



**HAL**  
open science

# Approximating Highly Inapproximable Problems on Graphs of Bounded Twin-Width

Pierre Bergé, Edouard Bonnet, Hugues Déprés, Rémi Watrigant

► **To cite this version:**

Pierre Bergé, Edouard Bonnet, Hugues Déprés, Rémi Watrigant. Approximating Highly Inapproximable Problems on Graphs of Bounded Twin-Width. STACS 2023, Mar 2023, Hambourg, Germany. hal-04292968

**HAL Id: hal-04292968**

**<https://hal.science/hal-04292968v1>**

Submitted on 17 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Approximating Highly Inapproximable Problems on Graphs of Bounded Twin-Width

Pierre Bergé ✉

Univ Lyon, CNRS, ENS de Lyon, Université Claude Bernard Lyon 1, LIP UMR5668, France

Édouard Bonnet ✉ 🏠 

Univ Lyon, CNRS, ENS de Lyon, Université Claude Bernard Lyon 1, LIP UMR5668, France

Hugues Déprés ✉ 🏠

Univ Lyon, CNRS, ENS de Lyon, Université Claude Bernard Lyon 1, LIP UMR5668, France

Rémi Watrigant ✉

Univ Lyon, CNRS, ENS de Lyon, Université Claude Bernard Lyon 1, LIP UMR5668, France

---

## Abstract

---

For any  $\varepsilon > 0$ , we give a polynomial-time  $n^\varepsilon$ -approximation algorithm for MAX INDEPENDENT SET in graphs of bounded twin-width given with an  $O(1)$ -sequence. This result is derived from the following time-approximation trade-off: We establish an  $O(1)^{2^q-1}$ -approximation algorithm running in time  $\exp(O_q(n^{2^{-q}}))$ , for every integer  $q \geq 0$ . Guided by the same framework, we obtain similar approximation algorithms for MIN COLORING and MAX INDUCED MATCHING. In general graphs, all these problems are known to be highly inapproximable: for any  $\varepsilon > 0$ , a polynomial-time  $n^{1-\varepsilon}$ -approximation for any of them would imply that P=NP [Håstad, FOCS '96; Zuckerman, ToC '07; Chalermsook et al., SODA '13]. We generalize the algorithms for MAX INDEPENDENT SET and MAX INDUCED MATCHING to the independent (induced) packing of any fixed connected graph  $H$ .

In contrast, we show that such approximation guarantees on graphs of bounded twin-width given with an  $O(1)$ -sequence are very unlikely for MIN INDEPENDENT DOMINATING SET, and somewhat unlikely for LONGEST PATH and LONGEST INDUCED PATH. Regarding the existence of better approximation algorithms, there is a (very) light evidence that the obtained approximation factor of  $n^\varepsilon$  for MAX INDEPENDENT SET may be best possible. This is the first in-depth study of the approximability of problems in graphs of bounded twin-width. Prior to this paper, essentially the only such result was a polynomial-time  $O(1)$ -approximation algorithm for MIN DOMINATING SET [Bonnet et al., ICALP '21].

**2012 ACM Subject Classification** Theory of computation  $\rightarrow$  Graph algorithms analysis; Theory of computation  $\rightarrow$  Design and analysis of algorithms

**Keywords and phrases** Approximation algorithms, bounded twin-width

**Acknowledgements** We thank Colin Geniet, Eunjung Kim, and Stéphan Thomassé for useful discussions.

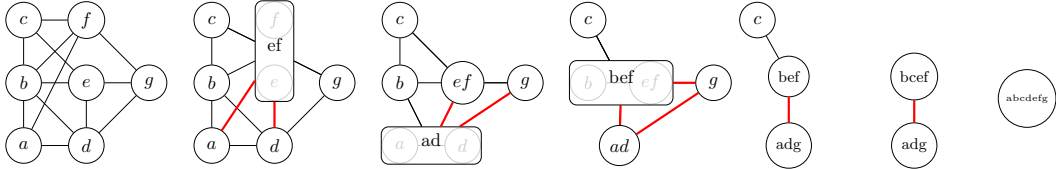
## 1 Introduction

Twin-width is a graph parameter introduced by Bonnet, Kim, Thomassé, and Watrigant [10]. Its definition involves the notions of *trigraphs* and of *contraction sequences*. A *trigraph* is a graph with two types of edges: black (regular) edges and red (error) edges. A (vertex) *contraction* consists of merging two (non-necessarily adjacent) vertices, say,  $u, v$  into a vertex  $w$ , and keeping every edge  $wz$  black if and only if  $uz$  and  $vz$  were previously black edges. The other edges incident to  $w$  become red (if not already), and the rest of the trigraph remains the same. A *contraction sequence* of an  $n$ -vertex<sup>1</sup> graph  $G$  is a sequence of trigraphs

---

<sup>1</sup> In this introduction, we might implicitly use  $n$  to denote the number of vertices, and  $m$ , the number of edges of the graph at hand.

42  $G = G_n, \dots, G_1 = K_1$  such that  $G_i$  is obtained from  $G_{i+1}$  by performing one contraction.  
 43 A  $d$ -sequence is a contraction sequence in which every vertex of every trigraph has at most  
 44  $d$  red edges incident to it. The *twin-width* of  $G$ , denoted by  $\text{tw}(G)$ , is then the minimum  
 45 integer  $d$  such that  $G$  admits a  $d$ -sequence. Figure 1 gives an example of a graph with a  
 46 2-sequence, i.e., of twin-width at most 2. Twin-width can be naturally extended to matrices  
 47 (with unordered [10] or ordered [8] row and column sets) over a finite alphabet, and thus to  
 48 binary structures.



■ **Figure 1** A 2-sequence witnessing that the initial graph has twin-width at most 2.

49 An equivalent viewpoint that will be somewhat more convenient is to consider a  $d$ -sequence  
 50 as a sequence of partitions  $\mathcal{P}_n := \{\{v\} : v \in V(G)\}, \mathcal{P}_{n-1}, \dots, \mathcal{P}_1 := \{V(G)\}$  of  $V(G)$ ,  
 51 such that for every integer  $1 \leq i \leq n-1$ ,  $\mathcal{P}_i$  has  $i$  parts and is obtained by merging two  
 52 parts of  $\mathcal{P}_{i+1}$  into one. Now the *red degree* of a part  $P \in \mathcal{P}_i$  is the number of other parts  
 53  $Q \in \mathcal{P}_i$  such that there is in  $G$  at least one edge and at least one non-edge between  $P$  and  $Q$ .  
 54 A  $d$ -sequence is such that no part of no partition of the sequence has red degree more than  $d$ .  
 55 In that case the *maximum red degree* of each partition is at most  $d$ . And we similarly get  
 56 the twin-width of  $G$  as the minimum integer  $d$  such that  $G$  admits a (partition)  $d$ -sequence.  
 57 The *quotient trigraph*  $G/\mathcal{P}_i$  is the trigraph  $G_i$ , if the (contraction)  $d$ -sequence  $G_n, \dots, G_1$   
 58 and the (partition)  $d$ -sequence  $\mathcal{P}_n, \dots, \mathcal{P}_1$  correspond.

59 Classes of binary structures with bounded twin-width include graph classes with bounded  
 60 treewidth, and more generally bounded clique-width, proper minor-closed classes, posets  
 61 with antichains of bounded size, strict subclasses of permutation graphs, as well as  $\Omega(\log n)$ -  
 62 subdivisions of  $n$ -vertex graphs [10], and some classes of (bounded-degree) expanders [5].  
 63 A notable variety of geometrically defined graph classes have bounded twin-width such as map  
 64 graphs, bounded-degree string graphs [10], classes with bounded queue number or bounded  
 65 stack number [5], segment graphs with no  $K_{t,t}$  subgraph, visibility graphs of 1.5D terrains  
 66 without large half-graphs, visibility graphs of simple polygons without large independent  
 67 sets [4].

68 For every class  $\mathcal{C}$  mentioned so far,  $O(1)$ -sequences can be computed in polynomial time<sup>2</sup>  
 69 on members of  $\mathcal{C}$ . For classes of binary structures including a binary relation interpreted as  
 70 a linear order on the domain (called *ordered binary structures*), there is a fixed-parameter  
 71 approximation algorithm for twin-width [8]. More precisely, given a graph  $G$  and an integer  $k$ ,  
 72 there are computable functions  $f$  and  $g$  such that one can output an  $f(k)$ -sequence of  $G$   
 73 or correctly report that  $\text{tw}(G) > k$  in time  $g(k)n^{O(1)}$ . Such an approximation algorithm  
 74 is currently missing for classes of general (not necessarily ordered) binary structures, and  
 75 in particular for the class of all graphs. We also observe that deciding if the twin-width of  
 76 a graph is at most 4 is an NP-complete task [3].

77 We will therefore assume that the input graph is given with a  $d$ -sequence, and treat  $d$   
 78 as a constant (or that the input comes from any of the above-mentioned classes). Thus  
 79 far, this is the adopted setting when designing faster algorithms on bounded twin-width

<sup>2</sup> Admittedly, for the geometric classes, a representation is (at least partially) needed.

80 graphs [10, 7, 33, 30, 19]. From the inception of twin-width [10] –actually already from the  
 81 seminal work of Guillemot and Marx [21]– it was clear that structures wherein this invariant  
 82 is bounded may *often* allow the design of parameterized algorithms. More concretely, it was  
 83 shown [10] that, on graphs  $G$  given with a  $d$ -sequence, model checking a first-order sentence  
 84  $\varphi$  is fixed-parameter tractable –it can be solved in time  $f(d, \varphi) \cdot n$ –, the special cases of,  
 85 say,  $k$ -INDEPENDENT SET or  $k$ -DOMINATING SET admit single-exponential parameterized  
 86 algorithms [7], an effective data structure almost linear in  $n$  can support constant-time edge  
 87 queries [33], the triangles of  $G$  can be counted in time  $O(d^2n + m)$  [30].

88 So far, however, the connection between *having bounded twin-width* and *enjoying enhanced*  
 89 *approximation factors* was tenuous. The only such result concerned MIN DOMINATING SET,  
 90 known to be inapproximable in polynomial-time within factor  $(1 - o(1)) \ln n$  unless  $P=NP$  [16],  
 91 but yet admits a constant-approximation on graphs of bounded twin-width given with an  
 92  $O(1)$ -sequence [7]. We start filling this gap by designing approximation algorithms on graphs  
 93 of bounded twin-width given with an  $O(1)$ -sequence for notably MAX INDEPENDENT SET  
 94 (MIS, for short), MAX INDUCED MATCHING, and COLORING. Getting better approximation  
 95 algorithms for MIS and COLORING in that particular scenario was raised as an open  
 96 problem [7]. Before we describe our results and elaborate on the developed techniques, let us  
 97 briefly present the notorious inapproximability of these problems in general graphs.

98 MIS and COLORING are NP-hard [20], and very inapproximable: for every  $\varepsilon > 0$ , it is  
 99 NP-hard to approximate these problems within ratio  $n^{1-\varepsilon}$  [23, 34]. The same was shown to  
 100 hold for MAX INDUCED MATCHING [13]. Besides, there is only little room to improve over  
 101 the brute-force algorithm in  $2^{O(n)}$ : Unless the Exponential Time Hypothesis<sup>3</sup> [25] (ETH)  
 102 fails, no algorithm can solve MIS in time  $2^{o(n)}$  [26] (nor the other two problems). For any  $r$   
 103 (possibly a function of  $n$ ) WMIS can be  $r$ -approximated in time  $2^{O(n/r)}$  [15, 12]. Bansal  
 104 et al. [2] essentially shaved a  $\log^2 r$  factor to the latter exponent. It is known though that  
 105 polynomial shavings are unlikely. Chalermsook et al. [14] showed that, for any  $\varepsilon > 0$  and  
 106 sufficiently large  $r$  (again  $r$  can be function of  $n$ ), an  $r$ -approximation for MIS and MAX  
 107 INDUCED MATCHING cannot take time  $2^{O(n^{1-\varepsilon}/r^{1+\varepsilon})}$ , unless the ETH fails. For instance,  
 108 investing time  $2^{O(\sqrt{n})}$ , one cannot hope for significantly better than a  $\sqrt{n}$ -approximation.

## 109 Contributions and techniques

110 Our starting point is a constant-approximation algorithm for MIS running in time  $2^{O(\sqrt{n})}$   
 111 when presented with an  $O(1)$ -sequence, which is very unlikely to hold in general graphs by  
 112 the result of Chalermsook et al. [14].

113 ► **Theorem 1.** *On  $n$ -vertex graphs given with a  $d$ -sequence MAX INDEPENDENT SET can be*  
 114  *$O_d(1)$ -approximated in time  $2^{O_d(\sqrt{n})}$ .*

115 Our algorithm builds upon the functional equivalence between twin-width and the so-called  
 116 *versatile twin-width* [5]. We defer the reader to Section 2 for a formal definition of versatile  
 117 twin-width. For our purpose, one only needs to know the following useful consequence of that  
 118 equivalence. From a  $d'$ -sequence of  $G$ , we can compute in polynomial time another partition  
 119 sequence  $\mathcal{P}_n, \dots, \mathcal{P}_1$  of  $G$  of width  $d := f(d')$ , for some computable function  $f$ , such that for  
 120 every integer  $1 \leq i \leq n$ , all the  $i$  parts of  $\mathcal{P}_i$  have size at most  $d \cdot \frac{n}{i}$ . Even if some parts of  $\mathcal{P}_i$   
 121 can be very small, this partition is balanced in the sense that no part can be larger than  $d$   
 122 times the part size in a perfectly balanced partition. Of importance to us is  $\mathcal{P}_{\lfloor \sqrt{n} \rfloor}$  when the

<sup>3</sup> That is, the assumption that there is a  $\delta > 0$  such that  $n$ -variable 3-SAT cannot be solved in time  $\delta^n$ .

123 number of parts ( $\lfloor \sqrt{n} \rfloor$ ) and the size of a larger part in the partition (at most  $d \frac{n}{\lfloor \sqrt{n} \rfloor} \approx d\sqrt{n}$ )  
 124 are somewhat level.

125 We can then properly color the red graph (made by the red edges on the vertex set  $\mathcal{P}_{\lfloor \sqrt{n} \rfloor}$ )  
 126 with  $d + 1$  colors. Any color class  $X$  is a subset of parts of  $\mathcal{P}_{\lfloor \sqrt{n} \rfloor}$  such that between two  
 127 parts there are either all edges (black edge) or no edge at all (non-edge). In graph-theoretic  
 128 terms, the subgraph  $G_X$  of  $G$  induced by all the vertices of all the parts of  $X$  have a simple  
 129 modular decomposition: a partition of at most  $\sqrt{n}$  modules each of size at most  $d\sqrt{n}$ . It is  
 130 thus routine to compute a largest independent set of  $G_X$  essentially in time exponential in  
 131 the maximum between the number of modules and the maximum size of a module, that is,  
 132 in at most  $d\sqrt{n}$ . As one color class  $X^*$  contains more than a  $\frac{1}{d+1}$  fraction of the optimum,  
 133 we get our  $d + 1$ -approximation when computing a largest independent set of  $G_{X^*}$ . Figure 2  
 134 on page 13 serves as a visual summary of what we described so far.

135 The next step is to substitute recursive calls of our approximation algorithm to exact  
 136 exponential algorithms on induced subgraphs of size  $O_d(\sqrt{n})$ . Following this inductive process  
 137 at depth  $q = 2, 3, 4, \dots$ , we degrade the approximation ratio to  $(d + 1)^3, (d + 1)^7, (d + 1)^{15}$ ,  
 138 etc. but meanwhile we boost the running time to  $2^{O_d(n^{1/4})}, 2^{O_d(n^{1/8})}, 2^{O_d(n^{1/16})}$ , etc. In effect  
 139 we show by induction that:

140 ► **Theorem 2.** *On  $n$ -vertex graphs given with a  $d$ -sequence MAX INDEPENDENT SET has an*  
 141  *$O_d(1)^{2^q - 1}$ -approximation algorithm running in time  $2^{O_{d,q}(n^{2^{-q}})}$ , for every integer  $q \geq 0$ .*

142 The following polynomial-time algorithm is a corollary of Theorem 2 choosing  $q =$   
 143  $O_{d,\varepsilon}(\log \log n)$ .

144 ► **Theorem 3.** *For every  $\varepsilon > 0$ , MAX INDEPENDENT SET can be  $n^\varepsilon$ -approximated in*  
 145 *polynomial-time  $O_{d,\varepsilon}(1) \cdot \log^{O_d(1)} n \cdot n^{O(1)}$  on  $n$ -vertex graphs given with a  $d$ -sequence.*

146 Note that the exponent of the polynomial factor is an absolute constant (not depending on  $d$   
 147 nor on  $\varepsilon$ ).

148 We then apply our framework to COLORING and MAX INDUCED MATCHING.

149 ► **Theorem 4.** *For every  $\varepsilon > 0$ , COLORING and MAX INDUCED MATCHING admit polynomial-*  
 150 *time  $n^\varepsilon$ -approximation algorithms on  $n$ -vertex graphs of bounded twin-width given with an*  
 151  *$O(1)$ -sequence.*

152 The main additional difficulty for COLORING is that one cannot satisfactorily solve/ap-  
 153 proximate that problem on a modular decomposition by simply coloring its modules and  
 154 its quotient graph. One needs to tackle a more general problem called SET COLORING.  
 155 Fortunately this generalization is the fixed point we are looking for: approximating SET  
 156 COLORING can be done in our framework by mere recursive calls (to itself).

157 For MAX INDUCED MATCHING, we face a new kind of obstacle. It can be the case that  
 158 no decent solution is contained in any color class  $X$  –in the chosen  $d + 1$ -coloring of the  
 159 red graph  $G/\mathcal{P}_{\lfloor \sqrt{n} \rfloor}$ . For instance, it is possible that any such color class  $X$  induces in  $G$   
 160 an edgeless graph, while very large induced matchings exist with endpoints in two distinct  
 161 color classes. We thus need to also find large induced matchings within the black edges  
 162 and within the red edges of  $G/\mathcal{P}_{\lfloor \sqrt{n} \rfloor}$ . This leads to a more intricate strategy intertwining  
 163 the coloring of bounded-degree graphs (specifically the red graph and the square of its line  
 164 graph) and recursive calls to induced subgraphs of  $G$ , and to special induced subgraphs of  
 165 the total graph (i.e., made by both the red and black edges) of  $G/\mathcal{P}_{\lfloor \sqrt{n} \rfloor}$ . Although this is  
 166 not necessary, one can observe that the latter graphs are also induced subgraphs of  $G$  itself.

167 We then explore the limits of our results and framework in terms of amenable problems.  
 168 We give the following technical generalization to the approximation algorithms for MIS and  
 169 MAX INDUCED MATCHING.

170 ► **Theorem 5.** *For every connected graph  $H$  and  $\varepsilon > 0$ , MUTUALLY INDUCED  $H$ -PACKING  
 171 admits a polynomial-time  $n^\varepsilon$ -approximation algorithms on  $n$ -vertex graphs of bounded twin-  
 172 width given with an  $O(1)$ -sequence.*

173 In this problem, one seeks for a largest induced subgraph that consists of a disjoint  
 174 union of copies of  $H$ . All the previous technical issues are here combined. We try all the  
 175 possibilities of batching the vertices of  $H$  into at most  $|V(H)|$  parts of  $G/\mathcal{P}_{\lfloor\sqrt{n}\rfloor}$ , based on  
 176 the trigraph that these parts define. For instance with  $H = K_2$  (an edge), i.e., the case  
 177 of MAX INDUCED MATCHING, the three possible trigraphs are the 1-vertex trigraph, two  
 178 vertices linked by a red edge, and two vertices linked by a black edge. In the general case,  
 179 the problem generalization is quite delicate to find. We have to keep some partitions of  
 180  $V(G)$  and  $V(H)$  to enforce that the copies of  $H$  in  $G$  follow a pattern that the algorithm  
 181 committed to higher up in the recursion tree, and a weight function on  $|V(H)|$ -tuples of  
 182 vertices of  $G$ , not to forget how many mutually induced copies of  $H$  can be packed *within*  
 183 these vertices. The other novelty is that some recursive calls are on induced subgraphs of  
 184 the total graph of  $G/\mathcal{P}_{\lfloor\sqrt{n}\rfloor}$  that are *not* induced subgraphs of  $G$ . Fortunately, these graphs  
 185 keep the same bound of versatile twin-width, and thus our framework allows it.

