance of ANNOTATED LIST P_4 -COLORING. Verify that G is bipartite with bipartition $V_1 = P \cup Y \cup \hat{Y} \cup Z \cup \hat{Z} \cup \{s_{e, f}^i, t_{e, f}^i \mid f \in \binom{[k]}{2}, e \in [n]^2, i \in [t']\}$ and $V_2 = \{q_{e, f}^i, r_{e, f}^i, \hat{q}_{e, f}^i, \hat{r}_{e, f}^i \mid f \in \binom{[k]}{2}, e \in [n]^2, i \in [t']\}$. Hence, V_1 contains all vertices whose lists are a subset of $\{a, c\}$ and V_2 contains all remaining vertices, and it can be verified that the lists of these vertices are a subset of $\{b, d\}$. Observe that indeed each set in F is a subset of either V_1 or V_2 , and each set in \mathcal{S} is a subset of V_1 .

Furthermore, it is straightforward to verify that $|F| \leq |V(G)|$ as promised for ANNOTATED LIST P_4 -COLORING (note that we only add elements to F in Steps 3 and 7). We can also verify that

$$\sum_{S \in \mathcal{S}} |S| \leq \underbrace{2 \cdot t'}_{\text{Step 7}} + \underbrace{t' \cdot k \cdot (n+1)}_{\text{Step 8}} + \underbrace{t' \cdot n^2 \cdot \binom{k}{2} \cdot 3}_{\text{Step 9}} \leq 3|V(G)|.$$

As such, we have created a valid instance of ANNOTATED LIST P_4 -COLORING. The next two claims show that the constructed graph G indeed acts as the logical OR of the given input instances.

▷ **Claim 8.** If some input graph G_{i^*, j^*} has a clique of size k , then G is annotated P_4 -colorable.

Proof. Let such $i^*, j^* \in [t']$ be given, we create an annotated P_4 -coloring $h: V(G) \rightarrow \{a, b, c, d\}$ for G . First of all, for all $j \neq j^*$ with $j \in [t']$, let $h(y_j) := a$ and let $h(\hat{y}_j) := c$. Let $h(y_{j^*}) := c$ and let $h(\hat{y}_{j^*}) := a$. Similarly, for $i \neq i^*$ we let $h(z_i) := a$ and let $h(\hat{z}_i) := c$. Furthermore define $h(z_{i^*}) := c$ and $h(\hat{z}_{i^*}) := a$. Hereby, not all vertices in \hat{Y} have color c , and not all vertices in \hat{Z} have color c , such that we satisfy the sets added to \mathcal{S} in Step 7 of the construction.

For all $p \in P_j$ for $j \neq j^*$, let $h(p) := c$. Furthermore, for all $e \in [n]^2$, $f \in \binom{[k]}{2}$ and $i \neq i^*$ with $i \in [t']$, we define $h(q_{e,f}^i) := h(r_{e,f}^i) = b$, $h(\hat{q}_{e,f}^i) := h(\hat{r}_{e,f}^i) = d$, and $h(s_{e,f}^i) := h(t_{e,f}^i) = c$.

It remains to color the vertices in P_{j^*} and Q_{i^*} . Let $K = \{x_{i_1}, \dots, x_{i_k}\}$ be a clique in G_{i^*, j^*} of size k . For $m \in [k]$, $\ell \in [n]$ let $h(p_{\ell, m}^{j^*}) := a$ if $i_m = \ell$. Otherwise, let $h(p_{\ell, m}^{j^*}) := c$. In this way, for each $m \in [k]$, the set $\{y_{j^*}\} \cup \{p_{\ell, m}^{j^*} \mid \ell \in [n]\}$ contains a vertex that receives color a , as desired. We now extend this coloring to Q_{i^*} . Let $e = (e_1, e_2) \in [n]^2$ and let $f \in \binom{[k]}{2}$ such that $f = \{f_1, f_2\}$ for $f_1 < f_2$. Let $h(q_{e,f}^{i^*}) := b$ if the unique neighbor of $q_{e,f}^{i^*}$ in P_{j^*} has color a . Otherwise, let $h(q_{e,f}^{i^*}) := d$. We color $r_{e,f}^{i^*}$ in the same way, thus $h(r_{e,f}^{i^*}) := b$ if its unique neighbor in P_{j^*} has color a , and $h(r_{e,f}^{i^*}) := d$ otherwise. Color $\hat{q}_{e,f}^{i^*}$ with the only color in $\{b, d\} \setminus \{h(q_{e,f}^{i^*})\}$ and similarly color $\hat{r}_{e,f}^{i^*}$ with the only color in $\{b, d\} \setminus \{h(r_{e,f}^{i^*})\}$. Finally, let $h(s_{e,f}^{i^*}) := c$ if $h(\hat{q}_{e,f}^{i^*}) = d$ and let $h(s_{e,f}^{i^*}) := a$ otherwise. Similarly, let $h(t_{e,f}^{i^*}) := c$ if $h(\hat{r}_{e,f}^{i^*}) = d$ and let $h(t_{e,f}^{i^*}) := a$ otherwise. This concludes the definition of h . It remains to show that h is a valid annotated P_4 -coloring of G . We split this into three parts.

