DIMACS Series in Discrete Mathematics
and Theoretical Computer Science

Approximation Results for the Optimum Cost Chromatic Partition Problem

Klaus Jansen

Abstract. In this paper, we study the optimum cost chromatic partition (OCCP) problem for several graph classes. The OCCP problem is the problem of coloring the vertices of a graph such that adjacent vertices get different colors and that the total coloring costs are minimum.

We prove several approximation results for the OCCP problem restricted to bipartite, chordal, comparability, interval, permutation, split and unimodular graphs. We prove that there exists no polynomial approximation algorithm with ratio $O(|V|^{0.5-\epsilon})$ for the OCCP problem restricted to bipartite and interval graphs, unless $P = NP$.

Furthermore, we propose approximation algorithms with ratio $O(|V|^{0.5})$ for bipartite, interval and unimodular graphs. Finally, we prove that there exists no polynomial approximation algorithm with ratio $O(|V|^{1-\epsilon})$ for the OCCP problem restricted to split, chordal, permutation and comparability graphs, unless $P = NP$.

1. Introduction

In this paper, we study the optimum cost chromatic partition (OCCP) problem for several graph classes. The OCCP problem can be described as follows: Given a graph $G = (V, E)$ with n vertices and a sequence of coloring costs $(k_1, \ldots, k_n)$, find a feasible coloring $f(v)$ for each vertex $v \in V$ such that the total coloring costs $\sum_{v \in V} k_{f(v)}$ are minimum. A coloring $f : V \to \{1, \ldots, n\}$ is feasible if adjacent vertices have different colors. Alternatively, the OCCP problem can be formulated as follows: Given a graph $G = (V, E)$ with n vertices and a sequence of coloring costs $(k_1, \ldots, k_n)$, find a partition into independent sets $U_1, \ldots, U_s$ such that $\sum_{c=1}^{s} k_c \cdot |U_c|$ is minimum. We may assume that $k_c \leq k_d$ whenever $c < d$.

The OCCP problem introduced by Supowit [**23**] corresponds to a VLSI layout problem. A net is a group of two terminals (e.g. points on a circle or points on two opposite parallel lines) that need to be electrically connected. A routing segment is a line that connects two terminals of a net. Given a netlist and an associated cost value for each layer, the objective is to partition the net list and place them on different layers in such a way that routing segments do not cross on the same layer and that the total costs are minimum. Therefore, this layout problem is equivalent to the OCCP problem (restricted e.g. to circle graphs or to permutation graphs).

1991 *Mathematics Subject Classification.* 05 C 15.

Another application is given by Kroon et al. [**19**]. The OCCP problem for interval graphs is equivalent to the Fixed Interval Scheduling Problem (FISP) with machine dependent processing costs. In this scheduling problem each job $j \in J$ must be executed during a given time interval (s_j, f_j). We assume that a sufficient number of machines is available and that each job must be executed by one of the machines. If job j is executed by machine m, then the associated processing costs are k_m. The objective is to find a feasible schedule for all jobs with minimum total processing costs. Other variants of the FISP have been considered in the literature (e.g. by Arkin and Silverberg [**1**], Doneti and Emmons [**6**], Fischetti, Martello and Toth [**7, 8, 9**] and Kolen and Kroon [**17, 18**]).

It is not difficult to see that the OCCP problem is NP-complete for arbitrary graphs. Sen et al. [**21**] proved that the OCCP problem for circle graphs is NP-complete. Moreover, they considered an integer linear program formulation of the OCCP problem. The polytope corresponding to the constraints of this ILP contains only integral (0/1) vertices if the cartesian product $G \times K_{|V|}$ is a perfect graph. Sen et al. [**21**] proved that $G \times K_{|V|}$ is perfect if G is a tree and that there exists a perfect graph G such that $G \times K_{|V|}$ is not perfect.