186 Defining, for a family of graphs  $\mathcal{H}$ , MUTUALLY INDUCED  $\mathcal{H}$ -PACKING as the same  
 187 problem where the connected components of the induced subgraph should all be in  $\mathcal{H}$ ,  
 188 we get a similar approximation factor when  $\mathcal{H}$  is a finite set of connected graphs. (Note  
 189 that MUTUALLY INDUCED  $H$ -PACKING is sometimes called INDEPENDENT INDUCED  $H$ -  
 190 PACKING.) In particular, we can similarly approximate INDEPENDENT  $H$ -PACKING, which is  
 191 the same problem but the copies of  $H$  need not be induced. (Our approximation algorithms  
 192 could extend to other  $H$ -packing variants without the independence requirement, but these  
 193 problems can straightforwardly be  $O(1)$ -approximated in general graphs.)

194 We can handle some cases when  $\mathcal{H}$  is infinite, too. For instance, by slightly adapting the  
 195 case of MIS, we can get an  $n^\varepsilon$ -approximation when  $\mathcal{H}$  is the set of all cliques. We show this  
 196 more involved example, also expressible as MUTUALLY INDUCED  $\mathcal{H}$ -PACKING for  $\mathcal{H}$  the set  
 197 of all trees or the set all stars.

198 ► **Theorem 6.** *For every  $\varepsilon > 0$ , finding the induced (star) forest with the most edges admits  
 199 a polynomial-time  $n^\varepsilon$ -approximation algorithms on  $n$ -vertex graphs of bounded twin-width  
 200 given with an  $O(1)$ -sequence.*

201 As we already mentioned, our framework is exclusively useful for problems that are very  
 202 inapproximable in general graphs; at least for which an  $n^\varepsilon$ -approximation algorithm is not  
 203 known for every  $\varepsilon > 0$ . Are there natural such problems that cannot be approximated better  
 204 in graphs of bounded twin-width? We answer this question positively with the example of  
 205 MIN INDEPENDENT DOMINATING SET.

206 ► **Theorem 7.** *For every  $\varepsilon > 0$ , MIN INDEPENDENT DOMINATING SET does not admit an  
 207  $n^{1-\varepsilon}$ -approximation algorithm in  $n$ -vertex graphs given with an  $O(1)$ -sequence, unless  $P=NP$ .*

208 The reduction is the same as the one for general graphs [22], but performed from a planar  
 209 variant of 3-SAT. The obtained instances are not planar but can be contracted to planar  
 210 trigraphs, hence overall have bounded twin-width.

211 Finally the case of LONGEST PATH and LONGEST INDUCED PATH is interesting. The  
 212 best approximation factor for the former [18] is worse than  $n^{0.99}$ , while the latter is known



213 to have the same inapproximability as MIS [31]. However an  $n^\varepsilon$ -approximation algorithm  
 214 (for every  $\varepsilon > 0$ ) is not excluded for LONGEST PATH. We show that the property of bounded  
 215 twin-width is unlikely to help for these two problems, as it would lead to better approximation  
 216 algorithms for LONGEST PATH in general graphs. This is mainly because subdividing at least  
 217  $2 \log n$  times every edge of any  $n$ -vertex graph gives a graph with twin-width at most 4 [3].

218 ► **Theorem 8.** *For any  $r = \omega(1)$ , an  $r$ -approximation for LONGEST INDUCED PATH  
 219 or LONGEST PATH on graphs given with an  $O(1)$ -sequence would imply a  $(1 + o(1))r$ -  
 220 approximation for LONGEST PATH in general graphs.*

221 In turn, this can be used to exhibit a family  $\mathcal{H}$  with an infinite antichain for the *induced*  
 222 *subgraph* relation such that MUTUALLY INDUCED  $\mathcal{H}$ -PACKING is *hard* to  $n^\varepsilon$ -approximate on  
 223 graphs of bounded twin-width. The family  $\mathcal{H}$  is simply the set of all paths terminated by  
 224 triangles at both ends.

225 ► **Theorem 9.** *There is an infinite family  $\mathcal{H}$  of connected graphs such that if for every  $\varepsilon > 0$ ,  
 226 MUTUALLY INDUCED  $\mathcal{H}$ -PACKING admits an  $n^\varepsilon$ -approximation algorithm on  $n$ -vertex graphs  
 227 given with an  $O(1)$ -sequence, then so does LONGEST PATH on general graphs.*

228 Table 1 summarizes our results and hints at future work.

Problem name	lower bound general graphs	upper bound bounded tww	lower bound bounded tww
MAX INDEPENDENT SET	$n^{1-\varepsilon}$	$n^\varepsilon$	?, self-improvement
COLORING	$n^{1-\varepsilon}$	$n^\varepsilon$	$4/3 - \varepsilon$
MAX INDUCED MATCHING	$n^{1-\varepsilon}$	$n^\varepsilon$	?
MUT. IND. $H$ -PACKING	$n^{1-\varepsilon}$	$n^\varepsilon$ ( $H$ connected)	?
MUT. IND. $\mathcal{H}$ -PACKING	$n^{1-\varepsilon}$	$n^\varepsilon$ for some $\mathcal{H}$	LONGEST PATH-hard
MIN IND. DOM. SET	$n^{1-\varepsilon}$	$n/\text{polylog}(n)$	$n^{1-\varepsilon}$
LONGEST PATH	$2^{\log^{1-\varepsilon} n}$	$n/\exp(\Omega(\sqrt{\log n}))$	LONGEST PATH-hard
LONGEST INDUCED PATH	$n^{1-\varepsilon}$	$n/\text{polylog}(n)$	LONGEST PATH-hard
MIN DOMINATING SET	$(1 - \varepsilon) \ln n$	$O(1)$	?

► **Table 1** Approximability status of graph problems in general graphs and in graphs of bounded twin-width given with an  $O(1)$ -sequence. Everywhere “ $\varepsilon$ ” should be read as “ $\forall \varepsilon > 0$ ”. Our results are enclosed by boxes. “LONGEST PATH-hard” means that getting an  $r$ -approximation would yield essentially the same ratio for LONGEST PATH in general graphs. The other lower bounds are under standard complexity-theoretic assumptions, mostly  $P \neq NP$ . Not to clutter the table, we do not put the references, which can all be found in the paper.

229 For the main highly inapproximable graph problems, we either obtain an  $n^\varepsilon$ -approximation  
 230 algorithm on graphs of bounded twin-width given with an  $O(1)$ -sequence, or a conditional  
 231 obstruction to such an algorithm. In the former case, can we improve further the approxima-  
 232 tion factor? The next theorem was observed using the self-improvement reduction of Feige  
 233 et al. [17], which preserves the twin-width bound. This reduction consists of going from a  
 234 graph  $G$  to the lexicographic product  $G[G]$ , where every vertex of  $G$  is replaced by a module  
 235 inducing a copy of  $G$  (and iterating this trick).

236 ► **Theorem 10** ([7]). *Let  $r : \mathbb{N} \rightarrow \mathbb{R}$  be any non-decreasing function such that for every  
 237  $\varepsilon > 0$ ,  $r(n) = o(n^\varepsilon)$ . If MAX INDEPENDENT SET admits an  $r(n)$ -approximation algorithm*

238 on  $n$ -vertex graphs of bounded twin-width given with an  $O(1)$ -sequence, then it further admits  
 239 an  $r(n)^\varepsilon$ -approximation.

240 To our knowledge, the application of the self-improvement trick is always to strengthen  
 241 a lower bound, and never to effortlessly obtain a better approximation factor. Therefore, we  
 242 may take Theorem 10 as a weak indication that our approximation ratio is best possible.  
 243 Still, not even a polynomial-time approximation scheme (PTAS) is ruled out for MIS (nor  
 244 for MAX INDUCED MATCHING, MIN DOMINATING SET, etc.) and we would like to see better  
 245 approximation algorithms. For COLORING, as was previously observed [7], a PTAS is ruled  
 246 out by the NP-hardness of deciding if a planar graph is 3-colorable or 4-chromatic, since  
 247 planar graphs have twin-width at most 9 and a 9-sequence can be found in linear time [24].

## 248 2 Preliminaries

249 For  $i$  and  $j$  two integers, we denote by  $[i, j]$  the set of integers that are at least  $i$  and at  
 250 most  $j$ . For every integer  $i$ ,  $[i]$  is a shorthand for  $[1, i]$ .

### 251 2.1 Handled graph problems

252 We will consider several problems throughout the paper. We recall here the definition of the  
 253 most central ones. Some technical problem generalizations will be defined along the way.

WEIGHTED MAX INDEPENDENT SET (WMIS, for short)

254 **Input:** A graph  $G$  and a weight function  $V(G) \rightarrow \mathbb{Q}$ .

**Output:** A set  $S \subseteq V(G)$  such that  $\forall u, v \in S, uv \notin E(G)$  maximizing  $w(S) := \sum_{v \in S} w(v)$ .

255 A feasible solution to WMIS is called an *independent set*. The MAX INDEPENDENT SET  
 256 (MIS, for short) problem is the particular case with  $w(v) = 1, \forall v \in V(G)$ . We may denote  
 257 by  $\alpha(G)$ , the *independence number*, that is the optimum value of WMIS on graph  $G$ .

COLORING

258 **Input:** A graph  $G$ .

**Output:** A partition  $\mathcal{P}$  of  $V(G)$  into independent sets minimizing the cardinality of  $\mathcal{P}$ .

259 Equivalently, COLORING can be expressed as finding an integer  $k$  and a map  $c : V(G) \rightarrow [k]$   
 260 such that for every  $uv \in E(G)$ ,  $c(u) \neq c(v)$ , while minimizing  $k$ .

MAX INDUCED MATCHING

261 **Input:** A graph  $G$ , possibly together with a weight function  $w : E(G) \rightarrow \mathbb{Q}$ .

**Output:** A set  $S \subseteq E(G)$  such that  $\forall uv \neq u'v' \in S, \{u, v\} \cap \{u', v'\} = \emptyset$  and  
 $G[\{u, v, u', v'\}]$  has exactly two edges, maximizing  $w(S) := \sum_{e \in S} w(e)$ .

262 An *induced matching* is a pairwise disjoint set of edges (i.e., a matching) with no edge  
 263 bridging them. We now give a common generalization of WMIS and MAX INDUCED  
 264 MATCHING.

MUTUALLY INDUCED  $\mathcal{H}$ -PACKING

265 **Input:** A graph  $G$ , possibly together with a weight function  $w : V(G) \rightarrow \mathbb{Q}$ .

**Output:** A set  $S \subseteq V(G)$  such that  $G[S]$  is a disjoint union of graphs each isomorphic  
 to a graph in  $\mathcal{H}$ , maximizing  $w(S) := \sum_{v \in S} w(v)$ .

266 When  $\mathcal{H}$  consists of a single graph, say  $H$ , we simply denote the former problem MUTUALLY  
 267 INDUCED  $H$ -PACKING. WMIS and MAX INDUCED MATCHING are the special cases when  $H$   
 268 is a vertex and an edge, respectively.



## 2.2 The contraction and partition viewpoints of twin-width

A *trigraph*  $G$  has vertex set  $V(G)$ , black edge set  $E(G)$ , red edge set  $R(G)$  such that  $E(G) \cap R(G) = \emptyset$  (and  $E(G), R(G) \subseteq \binom{V(G)}{2}$ ). A *contraction* in a trigraph  $G$  replaces a pair of (non-necessarily adjacent) vertices  $u, v \in V(G)$  by one vertex  $w$  that is linked to  $G - \{u, v\}$  in the following way to form a new trigraph  $G'$ . For every  $z \in V(G) \setminus \{u, v\}$ ,  $wz \in E(G')$  whenever  $uz, vz \in E(G)$ ,  $wz \notin E(G') \cup R(G')$  whenever  $uz, vz \notin E(G) \cup R(G)$ , and  $wz \in R(G')$ , otherwise. The *red graph*  $(V(G), R(G))$  will be denoted by  $\mathcal{R}(G)$ . We denote by  $\mathcal{T}(G)$  the *total graph* of  $G$  defined as  $(V(G), E(G) \cup R(G))$ . An *induced subtrigraph* of a trigraph  $G$  is obtained by removing vertices (but no edges) to  $G$ , analogously to induced subgraphs. A partial contraction sequence of an  $n$ -vertex (tri)graph  $G$  (to a trigraph  $H$ ) is a sequence of trigraphs  $G = G_n, \dots, G_t = H$  for some  $t \in [n]$  such that  $G_i$  is obtained from  $G_{i+1}$  by performing one contraction. A (complete) contraction sequence is such that  $t = 1$ , that is,  $H$  is the 1-vertex trigraph. A *d-sequence*  $\mathcal{S}$  of  $G$  is a contraction sequence of  $G$  in which the red graph of every trigraph of  $\mathcal{S}$  has maximum degree at most  $d$ .

Assume that there is a partial contraction sequence from a (tri)graph  $G$  to a trigraph  $H$ . If  $u$  is a vertex of  $H$ , then  $u(G) \subseteq V(G)$  denotes the set of vertices eventually contracted into  $u$  in  $H$ . We denote by  $\mathcal{P}(H)$  the partition  $\{u(G) : u \in V(H)\}$  of  $V(G)$ . If  $G$  is clear from the context, we may refer to a *part* of  $H$  as any set in  $\{u(G) : u \in V(H)\}$ . We will mostly see  $d$ -sequences as sequences of partitions, that is,  $\mathcal{P}_n, \dots, \mathcal{P}_t$  with  $\mathcal{P}_i := \{u(G) : u \in V(G_i)\}$  when  $G_n, \dots, G_t$  is a partial (contraction)  $d$ -sequence.

Given a graph  $G$  and a partition  $\mathcal{P}$  of  $V(G)$ , the *quotient graph* of  $G$  with respect to  $\mathcal{P}$  is the graph with vertex set  $\mathcal{P}$ , where  $PP'$  is an edge if there is  $u \in P$  and  $v \in P'$  such that  $uv \in E(G)$ . Given a (tri)graph  $G$  and a partition  $\mathcal{P}$  of  $V(G)$ , the *quotient trigraph*  $G/\mathcal{P}$  is the trigraph with vertex set  $\mathcal{P}$ , where  $PP'$  is a black edge if these two parts are fully adjacent – for every  $u \in P$  and every  $v \in P'$ ,  $uv \in E(G)$  –, and a red edge if either there is  $u \in P$  and  $v \in P'$  such that  $uv \in R(G)$ , or there is  $u_1, u_2 \in P$  and  $v_1, v_2 \in P'$  such that  $u_1v_1 \in E(G)$  and  $u_2v_2 \notin E(G)$ .

A trigraph  $H$  is a *cleanup* of another trigraph  $G$  if  $V(H) = V(G)$ ,  $R(H) \subseteq R(G)$ , and  $E(G) \subseteq E(H) \subseteq E(G) \cup R(G)$ . That is,  $H$  is obtained from  $G$  by turning some of its red edges into black edges or non-edges. We further say that  $H$  is *full cleanup* of  $G$  if  $H$  has no red edge, and thus, is considered as a graph. Note that the total graph  $\mathcal{T}(G)$  and the *black graph*  $(V(G), E(G))$  of a trigraph  $G$  are extreme examples of full cleanups of  $G$ .

## 2.3 Balanced partition sequences

The notion of *versatile twin-width* is a crucial opening step to our algorithms; see [5]. Let us call *d-contraction* a contraction between two trigraphs of maximum red degree at most  $d$ . A *tree of d-contractions* of a trigraph  $G$  (of maximum red degree at most  $d$ ) is a rooted tree, whose root is labeled by  $G$ , whose leaves are all labeled by 1-vertex trigraphs  $K_1$ , and such that one can go from any parent to any of its children by performing a single  $d$ -contraction. Observe that  $d$ -sequences coincide with trees of  $d$ -contractions that are paths. A trigraph  $G$  has *versatile twin-width*  $d$  if  $G$  admits a tree of  $d$ -contractions in which every internal node, labeled by, say,  $F$ , has at least  $|V(F)|/d$  children each obtained by contracting one of a list of  $|V(F)|/d$  pairwise disjoint pairs of vertices of  $F$ .

It was shown that twin-width and versatile twin-width are functionally equivalent [5]. The relevant consequence for our purposes is that every graph  $G$  with a  $d'$ -sequence admits a *balanced d-sequence*, where  $d = h(d')$  depends only on  $d'$ , i.e., one for which the partitions  $\mathcal{P}_n, \dots, \mathcal{P}_1$  are such that for every  $i \in [n]$  and  $P \in \mathcal{P}_i$ ,  $|P| \leq d \cdot \frac{n}{i}$ . As we will resort to

315 recursion on induced subtrigraphs and quotient trigraphs, we need to keep more information  
 316 on those substances that the mere fact that they have twin-width at most  $d$  (otherwise the  
 317 twin-width bound could quickly diverge).

318 This will be done by opening up the proof in [5], and handling divided  $0, 1, r$ -matrices  
 319 with some specific properties. Thus we need to recall the relevant definitions.

320 Given two partitions  $\mathcal{P}, \mathcal{P}'$  of the same set, we say that  $\mathcal{P}'$  is a *coarsening* of  $\mathcal{P}$  if every  
 321 part of  $\mathcal{P}$  is contained in a part of  $\mathcal{P}'$ , and  $\mathcal{P}, \mathcal{P}'$  are distinct. Given a matrix  $M$ , we call *row*  
 322 *division* (resp. *column division*) a partition of the rows (resp. columns) of  $M$  into parts of  
 323 consecutive rows (resp. columns). A  $(k, \ell)$ -*division*, or simply *division*, of a matrix  $M$  is a pair  
 324  $(\mathcal{R} = \{R_1, \dots, R_k\}, \mathcal{C} = \{C_1, \dots, C_\ell\})$  where  $\mathcal{R}$  is a row division and  $\mathcal{C}$  is a column division.  
 325 In a matrix division  $(\mathcal{R}, \mathcal{C})$ , each part  $R \in \mathcal{R}$  is called a *row part*, and each part  $C \in \mathcal{C}$  is  
 326 called a *column part*. Given a subset  $R$  of rows and a subset  $C$  of columns in a matrix  $M$ , the  
 327 *zone*  $M[R, C]$  denotes the submatrix of all entries of  $M$  at the intersection between a row of  $R$   
 328 and a column of  $C$ . A *zone* of a matrix partitioned by  $(\mathcal{R}, \mathcal{C}) = (\{R_1, \dots, R_k\}, \{C_1, \dots, C_\ell\})$   
 329 is any  $M[R_i, C_j]$  for  $i \in [k]$  and  $j \in [\ell]$ . A zone is *constant* if all its entries are identical,  
 330 *horizontal* if all its columns are equal, and *vertical* if all its rows are equal. A *0,1-corner* is a  
 331  $2 \times 2$   $0, 1$ -matrix which is neither horizontal nor vertical.

332 Unsurprisingly,  $0, 1, r$ -matrices are such that each entry is in  $\{0, 1, r\}$  where  $r$  is an error  
 333 symbol that should be understood as a red edge. A *neat division* of a  $0, 1, r$ -matrix is a  
 334 division for which every zone either contains only  $r$  entries or contains no  $r$  entry and is  
 335 horizontal or vertical (or both, i.e., constant). Zones filled with  $r$  entries are called *mixed*.  
 336 A *neatly divided matrix* is a pair  $(M, (\mathcal{R}, \mathcal{C}))$  where  $M$  is a  $0, 1, r$ -matrix and  $(\mathcal{R}, \mathcal{C})$  is a neat  
 337 division of  $M$ . A  *$t$ -mixed minor* in a neatly divided matrix is a  $(t, t)$ -division which coarsens  
 338 the neat subdivision, and contains in each of its  $t^2$  zones at least one mixed zone (i.e., filled  
 339 with  $r$  entries) or a  $0, 1$ -corner. A neatly divided matrix is said  *$t$ -mixed free* if it does not  
 340 admit a  $t$ -mixed minor.