First of all, we verify that each $S \in \mathcal{S}$ contains a vertex that does not get color c . For \hat{Y} and \hat{Z} this was verified before. Consider a set $\{y_j\} \cup \{p_{\ell, m}^{j^*} \mid \ell \in [n]\}$ added in Step 8. Observe that if $j \neq j^*$ then y_j has color a and we are done. Otherwise, by definition, we have $h(p_{\ell, m}^{j^*}) := a$ and thus indeed this set has a vertex of color a . Now consider a set $\{s_{e,f}^i, t_{e,f}^i, z_i\}$ added in Step 9. If $i \neq i^*$, vertex z_i has color a and we are done. Otherwise if $i = i^*$, we claim that it cannot be the case that $h(s_{e,f}^{i^*}) = h(t_{e,f}^{i^*}) = c$. Suppose towards a contradiction that indeed both these vertices have color c . By the choice of our coloring, this implies that $h(\hat{q}_{e,f}^{i^*}) = h(\hat{r}_{e,f}^{i^*}) = d$ and thus $h(q_{e,f}^{i^*}) = h(r_{e,f}^{i^*}) = b$. Letting $e = (e_1, e_2) \in [n]^2$ and $f = \{f_1, f_2\}$ for $f_1 < f_2$, that means that $q_{e,f}^{i^*}$ and $r_{e,f}^{i^*}$ have their unique neighbor in P_{j^*} of color a , implying $h(p_{e_1, f_1}^{j^*}) = h(p_{e_2, f_2}^{j^*}) = a$. So these edges were constructed in Step 4, implying $x_{e_1} x_{e_2} \notin E(G_{i^*, j^*})$. Since $x_{e_1} \in K$ and $x_{e_2} \in K$, this contradicts that K is a clique.

Secondly, verify that for all pairs in $\{u, v\} \in F$, $h(u) \neq h(v)$: we only add sets to F in Steps 3 and 7. We always ensure in the construction that if $\{u, v\} \in F$, the two vertices get different colors.

Thirdly, we verify the coloring of endpoints of edges in G . First of all, consider the edges added in Step 3 and observe that we always color the endpoints properly in the description above: if $\hat{q}_{e,f}^i$ gets color d , we color $s_{e,f}^i$ with c which is allowed; if $\hat{q}_{e,f}^i$ has color b , we use color a in $s_{e,f}^i$ which is again fine. One may verify that the same holds for edges $\hat{r}_{e,f}^i t_{e,f}^i$. Now consider the edges between a vertex $u \in P$ and $v \in Q$. If $u \notin P_{i^*}$ it follows that $h(u) = c$. Since by the lists, $h(v) \in \{b, d\}$ this implies that this edge is properly colored. Similarly, if $v \notin Q_{j^*}$ we obtain $h(v) = b$ and since $h(u) \in \{a, c\}$ we are again done. If $u \in P_{j^*}$ and $v \in P_{i^*}$ one may observe that the edge uv is properly colored by definition: v has color d only if it has no neighbors of color a (and $h(u) \in \{a, c\}$ thus implies $h(u) = c$), and otherwise v has color b such that the edge is again properly colored by $h(u) \in \{a, c\}$. \triangleleft

\triangleright Claim 9. If G has an annotated P_4 -coloring h , then there exist $i^*, j^* \in [t']$ such that G_{i^*, j^*} has a clique of size k .

Proof. Since $\hat{Y}, \hat{Z} \in \mathcal{S}$, there exist $i^*, j^* \in [t']$ such that $h(\hat{y}_{j^*}) \neq c$ and $h(\hat{z}_{i^*}) \neq c$, implying by the lists that $h(\hat{y}_{j^*}) = h(\hat{z}_{i^*}) = a$. Since $\{\hat{y}_{j^*}, y_{j^*}\} \in F$ and $\{\hat{z}_{i^*}, z_{i^*}\} \in F$ (by Step 7) we obtain that $h(y_{j^*}) = h(z_{i^*}) = c$. Now since $\{y_{j^*}\} \cup \{p_{\ell, m}^{j^*} \mid \ell \in [n]\} \in \mathcal{S}$ for all $m \in [k]$, it follows that for all $m \in [k]$, there exists $i_m \in [n]$ such that $h(p_{i_m, m}^{j^*}) = a$. Let x_1, \dots, x_n be the vertices of G_{i^*, j^*} , define $K := \{x_{i_1}, \dots, x_{i_k}\}$. We show that K is a size- k clique in G_{i^*, j^*} by showing that $x_{i_m} x_{i_{m'}}$ is an edge for all $m \neq m'$. Observe that this then also proves that all selected vertices are distinct as the input graphs have no self-loops.