Kroon et al. [**19**] studied the OCCP problem for interval graphs and trees. They showed that the problem restricted to trees can be solved in linear time and that the problem restricted to interval graphs is NP-complete even if there are only four different values for the coloring costs. If there are only two different values for the coloring costs, then the OCCP problem is equivalent to the maximum q-colorable subgraph problem. Suppose that the first q costs are equal and that the last $n-q$ costs are equal ($k_1 = \ldots = k_q < k_{q+1} = \ldots = k_n$). Then, we get an optimum solution if the maximum q-colorable subgraph is colored with the colors $1, \ldots, q$ and if the other vertices are colored with the remaining colors. The maximum q-colorable subgraph problem has been studied extensively by Frank [**10**], Gavril [**12**], Yannakakis and Gavril [**25**], Jansen et al. [**15**] and Chang et al. [**4**]. If for each graph $G \in \mathcal{G}$ a minimum coloring of G can be computed in polynomial time, then we obtain a simple approximation algorithm for the OCCP problem restricted to $\mathcal{G}$ directly with worst case ratio $O(|V|)$.

We give several approximation results for the OCCP problem restricted to bipartite, chordal, comparability, interval, permutation, unimodular and split graphs. We prove that there exists no polynomial approximation algorithm with ratio $O(|V|^{0.5-\epsilon})$ for the OCCP problem restricted to bipartite and interval graphs, unless $P = NP$. Furthermore, we propose approximation algorithms with ratio $O(|V|^{0.5})$ for both graph classes and for unimodular graphs. Finally, we prove that there exists no polynomial approximation algorithm with ratio $O(|V|^{1-\epsilon})$ for the OCCP problem restricted to split, chordal, permutation and comparability graphs, unless $P = NP$.

Definitions In the following, we give the definitions of some used graph classes. For graph classes not defined in this paper we refer to [**13**].

A graph $G = (V, E)$ is *bipartite* if its vertex set V can be partitioned into two independent sets $V = U_1 \cup U_2$ (i.e. each edge has one endpoint in U_1 and the other in U_2). A graph $G = (V, E)$ is an *interval graph* if its vertices can be represented by open intervals on the real line such that two vertices are adjacent if and only if the corresponding intervals intersect each other. Alternatively, a graph $G = (V, E)$ with $V = \{1, \ldots, n\}$ is an *interval graph* if there are positive integers

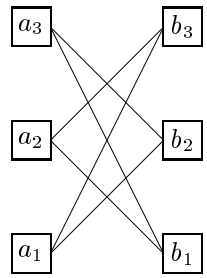

FIGURE 1. The control graph.

$0 \leq a_i < b_i \leq n$ such that $\{i, j\} \in E$ (for $i \neq j$) if and only if the intervals $(a_i, b_i]$ and $(a_j, b_j]$ intersect.

A graph G is a *split graph* if there exists a partition $V = U \cup C$ of its vertex set V into an independent set U and a clique C. Let $\Pi = (i_1, \ldots, i_n)$ be a permutation of $\{1, \ldots, n\}$. We denote by $\Pi^{-1}(i)$ the position of i in Π. A graph $G = (V, E)$ with $V = \{1, \ldots, n\}$ is a *permutation graph* if there is a permutation Π such that $\{i, j\} \in E$ if and only if $i > j$ and $\Pi^{-1}(i) < \Pi^{-1}(j)$. A graph G is *unimodular* if its clique matrix (a row for each maximal clique) is totally unimodular. A matrix A is totally unimodular if each square submatrix of A has determinant equal to 0, $+1$ or -1. Unimodular graphs include bipartite graphs and their line graphs, interval graphs and the class of graphs which do not contain an odd circuit of length at least five, the complement of such a circuit, or a $K_4 - e$ as an induced subgraph.

2. Bipartite graphs

In this section we prove that OCCP is hard to approximate $O(|V|^{0.5-\epsilon})$ for bipartite graphs. After that, we propose an approximation algorithm with ratio $O(|V|^{0.5})$.

2.1. Non-Approximability result. We use the precoloring extension problem that is NP-complete for bipartite graphs proved by Bodlaender, Jansen and Woeginger [**3**].

1-PrExt for bipartite graphs:

Given: A bipartite graph $G = (V, E)$ with vertex set $V = A \cup B$ and edge set $E \subset \{\{v, w\} | v \in A, w \in B\}$, and three specified vertices $a_1, a_2, a_3 \in A$,

Question: Does there exist a 3-coloring of G with $f(a_1) = 1$, $f(a_2) = 2$ and $f(a_3) = 3$?

First, we show the NP-completeness of the OCCP problem using an integer parameter K. Later, we specify the parameter K to achieve our non - approximability result.