341 A *mixed cut of a row part*  $R \in \mathcal{R}$  of a neatly divided matrix  $(M, (\mathcal{R}, \mathcal{C} = \{C_1, C_2, \dots\}))$  is  
 342 an index  $i$  such that both  $M[R, C_i]$  and  $M[R, C_{i+1}]$  are not mixed, and there is a  $0, 1$ -corner  
 343 in the  $2$ -by- $|R|$  zone defined by the last column of  $C_i$ , the first column of  $C_{i+1}$ , and  $R$ . The  
 344 *mixed value of a row part*  $R \in \mathcal{R}$  of a neatly divided matrix  $(M, (\mathcal{R}, \mathcal{C} = \{C_1, C_2, \dots\}))$  is  
 345 the number of mixed zones  $M[R, C_j]$  plus the number of mixed cuts between two (adjacent  
 346 non-mixed) zones  $M[R, C_j]$  and  $M[R, C_{j+1}]$ . We similarly define the mixed value of a column  
 347 part  $C \in \mathcal{C}$ . The *mixed value of a neat division* of a  $0, 1, r$ -matrix is the maximum of  
 348 the mixed values taken over every part. The *part size* of a division  $(\mathcal{R}, \mathcal{C})$  is defined as  
 349  $\max(\max_{R \in \mathcal{R}} |R|, \max_{C \in \mathcal{C}} |C|)$ . A division is *symmetric* if the largest row index of each row  
 350 part and the largest column index of each column part define the same set of integers. We  
 351 call *symmetric fusion* of a symmetric division the fusion of two consecutive parts in  $\mathcal{C}$  and  
 352 of the two corresponding parts in  $\mathcal{R}$ . A symmetric fusion on a symmetric division yields  
 353 another symmetric division. A matrix  $A := (a_{i,j})_{i,j}$  is said *symmetric* in the usual sense,  
 354 namely, for every entry  $a_{i,j}$  of  $A$ ,  $a_{i,j} = a_{j,i}$ .

355 In what follows, we set  $c_d := 8/3(t+1)^2 2^{4t}$ . The following definition is key.

356 ► **Definition 11.** Let  $\mathcal{M}_{n,d}$  be the class of the neatly divided  $n \times n$  symmetric  $0, 1, r$ -matrices  
 357  $(M, (\mathcal{R}, \mathcal{C}))$ , such that  $(\mathcal{R}, \mathcal{C})$  is symmetric and has:

- 358 ■ mixed value at most  $4c_d$ ,
- 359 ■ part size at most  $2^{4c_d+2}$ , and
- 360 ■ no  $d$ -mixed minor.

361 The *red number* of a matrix is the maximum number of  $r$  entries in a single column or  
 362 row of the matrix.

363 ► **Lemma 12.** *Let  $(M, (\mathcal{R}, \mathcal{C})) \in \mathcal{M}_{n,d}$ . The red number of  $M$  is at most  $c_d \cdot 2^{4c_d+4}$ . Thus,*  
 364 *the trigraph whose adjacency matrix is  $M$  has maximum red degree at most  $c_d \cdot 2^{4c_d+4}$ .*

365 **Proof.** Any row or column intersects at most  $4c_d$  mixed zones (filled with  $r$  entries). Each  
 366 mixed zone has width and length bounded by the part size  $2^{4c_d+2}$ . Hence the maximum  
 367 total number of  $r$  entries on a single row or column is at most  $4c_d \cdot 2^{4c_d+2} = c_d \cdot 2^{4c_d+4}$ . ◀

368 A *coarsening* of a neatly divided matrix  $(M, (\mathcal{R}, \mathcal{C}))$  is a neatly divided matrix  $(M', (\mathcal{R}', \mathcal{C}'))$   
 369 such that  $(\mathcal{R}', \mathcal{C}')$  is a coarsening of  $(\mathcal{R}, \mathcal{C})$ , and  $M'$  is obtained from  $M$  by setting to  $r$  all  
 370 entries that lie, in  $M$  divided by  $(\mathcal{R}', \mathcal{C}')$ , in a zone with at least one  $r$  entry or a 0,1-corner.  
 371 We also refer to the process of going from  $(M, (\mathcal{R}, \mathcal{C}))$  to  $(M', (\mathcal{R}', \mathcal{C}'))$  as *coarsening operation*.  
 372 A coarsening operation from  $(M, (\mathcal{R}, \mathcal{C})) \in \mathcal{M}_{n,d}$  to  $(M', (\mathcal{R}', \mathcal{C}'))$  is said *invariant-preserving*  
 373 if  $(M', (\mathcal{R}', \mathcal{C}')) \in \mathcal{M}_{n,d}$ .

374 The following lemma is the crucial building block of the current section.

375 ► **Lemma 13** ([6, Lemma 18]). *We set  $s := 2^{4c_d+4}$ . Every neatly divided matrix  $(M, (\mathcal{R}, \mathcal{C})) \in$   
 376  $\mathcal{M}_{n,d}$  has an invariant-preserving coarsening  $(M', (\mathcal{R}', \mathcal{C}')) \in \mathcal{M}_{n,d}$  with  $\lfloor n/s \rfloor$  disjoint pairs  
 377 of identical columns. Given  $(M, (\mathcal{R}, \mathcal{C}))$ , both  $(M', (\mathcal{R}', \mathcal{C}'))$  and the pairs of columns can be  
 378 computed in  $n^{O(1)}$  time.*

379 In [6], it is not explicitly stated that the invariant-preserving coarsening (hence the pairs  
 380 of identical columns) can be found in polynomial time. However it is easy to check that  
 381 the proof is effective, since it greedily symmetrically fuses two consecutive parts, provided  
 382 the resulting divided matrix remains in  $\mathcal{M}_{n,d}$ . A special case of the following observation is  
 383 shown in [6, Lemma 19].

384 ► **Lemma 14.** *Let  $(M, (\mathcal{R}, \mathcal{C})) \in \mathcal{M}_{n,d}$  be a neatly divided matrix. Removing a set of  
 385  $h$  columns and the  $h$  corresponding rows, and possibly removing from the division the parts  
 386 that are now empty, results in a neatly divided matrix in  $\mathcal{M}_{n-h,d}$ .*

387 **Proof.** By construction, the new matrix and division are symmetric. The new neatly divided  
 388 matrix remains  $d$ -mixed free. The part size and the mixed value can only decrease. ◀

389 ► **Lemma 15** ([6, Beginning of Lemma 20]). *Given any graph  $G$  with a  $d$ -sequence, one  
 390 can find in polynomial-time an adjacency matrix  $M$  of  $G$ , such that  $(M, (\mathcal{R}, \mathcal{C}))$  is a neatly  
 391 divided matrix of  $\mathcal{M}_{n,2d+2}$  with  $(\mathcal{R}, \mathcal{C})$  the finest division of  $M$  (i.e., the one where all parts  
 392 are of size 1).*

393 The adjacency matrix of a trigraph extends the one of a graph by putting  $r$  symbols  
 394 when the vertices of the corresponding row and column are linked by a red edge. A neatly  
 395 divided matrix  $(M, (\mathcal{R}, \mathcal{C}))$  is said *conform* to a trigraph  $G$  if  $M$  is the adjacency matrix of  
 396 a trigraph  $G'$  such that  $G$  is a cleanup of  $G'$ . Furthermore, we assume (and keep implicit)  
 397 that we know the one-to-one correspondence between each row (and corresponding column)  
 398 of  $M$  and vertex of  $G$ .

399 ► **Lemma 16.** *Let  $d$  be a natural,  $s := 2^{4c_d+4}$ , and  $d' := c_d \cdot 2^{4c_d+4}$ . Let  $G$  be an  $n$ -vertex  
 400 trigraph given with a neatly divided matrix  $(M, (\mathcal{R}, \mathcal{C})) \in \mathcal{M}_{n,d}$  conform to  $G$ . A partial  
 401  $d'$ -sequence  $\mathcal{S}$  from  $G$  to a trigraph  $H$  satisfying*

402 ■  $|V(H)| = \lfloor \sqrt{n} \rfloor$ , and

403 ■  $\forall u \in V(H), |u(G)| \leq s\sqrt{n}$ ,

404 *and a neatly divided matrix  $(M', (\mathcal{R}', \mathcal{C}')) \in \mathcal{M}_{\lfloor \sqrt{n} \rfloor, d}$  conform to  $H$  can be computed in time  
 405  $n^{O(1)}$ .*

406 **Proof.** This is a consequence of Lemmas 13 and 14; see the proof of the more general  
407 Lemma 18.  $\blacktriangleleft$

408 Combining Lemmas 15 and 16, one obtains the following.

409 **► Lemma 17.** *Let  $d$  be a natural,  $s := 2^{4c_d+4}$ , and  $d' := c_d \cdot 2^{4c_d+4}$ . Given an  $n$ -vertex graph  
410  $G$  with a  $d$ -sequence, one can compute in time  $n^{O(1)}$  a partition  $\mathcal{P} = \{P_1, P_2, \dots, P_{\lfloor \sqrt{n} \rfloor}\}$  of  
411  $V(G)$  satisfying*

- 412  $\blacksquare$  *for every integer  $1 \leq i \leq \lfloor \sqrt{n} \rfloor$ ,  $|P_i| \leq s\sqrt{n} \leq d'\sqrt{n}$ , and*
- 413  $\blacksquare$  *the red graph of  $G/\mathcal{P}$  has maximum degree at most  $d'$ .*

414 We will need a stronger inductive form of Lemma 17, also a consequence of Lemmas 15  
415 and 16.

416 **► Lemma 18.** *Let  $\hat{d}$  be a natural,  $d = 2\hat{d} + 2$ , and set  $s := 2^{4c_d+4}$ , and  $d' := c_d \cdot 2^{4c_d+4}$ .  
417 Given an  $n$ -vertex graph  $G$  given with a  $\hat{d}$ -sequence, or an  $n$ -vertex trigraph  $G$  with a neatly  
418 divided matrix  $(M, (\mathcal{R}, \mathcal{C})) \in \mathcal{M}_{n,d}$  such that  $M$  is conform to  $G$ , one can compute in time  
419  $n^{O(1)}$  a partition  $\mathcal{P} = \{P_1, P_2, \dots, P_{\lfloor \sqrt{n} \rfloor}\}$  of  $V(G)$  with maximum red degree at most  $d'$   
420 satisfying that, for every integer  $1 \leq i \leq \lfloor \sqrt{n} \rfloor$ ,  $|P_i| \leq s\sqrt{n} \leq d'\sqrt{n}$ , and for any trigraph  $H$   
421 that is*

- 422  $\blacksquare$  *a cleanup of an induced subtrigraph of  $G/\mathcal{P}$ , or*
- 423  $\blacksquare$  *an induced subtrigraph  $G[\bigcup_{i \in J \subseteq [\lfloor \sqrt{n} \rfloor]} P_i]$ ,*

424 *a neatly divided matrix  $(M', (\mathcal{R}', \mathcal{C}')) \in \mathcal{M}_{|V(H)|,d}$  conform to  $H$  can be computed in time  
425  $n^{O(1)}$ .*

426 **Proof.** If we are given a graph  $G$  with a  $\hat{d}$ -sequence, we immediately compute a neatly  
427 divided matrix  $(M, (\mathcal{R}, \mathcal{C})) \in \mathcal{M}_{n,d}$  conform to  $G$ , by Lemma 15. We then proceed as if we  
428 received the second kind of input.

429 We will build iteratively the partition  $\mathcal{P} = \{P_1, P_2, \dots, P_{\lfloor \sqrt{n} \rfloor}\}$  starting from the finest  
430 partition. At each step we merge two parts, until the number of parts is  $\lfloor \sqrt{n} \rfloor$ . At this point,  
431 we have the desired partition  $\mathcal{P}$ .

432 We iteratively maintain a trigraph  $G^z$  and a neatly divided matrix  $(M^z, (\mathcal{R}^z, \mathcal{C}^z)) \in$   
433  $\mathcal{M}_{n-z+1,d}$  conform to it. The maintained partition is just the one corresponding to the  
434 parts of  $G^z$ . Initially,  $G^1$  is  $G$ , and  $(M^1, (\mathcal{R}^1, \mathcal{C}^1)) = (M, (\mathcal{R}, \mathcal{C})) \in \mathcal{M}_{n,d}$ . At step  $z$   
435 we do the following. We apply Lemma 13 on  $(M^z, (\mathcal{R}^z, \mathcal{C}^z)) \in \mathcal{M}_{n-z+1,d}$  and obtain, in  
436 polynomial-time, an invariant-preserving coarsening  $(M'^z, (\mathcal{R}'^z, \mathcal{C}'^z)) \in \mathcal{M}_{n-z+1,d}$ , and  $h :=$   
437  $\lfloor (n-z+1)/s \rfloor$  disjoint pairs of equal columns  $\{c_1, c'_1\}, \dots, \{c_h, c'_h\}$  in  $(M'^z, (\mathcal{R}'^z, \mathcal{C}'^z))$ . Let  
438  $\{r_1, r'_1\}, \dots, \{r_h, r'_h\}$  be the corresponding rows, and  $\{v_1, v'_1\}, \dots, \{v_h, v'_h\}$  the corresponding  
439 vertices. Observe that a coarsening of a neatly divided matrix conform to a trigraph is still  
440 conform to that trigraph, since the new matrix may only have some  $r$  entries in place of  
441 some previously 0 or 1 entries. In particular,  $(M'^z, (\mathcal{R}'^z, \mathcal{C}'^z))$  is conform to  $G^z$ .

442 There is at least one pair  $\{v_i, v'_i\}$  whose contraction forms a part of size at most  $n/h$ .  
443 Indeed, otherwise the union of the parts corresponding to  $v_1, v'_1, \dots, v_h, v'_h$  is larger than  $n$ .  
444 We remove  $c'_i$  and  $r'_i$  from  $(M'^z, (\mathcal{R}'^z, \mathcal{C}'^z))$ . By Lemma 14, we obtain a neatly divided matrix  
445 of  $M_{n-z,d}$  that we denote by  $(M^{z+1}, (\mathcal{R}^{z+1}, \mathcal{C}^{z+1}))$ . As we stop when  $n-z+1 = \lfloor \sqrt{n} \rfloor$ , it  
446 means that the maximum size of a part of our partition is at most  $n/h \leq sn/\sqrt{n} = s\sqrt{n}$ .  
447 The bound on the maximum red degree of the obtained partition (actually of all maintained  
448 partitions) is given by Lemma 12.

449 We now show to find, for any cleanup  $H$  of an induced subtrigraph of  $G/\mathcal{P}$ , a neatly  
450 divided matrix  $(M', (\mathcal{R}', \mathcal{C}')) \in \mathcal{M}_{|V(H)|,d}$  conform to  $H$ . We first observe, as a consequence  
451 of Lemmas 13 and 14, that  $(M^{\lfloor \sqrt{n} \rfloor}, (\mathcal{R}^{\lfloor \sqrt{n} \rfloor}, \mathcal{C}^{\lfloor \sqrt{n} \rfloor})) \in \mathcal{M}_{\lfloor \sqrt{n} \rfloor, d}$  is conform to  $G/\mathcal{P}$ . Taking

452 an induced subgraph  $H'$  of  $G/\mathcal{P}$  (i.e., removing vertices from it), we get, by removing  
 453 the corresponding rows and columns in  $(M^{\lfloor \sqrt{n} \rfloor}, (\mathcal{R}^{\lfloor \sqrt{n} \rfloor}, \mathcal{C}^{\lfloor \sqrt{n} \rfloor}))$  a neatly divided matrix  
 454  $(M', (\mathcal{R}', \mathcal{C}')) \in \mathcal{M}_{|V(H')|, d}$  conform to  $H'$ , by Lemma 14. Note finally that taking a cleanup  
 455  $H$  of  $H'$ , we can simply keep  $(M', (\mathcal{R}', \mathcal{C}'))$  as a neatly divided matrix of  $\mathcal{M}_{|V(H)|, d}$  conform  
 456 to  $G$ . The second item, concerning induced subtrigraphs  $G[\bigcup_{i \in J \subseteq [\lfloor \sqrt{n} \rfloor]} P_i]$  is a simple  
 457 application of Lemma 14, and works more generally for any induced subtrigraph of  $G$ . ◀

458 In effect, we will only apply Lemma 18 for graphs  $G$  and  $H$ , i.e., when  $H$  is an induced  
 459 subgraph of  $G$  or a full cleanup of an induced subtrigraph of  $G/\mathcal{P}$ . Indeed, the structures  
 460  $H$  will correspond to subinstances. We want those to be graphs, so that the tackled graph  
 461 problem is well-defined on them.

### 462 3 Approximation algorithms for Max Independent Set

463 We naturally start our study with MAX INDEPENDENT SET, a central problem that is very  
 464 inapproximable [23, 34], and yet constitutes the textbook example of our approach.

#### 465 3.1 Subexponential-time constant-approximation algorithm

466 We present a subexponential-time  $O_d(1)$ -approximation for WMIS on graphs given with  
 467 a  $d$ -sequence, which we recall, is unlikely to exist in general graphs [14].

468 ▶ **Lemma 19.** *Let  $d'$  be a natural,  $s := 2^{4c_{d'}+4}$ , and  $d := c_{d'} \cdot 2^{4c_{d'}+4}$ . Assume  $n$ -vertex inputs  
 469  $G$ , vertex-weighted by  $w$ , are given with a  $d'$ -sequence. WEIGHTED MAX INDEPENDENT SET  
 470 can be  $(d+1)$ -approximated in time  $2^{O_d(\sqrt{n})}$  on these inputs.*

471 **Proof.** By Lemma 17, we compute in polynomial time a partition  $\mathcal{P} = \{P_1, \dots, P_{\lfloor \sqrt{n} \rfloor}\}$   
 472 of  $V(G)$  whose parts have size at most  $s\sqrt{n}$  and such that  $\mathcal{R}(G/\mathcal{P})$  has maximum degree at  
 473 most  $d$ .

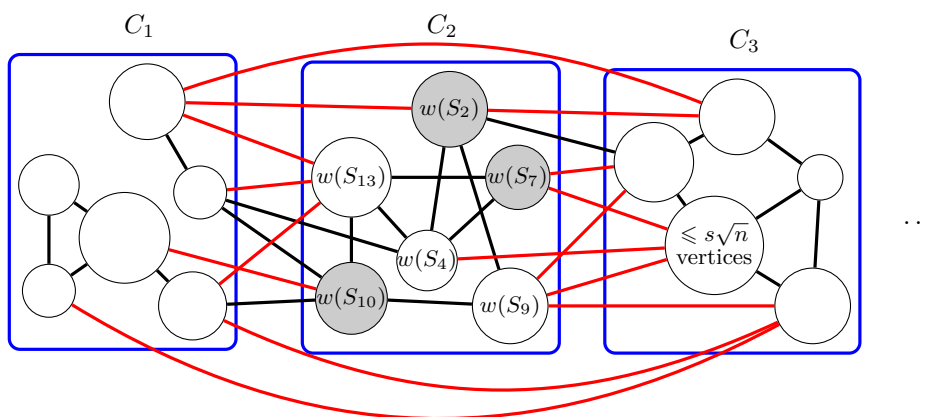
474 For every integer  $1 \leq i \leq \lfloor \sqrt{n} \rfloor$ , we compute a heaviest independent set in  $G[P_i]$ , say  $S_i$ .  
 475 Even with an exhaustive algorithm, this takes time  $\sqrt{n} \cdot s^2 \sqrt{n} \cdot 2^{s\sqrt{n}} = 2^{O_d(\sqrt{n})}$ . We then  
 476  $(d+1)$ -color (in linear time)  $\mathcal{R}(G/\mathcal{P})$ , which is possible since this graph has maximum degree  
 477 at most  $d$ . This defines a coarsening of  $\mathcal{P}$  in  $d+1$  parts  $\mathcal{Q} = \{C_1, \dots, C_{d+1}\}$ . Thus,  $\mathcal{Q}$  is  
 478 a partition of  $V(G)$  such that  $C_j$  consists of all the parts  $P_i \in \mathcal{P}$  receiving color  $j$  in the  
 479  $(d+1)$ -coloring of  $\mathcal{R}(G/\mathcal{P})$ .

480 For every  $j \in [d+1]$ , let  $H_j$  be the graph  $(G/\mathcal{P})[C_j]^4$  vertex-weighted by  $P_i \subseteq C_j \mapsto w(S_i)$ .  
 481 Note that  $(G/\mathcal{P})[C_j]$  can indeed be assimilated to a graph, since it has, by design, no red edge.  
 482 We compute a heaviest independent set in  $H_j$ , say  $R_j$ . This takes time  $(d+1) \cdot n \cdot 2^{s\sqrt{n}} = 2^{O_d(\sqrt{n})}$ .  
 483 We output  $\bigcup_{P_i \subseteq R_j} S_i$  for the index  $j \in [d+1]$  maximizing  $\sum_{P_i \subseteq R_j} w(S_i)$ .

484 This finishes the description of the algorithm. We already argued that its running time is  
 485  $2^{O_d(\sqrt{n})}$ . We shall justify that it does output an independent set of weight at least a  $\frac{1}{d+1}$   
 486 fraction of the optimum  $\alpha(G)$ .