Let $m, m' \in [k]$. Without loss of generality let $m < m'$. Suppose towards a contradiction that $x_{i_m} x_{i_{m'}} \notin E(G_{i^*, j^*})$. Then, in Step 4, we added the edges $p_{i_m, m}^{j^*} q_{(i_m, i_{m'}), \{m, m'\}}^{i^*}$ and $p_{i_{m'}, m'}^{j^*} r_{(i_m, i_{m'}), \{m, m'\}}^{i^*}$. Note that since we choose $x_{i_m}, x_{i_{m'}} \in K$, it must hold that $h(p_{i_m, m}^{j^*}) = h(p_{i_{m'}, m'}^{j^*}) = a$. Since b is the only neighbor of a in the P_4 , we get $h(q_{(i_m, i_{m'}), \{m, m'\}}^{i^*}) = h(r_{(i_m, i_{m'}), \{m, m'\}}^{i^*}) = b$. Since in Step 3 we added $\{q_{(i_m, i_{m'}), \{m, m'\}}^{i^*}, \hat{q}_{(i_m, i_{m'}), \{m, m'\}}^{i^*}\}$ and $\{r_{(i_m, i_{m'}), \{m, m'\}}^{i^*}, \hat{r}_{(i_m, i_{m'}), \{m, m'\}}^{i^*}\}$ to F , we obtain $h(\hat{q}_{(i_m, i_{m'}), \{m, m'\}}^{i^*}) = h(\hat{r}_{(i_m, i_{m'}), \{m, m'\}}^{i^*}) = d$. Since $\hat{q}_{(i_m, i_{m'}), \{m, m'\}}^{i^*} s_{(i_m, i_{m'}), \{m, m'\}}^{i^*}$ and $\hat{r}_{(i_m, i_{m'}), \{m, m'\}}^{i^*} t_{(i_m, i_{m'}), \{m, m'\}}^{i^*}$ are edges in G (also added in Step 3), we get that $h(s_{(i_m, i_{m'}), \{m, m'\}}^{i^*}) = h(t_{(i_m, i_{m'}), \{m, m'\}}^{i^*}) = c$. However, note that $\{r_{(i_m, i_{m'}), \{m, m'\}}^{i^*}, \hat{r}_{(i_m, i_{m'}), \{m, m'\}}^{i^*}, z_{i^*}\} \in \mathcal{S}$, by Step 9. These three vertices all have color c , contradicting that h is a valid annotated P_4 -coloring of G . \triangleleft

Using the claims above and the bound on the size of $V(G)$ computed earlier, we conclude that we have given a degree-2 cross-composition to annotated P_4 -coloring, such that the lower bound follows from Theorem 5. \blacktriangleleft

4 Gadgets in hard graphs for List H -Coloring

Now we are going back to investigating the LIST H -COLORING problem, for fixed graphs H . To transfer the lower bound of Lemma 7 to LIST H -COLORING for all graphs H which are not bi-arc graphs, we use a two-step process. First we use an idea of Feder, Hell, and Huang [14] which allows us to efficiently reduce so-called *consistent* instances of the LIST H^* -COLORING problem, where H^* is a (simple) bipartite graph naturally associated to H , to equivalent instances of LIST H -COLORING on the same vertex set. This implies that LIST H -COLORING is at least as hard to sparsify as consistent instances LIST H^* -COLORING, where H^* is a bipartite graph. Then we will develop a number of gadgets to reduce ANNOTATED LIST P_4 -COLORING to LIST H^* -COLORING on consistent instances, in a way that preserves sparsification lower bounds. Together, this chain of reductions will prove Theorem 1.

4.1 Bi-arc graphs, associated bipartite graphs, and consistent instances

Recall that the complexity dichotomy for LIST H -COLORING was proven in three steps:

1. for reflexive H , the polynomial cases appear to be interval graphs [12],
2. for irreflexive H , the polynomial cases appear to be bipartite co-circular-arc graphs [13],
3. for general graphs, the polynomial cases are the so-called bi-arc graphs [14].

The main idea of showing the final step of the dichotomy was a reduction to the bipartite case. For a graph H , by H^* we denote the *associated bipartite graph*, defined as follows. The vertex set of H^* is the union of two independent sets: $V_1 := \{x' \mid x \in V(H)\}$ and $V_2 := \{x'' \mid x \in V(H)\}$. The vertices $x' \in V_1$ and $y'' \in V_2$ are adjacent if and only if $xy \in E$. Note that the edges of type $x'x''$ in H^* correspond to loops in H .

As we mentioned in the introduction, bi-arc graphs are defined in terms of a certain geometric representation, but for us much more convenient will be to use the following characterization in terms of the associated bipartite graph.

► **Theorem 10** (Feder, Hell, and Huang [14]). *Let H be an undirected graph, possibly with loops. The following are equivalent.*

1. H is a bi-arc graph.
2. H^* is the complement of a circular-arc graph.

Thus the graphs H for which LIST H -COLORING is NP-hard, are precisely those for which LIST H^* -COLORING is NP-hard: when H^* is not the complement of a circular-arc graph.

Now let us explain how showing the hardness of LIST H -COLORING can be reduced to showing the hardness of LIST H^* -COLORING. Here we need the notion of a *consistent* instance of the problem.

► **Definition 11.** *Let F be a connected bipartite graph with bipartition classes X and Y . An instance (G, L) of LIST F -COLORING is consistent, if G is bipartite and has a bipartition into classes $A, B \subseteq V(G)$, such that $L(a) \subseteq X$ for all $a \in A$, and $L(b) \subseteq Y$ for all $b \in B$.*

The following Proposition follows from the idea of Feder, Hell, and Huang [14], and provides a reduction from LIST H^* -COLORING to LIST H -COLORING that preserves the vertex set of G . Its exact statement comes from recent work by an overlapping set of authors [34, 35].