THEOREM 2.1. *The OCCP problem for bipartite graphs is NP-complete if there are at least four different cost values.*

PROOF. The theorem is proved by a reduction from 1-PrExt restricted to bipartite graphs. We may assume that $G = (A \cup B, E)$ contains three further vertices $b_1, b_2, b_3 \in B$ with $\{a_i, b_j\} \in E$ for $1 \leq i \neq j \leq 3$ (see Figure 1). Let n be the number of vertices in G. Clearly, the number n is greater than 6.

vertices $v' \in V'$	$v_{1,j}, v_{2,j}, v_6$	$v_{3,j'}, v_{4,j'}$	v_5
coloring f	1	2	3

TABLE 1. A good coloring for G'

Let I be an instance of 1-PrExt containing the bipartite graph $G = (A \cup B, E)$ with $a_1, a_2, a_3 \in A$ and $b_1, b_2, b_3 \in B$ as described above. Let K be a positive integer with $K \geq 1$. An instance I' of the OCCP problem is constructed as follows. First, we define a bipartite graph $G' = (V', E')$ with vertex set

$$\begin{aligned} V' &= \{v_{1,j}, v_{2,j} | 1 \leq j \leq 2000K^2n\} \cup \\ &\quad \{v_{3,j'}, v_{4,j'} | 1 \leq j' \leq 100Kn\} \cup \{v_5, v_6\} \end{aligned}$$

and edge set

$$\begin{aligned} E' &= \{\{v_{1,j}, v_{3,j'}\}, \{v_{2,j}, v_{4,j'}\} | 1 \leq j \leq 2000K^2n, 1 \leq j' \leq 100Kn\} \cup \\ &\quad \{\{v_5, v_{3,j'}\}, \{v_6, v_{4,j'}\} | 1 \leq j' \leq 100Kn\} \cup \\ &\quad \{v_5, v_6\}. \end{aligned}$$

The bipartite graph G' illustrated in Figure 2 contains $4000K^2n + 200Kn + 2$ vertices. Then, we connect G and G' using the following edges:

$$\begin{aligned} \bar{E} &= \{\{a_1, v_{3,j'}\}, \{b_1, v_{4,j'}\}, \{b_1, v_5\}, \{a_2, v_{2,j}\}, \\ &\quad \{\{b_2, v_{1,j}\}, \{b_2, v_5\}, \{a_3, v_{3,j'}\}, \{a_3, v_{2,j}\}, \\ &\quad \{b_3, v_{1,j}\}, \{b_3, v_{4,j'}\} \mid 1 \leq j \leq 2000K^2n, 1 \leq j' \leq 100Kn\}\}. \end{aligned}$$

In total, the bipartite graph $\bar{G}$ for I' is given by

$$\bar{G} = (A \cup B \cup V', E \cup E' \cup \bar{E}).$$

The cost values are $k_1 = 1$, $k_2 = 10K$, $k_3 = 100K^2$ and $k_4 = 15000K^3n$. A cheap coloring of $\bar{G}$ has to use only three colors; otherwise the costs would be more than $15000K^3n$.

If we use the 3-coloring f for G' as given in Table 1, the total costs are $(4000K^2n + 1) \cdot 1 + 200Kn \cdot 10K + 1 \cdot 100K^2 = 6000K^2n + 100K^2 + 1 \leq 6100K^2n$ (for $n > 6$).

We prove the following statement: I is a yes instance of 1-PrExt if and only if the minimum total costs of coloring all vertices in I' don't exceed $6100K^2n + 100K^2 + 1$.

Suppose, that f is a 3-coloring for G with $f(a_1) = f(b_1) = 1$, $f(a_2) = f(b_2) = 2$ and $f(a_3) = f(b_3) = 3$. Then, using the 3-coloring for G' as described above (in Table 1), the total costs for G' are equal to $6000K^2n + 100K^2 + 1$. This 3-coloring for G' is compatible with the solution of 1-PrExt for G (consider the connecting edges in $\bar{E}$). Furthermore, the costs of the 3-coloring for G can be bounded simply by $100K^2n$. Therefore, the total costs are at most $6100K^2n + 100K^2 + 1$.