487  **$I$  is indeed an independent set.** For any  $j \in [d+1]$ , consider two vertices  $x, y \in$   
 488  $\bigcup_{P_i \subseteq R_j} S_i$ . If  $\{x, y\} \in S_i$  for some  $i$ , then  $x$  and  $y$  are non-adjacent since  $S_i$  is an independent  
 489 set of  $G[P_i]$ . Else  $x \in S_i$  and  $y \in S_{i'}$  for some  $i \neq i'$ .  $P_i$  and  $P_{i'}$  are not linked by a black  
 490 edge in  $(G/\mathcal{P})[C_j]$  since  $R_j$  is an independent set in  $H_j$ , nor they can be linked by a red  
 491 edge (there are none in  $(G/\mathcal{P})[C_j]$ ). Thus again,  $x$  and  $y$  are non-adjacent in  $G$ .

<sup>4</sup> We use this notation as a slight abuse of notation for  $(G/\mathcal{P})[\{P_i : P_i \subseteq C_j\}]$ .



■ **Figure 2** The trigraph  $G/\mathcal{P}$  with its  $\lfloor \sqrt{n} \rfloor$  vertices, each corresponding to a subset of at most  $s\sqrt{n}$  vertices of  $G$ . The weights  $w(S_i)$  of heaviest independent sets  $S_i$  of  $G[P_i]$  for each part  $P_i$  of the color class  $C_2$  of the  $d + 1$ -coloring of  $\mathcal{R}(G/\mathcal{P})$ . A heaviest independent set in the so-weighted  $(G/\mathcal{P})[C_2]$  (shaded) corresponds to an optimum solution in  $G[\bigcup_{P_i \subseteq C_2} P_i]$ . One of these  $d + 1$  independent sets is a  $d + 1$ -approximation.

492  $I$  has weight at least  $\frac{\alpha(G)}{d+1}$ . We claim that  $\bigcup_{P_i \subseteq R_j} S_i$  is a heaviest independent set of  
 493  $G[C_j]$ . Note that the  $P_i$ s that are included in  $C_j$  (and partition it) form a module partition of  
 494  $G[C_j]$ . In particular, any heaviest independent set intersecting some  $P_i \subseteq C_j$  has to contain  
 495 a heaviest independent of  $G[P_i]$ . This is precisely what the algorithm computes. Then a  
 496 heaviest independent set in  $G[C_j]$  packs such subsolutions to maximize the total weight,  
 497 which is what is computed in  $H_j$ .

498 We conclude by the pigeonhole principle, since a heaviest independent set  $X$  of  $G$  is such  
 499 that  $w(X \cap C_j) \geq \frac{\alpha(G)}{d+1}$  for some  $j \in [d + 1]$ . ◀

### 500 3.2 Improving the approximation factor

501 We notice in this short section that the approximation factor of Lemma 19 can be improved  
 502 using the notion of *clustered coloring*. The *clustered chromatic number* of a class of graphs is  
 503 the smallest integer  $k$  such that there is a constant  $c$  for which all the graphs of the class can  
 504 be  $k$ -colored such that every color class induces a subgraph whose connected components  
 505 have size at most  $c$ . A proper coloring is a particular case of clustered coloring when  $c = 1$ .

506 Instead of properly coloring the red graph, as we did in the proof of Lemma 19, we could  
 507 use less colors and allow for small monochromatic components (in place of monochromatic  
 508 components of size 1). We use for that the following bound due to Alon et al.

509 ▶ **Theorem 20** ([1]). *The class of graphs of maximum degree at most  $d$  has clustered chromatic*  
 510 *number at most  $\lceil \frac{d+2}{3} \rceil$ .*

511 We can use this lemma to improve our approximation algorithms.

512 ▶ **Theorem 21.** *On inputs as in Lemma 19 with  $s := 2^{4c_{d'}+4}$ , and  $d := c_{d'} \cdot 2^{4c_{d'}+4}$ ,*  
 513 *WEIGHTED MAX INDEPENDENT SET further admits an  $\lceil \frac{d+2}{3} \rceil$ -approximation algorithm in*  
 514 *time  $2^{O_d(\sqrt{n})}$ .*

515 **Proof.** Again, we compute in polynomial time a partition  $\mathcal{P} = \{P_1, \dots, P_{\lfloor \sqrt{n} \rfloor}\}$  of  $V(G)$   
 516 whose parts have size at most  $s\sqrt{n}$  and such that  $\mathcal{R}(G/\mathcal{P})$  has maximum degree at most  $d$ ,



517 using Lemma 17. Let  $c$  be the constant such that  $\mathcal{R}(G/\mathcal{P})$  admits a clustered coloring using  
 518  $\lceil \frac{d+2}{3} \rceil$  colors such that each color class  $C_j$  (with  $j \in [\lceil \frac{d+2}{3} \rceil]$ ) is such that the connected  
 519 components  $C_j^1, C_j^2, \dots, C_j^{h_j} \subseteq C_j$  of  $\mathcal{R}(G/\mathcal{P})[C_j]$  have size at most  $c$  each. This coloring  
 520 is guaranteed to *exist* by Theorem 20. Due to the overall running time, we might as well  
 521 compute it by exhaustive search, in time  $2^{O_d(\sqrt{n})}$ .

522 For every  $j \in [\lceil \frac{d+2}{3} \rceil]$  and  $h \in [h_j]$ , we denote by  $P_1(C_j^h), \dots, P_{c(j,h)}(C_j^h)$  the  $c(j,h) \leq c$   
 523 parts  $P_i \in \mathcal{P}$  that are included in  $C_j^h$ . For every  $j \in [\lceil \frac{d+2}{3} \rceil]$ , every  $h \in [h_j]$ , and every  
 524  $J \subseteq [c(j,h)]$ , we compute a heaviest independent set in  $G[\bigcup_{z \in J} P_z(C_j^h)]$ , which we denote  
 525 by  $S_{j,h,J}$ . This takes time  $O(\sqrt{n} \cdot 2^c \cdot 2^{sc\sqrt{n}}) = 2^{O_d(\sqrt{n})}$  since  $|\bigcup_{z \in J} P_z(C_j^h)| \leq c \cdot s\sqrt{n}$ .

526 For each  $C_j$ , in time  $(2^c)^{\sqrt{n}} = 2^{c\sqrt{n}}$ , we exhaustively try all subsets  $X \subseteq \bigcup_{P_i \in C_j} P_i$  that  
 527 are unions of  $S_{j,h,J}$  filtering them out when  $G[X]$  is not edgeless, and keep a heaviest of  
 528 them, say  $R_j$ . Since there can only be black edges or non-edges between some  $P_i \in C_j^h$  and  
 529  $P_{i'} \in C_j^{h'}$  with  $h \neq h'$ , it is clear that a heaviest independent set of  $G[\bigcup_{P_i \in C_j} P_i]$  is indeed a  
 530 union of  $S_{j,h,J}$  (with fixed  $j$ ). We output a heaviest set among the  $R_j$ s, which is the desired  
 531  $\lceil \frac{d+2}{3} \rceil$ -approximation. The running time is as claimed.  $\blacktriangleleft$

### 532 3.3 Time-approximation trade-offs

533 Lemma 19 and Theorem 21 run exhaustive algorithms on induced subgraphs of size  $O_d(\sqrt{n})$ .  
 534 As such, the latter inputs keep the same twin-width upper bound. To speed up the algorithm  
 535 (admittedly while worsening the approximation factor) it is tempting to recursively call our  
 536 very algorithm. We show that this leads to a time-approximation trade-off parameterized  
 537 by an integer  $q = 0, \dots, O_d(\log \log n)$ . At one end of this discrete curve, one finds the exact  
 538 exponential algorithm ( $q = 0$ ), and more interestingly the  $d + 1$ -approximation in time  
 539  $2^{O_d(\sqrt{n})}$  ( $q = 1$ ), while at the other end lies a polynomial-time algorithm with approximation  
 540 factor  $n^\varepsilon$ , where  $\varepsilon > 0$  can be made as small as desired.

541 As we will deal with the same kind of recursions for several problems, we show the  
 542 following generic abstraction.

543 **► Lemma 22.** *Let  $\hat{d}$  be a natural,  $d' = 2\hat{d} + 2$ , and  $d := c_{d'} \cdot 2^{4c_{d'} + 4}$ . Let  $\Pi$  be an optimization  
 544 graph problem where inputs come with a  $\hat{d}$ -sequence of their  $n$ -vertex graph  $G$ , or with a neatly  
 545 divided matrix  $(M, (\mathcal{R}, \mathcal{C})) \in \mathcal{M}_{n,d'}$  conform to  $G$ . Let  $\mathcal{P}$  be the partition of  $V(G)$  given  
 546 by Lemma 18. Assume that*

- 547 1.  *$\Pi$  can be exactly solved in time  $2^{O(n)}$ , and there are constants  $c_1, c_2, c_3$ , and a function*  
 548  *$f \geq 1$  such that*
- 549 2. *a  $d^{c_3} r^2$ -approximation of  $\Pi$  on  $G$  can be built in time  $n^{c_2}$  by using at most  $n^{c_1}$  calls*  
 550 *to an  $r$ -approximation of  $\Pi$  –or another optimization problem  $\Pi'$  already satisfying the*  
 551 *conclusion of the lemma– on an induced subgraph of  $G$  with at most  $f(d)\sqrt{n}$  vertices or*  
 552 *a full cleanup of an induced subtrigraph of  $G/\mathcal{P}$  (on at most  $\sqrt{n}$  vertices).*

Then  $\Pi$  can be  $d^{c_3(2^q-1)}$ -approximated in time

$$(f(d)^q n)^{(2-2^{-q})(c_1+c_2)} \cdot 2^{f(d)^{2(1-2^{-q})} n^{2^{-q}}},$$

553 for any non-negative integer  $q$ .

554 **Proof.** The proof is by induction on  $q$ . The case  $q = 0$  is implied by Item 1. The case  $q = 1$ ,  
 555 and the induction step in general, is nothing more than an abstraction of Lemma 19, where  
 556 exhaustive algorithms are replaced by recursive calls.

557 For any  $q \geq 0$ , we assume that  $\Pi$  can  $d^{c_3(2^q-1)}$ -approximated in the claimed running time,  
 558 and show the same statement for the value  $q + 1$ . Following Item 2, we run this algorithm –or

559 one for another optimization problem  $\Pi'$  satisfying the conclusion of the lemma— at most  $n^{c_1}$   
 560 times on  $f(d)\sqrt{n}$ -vertex induced subgraphs of the input graph  $G$  or on full cleanups of induced  
 561 subtrigraphs of  $G/\mathcal{P}$ . The latter graphs have at most  $\sqrt{n} \leq f(d)\sqrt{n}$  vertices. By Lemma 18,  
 562 we can compute in polynomial time a neatly divided matrix  $(M', (\mathcal{R}', \mathcal{C}')) \in \mathcal{M}_{|V(H)|, d'}$   
 563 conform to  $H$ , for each graph  $H$  of a recursive call; hence the induction applies.

Overall this takes time at most

$$\begin{aligned} & n^{c_1} + n^{c_2} \cdot \left( (f(d)^q \cdot f(d)\sqrt{n})^{(2-2^{-q})(c_1+c_2)} \cdot 2^{f(d)^{2(1-2^{-q})}(f(d)\sqrt{n})^{2^{-q}}} \right) \\ & \leq (f(d)^{q+1}n)^{c_1+c_2+\frac{1}{2}(2-2^{-q})(c_1+c_2)} \cdot 2^{f(d)^{2(1-2^{-q})+2^{-q}} n^{\frac{2^{-q}}{2}}} \\ & = (f(d)^{q+1}n)^{(2-\frac{2^{-q}}{2})(c_1+c_2)} \cdot 2^{f(d)^{2-2^{-q}+1+2^{-q}} n^{2^{-(q+1)}}} \\ & = (f(d)^{q+1}n)^{(2-2^{-(q+1)})(c_1+c_2)} \cdot 2^{f(d)^{2(1-2^{-(q+1)})} n^{2^{-(q+1)}}} \end{aligned}$$

564 For the first inequality, we assume that the two summands are larger than 2, so their  
 565 sum can be bounded by their product.

566 Besides we get an approximation of factor at most  $(d^{c_3(2^q-1)})^2 d^{c_3} = d^{c_3(2^{q+1}-1)}$ . ◀

567 In more legible terms we have proved that:

568 ▶ **Lemma 23.** *Problems  $\Pi$  satisfying the assumptions of Lemma 22 can be  $d^{O(1)(2^q-1)}$ -*  
 569 *approximated in time  $2^{O_{d,q}(\sqrt[q]{n})}$ , for any non-negative integer  $q$ .*

570 If most graph problems admit single-exponential algorithms, we will deal with such  
 571 a problem that is only known to be solvable in time  $2^{O(n \log n)}$ . Therefore we prove a variant  
 572 of Lemma 22 with a slightly worse running time.

▶ **Lemma 24.** *Let  $\Pi$  be solvable in time  $2^{O(n \log n)}$  and satisfy the second item of Lemma 22. Then  $\Pi$  can be  $d^{c_3(2^q-1)}$ -approximated in time*

$$2^{((c_1+c_2)(2-2^{-q}) \log f(d) + f(d)^{2(1-2^{-q})} n^{2^{-q}}) \log n},$$

573 for any non-negative integer  $q$ .

**Proof.** We follow the proof of Lemma 22 when the induction now gives a running time of

$$\begin{aligned} & n^{c_2} + n^{c_1} \cdot 2^{((c_1+c_2)(2-2^{-q}) \log f(d) + f(d)\sqrt{n})^{2^{-q}} \log(f(d)\sqrt{n})} \\ & \leq 2^{((c_1+c_2)(2-2^{-(q+1)}) \log f(d) + f(d)^{2(1-2^{-(q+1)})} n^{2^{-(q+1)}}) \log n} \end{aligned}$$

574

575 Again the previous lemma can be rewritten as:

576 ▶ **Lemma 25.** *Problems  $\Pi$  satisfying the assumptions of Lemma 24 can be  $d^{O(1)(2^q-1)}$ -*  
 577 *approximated in time  $2^{O_{d,q}(\sqrt[q]{n \log n})}$ , for any non-negative integer  $q$ .*

578 We derive from Lemma 24 the following notable regimes.

579 ▶ **Theorem 26.** *Problems  $\Pi$  satisfying the assumptions of Lemma 24 admit polynomial-time*  
 580  *$n^\varepsilon$ -approximation algorithms, for any  $\varepsilon > 0$ .*

581 **Proof.** This is the particular case  $q = \lceil \log \frac{\varepsilon \log n}{2c_3 \log d} \rceil$ .

Indeed the approximation factor is then at most  $d^{c_3(2^q-1)} \leq d^{2c_3 \frac{\varepsilon \log n}{2c_3 \log d}} = 2^{\varepsilon \log n} = n^\varepsilon$ , while the running time is at most

$$\begin{aligned} 2^{((c_1+c_2)(2-2^{-q}) \log f(d) + f(d)^{2(1-2^{-q})} n^{2^{-q}}) \log n} &\leq 2^{\left(2(c_1+c_2) \log f(d) + f(d)^2 n^{\frac{2c_3 \log d}{\varepsilon \log n}}\right) \log n} \\ &= n^{2(c_1+c_2) \log f(d) + f(d)^2 d^{\frac{2c_3}{\varepsilon}}}. \end{aligned}$$

582 If further  $\Pi$  can be solved exactly in time  $2^{O(n)}$  (hence satisfies the assumptions  
583 of Lemma 22), one obtains a better running time, where the exponent of  $n$  does not  
584 depend on  $\varepsilon$ . Indeed,

$$585 \quad (f(d)^q n)^{(2-2^{-q})(c_1+c_2)} 2^{f(d)^{2(1-2^{-q})} n^{2^{-q}}} \leq \left(\frac{\varepsilon \log n}{c_3 \log d}\right)^{2(c_1+c_2) \log f(d)} 2^{f(d)^2 d^{\frac{2c_3}{\varepsilon}}} n^{2(c_1+c_2)}. \blacktriangleleft$$

586 **► Theorem 27.** *Problems  $\Pi$  satisfying the assumptions of Lemma 22, resp. Lemma 24,*  
587 *admit a  $\log n$ -approximation algorithm running in time  $2^{O_d(n^{\frac{1}{\log \log n}})}$ , resp.  $2^{O_d(n^{\frac{1}{\log \log n}} \log n)}$ .*

588 **Proof.** This is the particular case  $q = \lceil \log \left(\frac{\log \log n}{c_3 \log d} + 1\right) \rceil$ .

589 This value is computed such that the approximation factor  $d^{c_3(2^q-1)}$  is at most  $\log n$ . It  
590 can be easily checked that the running times are as announced.  $\blacktriangleleft$

591 We derive the following for WEIGHTED MAX INDEPENDENT SET.

592 **► Theorem 28.** *WEIGHTED MAX INDEPENDENT SET on  $n$ -vertex graphs  $G$  (vertex-weighted  
593 by  $w$ ) given with a  $d'$ -sequence satisfies the assumptions of Lemma 22. In particular, this  
594 problem admits*

- 595  $\blacksquare$  a  $(d+1)^{2^q-1}$ -approximation in time  $2^{O_{d,q}(n^{2^{-q}})}$ , for every integer  $q \geq 0$ ,
- 596  $\blacksquare$  an  $n^\varepsilon$ -approximation in polynomial-time  $O_{d,\varepsilon}(1) \log^{O_d(1)} n \cdot n^{O(1)}$ , for any  $\varepsilon > 0$ , and
- 597  $\blacksquare$  a  $\log n$ -approximation in time  $2^{O_d(n^{\frac{1}{\log \log n}})}$ ,
- 598 with  $d := c_{2d'+2} \cdot 2^{4c_{2d'+2}+4}$ .

599 **Proof.** Even the exhaustive algorithm exactly solves WMIS in time  $2^{O(n)}$ . We thus focus on  
600 showing that WMIS satisfies the second item of Lemma 22. We set  $c_1 \geq 1$  as the required  
601 exponent to turn a  $d'$ -sequence into a neatly divided matrix of  $\mathcal{M}_{n,2d'+2}$  conform to  $G$ ,  
602  $c_2 = \frac{1}{2} + \eta$  for any fixed  $\eta > 0$ , the appropriate  $1 < c_3 \leq 2$ , and  $f(d) = d \geq 1$ .

603 The algorithm witnessing the second item is simply the proof of Lemma 19. We first  
604 check that this algorithm makes  $\lfloor \sqrt{n} \rfloor + d + 1$  recursive calls on induced subgraphs of the  
605 input  $G$ : each of the  $\lfloor \sqrt{n} \rfloor$  graphs  $G[P_i]$  where  $P_i$  has indeed size at most  $O_d(\sqrt{n})$ , and each  
606 of the  $d + 1$  graphs  $(G/\mathcal{P})[C_j]$  (indeed an induced subgraph of  $G$  by definition of the black  
607 graph of a trigraph) on at most  $\sqrt{n}$  vertices.

608 We finally assume that each recursive call outputs an  $r$ -approximation of WMIS. Let  
609  $j \in [d+1]$  be such that  $w(C_j \cap I) \geq \frac{1}{d+1} w(I)$  for  $I$  a heaviest independent set of  $G$  vertex-  
610 weighted by  $w$ . Let  $J \subseteq [\lfloor \sqrt{n} \rfloor]$  be the indices of the  $P_i$ s that are intersected by  $C_j \cap I$ , that  
611 is,  $J = \{i : P_i \cap (C_j \cap I) \neq \emptyset\}$ . For every  $i \in J$ , set  $w_i = w(P_i \cap I)$ . Each recursive call on  
612 some  $P_i$  with  $i \in J$ , yields an independent set of weight at least  $\frac{w_i}{r}$ , by assumption. Thus  
613 the weights that our algorithm puts on  $(G/\mathcal{P})[C_j]$  are such that it has an independent set of  
614 weight at least  $\sum_{i \in J} \frac{w_i}{r} = \frac{w(C_j \cap I)}{r}$ . As we run an  $r$ -approximation on this graph, we get an  
615 independent set of weight at least  $\frac{w(C_j \cap I)}{r^2} \geq \frac{w(I)}{(d+1)r^2}$ . Thus WMIS satisfies the assumptions  
616 of Lemma 22, and we conclude.  $\blacktriangleleft$

#### 4 Finding the suitable generalization: the case of Coloring

In this section, we deal with the COLORING problem. Unlike for WMIS, we cannot solely resort to recursively calling our COLORING algorithm on smaller graphs. The right problem generalization needs to be found for the inductive calls to work through, and it happens to be SET COLORING.