► **Proposition 12** (Okrasa et al. [34, 35]). *Let H be a graph and let (G, L) be a consistent instance of LIST H^* -COLORING. Define $L' : V(G) \rightarrow 2^{V(H)}$ as $L'(x) := \{u \mid \{u', u''\} \cap L(x) \neq \emptyset\}$. Then $(G, L) \rightarrow H^*$ if and only if $(G, L') \rightarrow H$.*

4.2 Hard bipartite graphs H

The following notion was introduced by Feder, Hell, and Huang [13].

► **Definition 13.** *Let $k \geq 1$ and let H be a bipartite graph with bipartition classes X, Y . Let $U = \{u_0, \dots, u_{2k}\} \subseteq X$ and $V = \{v_0, \dots, v_{2k}\} \subseteq Y$ be ordered sets of vertices such that $\{u_0v_0, u_1v_1, \dots, u_{2k}v_{2k}\}$ is a set of edges of H . We say that (U, V) is a special edge asteroid (or, in short, an asteroid) of order $2k + 1$, if for every $i \in \{0, \dots, 2k\}$ there exists a u_i - u_{i+1} -path $P_{i,i+1}$ in H (indices are computed modulo $2k + 1$), such that*

- (a) *there are no edges between $\{u_i, v_i\}$ and $\{v_{i+k}, v_{i+k+1}\} \cup V(P_{i+k, i+k+1})$ and*
- (b) *there are no edges between $\{u_0, v_0\}$ and $\{v_1, \dots, v_{2k}\} \cup \bigcup_{i=1}^{2k-1} V(P_{i, i+1})$.*

Feder, Hell, and Huang showed the following characterization of hard bipartite cases of LIST H -COLORING, i.e., bipartite graphs H , whose complement is not a circular-arc graph.

► **Theorem 14** (Feder et al. [13]). *A bipartite graph H is not the complement of a circular-arc graph if and only if H contains an induced cycle with at least 6 vertices or an asteroid.*

While induced cycles of length at least 6 and asteroids suffice to prove NP-completeness of LIST H -COLORING, to prove sparsification lower bounds via ANNOTATED LIST P_4 -COLORING we need a more local structure. We therefore introduce the following notion.

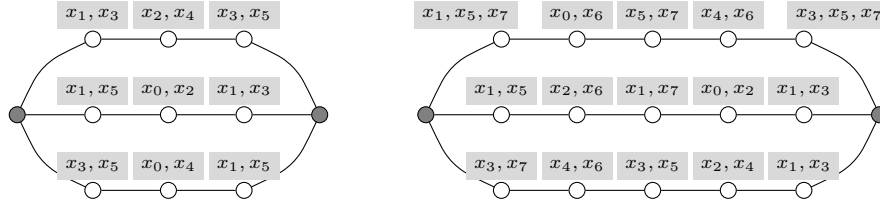
► **Definition 15.** An extended P_4 gadget in an undirected simple graph H is a tuple (a, b, c, d, e) of distinct vertices in H , such that all of the following hold:

1. $H[\{a, b, c, d\}]$ is isomorphic to P_4 ,
2. the sets $N_H(a), N_H(c), N_H(e)$ are pairwise incomparable, and
3. the sets $N_H(b), N_H(d)$ are pairwise incomparable.

Intuitively, if H contains an extended P_4 gadget, then the P_4 on (a, b, c, d) allows a LIST H -COLORING instance to express a homomorphism problem to P_4 , while the presence of vertex e and the incomparability of the neighborhoods allows gadgets to be constructed to enforce the semantics of the set F and the sequence \mathcal{S} in the definition of ANNOTATED LIST P_4 -COLORING, thereby allowing a reduction from that problem to LIST H -COLORING. The gadgets needed to simulate the pairwise constraints from F are given by the next lemma. Its proof can be found in the full version of the paper [4].

► **Lemma 16 (♠).** Let H be a bipartite graph which contains an induced cycle of at least 6 vertices or an asteroid. Then there exists an extended P_4 gadget (a, b, c, d, e) in H . Moreover, for every $Q \in \{\{a, c, e\}, \{b, d\}\}$ there is a consistent LIST H -COLORING instance (G_Q, L) containing two distinguished vertices γ_1, γ_2 such that a mapping $f: \{\gamma_1, \gamma_2\} \rightarrow Q$ can be extended to a proper list H -coloring of (G_Q, L) if and only if $f(\gamma_1) \neq f(\gamma_2)$.

Let us present the high-level idea of the proof of Lemma 16. It is straightforward to observe that if H contains an induced cycle with consecutive vertices x_0, x_1, \dots, x_{k-1} , where $k \geq 6$, then we can select a, b, c, d, e to be the consecutive vertices of this cycle, say x_0, x_1, x_2, x_3, x_4 . Furthermore, it is quite easy to construct (G_Q, L) for $Q \in \{\{a, c, e\}, \{b, d\}\}$, see Figure 2.



■ **Figure 2** A LIST H -COLORING instance $(G_{\{x_0, x_2, x_4\}}, L)$ satisfying the statement of Lemma 16 in case that H contains an induced C_6 (left) or an induced C_8 (right). Vertices γ_1, γ_2 are marked gray, and $L(\gamma_1) = L(\gamma_2) = \{x_0, x_2, x_4\}$.