Conversely, suppose that the minimum costs of coloring all vertices in I' don't exceed $6100K^2n + 100K^2 + 1$. Let f be such a coloring. First, we prove that $f(a_1) = f(b_1) = 1$, $f(a_2) = f(b_2) = 2$ and $f(a_3) = f(b_3) = 3$. The key idea is to show that otherwise the minimum costs exceed $6100K^2n + 100K^2 + 1$.

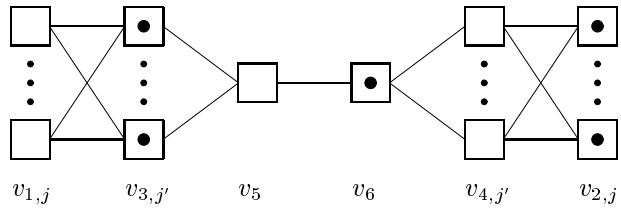

FIGURE 2. The constructed graph G' and a feasible 2-coloring of G'.

Case 1: $f(a_3) = 1$. This implies that $f(v_{3,j'}), f(v_{2,j}) \geq 2$. The cheapest coloring (see Figure 2) of the vertices in G' has costs at least

$$\begin{aligned} & (2000K^2n + 100Kn + 1) \cdot 1 + (2000K^2n + 100Kn + 1) \cdot 10K \\ = \; & 20000K^3n + 3000K^2n + 100Kn + 10K + 1. \end{aligned}$$

In this case, we obtain coloring costs of at least $20000K^3n > 6100K^2n + 100K^2 + 1$.

Case 2: $f(a_3) = 2$. This implies that $f(v_{3,j'}), f(v_{2,j}) \in \{1, 3\}$.

Case 2.1: $f(v_{3,j'}) = 3$ for each $1 \leq j' \leq 100Kn$. The cheapest coloring of the vertices in G' has costs of at least

$$(4000K^2n+1)\cdot 1+(100Kn+1)\cdot 10K+100Kn\cdot 100K^2=10000K^3n+5000K^2n+10K+1.$$

In this case, the coloring costs are at least $10000K^3n > 6100K^2n + 100K^2 + 1$.

Case 2.2: $f(v_{3,j'}) = 1$ for at least one $j' \in \{1, \ldots, 100Kn\}$. This case gives (as in case 1) coloring costs of at least $20000K^3n$.

Using the symmetry of the graph G', $f(b_3) \in \{1, 2\}$ implies also coloring costs of at least $10000K^3n$. Therefore, we must have $f(a_3) = f(b_3) = 3$.

Case 3: $f(a_2) = 1$. This implies that $f(v_{2,j}) \geq 2$ for each $1 \leq j \leq 2000K^2n$. Again, as in case 1 we obtain coloring costs of at least $20000K^3n$.

Case 4: $f(a_2) = 3$. This case is not possible, since $f(b_3) = 3$ and $\{a_2, b_3\} \in E$.

Case 5: $f(b_2) = 1$. This implies that $f(v_{1,j}) \geq 2$ for each $1 \leq j \leq 2000K^2n$ and that $f(v_5) \geq 2$. In this case, the cheapest coloring of the vertices in G' has costs of at least $20000K^3n+3000K^2n+100Kn+10K+1$. Again, this gives coloring costs of at least $20000K^3n > 6100K^2n + 100K^2 + 1$.

Case 6: $f(b_2) = 3$. This case is not possible, since $f(a_3) = 3$ and $\{a_3, b_2\} \in E$.

This case analysis shows that a_2 and b_2 are colored with color 2. Using the control graph (see Figure 1), we obtain that $f(a_1) = f(b_1) = 1$. Using this 3-coloring of the control graph and the edges connecting the graph G with G', the 3-coloring f satisfies the following properties:

(1) $f(v_{3,j'}) \neq 1$, $f(v_{4,j'}) \neq 1$ and $f(v_5) \neq 1$.
(2) $f(v_{2,j}) \neq 2$, $f(v_{1,j}) \neq 2$ and $f(v_5) \neq 2$.
(3) $f(v_{3,j'}) \neq 3$, $f(v_{2,j}) \neq 3$, $f(v_{1,j}) \neq 3$ and $f(v_{4,j'}) \neq 3$.