In the SET COLORING problem, the input is a couple  $(G, b)$  where  $G$  is a graph, and  $b$  is a function assigning a positive integer to each vertex of  $G$ . The goal is to find, for each  $v \in V(G)$ , a set  $S_v$  of at least  $b(v)$  colors such that  $S_u \cap S_v = \emptyset$  whenever  $uv \in E(G)$ , and minimizing  $|\cup_{v \in V(G)} S_v|$ . Let  $\chi_b(G)$  be the optimal value of SET COLORING for  $(G, b)$ . Observe that COLORING corresponds to the case where  $b(v) = 1$  for every  $v \in V(G)$ .

► **Theorem 29.** *SET COLORING (and hence COLORING) on  $n$ -vertex graphs  $G$  given with a  $d'$ -sequence satisfies the assumptions of Lemma 24. In particular, this problem admits*

- a  $(d+1)^{2^q-1}$ -approximation in time  $2^{O_{d,q}(n^{2^{-q}} \log n)}$ , for every integer  $q \geq 0$ , and
- an  $n^\varepsilon$ -approximation in polynomial-time for any  $\varepsilon > 0$ .

with  $d := c_{2d'+2} \cdot 2^{4c_{2d'+2}+4}$ .

**Proof.** It is known [32] that SET COLORING can be solved using the inclusion-exclusion principle in time  $O^*(\max_{v \in V(G)} b(v)^n) = 2^{O(n \log n)}$ . We now prove that it satisfies the second item of Lemma 22. We denote by  $\mathcal{A}$  the  $r$ -approximation algorithm of the statement, which we will use on instances of SET COLORING. In particular, we will call it at most  $\sqrt{n} + 1$  times, and will obtain at the end a  $(d+1)r^2$ -approximation on our input  $(G, b)$  in polynomial time.

We first apply Lemma 18 to get, in polynomial-time, a partition  $\mathcal{P} = \{P_1, \dots, P_{\lfloor \sqrt{n} \rfloor}\}$  of  $V(G)$  whose parts have size at most  $d\sqrt{n}$  and such that  $R(G/\mathcal{P})$  has maximum degree at most  $d$ . For every  $i \in [\lfloor \sqrt{n} \rfloor]$ , we use  $\mathcal{A}$  to compute an  $r$ -approximated solution  $c_{P_i}$  of  $(G[P_i], b|_{P_i})$ . We denote by  $b'$  the function which assigns, to each  $P_i$ , the number of colors of  $c_{P_i}$ . We now compute, in polynomial-time, a proper  $(d+1)$ -coloring of  $R(G/\mathcal{P})$ , which defines the sets  $C_1, \dots, C_{d+1}$ . For each  $j \in [d+1]$ , we construct another SET COLORING instance consisting of the graph  $H_j = (G/\mathcal{P})[C_j]$  (recall that this trigraph has no red edge, and can thus be seen as a graph), together with the function  $b'_{C_j}$ . Again we use  $\mathcal{A}$  to compute an  $r$ -approximated solution on  $(H_j, b'_{C_j})$ . We denote by  $c_H$  this solution. Let  $G_j$  be the subgraph of  $G$  induced by  $\cup_{P_i \in C_j} P_i$ , and  $b_j$  the restriction of  $b$  to  $V(G_j)$ . We now show how to construct a solution  $c_j$  of SET COLORING to  $(G_j, b_j)$  from  $c_H$  and all  $c_{P_i}$ . Recall that for every  $P_i \in C_j$ , every  $v \in P_i$ , we have that  $c_{P_i}(v)$  is a subset of  $\{1, \dots, b'(P_i)\}$  of size at least  $b(v)$ , and that  $c_H(P_i)$  is a subset of size at least  $b'(P_i)$ . Hence, for each  $P_i \in C_j$ , one can choose an arbitrary bijection  $\tau$  from  $\{1, \dots, b'(P_i)\}$  to  $c_H(P_i)$ , and define to each vertex  $v \in P_i$  the set  $c_j(v)$  as  $\{\tau(x) : x \in c_{P_i}(v)\}$ .

By construction, this solution is a feasible one for the instance  $(G_j, b_j)$ . Let us prove that it is an  $r^2$ -approximation of  $\chi_{b_j}(G_j)$ . First, by definition of  $c_H$ , our solution uses at most  $r \cdot \chi_{b'_{C_j}}(H_j)$  colors. Then, by definition of  $c_{P_i}$  for every  $P_i \in C_j$ , we have  $b'_{C_j}(P_i) \leq r \cdot \chi_{b|_{P_i}}(G[P_i])$ . Now, denote by  $\Gamma$  the function which assigns to each  $P_i \in C_j$  the number  $\chi_{b|_{P_i}}(G[P_i])$ . We now use the following claim, whose proof is left to the reader.

▷ **Claim 30.** Let  $(G, b)$  be an instance of SET COLORING, and  $r \in \mathbb{R}_+$ . It holds that  $\chi_{r \cdot b}(G) \leq r \cdot \chi_b(G)$ , where  $r \cdot b$  is the function which assigns  $r \cdot b(v)$  to each  $v \in V(G)$ .

This implies  $\chi_{b'_{C_j}}(H_j) \leq r \cdot \chi_\Gamma(H_j)$ , and thus our solution uses at most  $r^2 \cdot \chi_\Gamma(H_j)$  colors. We now prove the following claim.

662  $\triangleright$  **Claim 31.**  $\chi_\Gamma(H_j) \leq \chi_{b_j}(G_j)$ .

663 **Proof of the claim.** Let  $c$  be an optimal solution for  $(G_j, b_j)$ . For every distinct  $P_i, P_{i'} \in C_j$   
 664 such that  $P_i P_{i'}$  is an edge of  $H_j$ , it holds that there are all possible edges between  $P_i$  and  
 665  $P_{i'}$  in  $G_j$  (by definition of the coloring  $C_1, \dots, C_{d+1}$ ), hence it holds that  $\bigcup_{v \in P_i} c(v)$  and  
 666  $\bigcup_{v \in P_{i'}} c(v)$  have empty intersection. Moreover, by definition of  $\Gamma$ , we have that  $\bigcup_{v \in P_i} c(v)$   
 667 is of size at least  $\Gamma(P_i)$ , hence the function which assigns  $\bigcup_{v \in P_i} c(v)$  to each  $P_i$  is a feasible  
 668 solution for  $(H_j, \Gamma)$  using at most  $\chi_{b_j}(G_j)$  colors.  $\blacktriangleleft$

669 We now have in hand an  $r^2$ -approximated solution of  $(G_j, b_j)$  for every  $j \in [d+1]$ , which  
 670 can be turned into a  $(d+1)r^2$ -approximated solution of  $(G, b)$ , as desired.  $\blacktriangleleft$

## 671 **5** Edge-based problems: the case of Max Induced Matching

672 So far, we only considered problems where approximated solutions in each part  $P_i$  of a partition  
 673  $\mathcal{P}$  of  $V(G)$  of small width, and in some selected induced subgraphs of  $(V(G/\mathcal{P}), E(G/\mathcal{P}))$ ,  
 674 were enough to build an approximated solution for  $G$ .<sup>5</sup> We now handle problems for which  
 675 a number of edges is to be optimized. Now all competitive solutions can integrally lie in  
 676 between pairs of parts  $P_i, P_j$  linked by a black or a red edge in  $G/\mathcal{P}$ . This complicates  
 677 matters, and forces us to be competitive there as well, naturally splitting the algorithm into  
 678 three subroutines.

679 We present the algorithms for MAX SUBSET INDUCED MATCHING where one is given, in  
 680 addition to the input graph  $G$  (possibly with edge weights), a subset  $Y \subseteq E(G)$ , and the  
 681 goal is to find a heaviest induced matching  $S$  of  $G$  such that  $S \subseteq Y$ . Then MAX INDUCED  
 682 MATCHING is the particular case when  $Y = E(G)$ . Of course, we could solely use the edge  
 683 weights to emulate  $Y$  (by giving negative weights to all the edges in  $E(G) \setminus Y$ ). We believe  
 684 this formalism is slightly more convenient for the reader to quickly and explicitly identify  
 685 where our algorithm is seeking mutually induced edges.

686 Since the case of MAX INDUCED MATCHING is more involved than were the treatment of  
 687 MIS and COLORING, we again split the arguments into the design of a subexponential-time  
 688 constant-approximation algorithm (Lemma 34) followed by how this algorithm meets the  
 689 requirements of Lemma 22 (Theorem 33).

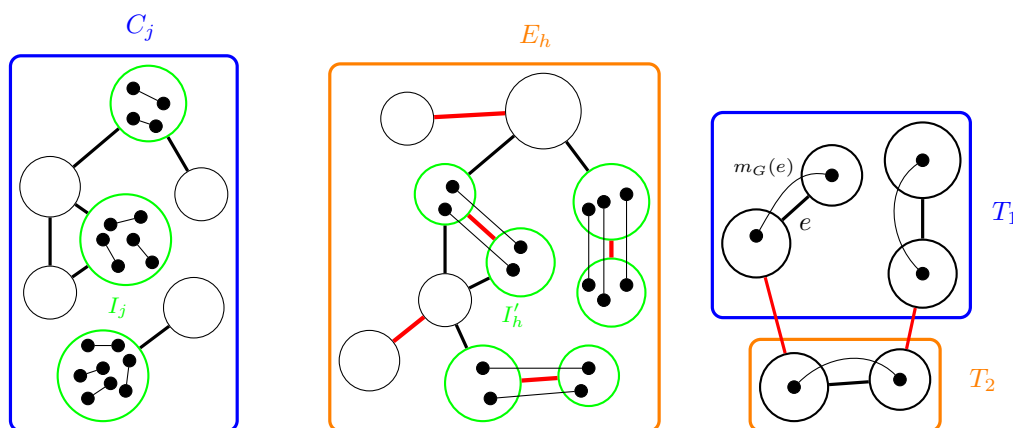
690  $\blacktriangleright$  **Lemma 32.** *Assume every input  $(G, Y)$  is given with a  $d'$ -sequence of the  $n$ -vertex, edge-*  
 691 *weighted by  $w$ , graph  $G$ . We set  $d := c_{d'} \cdot 2^{4c_{d'}+4}$ , and  $s := 2^{4c_{d'}+4}$ . MAX SUBSET INDUCED*  
 692 *MATCHING can be  $O(d^2)$ -approximated in time  $2^{O_d(\sqrt{n})}$  on these inputs.*

693 **Proof.** Again, by Lemma 17, we start by computing in polynomial time a partition of  $V(G)$ ,  
 694  $\mathcal{P} = \{P_1, \dots, P_{\lfloor \sqrt{n} \rfloor}\}$ , of parts with size at most  $s\sqrt{n}$  and such that  $\mathcal{R}(G/\mathcal{P})$  has maximum  
 695 degree at most  $d$ .

696 We  $(d+1)$ -color  $\mathcal{R}(G/\mathcal{P})$ , which defines a coarsening  $\{C_1, \dots, C_{d+1}\}$  of  $\mathcal{P}$ . We also  
 697 distance-2-edge-color  $\mathcal{R}(G/\mathcal{P})$  with  $z = 2(d-1)d + 1$  colors, that is, properly (vertex-)color  
 698 the square of its line graph. Observe that  $z - 1$  upperbounds the maximum degree of the  
 699 square of the line graph of  $\mathcal{R}(G/\mathcal{P})$ . This partitions the edges of  $\mathcal{R}(G/\mathcal{P})$  into  $\{E_1, \dots, E_z\}$ .  
 700 For each red edge  $e = P_i P_j \in R(G/\mathcal{P})$ , we denote by  $p(e)$  the set  $P_i \cup P_j$ . We also set  
 701  $X_h = p(E_h) = \bigcup_{e \in E_h} p(e)$  for each  $h \in [z]$ .

702 Let  $M \subseteq Y$  be a fixed (unknown) heaviest induced matching of  $G$  contained in  $Y$ .  
 703 Let  $M_v, M_r, M_b$  partition  $M$ , where  $M_v$  (as vertex) consists of the edges of  $M$  with both

<sup>5</sup> The improvement based on clustered coloring slightly departed from that simple scheme.



(a) Computing  $R_j$  consists first of determining the heaviest induced matching in each part  $P_i$  and then, for color  $C_j$ , to compute the maximum independent set  $I_j$  (in green) weighted by the size of the matchings.

(b) Color  $E_h$  reveals a set of red edges from trigraph  $G/\mathcal{P}$ . Set  $R'_h$  corresponds to the heaviest matching among these edges which is mutually induced regarding the black edges. The weight of the red edges  $e$  is  $w(S'_e)$ .

(c) An example of set  $S$  of size 3 with two colors  $T_1$  and  $T_2$ . The induced matching  $R'_i$  for color  $T_i$  is obtained by considering the maximum-weighted edge  $m_G(e)$  between the two parts of  $e$ .

■ **Figure 3** Illustration of how to determine the induced matching  $N_v$ ,  $N_r$ , and  $N_b$  (in that order, from left to right).

704 endpoints in a same  $P_i$ ,  $M_r$  (as red) corresponds to edges of  $M$  between some  $P_i$  and  $P_j$   
 705 with  $P_i P_j \in R(G/\mathcal{P})$ , and  $M_b$  (as black), the edges of  $M$  between some  $P_i$  and  $P_j$  with  
 706  $P_i P_j \in E(G/\mathcal{P})$ . We compute three induced matchings  $N_v, N_r, N_e \subseteq Y$  of  $G$ , capturing  
 707 a positive fraction of  $M_v, M_r, M_e$ , respectively. Figure 3 gives the intuition of the procedures  
 708 which determine each of these approximated solutions.

709 **Computing  $N_v$ .** For every integer  $1 \leq i \leq \lceil \sqrt{n} \rceil$ , we compute a heaviest induced  
 710 matching in  $G[\mathcal{P}]$  contained in  $Y$ , say  $S_i$ , in time  $2^{O_d(\sqrt{n})}$ . For each  $j \in [d+1]$ , let  $H_j$  be  
 711 the graph  $(G/\mathcal{P})[C_j]$  with every vertex  $P_i \in C_j$  weighted by  $w(S_i)$ . We compute a heaviest  
 712 independent set  $I_j$  in  $H_j$ , also in time  $2^{O_d(\sqrt{n})}$ .

713 Let  $R_j$  be the induced matching  $\{e \in S_i : P_i \in I_j\}$ . It is indeed an induced matching  
 714 in  $G$  contained in  $Y$ , since each  $S_i$  is so, there is no red edge in  $(G/\mathcal{P})[C_j]$ , and  $I_j$  is an  
 715 independent set of  $H_j$ . The solution  $N_v$  is then a heaviest among the  $R_j$ s.

716 **Computing  $N_r$ .** For each  $e = P_i P_j \in R(G/\mathcal{P})$ , we compute a heaviest induced match-  
 717 ing  $S'_e$  in  $G[p(e)] = G[P_i \cup P_j]$  among those that are included in  $Y$  and have only edges with one  
 718 endpoint in  $P_i$  and the other endpoint in  $P_j$ . This takes times at most  $\frac{\sqrt{nd}}{2} \cdot 2^{O_d(\sqrt{n})} = 2^{O_d(\sqrt{n})}$   
 719 by trying out all vertex subsets, since  $|P_i \cup P_j| \leq 2s\sqrt{n}$ . For each  $h \in [z]$ , let  $H'_h$  be the  
 720 graph  $(G/\mathcal{P})[\{P_i : P_i \text{ is incident to an edge } e \in E_h\}]$  and the red edges  $e \in E_h$  are turned  
 721 black and get weight  $w(S'_e)$ . We compute a heaviest induced matching  $I'_h$  in  $H'_h$  among those  
 722 included in  $E_h$ , in time  $2^{O_d(\sqrt{n})}$ . Note here that we changed the prescribed set of edges  $Y$  to  
 723  $E_h$ .

724 Let  $R'_h$  be the induced matching  $\{f \in S'_e : e \in I'_h\} \subseteq Y$  of  $G$ . Indeed, each  $S'_e \subseteq Y$  is an  
 725 induced matching, and there is no red edge between an endpoint of  $e \in I'_h$  and an endpoint  
 726 of  $e' \neq e \in I'_h$  (since  $E_h$  is a color class in a distance-2-edge-coloring of  $\mathcal{R}(G/\mathcal{P})$ ), nor a black  
 727 edge (by virtue of  $I'_h$  being an induced matching of  $H'_h$ ). The solution  $N_r$  is then a heaviest  
 728 among the  $R'_h$ s.



729 **Computing  $N_b$ .** Observe first that an induced matching of  $G$  can only contain at most  
 730 one edge between  $P_i$  and  $P_j$  when  $P_i P_j \in E(G/\mathcal{P})$ . Thus in the graph  $(V(G/\mathcal{P}), E(G/\mathcal{P}))$ ,  
 731 we give weight  $\max\{w(f) : f = uv \in Y, u \in P_i, v \in P_j\}$ , with the convention that  $\max \emptyset = -1$ ,  
 732 to each edge  $e = P_i P_j \in E(G/\mathcal{P})$ , call  $G'$  the resulting edge-weighted graph, and denote by  
 733  $m_G(e)$  an edge  $f \in Y$  realizing this maximum. We compute a heaviest induced matching  $S$   
 734 of  $G'$  included in  $E(G')$ , in time  $2^{O_d(\sqrt{n})}$ . Let  $H_S$  be the graph with vertex set  $S$ , and an  
 735 edge between  $e$  and  $e'$  whenever there is a red edge in  $G/\mathcal{P}$  between an endpoint of  $e$  and an  
 736 endpoint of  $e'$ . As  $H_S$  has degree at most  $2d$ , it can be  $2d+1$ -colored; let  $T_1, \dots, T_{2d+1}$  the  
 737 corresponding color classes.

738 For each  $i \in [2d+1]$ , let  $R_i''$  be the induced matching  $\{m_G(e) : e \in T_i\} \subseteq Y$  of  $G$ . Indeed,  
 739  $S$  is an induced matching in the black graph of  $G/\mathcal{P}$ , and the underlying vertices of  $T_i$  do not  
 740 induce any red edge in  $G/\mathcal{P}$ , by design. The solution  $N_r$  is then a heaviest among the  $R_i''$ s.

741 We finally output a heaviest set among  $N_v, N_r, N_b$ . The overall running time is  $2^{O_d(\sqrt{n})}$   
 742 as we make a polynomial number of calls to (exhaustive) subroutines on graphs with  $O_d(\sqrt{n})$   
 743 vertices, and color in linear time  $O(n)$ -vertex graphs of maximum degree  $\Delta$  with  $\Delta+1$   
 744 colors. We already argued that  $N_v, N_r, N_b \subseteq Y$  are all induced matchings in  $G$ , thus so is  
 745 our output.

746 We shall just show that we meet the claimed approximation factor. First, one can observe  
 747  $w(N_v) \geq \frac{w(M_v)}{d+1}$ . Second, at least a  $\frac{1}{z}$  fraction of the weight of  $M_r$  intersects some fixed  $E_i$   
 748 (with  $i \in [z]$ ). Let  $\mathcal{J}$  be the parts of  $\mathcal{P}$  intersected by  $M_r \cap X_i$ . As there cannot be a black  
 749 edge between two parts of  $\mathcal{J}$  (otherwise  $M_r$  is not an induced matching as defined), our  
 750 algorithm indeed computes an induced matching of  $G[X_i]$  included in  $Y$  of weight at least  
 751  $w(M_r \cap X_i)$ . Hence  $w(N_r) \geq \frac{w(M_r)}{z}$ .

752 Third, we already argued that an induced matching in  $G'$  corresponds to an induced  
 753 matching in the black graph of  $G/\mathcal{P}$ . Thus at least one of the  $R_i''$  (with  $i \in [2d+1]$ ) contains  
 754 at least a  $\frac{1}{2d+1}$  fraction of the weight of  $M_b$ . Therefore  $w(N_b) \geq \frac{w(M_b)}{2d+1}$ .