So let us assume that H contains an asteroid (U, V) , where $U = \{u_0, u_1, \dots, u_{2k}\}$ and $V = \{v_0, v_1, \dots, v_{2k}\}$. First, we observe that without loss of generality we can assume that certain minimality conditions concerning (U, V) and the paths $P_{i, i+1}$ of the asteroid are satisfied. In particular, we consider an inclusion-wise minimal asteroid in H , and each path $P_{i, i+1}$ is induced. We observe that if $V(P_{0,1}) \setminus \{u_0, v_0, u_1, v_1\}$ contains at least two vertices, then we can select a, b, c, d to be the middle four vertices of $P_{0,1}$, and e to be u_{k+1} . So in the remaining case the set $V(P_{0,1}) \cup \{v_0, v_1\}$ induces a path with 5 vertices (note that by the definition of an asteroid there are no edges between $\{u_0, v_0\}$ and $\{u_1, v_1\}$). Now we employ some case analysis and consider possible edges and non-edges that might exist in H . It turns out that we can always find a suitable extended P_4 gadget.

Now let us show an idea how to find $(G_{\{a, c, e\}}, L)$ (the construction of $G_{\{b, d\}}$ is analogous). In each case, when finding an extended P_4 gadget, we are also able to construct an auxiliary gadget (G', L) with two special vertices β_1, β_2 , such that $L(\beta_1) = \{a, c, e\}$ and $L(\beta_2) = \{u_0, u_1, u_{k+1}\}$, and there is a bijection $\sigma: \{a, c, e\} \rightarrow \{u_0, u_1, u_{k+1}\}$ such that:

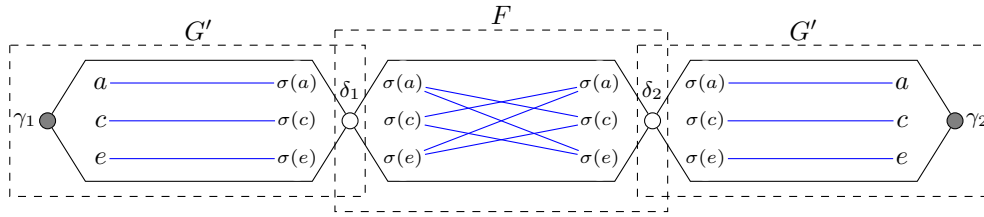


Figure 3 The construction of $(G_{\{a,c,e\}}, L)$ as a composition of two copies of G' and a copy of F . We have $L(\gamma_1) = L(\gamma_2) = \{a, c, e\}$ and $L(\delta_1) = L(\delta_2) = \{\sigma(a), \sigma(c), \sigma(e)\}$. Blue lines denote which mappings of $\gamma_1, \delta_1, \delta_2, \gamma_2$ to the vertices on their lists can be extended to a list homomorphism of particular gadgets.

- (i) in any list H -coloring f of (G', L) , if $f(\beta_1) = x$ (for $x \in \{a, c, e\}$), then $f(\beta_2) = \sigma(x)$,
- (ii) for any $x \in \{a, c, e\}$ there is a list H -coloring f of (G', L) , such that $f(\beta_1) = x$ and $f(\beta_2) = \sigma(x)$.

Intuitively, (G', L) can be used to uniquely translate any $x \in \{a, c, e\}$ appearing as a color of β_1 , to $\sigma(x)$ appearing as a color of β_2 .

The last building block of (G, L) is a graph built by Feder, Hell, and Huang [13, Fig. 3], which we call the *unequal gadget*. The unequal gadget is an instance (F, L) of LIST H -COLORING with two distinguished vertices δ_1, δ_2 , such $L(\delta_1) = L(\delta_2) = \{u_0, u_1, u_{k+1}\}$ and any function $f: \{\delta_1, \delta_2\} \rightarrow \{u_0, u_1, u_{k+1}\}$ can be extended to a proper list H -coloring of (F, L) if and only if $f(\delta_1) \neq f(\delta_2)$.

To build (G, L) , we introduce two copies of (G', L) and one copy of (F, L) , and unify the appropriate vertices, as shown in Figure 3. Note that the lists of unified vertices agree, so this operation is safe.

From the gadgets of Lemma 16, we can also make efficient larger gadgets to enforce that in a large group of vertices, at least one vertex is not colored c . The construction is an adaptation of a gadget due to Jaffke and Jansen [24]; again the proof can be found in the full version of the paper [4].

► **Lemma 17** (♠). *Let H be a bipartite graph which contains an induced cycle of at least 6 vertices or an asteroid, and let (a, b, c, d, e) be an extended P_4 gadget in H as guaranteed by Lemma 16. For any $k \geq 2$ one can construct a consistent LIST H -COLORING instance (G, L) in polynomial time containing k distinguished vertices $\gamma_1, \dots, \gamma_k$ such that $|V(G)| \in \mathcal{O}(k)$, and such that a mapping $f: \{\gamma_1, \dots, \gamma_k\} \rightarrow \{a, c, e\}$ can be extended to a proper list H -coloring of (G, L) if and only if there exists an $i \in [k]$ with $f(\gamma_i) \neq c$.*

Using these gadgets in the two-step process described in the beginning of Section 4, we now obtain the following.