These properties imply:

(1) $f(v_{1,j}) = f(v_{2,j}) = 1$ for each $1 \leq j \leq 2000K^2n$,
(2) $f(v_{3,j'}) = f(v_{4,j'}) = 2$ for each $1 \leq j \leq 100Kn$ and
(3) $f(v_5) = 3$ and $f(v_6) = 1$.

The costs of the coloring f for G' are equal to $6000K^2n + 100K^2 + 1$. For the graph G, coloring costs of at most $100K^2n$ remain. Since we have only three colors in G (note that $k_4 = 15000K^3n$) and $f(a_1) = f(b_1) = 1$, $f(a_2) = f(b_2) = 2$ and

$f(a_3) = f(b_3) = 3$, the coloring f restricted to G is a solution of the 1-PrExt problem. □

If there is no solution of the 1-PrExt problem, then we have either four colors in $\bar{G}$ with coloring costs of at least $k_4 = 15000K^3n$ or a 3-coloring with coloring costs of at least $10000K^3n$ (see the cases $1-6$).

THEOREM 2.2. *For each $\epsilon < \frac{1}{2}$, there exists no polynomial approximation algorithm with ratio $O(|V|^{0.5-\epsilon})$ for the OCCP problem restricted to bipartite graphs, unless $P = NP$.*

PROOF. Let H be an approximation algorithm for the OCCP problem that computes a coloring with costs $H(I') \leq c|V|^{0.5-\epsilon}OPT(I')$, where c is a constant and $OPT(I')$ are the minimum costs of a solution of instance I'.

We construct for an instance I of the 1-PrExt problem an instance I' of the OCCP problem as described in the proof above. We obtain a graph with at most $4300K^2n$ vertices. If there exists a solution of the 1-PrExt problem, the optimum solution of the OCCP instance I' has costs of at most $6200K^2n$. In this case, our approximation algorithm produces the value $H(I') \leq 6200K^2cn|V|^{0.5-\epsilon}$. Since the number of vertices in I' is at most $4300K^2n$, we have

$$c|V|^{0.5-\epsilon} \leq (4300)^{0.5}K^{1-2\epsilon}cn^{0.5}.$$

If there exists no solution of the 1-PrExt instance, then $OPT(I') > 10000K^3n$ and, therefore, algorithm H generates a solution with costs greater than $10000K^3n$. Next, we consider the inequality

$$10000K^3n > (4300)^{0.5}6200K^{3-2\epsilon}cn^{1.5}.$$

This inequality is satisfied if and only if

$$K^{2\epsilon} > \frac{c(4300)^{0.5}6200}{10000}n^{0.5}.$$

We define

$$K = \left\lceil \sqrt[2\epsilon]{\frac{c(4300)^{0.5}6200}{10000}n^{0.5}]} \right\rceil + 1.$$

Since c and ϵ are constant, K is a polynomial in n and, therefore, the instance I' can be constructed in polynomial time. If there exists no solution of the 1-PrExt problem, then H generates a solution with costs of at least

$$10000K^3n > (4300)^{0.5}6200K^{3-2\epsilon}cn^{1.5} \geq 6200K^2cn|V|^{0.5-\epsilon}.$$

Therefore, by using the polynomial time approximation algorithm H, we could decide the existence of a solution for the 1-PrExt problem, which would imply $P = NP$. □

2.2. Approximability result. Next, we consider an approximation algorithm A with ratio $O(|V|^{0.5})$ for the OCCP problem restricted to bipartite graphs $G = (V, E)$. We can compute a 2-coloring of each connected component Z of V such that at least $\frac{|Z|}{2}$ vertices are colored with the first color and that the other vertices are colored with the second color. The key idea of the approximation algorithm is to compute two colorings for the problem and to choose the cheaper one.

Algorithm A:

Given: Instance I of the OCCP problem containing a bipartite graph $G = (V, E)$ and cost vector $(k_1, \ldots, k_{|V|})$.

(1): Compute a 2-coloring of G with n_1 vertices colored with color 1 and $|V|-n_1$ vertices colored with color 2 such that n_1 is maximum and, therefore, $n_1 \geq \frac{|V|}{2}$. The costs of the first coloring are $A_1(I) = n_1k_1 + (|V| - n_1)k_2$.