755 Finally the output induced matching has at least weight

$$756 \frac{w(M)}{3 \cdot \max(d+1, z, 2d+1)} = \frac{w(M)}{3z} = \frac{w(M)}{3(2(d-1)d+1)}. \quad \blacktriangleleft$$

757 **► Theorem 33.** *MAX SUBSET INDUCED MATCHING on an  $n$ -vertex graph  $G$ , edge-weighted  
 758 by  $w$ , with prescribed set  $Y \subseteq E(G)$ , and given with a  $d'$ -sequence, satisfies the assumptions  
 759 of Lemma 22. In particular, this problem admits*

- 760 ■ a  $(d+1)^{2^q-1}$ -approximation in time  $2^{O_{d,q}(n^{2^{-q}})}$ , for every integer  $q \geq 0$ ,
  - 761 ■ an  $n^\varepsilon$ -approximation in polynomial-time  $O_{d,\varepsilon}(1) \log^{O_d(1)} n \cdot n^{O(1)}$ , for any  $\varepsilon > 0$ , and
  - 762 ■ a  $\log n$ -approximation in time  $2^{O_d(n^{\frac{1}{\log \log n}})}$ ,
- 763 with  $d := c_{2d'+2} \cdot 2^{4c_{2d'+2}+4}$ .

764 **Proof.** The exhaustive algorithm (trying out all vertex subsets and checking whether they  
 765 induce a matching included in  $Y$ ) solves MAX SUBSET INDUCED MATCHING in time  $2^{O(n)}$ .  
 766 Thus we show MAX SUBSET INDUCED MATCHING satisfies the second item of Lemma 22,  
 767 as witnessed by Lemma 34 where subcalls are dealt with recursively. We set  $c_2 \geq 1$  as  
 768 the required exponent to turn a  $d'$ -sequence into a neatly divided matrix of  $\mathcal{M}_{n,2d'+2}$ , and  
 769 compute the various needed colorings, the appropriate  $\frac{1}{2} < c_1 < 1$ , and  $2 < c_3 < 3$ , and  
 770  $f(d) = 2d \geq 1$  with  $s := 2^{4c_{d'}+4}$ .

771 In computing  $N_v$ , the algorithm makes  $\lfloor \sqrt{n} \rfloor$  recursive calls and  $d+1$  calls to WEIGHTED  
 772 MAX INDEPENDENT SET on induced subgraphs of  $G$ . All of these induced subgraphs are  
 773 on less than  $f(d)\sqrt{n}$  vertices. Computing  $N_r$  makes at most  $\frac{\sqrt{nd}}{2}$  recursive calls on induced

774 subgraphs of  $G$  with at most  $f(d)\sqrt{n}$  vertices, followed by at most  $2(d-1)d+1$  recursive  
 775 calls on full cleanups of induced subtrigraphs of  $G/\mathcal{P}$  with at most  $\sqrt{n}$  vertices (in fact, one  
 776 can observe that the latter recursive calls happen to also be on induced subgraphs of  $G$ ).  
 777 Finally, computing  $N_b$  makes one recursive call to a full cleanup of  $G/\mathcal{P}$  on  $\lfloor\sqrt{n}\rfloor$  vertices.

778 In summary, we make  $O_d(\sqrt{n})$  recursive calls or calls to another problem WMIS (which  
 779 already satisfies Lemma 22 with better constants) on induced subgraphs of  $G$  or full cleanups  
 780 of (the whole)  $G/\mathcal{P}$ , each on  $O_d(\sqrt{n})$  vertices. Hence, by Lemma 18, the induction applies.

781 We check that getting  $r$ -approximations on every subcall allows to output a global  
 782  $3(2(d-1)d+1)r^2$ -approximation. For that we argue that  $N_v$  (resp.,  $N_r, N_b$ ) is a  $(2(d-1)d+$   
 783  $1)r^2$ -approximation of  $M_v$  (resp.,  $M_r, M_b$ ). The fact that  $N_v$  is a  $(d+1)r^2$ -approximation  
 784 (hence a  $(2(d-1)d+1)r^2$ -approximation, since we assume that  $d \geq 1$ ) of  $M_v$  directly  
 785 follows Theorem 28.

786 We now show that  $N_r$  is a  $(2(d-1)d+1)r^2$ -approximation of  $M_r$ . Let  $h \in [z] = [2(d-$   
 787  $1)d+1]$  be an index maximizing  $w(M_r \cap E(G[X_h]))$ . Thus  $w(M_r \cap E(G[X_h])) \geq \frac{w(M_r)}{2(d-1)d+1}$ .  
 788 Let  $F_h \subseteq E_h$  be the edges  $e = P_i P_j$  of  $\mathcal{R}(G/\mathcal{P})$  that are inhabited by  $M_r$  (i.e.,  $M_r$  contains  
 789 at least one edge between  $P_i$  and  $P_j$ ). Note that our algorithm makes an  $r$ -approximation  
 790 of the optimum such solutions on  $p(e)$  (selecting only edges between  $P_i$  and  $P_j$ ). Thus the  
 791  $r$ -approximation on  $H'_h$  yields the desired  $(2(d-1)d+1)r^2$ -approximation  $N_r$ .

792 Finally, one can easily see that  $N_b$  is a  $(2d+1)r$ -approximation of  $M_b$  (note, here, the  
 793 absence of a 2 in the exponent of  $r$ ). ◀

## 794 6 Technical generalizations

### 795 6.1 Mutually Induced $H$ -packing

796 In this section we present a far-reaching generalization of the approximation algorithms  
 797 for MAX INDEPENDENT SET and MAX INDUCED MATCHING. For any fixed graph  $H$ , let  
 798 MUTUALLY INDUCED  $H$ -PACKING be the problem where one seeks a largest collection of  
 799 mutually induced copies of  $H$  in the input graph  $G$ , that is, a largest set  $S$  such that  
 800  $G[S]$  is a disjoint union of (copies of) graphs  $H$ . We get similar approximation guarantees  
 801 for MUTUALLY INDUCED  $H$ -PACKING, for any connected graph  $H$ . Observe that MAX  
 802 INDEPENDENT SET and MAX INDUCED MATCHING are the special cases when  $H$  is a single  
 803 vertex and a single edge, respectively.

804 We in fact approximate a technical generalization that we call ANNOTATED MUTU-  
 805 ALLY INDUCED  $H$ -PACKING. The input is a tuple  $(G, w, z, \gamma, \gamma')$  where  $G$  is a graph,  
 806  $w : V(G)^{|V(H)|} \rightarrow \mathbb{Q}$  is a weight function over the tuples *without repetition* of  $V(G)$  of  
 807 size  $|V(H)|$  (that we will use to keep track of the number of mutually induced copies *within*  
 808 a given tuple of vertices of  $G$ ),  $z$  is an integer between 1 and  $|V(H)|$ ,  $\gamma : V(G) \rightarrow [z]$  is  
 809 a labeled partition of  $V(G)$  into  $z$  classes, and  $\gamma' : V(H) \rightarrow [z]$  is a labeled partition of  
 810  $V(H)$  into  $z$  classes. Note that the MUTUALLY INDUCED  $H$ -PACKING is obtained when  
 811  $w(Z) = [G[Z] \text{ is isomorphic to } H]$  (where  $[.]$  is the Iverson bracket, i.e., taking value 1 if the  
 812 property it surrounds is true, and 0 otherwise) and  $z = 1$  (which forces the value of  $\gamma$  and  
 813  $\gamma'$ ). The goal is to find a subset  $S$  such that

- 814 ■  $G[S]$  is a disjoint union of copies of  $H$ ,
- 815 ■ there is an isomorphism between each copy  $C$  of  $H$  (in  $S$ ) and  $H$  which preserves  $\gamma, \gamma'$ ,  
 816 i.e., every vertex  $v$  of  $C$  is mapped to a vertex  $v' \in V(H)$  with  $\gamma(v) = \gamma'(v')$ , and
- 817 ■  $\sum_{C \text{ copy of } H \text{ in } S} w(V(C))$  is maximized.

818 We will need the notion of *compatible trigraphs* of a (labeled) graph. Given a graph  $H$ ,

819 we call *compatible trigraph of  $H$*  any trigraph on at most  $|V(H)|$  vertices obtained by turning  
 820 some (possibly none) black edges or non-edges of trigraph  $H/\mathcal{Q}$  (for any fixed choice of a  
 821 partition  $\mathcal{Q}$  of  $V(H)$ ) into red edges. In other words, a compatible trigraph  $H'$  of  $H$  is such  
 822 that there is a cleanup  $H''$  of  $H'$  that is also a quotient trigraph of  $H$ . Note that the number  
 823 of compatible trigraphs of an  $h$ -vertex graph  $H$  is upperbounded by  $B_h \cdot 2^{\binom{h}{2}} = 2^{O(h^2)}$ , where  
 824  $B_h$  is the  $h$ -th Bell number, which counts the number of partitions of a set of size  $h$ .

825 Given a graph  $G$  vertex-partitioned by  $\mathcal{P}$  and a trigraph  $H$ , a subset  $S \subseteq V(G)$  is said *cut*  
 826 *by  $\mathcal{P}$  along  $H$*  if  $G[S]/\mathcal{P}$  is isomorphic to  $H$ . By extension, the copy of  $G[S]$  in  $G$  (induced  
 827 by  $S$ ) is also said cut by  $\mathcal{P}$  along  $H$ .

828 **► Lemma 34.** *For any connected graph  $H$ , ANNOTATED MUTUALLY INDUCED  $H$ -PACKING,*  
 829 *when every input  $(G, w, z, \gamma, \gamma')$  is given with a  $d'$ -sequence of the  $n$ -vertex graph  $G$ , satisfies*  
 830 *the assumptions of Lemma 22. In particular, this problem admits*

- 831 ■ a  $d^{O_h(2^q)}$ -approximation in time  $2^{O_{d,h,q}(n^{2-q})}$ , for every integer  $q \geq 0$ ,
- 832 ■ an  $n^\varepsilon$ -approximation in polynomial-time  $O_\varepsilon(1) \cdot n^{O_{d,h}(1)}$ , for any  $\varepsilon > 0$ ,
- 833 with  $h = |V(H)|$ , and  $d := c_{2d'+2} \cdot 2^{4c_{2d'+2}+4}$ .

834 **Proof.** As the first item of Lemma 22 is satisfied, we describe an algorithm that fulfills the  
 835 requirement of its second item. We proceed by induction on the number of vertices of  $H$ .  
 836 Thus we can assume that ANNOTATED MUTUALLY INDUCED  $J$ -PACKING, with  $J$  a connected  
 837 graph on less vertices than  $H$ , satisfies Lemma 22. We already did the base case of the  
 838 induction, which was WEIGHTED MAX INDEPENDENT SET.

839 **Algorithm.** Again, by Lemma 18, we start by computing in polynomial time a partition  
 840 of  $V(G)$ ,  $\mathcal{P} = \{P_1, \dots, P_{\lfloor \sqrt{n} \rfloor}\}$ , of parts with size at most  $d\sqrt{n}$  and such that  $\mathcal{R}(G/\mathcal{P})$  has  
 841 maximum degree at most  $d$ . Let  $S$  be a fixed (unknown) heaviest (with respect to  $w$ ) mutually  
 842 induced  $H$ -Packing of  $G$  preserving  $\gamma, \gamma'$ .

843 For every compatible trigraph  $H'$  of  $H$ , we look for mutually induced copies of  $H$  in  $G$   
 844 cut by  $\mathcal{P}$  along  $H'$ , and preserving  $\gamma, \gamma'$ . As the number of compatible trigraphs of  $H$  is  
 845  $2^{O(h^2)}$ , a  $1/2^{O(h^2)}$  fraction of the weight of  $S$  is made of mutually induced copies of  $H$  which  
 846 are cut by  $\mathcal{P}$  along a fixed compatible trigraph  $H'$ . We now focus on this particular “run.”

847 We distinguish two cases:

- 848 ■ (A)  $H'$  has at least one black edge, or
- 849 ■ (B)  $H'$  has no black edge.

850 As  $H$  is connected, the total graph of  $H'$  is also connected. Indeed, switching some edges  
 851 or non-edge to red edges in the quotient trigraph of  $H$  cannot disconnect the total graph,  
 852 which can only gain edges. Thus in case (A), every red component of  $H'$  has at least one  
 853 incident black edge, and in case (B),  $H'$  has a single red component (and no black edge).

854 In general, we want to individually pack red components of  $H'$  (first type of recursive  
 855 calls in smaller induced subgraphs of  $G$ ), then combine those red components by connecting  
 856 them with the right pattern of black edges (second type of recursive calls in the total graph  
 857 of  $G/\mathcal{P}$ ). Handling both cases (A) and (B) in an unified way runs into the technical issue  
 858 that the weight function may destroy our combined solutions in an uncontrollable manner.  
 859 The case distinction works as a win-win argument. In case (A), due to the presence of a  
 860 black edge in  $H'$ , we can pack at most one mutually induced copy of  $H$  within any fixed  
 861 subtrigraph of  $G/\mathcal{P}$  matching  $H'$ . We thus exempt ourselves from the first type of recursive  
 862 calls. In case (B), we do need the two types of recursive calls (as in WMIS), but the first  
 863 type is done on the whole  $H$ . Thus the current weight function (on  $h$ -tuples) is informative  
 864 enough.

**Case (A).** The essential element here is to build a new weight function  $w'$  on the  $h'$ -tuples of the total graph  $\mathcal{T}(G/\mathcal{P})$ , with  $h' := |V(H')|$ . For every injective map  $\iota : V(H') \rightarrow \mathcal{P}$  inducing a trigraph isomorphism and preserving  $\gamma, \gamma'$ , for every ordering of  $\iota(V(H'))$  into an  $h'$ -tuple  $(P_1, \dots, P_{h'})$ , we set

$$w'(P_1, \dots, P_{h'}) := \max\{w(v_1^1, v_1^2, \dots, v_1^{a_1}, \dots, v_{h'}^1, v_{h'}^2, \dots, v_{h'}^{a_{h'}}) : v_1^1, v_1^2, \dots, v_1^{a_1} \in P_1, \dots, v_{h'}^1, v_{h'}^2, \dots, v_{h'}^{a_{h'}} \in P_{h'}, \text{ and } G[\{v_1^1, v_1^2, \dots, v_1^{a_1}, \dots, v_{h'}^1, v_{h'}^2, \dots, v_{h'}^{a_{h'}}\}] \text{ is isomorphic to } H\}.$$

865 Indeed as we previously observed, in case (A), at most one mutually induced copy of  $H$   
866 respecting the cut along  $H'$  can be packed in the subgraph of  $G$  induced by the vertices of  
867  $\iota(V(H'))$ . (In the definition of  $w'$ , we can further impose that  $a_i$  matches the number of  
868 vertices of  $H$  in the corresponding part of  $H'$  but this is not necessary.)

869 All the  $h'$ -tuples not getting an image by  $w'$  in the previous loop (realized in time  $n^{O(h)}$ )  
870 are assigned the value 0. We then make a recursive call to ANNOTATED MUTUALLY INDUCED  
871  $\mathcal{T}(H')$ -PACKING on input  $(\mathcal{T}(G/\mathcal{P}), w', 1, \gamma_0, \gamma'_0)$  where we recall that  $\mathcal{T}(\cdot)$  is the total graph,  
872 and  $\gamma_0, \gamma'_0$  are the constant 1 functions.

873 **Case (B).** For every injective map  $\iota : V(H') \rightarrow \mathcal{P}$  inducing a trigraph isomorphism and  
874 preserving  $\gamma, \gamma'$ , we make a recursive call to ANNOTATED MUTUALLY INDUCED  $H$ -PACKING  
875 with input  $(G_\iota = G[\bigcup_{P \in \iota(V(H'))} P], w, h, \gamma_\iota, \gamma'_\iota)$  where two vertices get the same label by  $\gamma_\iota$   
876 if and only if they have the same label by  $\gamma$  and lie in the same  $P \in \iota(V(H'))$ , and  $\gamma'_\iota$  gives  
877 to a vertex  $v' \in X \in V(H')$  of  $H$  the same label given to the vertices  $v \in \iota(X)$  such that  
878  $\gamma'(v') = \gamma(v)$ . Informally  $\gamma_\iota, \gamma'_\iota$  forces the recursive call to commit to the map  $\iota$  and the  
879 former functions  $\gamma, \gamma'$ .

880 Each such recursive call yields a mutually induced packing of  $H$ . Since the red graph  
881 of  $G/\mathcal{P}$  has degree at most  $d$ , we can color the (ordered) tuples of  $\mathcal{P}$  of length up to  $h$  and  
882 inducing a connected subgraph of  $\mathcal{R}(G/\mathcal{P})$  with at most  $p(h, d) = hd^{2h} \cdot d^{2h} \cdot h! + 1$  colors  
883 such that every color class consists of disjoint tuples pairwise not linked by a red edge in  
884  $G/\mathcal{P}$ . Indeed the claimed number of colors minus 1 upperbounds, in  $\mathcal{R}(G/\mathcal{P})$ , the number  
885 of connected tuples of length up to  $h$  that can touch (i.e., intersect or be adjacent to) a  
886 fixed connected tuple of length up to  $h$ . One color class contains a fraction  $1/p(h, d)$  of the  
887 weight of the optimal solution  $S$  (subject to the same constraints). Running through all  
888 color classes  $j$  (and focusing on one containing a largest fraction of the optimum), we define  
889 a weight function  $w'$  on the  $h'$ -tuples of  $\mathcal{T}(G/\mathcal{P})$ , with  $h' = |V(H')|$ , by giving to a tuple the  
890 weight returned by the corresponding recursive call whenever it is part of color class  $j$ , and  
891 weight 0 otherwise. We then make a recursive call to ANNOTATED MUTUALLY INDUCED  
892  $\mathcal{T}(H')$ -PACKING on input  $(\mathcal{T}(G/\mathcal{P}), w', 1, \gamma_0, \gamma'_0)$  where we recall that  $\mathcal{T}(\cdot)$  is the total graph,  
893 and  $\gamma_0, \gamma'_0$  are the constant 1 functions.

894 We output a heaviest solution among all runs. We now check that the algorithm is as  
895 prescribed by Lemma 22.

896 **Number of recursive calls.** We make at most  $2^{O(h^2)} \cdot h \cdot |V(G/\mathcal{P})|^h = n^{O_n(1)}$  recursive  
897 calls to ANNOTATED MUTUALLY INDUCED  $H$ -PACKING, and at most  $p(h, d) + 1 = O_{d,h}(1)$   
898 recursive calls to ANNOTATED MUTUALLY INDUCED  $\mathcal{T}(H')$ -PACKING. Hence there is a  
899 constant  $c_1$  (function of  $d$  and  $h$ ) such that the number of calls is bounded by  $n^{c_1}$ .

900 **Nature and size of the inputs of the recursive calls.** Both  $H$  and  $\mathcal{T}(H')$  have  
901 strictly less vertices than  $H$  or are equal to  $H$ . Thus the induction on  $h$  applies. Besides,  
902  $G[\bigcup_{P \in \iota(V(H))} P]$  is an induced subgraph of  $G$  of size at most  $h \cdot d\sqrt{n} = O_{d,h}(1) \cdot \sqrt{n}$ , and  
903  $\mathcal{T}(G/\mathcal{P})$  is a full cleanup of  $G/\mathcal{P}$  of size at most  $\lfloor \sqrt{n} \rfloor$ .

904 **Running time.** Outside of the recursive calls, one can observe that our algorithm takes  
 905 times  $O_{d,h}(1) \cdot n^{O_h(1)}$ . Hence there is a constant  $c_2$  (function of  $d$  and  $h$ ) such that the  
 906 running time of that part is bounded by  $n^{c_2}$ .

907 **Correctness and approximation guarantee.** As all the recursive calls are on induced  
 908 subgraphs of  $G$  or of the total graph  $\mathcal{T}(G/\mathcal{P})$ , we return a mutually induced collection of  
 909 graphs of the size of  $H$ . All these graphs are indeed induced copies of  $H$  since the weight  
 910 function prevents the false positives of copies of  $H$  in the total graph  $\mathcal{T}(G/\mathcal{P})$  but not in  $G$   
 911 (these tuples are given weight 0). Finally it can be checked that the returned solution has  
 912 weight a fraction  $(2^{O(h^2)} \cdot \max(r, p(h, d)r^2))^{-1}$  of the optimum, which can also be seen as  
 913 a  $d^{c_3}r^2$ -approximation for some constant  $c_3$  depending on  $d$  and  $h$ . ◀

## 914 6.2 Independent induced packing of stars and forests

915 The techniques employed to design approximations algorithms for MAX SUBSET INDUCED  
 916 MATCHING can be extended in order to tackle more general problems. In particular, we  
 917 show in this section a generalization of Theorem 33 for MAX EDGE INDUCED STAR FOREST  
 918 and MAX EDGE INDUCED FOREST. These two problems stand as the version of MUTUALLY  
 919 INDUCED  $\mathcal{H}$ -PACKING where  $\mathcal{H}$  is respectively either the infinite family of stars or trees.