► **Theorem 1.** *If H is an undirected graph that is not a bi-arc graph, possibly with loops, then LIST H -COLORING parameterized by the number of vertices n admits no generalized kernel of size $\mathcal{O}(n^{2-\epsilon})$ for any $\epsilon > 0$, unless $\text{NP} \subseteq \text{coNP/poly}$.*

Proof. We start by showing that for any bipartite graph H that is not a bi-arc graph, LIST H -COLORING allows no nontrivial sparsification. We use a linear-parameter transformation from ANNOTATED LIST P_4 -COLORING, such that the lower bound follows from Lemma 7.

Since H is bipartite and not a bi-arc graph, it is not the complement of a circular arc graph [14], and it follows from Theorem 14 that H has an induced cycle of length at least six or an asteroid. It then follows from Lemma 16 that H has an extended P_4

gadget on distinguished vertices (a, b, c, d, e) of H . Furthermore, there exist two relevant gadgets as described by Lemma 16. We call the gadget constructed for $Q = \{a, c, e\}$ the a, c, e -NOT-gadget, and the one constructed for $Q = \{b, d\}$ the b, d -NOT-gadget.

Let an instance (G, L, \mathcal{S}, F) of ANNOTATED LIST P_4 -COLORING be given, we show how to create an instance \tilde{G} of LIST H -COLORING. Initialize \tilde{G} as G (ignoring the annotations), where every vertex in \tilde{G} receives the same list it had in G , where now a, b, c, d, e refer to the vertices of the extended P_4 gadget present in H . For any $\{u, v\} \in F$, if $L(u) \subseteq \{a, c, e\}$ (implying also $L(v) \subseteq \{a, c, e\}$), add a new a, c, e -NOT-gadget to \tilde{G} . Otherwise, meaning that $L(u) \subseteq \{b, d\}$ and $L(v) \subseteq \{b, d\}$, we add a new b, d -NOT-gadget to \tilde{G} . Identify vertex γ_1 of the added gadget with u , and vertex γ_2 with v .

For every $S = \{s_1, \dots, s_m\} \in \mathcal{S}$, add a new gadget as described by Lemma 17 for $k = m$ to \tilde{G} . Note that such a gadget has $\mathcal{O}(m)$ vertices. Identify vertex γ_i of the gadget with vertex s_i for all $i \in [m]$.

It is easy to observe from the correctness of the added gadgets, that \tilde{G} is list H -colorable if and only if G had a coloring respecting the annotations.

We continue by bounding the number of vertices in \tilde{G} . Using that $\sum_{S \in \mathcal{S}} |S| \leq 3|V(G)|$ and $|F| \leq |V(G)|$ by definition of ANNOTATED LIST P_4 -COLORING, we get

$$|V(\tilde{G})| = \underbrace{|V(G)|}_{\text{init}} + \underbrace{|V(G)| \cdot \mathcal{O}(1)}_{\text{NOT-gadgets}} + \underbrace{\mathcal{O}(|V(G)|)}_{\text{Lemma 17}} = \mathcal{O}(|V(G)|),$$

which is properly bounded for a linear parameter transformation. By Lemma 7, ANNOTATED LIST P_4 -COLORING does not have a nontrivial sparsification; as we described in the preliminaries linear-parameter transformations transfer such lower bounds; so by Lemma 7 and this fact there is no nontrivial sparsification for bipartite graphs H that are not bi-arc graphs. Observe that the constructed graph \tilde{G} is consistent, such that the lower bound holds even for consistent instances of LIST H -COLORING.

It remains to show the result for non-bipartite graphs H . Let H be an undirected graph that is not a bi-arc graph, such that H is non-bipartite. Let H^* be the associated bipartite graph of H . Since H is not a bi-arc graph, it follows that H^* is not the complement of a circular arc graph [14, Proposition 3.1]. Since H^* is bipartite and irreflexive it follows that H^* is not a bi-arc graph.

As proven above, it follows that LIST H^* -COLORING does not have a generalized kernel of size $\mathcal{O}(n^{2-\varepsilon})$, unless $\text{NP} \subseteq \text{coNP}/\text{poly}$. Proposition 12 gives a straightforward linear-parameter transformation from LIST H^* -COLORING to LIST H -COLORING, showing that the same lower bound holds for LIST H -COLORING. ◀

5 Conclusion

A natural open question is whether analogous results can be obtained for the (non-list) H -COLORING problem. Despite the obvious similarity of H -COLORING and LIST H -COLORING, they appear to behave very differently when it comes to proving lower bounds. All hardness proofs for LIST H -COLORING [12, 13, 14, 15, 34], including the proofs in this paper, are purely combinatorial and focus on the *local* structure of H . In all of them, we first identify some “hard” substructure H' in H , and then prove the lower bound for H' . This can be done, as we can ignore vertices in $V(H) \setminus V(H')$ by not including them in the lists. On the other hand, all proofs for H -COLORING use some algebraic tools [3, 21, 36, 40] which allow capturing the *global* structure of H . We therefore expect similar difficulties in the case of proving sparsification lower bounds for H -COLORING. Some results on the standard H -COLORING problem are described in the fourth author’s PhD thesis [38], but they fall short of a full dichotomy.