(2): Compute a maximum independent set U in G with $\alpha(G)$ vertices and color the vertices in U with color 1. Then, compute a 2-coloring of $G[V \setminus U]$ with n_1' vertices colored with color 2 and $|V| - \alpha(G) - n_1'$ vertices colored with color 3 such that $n_1' \geq \frac{|V|-\alpha(G)}{2}$. The costs of the second coloring are $A_2(I) = \alpha(G)k_1 + n_1'k_2 + (|V| - \alpha(G) - n_1')k_3$.

(3): Choose the cheaper coloring among the two colorings.

We note that the costs of the second coloring are bounded by

$$\alpha(G)k_1 + \frac{|V| - \alpha(G)}{2}(k_2 + k_3) \leq \alpha(G)k_1 + (|V| - \alpha(G))k_3.$$

THEOREM 2.3. *Algorithm A computes a solution of the OCCP problem restricted to bipartite graphs with approximation ratio* $\leq |V|^{0.5}$.

PROOF. Let I be an instance of the OCCP problem containing a bipartite graph $G = (V, E)$ and cost vector $(k_1, \ldots, k_{|V|})$. Then, we have two lower bounds for the optimum value $OPT(I)$:

$$\begin{array}{ll} (1) & OPT(I) \geq |V|k_1, \\ (2) & OPT(I) \geq \alpha(G)k_1 + (|V| - \alpha(G))k_2. \end{array}$$

We consider two cases to prove that $A(I) \leq O(|V|^{0.5})OPT(I)$.

Case A: $k_3 \leq |V|^{0.5}k_2$. In this case, we consider the second coloring. The costs of this coloring $A_2(I)$ are bounded by

$$\begin{array}{l} \alpha(G)k_1 + \frac{|V|-\alpha(G)}{2}(k_2 + k_3) \leq \\ \alpha(G)k_1 + \frac{(|V|-\alpha(G))(1+|V|^{0.5})}{2}k_2 \leq \\ \frac{1+|V|^{0.5}}{2}(\alpha(G)k_1 + (|V| - \alpha(G))k_2) \leq \\ \frac{1+|V|^{0.5}}{2}OPT(I). \end{array}$$

Case B: $k_3 > |V|^{0.5}k_2$. In this case, we consider the first coloring computed by A.

Case B.1: The optimum solution needs only two colors. Then, the first computed coloring in A is optimal.

Case B.2: The optimum solution needs more than two colors. In this case, the optimum value $OPT(I)$ is greater than $k_3 > |V|^{0.5}k_2$. Then, the costs $A_1(I)$ are bounded by

$$\begin{array}{l} n_1k_1 + (|V| - n_1)k_2 \leq \frac{|V|}{2}k_1 + \frac{|V|}{2}k_2 \leq \\ \frac{OPT(I)}{2} + \frac{|V|^{0.5}|V|^{0.5}k_2}{2} \leq \frac{OPT(I)}{2} + \frac{|V|^{0.5}}{2}OPT(I) = \\ \frac{1+|V|^{0.5}}{2}OPT(I). \end{array}$$

In both cases, $A(I) \leq |V|^{0.5}OPT(I)$ for each instance I. □

3. Interval graphs

In this section we prove that the OCCP problem restricted to interval graphs is hard to approximate with ratio $O(|V|^{0.5-\epsilon})$. Furthermore, we propose an approximation algorithm with ratio $O(|V|^{0.5})$ for interval graphs and also for unimodular graphs.

interval	numbers
$(1,2]$	t times
$(2,A_i]$	$1 \le i \le t$
$(A_i, X_{ij}]$	$1 \le i,j \le t$
$(X_{ij}, 10+a_i+b_j]$	$1 \le i,j \le t$
$(0,B_j]$	$1 \le j \le t$
$(B_j, X_{ij}]$	$1 \le i,j \le t$
$(11-c_k, 13]$	$1 \le k \le t$
$(X_{ij}, 14]$	$1 \le i,j \le t$
$(\frac{100Kt^2}{2000K^2t^4}, 3]$	t^2-t times
$(12, 14-\frac{100Kt^2}{2000K^2t^4}]$	t^2-t times
$(3,B_j]$	$t-1$ times, $1 \le j \le t$
$(0,A_i]$	$t-1$ times, $1 \le i \le t$
$(\frac{\ell}{2000K^2t^4}, \frac{\ell+1}{2000K^2t^4}]$	t times, $0 \le \ell < 2000K^2t^4$
$(13+\frac{\ell}{2000K^2t^4}, 13+\frac{\ell+1}{2000K^2t^4}]$	t times, $0 \le \ell < 2000K^2t^4$
$(\frac{\ell}{2000K^2t^4}, \frac{\ell+1}{2000K^2t^4}]$	t^2-t times, $0 \le \ell < 100Kt^2$
$(14-\frac{\ell+1}{2000K^2t^4}, 14-\frac{\ell}{2000K^2t^4}]$	t^2-t times, $0 \le \ell < 100Kt^2$