920 On the one hand, MAX EDGE INDUCED STAR FOREST asks, given a graph  $G$  and a subset  
 921  $Y \subseteq E(G)$ , for a collection of induced stars on  $G$ , made up of edges of  $Y$  only, maximizing  
 922 the number of edges (or leaves).

MAX EDGE INDUCED STAR FOREST

**Input:** Graph  $G$ , subset  $Y \subseteq E(G)$

923 **Output:** Collection  $(A_i)_{i \in [k]}$  of induced stars on  $G$ , made up of edges in  $Y$  only, such  
 that there is no edge between  $A_i$  and  $A_j$ , for any  $i \neq j \in [k]$ , which maximizes the  
 number of edges.

924 On the other hand, given the same input, MAX EDGE INDUCED FOREST asks for an  
 925 induced forest  $F$  on  $G$  with the largest set of edges.

926 We would like to emphasize the fact that the objective function of both problems counts  
 927 the number of *edges* in the solution, instead of *vertices*, as it is often the case in the literature  
 928 when looking for a collection of stars or trees in a graph. The reason for this is because  
 929 an approximated solution for these vertex versions can be obtained from an approximated  
 930 solution of WEIGHTED MAX INDEPENDENT SET (since any independent set is a star forest,  
 931 and any forest is a bipartite graph).

932 Observe moreover that a solution of MAX EDGE INDUCED FOREST can be 3-approximated  
 933 with a solution of MAX EDGE INDUCED STAR FOREST. Indeed, the edge set of any tree  
 934 can be partitioned into three distance-2-edge colors, which consist of a collection of stars.  
 935 Therefore, the induced forest  $F$  can be partitioned into three collections of induced stars. In  
 936 the remainder, we design approximation algorithms for MAX EDGE INDUCED STAR FOREST,  
 937 and directly deduce results for MAX EDGE INDUCED FOREST.

938 In the remainder, we propose approximation algorithms for MAX EDGE INDUCED STAR  
 939 FOREST. We provide in particular a  $n^\varepsilon$ -approximation algorithm for MAX EDGE INDUCED  
 940 STAR FOREST, running in polynomial time.

941 We need to find the suitable generalization of MAX EDGE INDUCED STAR FOREST, as it  
 942 was done for COLORING in Section 4. We call this problem MAX LEAVES INDUCED STAR  
 943 FOREST. Now, a weight function on vertices is added to the input, and we seek a collection  
 944 of mutually induced stars with maximum weight, the weight of a star being the sum of the  
 945 weights of its leaves (that is, the weight of the root is omitted).



MAX LEAVES INDUCED STAR FOREST

**Input:** Graph  $G$ , weights  $w_V : V \rightarrow \mathbb{N}$ , subset  $Y \subseteq E(G)$

**Output:** Collection  $(A_i)_{i \in [k]}$  of induced stars on  $G$  with root  $r_i$ ,  $A_i = \{r_i, s_i^1, \dots, s_i^{L_i}\}$ , made up of edges in  $Y$  only, with no edge between  $A_i$  and  $A_j$ , for any  $i \neq j \in [k]$ , maximizing

$$\sum_{i=1}^k w_V(A_i) = \sum_{i=1}^k \sum_{\ell=1}^{L_i} w(s_i^\ell)$$

We prove that MAX LEAVES INDUCED STAR FOREST follows the framework proposed in Lemma 22. We begin with the design of a subexponential-time algorithm approximating a solution of MAX LEAVES INDUCED STAR FOREST with a ratio function of twin-width.

► **Lemma 35.** *Assume every input of MAX LEAVES INDUCED STAR FOREST is given with a  $d'$ -sequence of the  $n$ -vertex  $G$ , and  $d := c_{2d'+2} \cdot 2^{4c_{2d'+2}+4}$ . MAX LEAVES INDUCED STAR FOREST can be  $O(d^2)$ -approximated in time  $2^{O_d(\sqrt{n})}$  on these inputs.*

**Proof.** We compute in polynomial time a partition of  $V(G)$ ,  $\mathcal{P} = \{P_1, \dots, P_{\lfloor \sqrt{n} \rfloor}\}$ , of parts with size at most  $d\sqrt{n}$  and such that  $\mathcal{R}(G/\mathcal{P})$  has maximum degree at most  $d$ , by Lemma 17.

As in Lemma 19, we  $(d+1)$ -color  $\mathcal{R}(G/\mathcal{P})$ , which defines a coarsening  $\{C_1, \dots, C_{d+1}\}$  of  $\mathcal{P}$ . Moreover, we distance-2-edge-color  $\mathcal{R}(G/\mathcal{P})$  with  $z = 2(d-1)d + 1$  colors. This partitions the edges of  $\mathcal{R}(G/\mathcal{P})$  into  $\{E_1, \dots, E_z\}$ . For each red edge  $e = P_i P_j \in R(G/\mathcal{P})$ , we denote by  $p(e)$  the set  $P_i \cup P_j$ .

Let  $A = \bigcup_{i=1}^k A_i$  be the union of all stars present in an optimum solution of MAX LEAVES INDUCED STAR FOREST in  $G$ . We have  $A \subseteq Y$ . Let  $A_v, A_r, A_b$  partition  $A$ , where  $A_v$  contains the edges of  $A$  with both endpoints in a same  $P_i$ ,  $A_r$  corresponds to edges of  $A$  between some  $P_i$  and  $P_j$  with  $P_i P_j \in R(G/\mathcal{P})$ , and  $A_b$ , the edges of  $A$  between some  $P_i$  and  $P_j$  with  $P_i P_j \in E(G/\mathcal{P})$ . The set of edges  $A_v$  (resp.  $A_r, A_b$ ) still form a collection of mutually induced stars. At least one over the three solutions produced by the partition  $A_v, A_r, A_b$  gives us a 3-approximation for this problem. Our algorithm consists of computing three solutions for MAX LEAVES INDUCED STAR FOREST of  $G$ , capturing a positive fraction of  $A_v, A_r, A_b$ , respectively.

**Computing a  $d+1$ -approx for  $A_v$ .** *Construction.* For every integer  $1 \leq i \leq \lfloor \sqrt{n} \rfloor$ , we compute an optimum solution for MAX LEAVES INDUCED STAR FOREST in  $G[P_i]$  contained in  $Y$ , say  $S_i$ , in time  $2^{O_d(\sqrt{n})}$ . This can be achieved with guesses of the vertices in  $P_i$ , as  $|P_i| \leq d\sqrt{n}$ .

Then, we focus on each color  $C_j$  of  $\mathcal{R}(G/\mathcal{P})$ , for  $j \in [d+1]$ . There is no red edge in  $H_j = (G/\mathcal{P})[C_j]$ . We compute a heaviest independent set  $I_j$  in  $H_j$  where the parts  $P_i$  are weighted by the edge weight of  $S_i$ . Let  $R_j$  be the union of all optimum solutions for MAX LEAVES INDUCED STAR FOREST on all  $P_i$  belonging to  $I_j$ . The solution returned is the maximum over all  $R_j$ s.

*Approximation ratio.* Let  $A_v^j$  be the subset of  $A_v$  made up of edges belonging to parts of  $C_j$ . There is no red edge between two parts of  $C_j$ , therefore their neighborhood consists of either full adjacency or full non-adjacency. As a consequence, a maximum-weighted collection of stars in  $C_j$  with edges inside parts intersects parts which are pairwise non-adjacent in  $(G/\mathcal{P})[C_j]$ , otherwise the stars are not mutually induced. Consequently, this justifies that the set  $R_j$  returned for each  $C_j$  is a maximum-weighted collection of stars in  $C_j$  made up of edges inside parts. In summary, the weight of each collection  $R_j$  is greater than the weight of  $A_v^j$ . As  $j \in [d+1]$ , a heaviest collection among all  $R_j$ s is a  $d+1$ -approximation of  $A_v$ .

**Computing a  $O(d^2)$ -approx for  $A_r$ .** *Construction.* For each  $e = P_i P_j \in R(G/\mathcal{P})$ , we



986 compute an optimal solution for MAX LEAVES INDUCED STAR FOREST in  $G[p(e)] = G[P_i \cup P_j]$   
 987 among those that are included in  $Y$  and have only edges with one endpoint in  $P_i$  and the  
 988 other endpoint in  $P_j$ . Said differently, we determine a maximum-weighted collection of  
 989 induced stars in  $G[p(e)]$  over  $Y$  with a root on one side (for example,  $P_i$ ) and all leaves  
 990 on the other side ( $P_j$ ). This costs at most  $2^{O_d(\sqrt{n})}$  by trying out all vertex subsets, since  
 991  $|P_i \cup P_j| \leq 2d\sqrt{n}$ . The set of vertices of the solution returned on  $G[p(e)]$  is denoted by  
 992  $B_e \subseteq p(e)$ .

993 For each  $h \in [z]$ , let  $H'_h$  be the trigraph  $(G/\mathcal{P})[\{P_i : P_i \text{ is incident to an edge } e \in E_h\}]$ .  
 994 The red edges of  $H'_h$  form an induced matching on the red graph of  $H'_h$  as they are at distance  
 995 2 in  $G/\mathcal{P}$ . We associate with any edge  $e \in E_h$  the edge weight of  $B_e$ . Then, we turn the  
 996 red edges of  $H'_h$  in black: let  $H''_h$  be the graph obtained. We solve MAX SUBSET INDUCED  
 997 MATCHING on  $H''_h$  by restricting it to edges of  $E_h$  (which plays the role of  $Y$ ): this is achieved  
 998 in  $2^{O(\sqrt{n})}$  as  $|V(H''_h)| \leq \sqrt{n}$ . Let  $I''_h$  be a maximum-weighted induced matching obtained.  
 999 For each  $h \in [z]$ , we obtain the union  $R_h$  of all  $B_e$ ,  $e \in I''_h$ :  $R_h = \bigcup_{e \in I''_h} B_e$ . We return an  
 1000  $R_h$  which maximizes the total edge weight, among all  $h \in [z]$ .

1001 *Approximation ratio.* Let  $A_r^h$  be the subset of  $A_r$  made up of edges being part of red  
 1002 edges  $E_h$  in  $G/\mathcal{P}$ , for  $h \in [z]$ . As the edges of  $E_h$  form an induced matching in  $\mathcal{R}(G/\mathcal{P})$ ,  
 1003 the union of solutions of MAX LEAVES INDUCED STAR FOREST over graphs  $G[p(e)]$  with  
 1004  $e \in E_h$  can only be connected through black edges of  $G/\mathcal{P}$ . Furthermore, two collections of  
 1005 stars over  $G[p(e)]$  and  $G[p(f)]$  are necessarily not mutually induced if there is a black edge  
 1006 between an endpoint of  $e$  and an endpoint of  $f$ . Consequently,  $R_h$  gives a maximum-weighted  
 1007 collection of mutually induced stars over  $E_h$  and its weight is at least the weight of  $A_r^h$ . The  
 1008 maximum-weighted collection over all  $R_h$  gives a  $z$ -approximation, as  $h \in [z]$ .

1009 **Computing a  $2d + 1$ -approx for  $A_b$ .** *Construction.* For each part  $P_i$ , we solve  
 1010 WEIGHTED MAX INDEPENDENT SET on  $G[P_i]$  with weight function  $w_V$ . Let  $I(P_i)$  be the inde-  
 1011 pendent set returned and  $w(P_i)$  its weight. We focus now on graph  $G' = (V(G/\mathcal{P}), E(G/\mathcal{P}))$ ,  
 1012 made up of the black edges of  $G/\mathcal{P}$ , and solve MAX LEAVES INDUCED STAR FOREST on it  
 1013 with weights  $w(P_i)$ . As  $|V(G')| \leq \sqrt{n}$ , this is achieved in  $2^{O(\sqrt{n})}$ .

1014 Let  $(B_h)_{h \in [k]}$  be the collection of stars returned,  $B_h = \{R_h, S_h^1, \dots, S_h^{L_h}\}$  and  $B \in E(G')$   
 1015 be the set of edges belonging to this collection. Based on the bounded maximum red degree  
 1016 of  $G/\mathcal{P}$ , we determine a  $O(d)$ -partition of the edges of  $B$ , in order to produce collections of  
 1017 mutually induced stars. Let  $H^*$  be the graph where each edge  $e$  in the collection  $(B_h)_{h \in [k]}$  is  
 1018 represented with a vertex and two of them  $e, f$  are adjacent if and only if there is a red edge  
 1019 in  $G/\mathcal{P}$  connecting an endpoint of  $e$  with an endpoint of  $f$ . This graph has degree at most  
 1020  $2d$ , so it can be  $2d + 1$ -colored: let  $T_1, \dots, T_{2d+1}$  be the corresponding color classes. Any set  
 1021 of edges  $T_j$  gives us a collection of mutually induced stars on trigraph  $G/\mathcal{P}$ , in the sense that  
 1022 there is neither a black nor a red edge between two stars.

1023 We fix some color class: say  $T_1$  w.l.o.g. Let  $(B_h^*)$  be the collection of stars produced  
 1024 by  $T_1$ , where  $B_h^* = \{R_h^*, S_h^{1,*}, \dots, S_h^{L_h^*,*}\}$ . For the root  $R_h^* = P_i$  of each star  $B_h^*$ , we select  
 1025 an arbitrary vertex  $r_h \in P_i$ . Let  $(B_h^{**})_{h \in [k]}$  be the following collection of stars (which are  
 1026 mutually induced) on  $G$ :  $B_h^{**} = \{r_h\} \cup \bigcup_{\ell=1}^{L_h^*} I(S_h^{\ell,*})$ . In brief, the collection  $(B_h^{**})_{h \in [k]}$  is  
 1027 made up of an arbitrary vertex of each root of stars  $B_h^*$  and a maximum-weighted independent  
 1028 set of each leaf of  $B_h^*$ . Remember that we computed this collection of stars for  $T_1$ : we return  
 1029 a maximum-weighted collection  $(B_h^{**})_{h \in [k]}$  among all the ones determined for  $T_j$ ,  $j \in [2d + 1]$ .

1030 *Approximation ratio.* Any collection  $B_b$  with stars belonging only to black edges of  $G/\mathcal{P}$   
 1031 reveals a collection of stars on the quotient graph. Concretely, two black edges of  $G/\mathcal{P}$   
 1032 containing each a branch of  $B_b$  must be either non-adjacent or form an induced 3-vertex path  
 1033 on  $G' = (V(G/\mathcal{P}), E(G/\mathcal{P}))$ . Conversely, considering a collection  $B^*$  of mutually induced

1034 stars of  $G'$  and, for each  $e \in B^*$ , a collection  $B_e^*$  of mutually induced stars on  $G[p(e)]$   
 1035 produces a global collection of stars of  $G$ : then, we can partition its edges into  $2d + 1$  parts  
 1036 (as with  $T_1, \dots, T_{2d+1}$ ) such that each part contains mutually induced stars. As the collection  
 1037  $B$  computed above provides us with a heaviest collection of  $G'$ , a maximum-weighted  $B_h^{**}$   
 1038 over all  $T_j$  is a  $2d + 1$ -approximation for  $B$ , whose weight is at least the weight of  $A_b$ .

1039 **Conclusion of the proof.** We finally output a heaviest collection of mutually induced  
 1040 stars among the three approximating respectively  $A_v$ ,  $A_r$ , and  $A_b$ . The overall running time is  
 1041 in  $2^{O_d(\sqrt{n})}$ . An upper bound for the approximation ratio of this algorithm is  $3z = O(d^2)$ . ◀

1042 As for the other problems treated in this article, we apply to MAX LEAVES INDUCED  
 1043 STAR FOREST the time-approximation trade-off proposed in Lemma 22.

1044 ▶ **Theorem 36.** MAX LEAVES INDUCED STAR FOREST on an  $n$ -vertex graph  $G$ , weight  
 1045 function  $w_V$ , with prescribed set  $Y \subseteq E(G)$ , and given with a  $d'$ -sequence, satisfies the  
 1046 assumptions of Lemma 22. In particular, this problem admits

- 1047 ■ a  $(d + 1)^{2^q - 1}$ -approximation in time  $2^{O_{d,q}(n^{2^{-q}})}$ , for every integer  $q \geq 0$ ,
  - 1048 ■ an  $n^\varepsilon$ -approximation in polynomial-time  $O_{d,\varepsilon}(1) \log^{O_d(1)} n \cdot n^{O(1)}$ , for any  $\varepsilon > 0$ , and
  - 1049 ■ a  $\log n$ -approximation in time  $2^{O_d(n^{\frac{1}{\log \log n}})}$ ,
- 1050 with  $d := c_{2d'+2} \cdot 2^{4c_{2d'+2}+4}$ .

1051 **Proof.** The exhaustive algorithm (trying out all vertex subsets and checking whether they  
 1052 induce a collection of mutually induced stars in  $Y$ ) solves MAX LEAVES INDUCED STAR  
 1053 FOREST in time  $2^{O(n)}$ . Thus we show MAX LEAVES INDUCED STAR FOREST satisfies the  
 1054 second item of Lemma 22. We set  $c_2 \geq 1$  as the required exponent to turn a  $d'$ -sequence  
 1055 into a neatly divided matrix of  $\mathcal{M}_{n,2d'+2}$  conform to  $G$ , and compute the various needed  
 1056 colorings, the appropriate  $\frac{1}{2} < c_1 < 1$ , and  $2 < c_3 < 3$ , and  $f(d) = 2d \geq 1$ .

1057 **Approximating  $A_v$ .** The algorithm makes  $\lfloor \sqrt{n} \rfloor$  recursive calls to solve MAX LEAVES  
 1058 INDUCED STAR FOREST on parts  $P_i$ . Furthermore,  $d + 1$  calls to WMIS are needed on  
 1059 induced subgraphs of  $G/\mathcal{P}$ . All of these induced subgraphs are on at most  $d\sqrt{n}$  vertices.

1060 **Approximating  $A_r$ .** The algorithm makes at most  $\frac{\sqrt{nd}}{2}$  recursive calls (one call per  
 1061 red edge of  $G/\mathcal{P}$ ) on induced subgraphs of  $G$  with at most  $2d\sqrt{n}$  vertices, followed by at  
 1062 most  $2(d - 1)d + 1$  calls of MAX SUBSET INDUCED MATCHING on full cleanups of induced  
 1063 subtrigraphs of  $G/\mathcal{P}$  with at most  $\sqrt{n}$  vertices.

1064 **Approximating  $A_b$ .** The algorithm makes  $\lfloor \sqrt{n} \rfloor$  calls to solve WMIS on parts  $P_i$  and  
 1065 one recursive call on a full cleanup of  $G/\mathcal{P}$  on  $\lfloor \sqrt{n} \rfloor$  vertices.

1066 In summary, we make  $O_d(\sqrt{n})$  recursive calls or calls to problems WMIS and MAX  
 1067 SUBSET INDUCED MATCHING (which already satisfy Lemma 22 with better constants) on  
 1068 induced subgraphs of  $G$  or full cleanups of (the whole)  $G/\mathcal{P}$ , each on  $O_d(\sqrt{n})$  vertices. Hence,  
 1069 by Lemma 18, the induction applies.

1070 Getting  $r$ -approximations on every subcall allows us to output a global  $3(2(d - 1)d + 1)r^2$ -  
 1071 approximation for MAX LEAVES INDUCED STAR FOREST:

- 1072 ■ collection  $A_v$  is approximated with ratio  $(d + 1)r^2$
- 1073 ■ collection  $A_r$  is approximated with ratio  $(2(d - 1)d + 1)r^2$
- 1074 ■ collection  $A_b$  is approximated with ratio  $(2d + 1)r^2$ .

1075 The extra factor 3 comes from the fact that we output the heaviest of these three solutions. ◀

1076 MAX EDGE INDUCED STAR FOREST is a particular case of MAX LEAVES INDUCED STAR  
 1077 FOREST with  $w_V(u) = 1$  for every vertex  $u \in V(G)$ . Furthermore, a solution of MAX EDGE  
 1078 INDUCED STAR FOREST is a 3-approximation of a solution of MAX EDGE INDUCED FOREST.  
 1079 These observations together with Theorem 36 allow us to state the following result.