References

- 1 Michael O. Albertson, Paul A. Catlin, and Luana Gibbons. Homomorphisms of 3-chromatic graphs. II. In *Proceedings of the 16th Southeastern international conference on combinatorics, graph theory and computing*, volume 47, pages 19–28, 1985.
- 2 Hans L. Bodlaender, Bart M. P. Jansen, and Stefan Kratsch. Kernelization lower bounds by cross-composition. *SIAM J. Discrete Math.*, 28(1):277–305, 2014. doi:10.1137/120880240.
- 3 Andrei A. Bulatov. H -coloring dichotomy revisited. *Theor. Comput. Sci.*, 349(1):31–39, 2005. doi:10.1016/j.tcs.2005.09.028.
- 4 Hubie Chen, Bart M. P. Jansen, Karolina Okrasa, Astrid Pieterse, and Paweł Rzażewski. Sparsification lower bounds for List H -Coloring. *CoRR*, abs/2009.08353, 2020. arXiv:2009.08353.
- 5 Hubie Chen, Bart M. P. Jansen, and Astrid Pieterse. Best-case and worst-case sparsifiability of boolean CSPs. In *Proceedings of the 13th International Symposium on Parameterized and Exact Computation*, volume 115 of *LIPICs*, pages 15:1–15:13. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018. doi:10.4230/LIPICs.IPEC.2018.15.
- 6 Víctor Dalmau, László Egri, Pavol Hell, Benoît Larose, and Arash Rafiey. Descriptive complexity of List H -Coloring problems in logspace: A refined dichotomy. In *30th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2015, Kyoto, Japan, July 6-10, 2015*, pages 487–498. IEEE Computer Society, 2015. doi:10.1109/LICS.2015.52.
- 7 Holger Dell and Dieter van Melkebeek. Satisfiability allows no nontrivial sparsification unless the polynomial-time hierarchy collapses. *J. ACM*, 61(4):23:1–23:27, 2014. doi:10.1145/2629620.
- 8 László Egri, Pavol Hell, Benoît Larose, and Arash Rafiey. Space complexity of list H -colouring: a dichotomy. In Chandra Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 349–365. SIAM, 2014. doi:10.1137/1.9781611973402.26.
- 9 László Egri, Andrei A. Krokhin, Benoît Larose, and Pascal Tesson. The complexity of the list homomorphism problem for graphs. *Theory Comput. Syst.*, 51(2):143–178, 2012. doi:10.1007/s00224-011-9333-8.
- 10 László Egri, Dániel Marx, and Paweł Rzażewski. Finding list homomorphisms from bounded-treewidth graphs to reflexive graphs: a complete complexity characterization. In Rolf Niedermeier and Brigitte Vallée, editors, *35th Symposium on Theoretical Aspects of Computer Science, STACS 2018, February 28 to March 3, 2018, Caen, France*, volume 96 of *LIPICs*, pages 27:1–27:15. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018. doi:10.4230/LIPICs.STACS.2018.27.
- 11 Louis Esperet, Mickaël Montassier, Pascal Ochem, and Alexandre Pinlou. A complexity dichotomy for the coloring of sparse graphs. *J. Graph Theory*, 73(1):85–102, 2013. doi:10.1002/jgt.21659.
- 12 Tomás Feder and Pavol Hell. List homomorphisms to reflexive graphs. *J. Comb. Theory, Ser. B*, 72(2):236–250, 1998. doi:10.1006/jctb.1997.1812.
- 13 Tomás Feder, Pavol Hell, and Jing Huang. List homomorphisms and circular arc graphs. *Combinatorica*, 19(4):487–505, 1999. doi:10.1007/s004939970003.
- 14 Tomás Feder, Pavol Hell, and Jing Huang. Bi-arc graphs and the complexity of list homomorphisms. *Journal of Graph Theory*, 42(1):61–80, 2003. doi:10.1002/jgt.10073.
- 15 Tomás Feder, Pavol Hell, and Jing Huang. List homomorphisms of graphs with bounded degrees. *Discrete Mathematics*, 307(3):386–392, 2007. Algebraic and Topological Methods in Graph Theory. doi:10.1016/j.disc.2005.09.030.
- 16 Fedor V. Fomin, Daniel Lokshtanov, Saket Saurabh, and Meirav Zehavi. *Kernelization: Theory of Parameterized Preprocessing*. Cambridge University Press, December 2018. doi:10.1017/9781107415157.
- 17 Anna Galluccio, Pavol Hell, and Jaroslav Nešetřil. The complexity of H -colouring of bounded degree graphs. *Discrete Math.*, 222(1-3):101–109, 2000. doi:10.1016/S0012-365X(00)00009-1.

- 18 Carla Groenland, Karolina Okrasa, Paweł Rzażewski, Alex D. Scott, Paul D. Seymour, and Sophie Spirkl. H -colouring P_t -free graphs in subexponential time. *Discrete Applied Mathematics*, 267:184–189, 2019. doi:10.1016/j.dam.2019.04.010.
- 19 Jiong Guo and Rolf Niedermeier. Invitation to data reduction and problem kernelization. *SIGACT News*, 38(1):31–45, 2007. doi:10.1145/1233481.1233493.
- 20 Geňa Hahn and Claude Tardif. *Graph homomorphisms: structure and symmetry*, pages 107–166. Springer Netherlands, 1997. doi:10.1007/978-94-015-8937-6_4.
- 21 Pavol Hell and Jaroslav Nešetřil. On the complexity of H -coloring. *J. Comb. Theory, Ser. B*, 48(1):92–110, 1990. doi:10.1016/0095-8956(90)90132-J.
- 22 Pavol Hell and Jaroslav Nešetřil. *Graphs and Homomorphisms*. Oxford Lecture Series in Mathematics. Oxford University Press, 2004.
- 23 Robert W. Irving. NP-completeness of a family of graph-colouring problems. *Discrete Appl. Math.*, 5(1):111–117, 1983. doi:10.1016/0166-218X(83)90020-3.
- 24 Lars Jaffke and Bart M. P. Jansen. Fine-grained parameterized complexity analysis of graph coloring problems. In *Proceedings of the 10th International Conference on Algorithms and Complexity*, pages 345–356, 2017. doi:10.1007/978-3-319-57586-5_29.
- 25 Bart M. P. Jansen. On sparsification for computing treewidth. *Algorithmica*, 71(3):605–635, 2015. doi:10.1007/s00453-014-9924-2.
- 26 Bart M. P. Jansen and Astrid Pieterse. Optimal data reduction for graph coloring using low-degree polynomials. In *12th International Symposium on Parameterized and Exact Computation, IPEC 2017, September 6-8, 2017, Vienna, Austria*, volume 89 of *LIPICs*, pages 22:1–22:12, 2017. doi:10.4230/LIPICs.IPEC.2017.22.
- 27 Bart M. P. Jansen and Astrid Pieterse. Sparsification upper and lower bounds for graph problems and not-all-equal SAT. *Algorithmica*, 79(1):3–28, 2017. doi:10.1007/s00453-016-0189-9.
- 28 Bart M. P. Jansen and Astrid Pieterse. Optimal data reduction for graph coloring using low-degree polynomials. *Algorithmica*, 81(10):3865–3889, 2019. doi:10.1007/s00453-019-00578-5.
- 29 Gábor Kun and Mario Szegedy. A new line of attack on the dichotomy conjecture. *Eur. J. Comb.*, 52:338–367, 2016. doi:10.1016/j.ejcb.2015.07.011.
- 30 Victor Lagerkvist and Magnus Wahlström. Kernelization of constraint satisfaction problems: A study through universal algebra. In *Proceedings of the 23rd International Conference on Principles and Practice of Constraint Programming*, volume 10416, pages 157–171. Springer, 2017. doi:10.1007/978-3-319-66158-2_11.
- 31 Gary MacGillivray and Mark H. Siggers. On the complexity of H -colouring planar graphs. *Discrete Math.*, 309(18):5729–5738, 2009. doi:10.1016/j.disc.2008.05.016.
- 32 H.A. Maurer, J.H. Sudborough, and E. Welzl. On the complexity of the general coloring problem. *Information and Control*, 51(2):128–145, 1981. doi:10.1016/S0019-9958(81)90226-6.
- 33 Jaroslav Nešetřil. Representations of graphs by means of products and their complexity. In *Proceedings of the 10th International Symposium on Mathematical Foundations of Computer Science*, volume 118 of *Lecture Notes in Computer Science*, pages 94–102, 1981. doi:10.1007/3-540-10856-4_76.
- 34 Karolina Okrasa, Marta Piczyk, and Paweł Rzażewski. Full complexity classification of the list homomorphism problem for bounded-treewidth graphs. In Fabrizio Grandoni, Grzegorz Herman, and Peter Sanders, editors, *28th Annual European Symposium on Algorithms, ESA 2020, September 7-9, 2020, Pisa, Italy (Virtual Conference)*, volume 173 of *LIPICs*, pages 74:1–74:24. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPICs.ESA.2020.74.
- 35 Karolina Okrasa, Marta Piczyk, and Paweł Rzażewski. Full complexity classification of the list homomorphism problem for bounded-treewidth graphs. *CoRR*, abs/2006.11155, 2020. arXiv:2006.11155.

- 36 Karolina Okrasa and Paweł Rzażewski. Fine-grained complexity of graph homomorphism problem for bounded-treewidth graphs. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 1578–1590. SIAM, 2020. doi:10.1137/1.9781611975994.97.
- 37 Karolina Okrasa and Paweł Rzażewski. Subexponential algorithms for variants of the homomorphism problem in string graphs. *J. Comput. Syst. Sci.*, 109:126–144, 2020. doi:10.1016/j.jcss.2019.12.004.
- 38 Astrid Pieterse. *Tight parameterized preprocessing bounds: sparsification via low-degree polynomials*. PhD thesis, Eindhoven University of Technology, Department of Mathematics and Computer Science, September 2019. Proefschrift.
- 39 M. Siggers. A new proof of the H -coloring dichotomy. *SIAM J. Discrete Math.*, 23(4):2204–2210, 2010. Corrected manuscript available from the author’s homepage. doi:10.1137/080736697.
- 40 Mark H. Siggers. Dichotomy for bounded degree H -colouring. *Discrete Appl. Math.*, 157(2):201–210, 2009. doi:10.1016/j.dam.2008.02.003.
- 41 William T. Trotter and John I. Moore. Characterization problems for graphs, partially ordered sets, lattices, and families of sets. *Discrete Mathematics*, 16(4):361–381, 1976. doi:10.1016/S0012-365X(76)80011-8.