1080 ► **Corollary 37.** *MAX EDGE INDUCED STAR FOREST and MAX EDGE INDUCED FOREST*  
 1081 *on an  $n$ -vertex graph  $G$ , with prescribed set  $Y \subseteq E(G)$ , and given with a  $d'$ -sequence, admit*  
 1082 *■ an  $n^\varepsilon$ -approximation in polynomial-time  $O_{d,\varepsilon}(1) \log^{O_d(1)} n \cdot n^{O(1)}$ , for any  $\varepsilon > 0$ , and*  
 1083 *■ a log  $n$ -approximation in time  $2^{O_d(n^{\frac{1}{\log \log n}})}$ ,*  
 1084 *with  $d := c_{2d'+2} \cdot 2^{4c_{2d'+2}+4}$ .*

## 1085 7 Limits

1086 We now discuss the limits of our framework. We give some examples of problems that are  
 1087 unlikely to have an  $n^\varepsilon$ -approximation algorithm on graphs of bounded twin-width. The  
 1088 first such problem is MIN INDEPENDENT DOMINATING SET, where one seeks a minimum-  
 1089 cardinality set which is both an independent set and a dominating set. In general  $n$ -vertex  
 1090 graphs, this problem cannot be  $n^{1-\varepsilon}$ -approximated in polynomial time unless  $P=NP$  [22],  
 1091 and cannot be  $r$ -approximated in time  $2^{o(n/r)}$  for any  $r = r(n)$ , unless the ETH fails [11].

1092 We show that MIN INDEPENDENT DOMINATING SET has the same polytime inapproximability in graphs of bounded twin-width.  
 1093

1094 ► **Theorem 38.** *For every  $\varepsilon > 0$ , MIN INDEPENDENT DOMINATING SET cannot be  $n^{1-\varepsilon}$ -*  
 1095 *approximated in polynomial time on  $n$ -vertex graphs of twin-width at most 9 given with a*  
 1096 *9-sequence, unless  $P=NP$ .*

1097 **Proof.** We perform the classic reduction of Halldórsson from SAT [22], but from PLANAR  
 1098 3-SAT where each literal has at most two occurrences, which remains NP-complete [29]. More  
 1099 precisely we add a triangle  $d_i, t_i, f_i$  for each variable  $x_i$  (with  $i \in [N]$ ), and an independent  
 1100 set  $I_j$  of size  $r$  for each 3-clause  $C_j$  (with  $j \in [M]$ ). We link  $t_i$  to all the vertices of  $I_j$   
 1101 whenever  $x_i$  appears positively in  $C_j$ , and we link  $f_i$  to all the vertices of  $I_j$  whenever  $x_i$   
 1102 appears negatively in  $C_j$ . This defines a graph  $G$  with  $n = 3N + rM$  vertices.

1103 It can be observed that if the PLANAR 3-SAT instance is satisfiable, then there is an  
 1104 independent dominating set of size  $N$ , whereas if the formula is unsatisfiable then any  
 1105 independent dominating set has size at least  $r$ . Setting  $r := N^{\frac{2-\varepsilon}{\varepsilon}}$ , the gap between  
 1106 positive and negative instances is  $\Theta_\varepsilon(1)n^{1-\varepsilon}$ , while preserving the fact that the reduction is  
 1107 polynomial.

1108 Let us now argue that  $G$  has twin-width at most 9, and that a 9-sequence of it can be  
 1109 computed in polynomial time. We can first contract each  $I_j$  into a single vertex without  
 1110 creating a red edge. Next we can contract every triangle  $d_i, t_i, f_i$  into a single vertex of red  
 1111 degree at most 4. At this point, the current trigraph is a planar graph of maximum degree at  
 1112 most 4. It was observed in [9] that planar trigraphs with maximum (total) degree at most 9  
 1113 have twin-width at most 9. This is because any planar graph has a pair of vertices on the  
 1114 same face with at most 9 neighbors (outside of themselves) combined [28]. Hence we get a  
 1115 9-sequence for  $G$  that can be computed in polynomial time. Incidentally the twin-width of  
 1116 planar graphs (that is, planar trigraphs without red edge) but no restriction on the maximum  
 1117 degree is also at most 9 [24]. ◀

1118 Another very inapproximable is LONGEST INDUCED PATH, which also does not admit  
 1119 a polytime  $n^{1-\varepsilon}$ -approximation algorithm unless  $P=NP$  [31], and cannot be  $r$ -approximated  
 1120 in time  $2^{o(n/r)}$  for any  $r = o(n)$ , unless the ETH fails [11]. The non-induced version, the  
 1121 LONGEST PATH problem, has a notoriously big gap between the best known approximation  
 1122 algorithm whose factor is  $n/\exp(\Omega(\sqrt{\log n}))$  [18], and the sharpest conditional lower bound  
 1123 which states that, for any  $\varepsilon > 0$ , a  $2^{\log^{1-\varepsilon} n}$ -approximation would imply that  $NP \subseteq QP$  [27].

1124 Despite being an open question for decades the existence or conditional impossibility of  
 1125 an approximation algorithm for LONGEST PATH with approximation factor, say,  $\sqrt{n}$  has not  
 1126 been settled. Nor do we know whether an  $n^\varepsilon$ -approximation for any  $\varepsilon > 0$  is possible. We  
 1127 now show that using our framework to obtain an  $n^\varepsilon$ -approximation for LONGEST INDUCED  
 1128 PATH of LONGEST PATH in graphs of bounded twin-width is unlikely to work, in the sense  
 1129 that it would immediately yield such an approximation factor for LONGEST PATH in general  
 1130 graphs.

1131 ► **Theorem 39.** *For any  $r = \omega(1)$ , an  $r$ -approximation for LONGEST INDUCED PATH or*  
 1132 *LONGEST PATH on graphs of twin-width at most 4 given with a 4-sequence would imply*  
 1133 *a  $(1 + o(1))r$ -approximation for LONGEST PATH in general graphs.*

1134 **Proof.** It was shown in [3] that any graph obtained by subdividing every edge of an  $n$ -vertex  
 1135 graph at least  $2 \log n$  has twin-width at most 4. Besides, a 4-sequence can then be computed  
 1136 in polynomial time.

1137 Let  $G$  be any graph with minimum degree at least 2 (note that this restriction does not  
 1138 make LONGEST PATH easier to approximate), and  $G'$  be obtained from  $G$  by subdividing  
 1139 each of its edges  $2 \lceil \log n \rceil$  times, and let  $s := 2 \lceil \log n \rceil + 1$ . Let us observe that  $G$  has a path  
 1140 of length  $\ell$  if and only if  $G'$  has a path of length  $(\ell + 2)s - 2$  if and only if  $G'$  has an induced  
 1141 path of length  $(\ell + 2)s - 4$ . Hence a polytime  $r$ -approximation for LONGEST INDUCED PATH  
 1142 or LONGEST PATH in graphs of bounded twin-width given a 4-sequence would translate into  
 1143 a  $(1 + o(1))r$ -approximation for LONGEST PATH in general graphs. ◀

1144 We can use Theorem 39 to get a similar weak obstruction to an  $n^\varepsilon$ -approximation for  
 1145 MUTUALLY INDUCED  $\mathcal{H}$ -PACKING in graphs of bounded twin-width, for some infinite family  
 1146 of connected graphs  $\mathcal{H}$ . Recall that by Lemma 34 such an approximation algorithm does  
 1147 exist when  $\mathcal{H}$  is a *finite* collection of connected graphs.

1148 Setting  $\mathcal{H}$  to be the set of all paths does not serve that purpose, since one can then  
 1149 use the approximation algorithm for MAX INDUCED MATCHING. Nevertheless this almost  
 1150 works. We just need to *decorate* the endpoints of the paths. For every positive integer  $n$ ,  
 1151 let  $D_n$  be the *decorated path* of length  $n$ , obtained from the  $n$ -vertex path  $P_n$  by adding for  
 1152 each endpoint  $u$  two adjacent vertices  $u', u''$  both adjacent to  $u$ . Informally,  $D_n$  is a path  
 1153 terminated by a triangle at each end.

1154 ► **Theorem 40.** *Let  $\mathcal{H} := \{D_n : n \in \mathbb{N}^+\}$  be the family of all decorated paths. If for*  
 1155 *every  $\varepsilon > 0$ , MUTUALLY INDUCED  $\mathcal{H}$ -PACKING admits an  $n^\varepsilon$  on  $n$ -vertex graphs of bounded*  
 1156 *twin-width given with a 4-sequence, then so does LONGEST PATH on general graphs.*

1157 **Proof.** Let  $G$  be any graph. For every pair  $u \neq v \in V(G)$ , define  $G_{uv}$  as the graph obtained  
 1158 from  $G$  by subdividing all its edges  $2 \lceil \log(n + 2) \rceil$  times, and adding two adjacent vertices  
 1159  $u', u''$  both adjacent to  $u$ , and two adjacent vertices  $v', v''$  both adjacent to  $v$ . Since there  
 1160 are only two triangles in  $G_{uv}$ , only one graph of  $\mathcal{H}$  can be present in a (mutually induced)  
 1161 packing. Thus MUTUALLY INDUCED  $\mathcal{H}$ -PACKING is now equivalent to finding a longest path  
 1162 between  $u$  and  $v$ . An  $n^\varepsilon$ -approximation algorithm for this problem would, by Theorem 39,  
 1163 give a similar approximation algorithm for LONGEST PATH in general graphs.

1164 Despite  $u', u'', v', v''$ ,  $G_{uv}$  still admits a 4-sequence. For instance, first contract  $u'$  and  $u''$ ,  
 1165 and contract  $v'$  and  $v''$ ; this does not create red edges, and has the same effect as deleting  
 1166  $u''$  and  $v''$ . The obtained graph is an induced subgraph of a  $2 \lceil \log(n + 2) \rceil$ -subdivision (of a  
 1167 graph on at most  $n + 2$  vertices). Hence it admits a polytime computable 4-sequence [3]. ◀

1168 — **References** —

- 1169 1 Noga Alon, Guoli Ding, Bogdan Oporowski, and Dirk Vertigan. Partitioning into graphs  
1170 with only small components. *J. Comb. Theory, Ser. B*, 87(2):231–243, 2003. doi:10.1016/  
1171 S0095-8956(02)00006-0.
- 1172 2 Nikhil Bansal, Parinya Chalermsook, Bundit Laekhanukit, Danupon Nanongkai, and Jesper  
1173 Nederlof. New tools and connections for exponential-time approximation. *Algorithmica*,  
1174 81(10):3993–4009, 2019. doi:10.1007/s00453-018-0512-8.
- 1175 3 Pierre Bergé, Édouard Bonnet, and Hugues Déprés. Deciding twin-width at most 4 is  
1176 NP-complete. In Mikolaj Bojanczyk, Emanuela Merelli, and David P. Woodruff, editors,  
1177 *49th International Colloquium on Automata, Languages, and Programming, ICALP 2022,*  
1178 *July 4-8, 2022, Paris, France*, volume 229 of *LIPICs*, pages 18:1–18:20. Schloss Dagstuhl -  
1179 Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPICs.ICALP.2022.18.
- 1180 4 Édouard Bonnet, Dibyayan Chakraborty, Eun Jung Kim, Noleen Köhler, Raul Lopes, and  
1181 Stéphan Thomassé. Twin-width VIII: delineation and win-wins. *CoRR*, abs/2204.00722, 2022.  
1182 arXiv:2204.00722, doi:10.48550/arXiv.2204.00722.
- 1183 5 Édouard Bonnet, Colin Geniet, Eun Jung Kim, Stéphan Thomassé, and Rémi Watrigant.  
1184 Twin-width II: small classes. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete*  
1185 *Algorithms (SODA)*, pages 1977–1996, 2021. doi:10.1137/1.9781611976465.118.
- 1186 6 Édouard Bonnet, Colin Geniet, Eun Jung Kim, Stéphan Thomassé, and Rémi Watrigant.  
1187 Twin-width II: small classes. *CoRR*, abs/2006.09877, 2020. URL: [http://arxiv.org/abs/](http://arxiv.org/abs/2006.09877)  
1188 [2006.09877](http://arxiv.org/abs/2006.09877), arXiv:2006.09877.
- 1189 In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages  
1190 1977–1996, 2021. doi:10.1137/1.9781611976465.118.
- 1191 7 Édouard Bonnet, Colin Geniet, Eun Jung Kim, Stéphan Thomassé, and Rémi Watrigant.  
1192 Twin-width III: max independent set, min dominating set, and coloring. In Nikhil Bansal,  
1193 Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata,*  
1194 *Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual*  
1195 *Conference)*, volume 198 of *LIPICs*, pages 35:1–35:20. Schloss Dagstuhl - Leibniz-Zentrum für  
1196 Informatik, 2021. doi:10.4230/LIPICs.ICALP.2021.35.
- 1197 8 Édouard Bonnet, Ugo Giocanti, Patrice Ossona de Mendez, Pierre Simon, Stéphan Thomassé,  
1198 and Szymon Torunczyk. Twin-width IV: ordered graphs and matrices. In Stefano Leonardi  
1199 and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory*  
1200 *of Computing, Rome, Italy, June 20 - 24, 2022*, pages 924–937. ACM, 2022. doi:10.1145/  
1201 3519935.3520037.
- 1202 9 Édouard Bonnet, Eun Jung Kim, Amadeus Reinald, and Stéphan Thomassé. Twin-width VI:  
1203 the lens of contraction sequences. In *Proceedings of the 2022 Annual ACM-SIAM Symposium*  
1204 *on Discrete Algorithms (SODA)*, pages 1036–1056. SIAM, 2022.
- 1205 10 Édouard Bonnet, Eun Jung Kim, Stéphan Thomassé, and Rémi Watrigant. Twin-width I:  
1206 tractable FO model checking. *J. ACM*, 69(1):3:1–3:46, 2022. doi:10.1145/3486655.
- 1207 11 Édouard Bonnet, Michael Lampis, and Vangelis Th. Paschos. Time-approximation trade-offs  
1208 for inapproximable problems. *J. Comput. Syst. Sci.*, 92:171–180, 2018. doi:10.1016/j.jcss.  
1209 2017.09.009.
- 1210 12 Nicolas Bourgeois, Bruno Escoffier, and Vangelis Th. Paschos. Approximation of max inde-  
1211 pendent set, min vertex cover and related problems by moderately exponential algorithms.  
1212 *Discret. Appl. Math.*, 159(17):1954–1970, 2011. doi:10.1016/j.dam.2011.07.009.
- 1213 13 Parinya Chalermsook, Bundit Laekhanukit, and Danupon Nanongkai. Graph products revisited:  
1214 Tight approximation hardness of induced matching, poset dimension and more. In Sanjeev  
1215 Khanna, editor, *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete*  
1216 *Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 1557–1576.  
1217 SIAM, 2013. doi:10.1137/1.9781611973105.112.
- 1218 14 Parinya Chalermsook, Bundit Laekhanukit, and Danupon Nanongkai. Independent set, induced  
1219 matching, and pricing: Connections and tight (subexponential time) approximation hardnesses.

- 1220 In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29*  
1221 *October, 2013, Berkeley, CA, USA*, pages 370–379, 2013. doi:10.1109/FOCS.2013.47.
- 1222 15 Marek Cygan, Lukasz Kowalik, Marcin Pilipczuk, and Mateusz Wykucz. Exponential-time  
1223 approximation of hard problems. *CoRR*, abs/0810.4934, 2008. URL: [http://arxiv.org/abs/](http://arxiv.org/abs/0810.4934)  
1224 [0810.4934](http://arxiv.org/abs/0810.4934), arXiv:0810.4934.
- 1225 16 Irit Dinur and David Steurer. Analytical approach to parallel repetition. In David B. Shmoys,  
1226 editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 -*  
1227 *June 03, 2014*, pages 624–633. ACM, 2014. doi:10.1145/2591796.2591884.
- 1228 17 Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Approximating  
1229 clique is almost NP-complete (preliminary version). In *32nd Annual Symposium on Foundations*  
1230 *of Computer Science, San Juan, Puerto Rico, 1-4 October 1991*, pages 2–12. IEEE Computer  
1231 Society, 1991. doi:10.1109/SFCS.1991.185341.
- 1232 18 Harold N. Gabow and Shuxin Nie. Finding a long directed cycle. *ACM Trans. Algorithms*,  
1233 4(1):7:1–7:21, 2008. doi:10.1145/1328911.1328918.
- 1234 19 Jakub Gajarský, Michal Pilipczuk, Wojciech Przybyszewski, and Szymon Torunczyk. Twin-  
1235 width and types. In Mikolaj Bojanczyk, Emanuela Merelli, and David P. Woodruff, editors,  
1236 *49th International Colloquium on Automata, Languages, and Programming, ICALP 2022,*  
1237 *July 4-8, 2022, Paris, France*, volume 229 of *LIPICs*, pages 123:1–123:21. Schloss Dagstuhl -  
1238 Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPICs.ICALP.2022.123.
- 1239 20 Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory*  
1240 *of NP-Completeness*. W. H. Freeman, 1979.
- 1241 21 Sylvain Guillemot and Dániel Marx. Finding small patterns in permutations in linear time.  
1242 In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms,*  
1243 *SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 82–101, 2014. doi:10.1137/1.  
1244 9781611973402.7.
- 1245 22 Magnús M. Halldórsson. Approximating the minimum maximal independence number. *Inf.*  
1246 *Process. Lett.*, 46(4):169–172, 1993. doi:10.1016/0020-0190(93)90022-2.
- 1247 23 Johan Håstad. Clique is hard to approximate within  $n^{1-\epsilon}$ . In *37th Annual Symposium on*  
1248 *Foundations of Computer Science, FOCS '96, Burlington, Vermont, USA, 14-16 October,*  
1249 *1996*, pages 627–636, 1996. doi:10.1109/SFCS.1996.548522.
- 1250 24 Petr Hliněný. Twin-width of planar graphs is at most 9, 2022. URL: [https://arxiv.org/](https://arxiv.org/abs/2205.05378)  
1251 [abs/2205.05378](https://arxiv.org/abs/2205.05378), doi:10.48550/ARXIV.2205.05378.
- 1252 25 Russell Impagliazzo and Ramamohan Paturi. On the Complexity of k-SAT. *J. Comput. Syst.*  
1253 *Sci.*, 62(2):367–375, 2001. doi:10.1006/jcss.2000.1727.
- 1254 26 Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. Which problems have strongly  
1255 exponential complexity? *J. Comput. Syst. Sci.*, 63(4):512–530, 2001. doi:10.1006/jcss.2001.  
1256 1774.
- 1257 27 David R. Karger, Rajeev Motwani, and G. D. S. Ramkumar. On approximating the longest  
1258 path in a graph. *Algorithmica*, 18(1):82–98, 1997. doi:10.1007/BF02523689.
- 1259 28 Anton Kotzig. Contribution to the theory of eulerian polyhedra. *Mat. Cas. SAV (Math.*  
1260 *Slovaca)*, 5:111–113, 1955.
- 1261 29 László Kozma. Minimum average distance triangulations. In Leah Epstein and Paolo Ferragina,  
1262 editors, *Algorithms - ESA 2012 - 20th Annual European Symposium, Ljubljana, Slovenia,*  
1263 *September 10-12, 2012. Proceedings*, volume 7501 of *Lecture Notes in Computer Science*, pages  
1264 695–706. Springer, 2012. doi:10.1007/978-3-642-33090-2\_60.
- 1265 30 Stefan Kratsch, Florian Nelles, and Alexandre Simon. On triangle counting parameterized  
1266 by twin-width. *CoRR*, abs/2202.06708, 2022. URL: <https://arxiv.org/abs/2202.06708>,  
1267 [arXiv:2202.06708](https://arxiv.org/abs/2202.06708).
- 1268 31 Carsten Lund and Mihalis Yannakakis. The approximation of maximum subgraph problems.  
1269 In Andrzej Lingas, Rolf G. Karlsson, and Svante Carlsson, editors, *Automata, Languages*  
1270 *and Programming, 20nd International Colloquium, ICALP93, Lund, Sweden, July 5-9, 1993*,



- 1271        *Proceedings*, volume 700 of *Lecture Notes in Computer Science*, pages 40–51. Springer, 1993.  
1272        doi:10.1007/3-540-56939-1\_60.
- 1273    **32**    Jesper Nederlof. Inclusion exclusion for hard problems, 2008.
- 1274    **33**    Michal Pilipczuk, Marek Sokolowski, and Anna Zych-Pawlewicz. Compact representation for  
1275        matrices of bounded twin-width. In Petra Berenbrink and Benjamin Monmege, editors, *39th*  
1276        *International Symposium on Theoretical Aspects of Computer Science, STACS 2022, March*  
1277        *15-18, 2022, Marseille, France (Virtual Conference)*, volume 219 of *LIPICs*, pages 52:1–52:14.  
1278        Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPICs.STACS.2022.  
1279        52.
- 1280    **34**    David Zuckerman. Linear degree extractors and the inapproximability of max clique and  
1281        chromatic number. *Theory of Computing*, 3(1):103–128, 2007.