TABLE 2. The intervals in the interval graph

3.1. Non-Approximability result. The NP-completeness proof uses a reduction from Numerical Three Dimensional Matching (N3DM) and is a modification of the pure NP-completeness proof of the OCCP problem given by Kroon et al. [**19**]. The N3DM problem that is NP-complete in the strong sense (see Garey and Johnson [**11**]) is defined as follows:

Numerical 3-Dimensional Matching:
Given: A positive integer t, and $3t$ rational numbers a_i, b_i and c_i such that $\sum_{i=1}^{t}(a_i+b_i+c_i) = t$ and $0 < a_i, b_i, c_i < 1$ for $1 \le i \le t$.
Question: Do there exist permutations ρ and δ of $\{1,\dots,t\}$ such that $a_i + b_{\rho(i)} + c_{\delta(i)} = 1$ for $1 \le i \le t$?

THEOREM 3.1. *For each $\epsilon < \frac{1}{2}$, there exists no polynomial approximation algorithm with ratio $O(|V|^{0.5-\epsilon})$ for the OCCP problem restricted to interval graphs, unless $P = NP$.*

PROOF. First, we give a reduction from N3DM with variable parameter $K \in \mathbb{N}$ and, later, we specify the parameter $K \in \mathbb{N}$ to achieve our non-approximability result. Let I_1 be an instance of N3DM with integer t and rational numbers a_i, b_i and c_i for $1 \le i \le t$. We choose further rational numbers A_i, B_j and X_{ij} such that all these numbers are different and that $4 < A_i < 5 < B_j < 6$ and $7 < X_{ij} < 9$ for $1 \le i,j \le t$. Next, we construct an instance I_2 of the OCCP problem. We use the intervals given in Table 2 for the interval graph.

Furthermore, there are t colors with costs 1, t^2-t colors with costs $10Kt^2$, t^2 colors with costs $100K^2t^4$ and all other colors have costs $20000K^3t^6$.

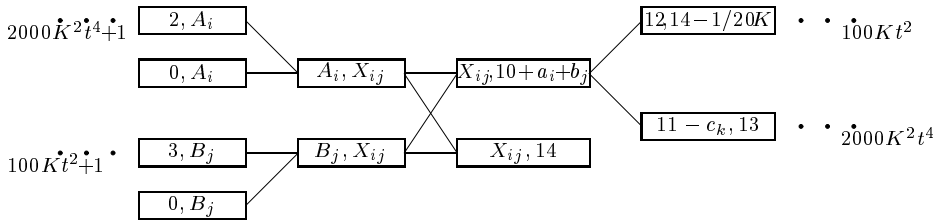

FIGURE 3. The construction of the intervals for the NP-completeness proof.

The **first claim** is to prove the following statement: I_1 is a yes instance of N3DM if and only if the minimum total costs of coloring all intervals of I_2 do not exceed

$$costs(K) := t \cdot (4000K^2t^4 + 5) \cdot 1 + (t^2 - t) \cdot (200Kt^2 + 5) \cdot 10Kt^2 + t^2 \cdot 3 \cdot 100K^2t^4 = 2000K^2t^5 + 5t + 2300K^2t^6 + 50Kt^4 - 50Kt^3.$$

If I_1 is a no instance of N3DM, then the total costs of coloring all intervals of I_2 are greater than $10000K^3t^6$. We notice that the value $costs(K)$ is bounded by $4355K^2t^6$.

Suppose that the minimum total costs of all intervals do not exceed $costs(K)$. Then, only the first $2t^2$ cheapest colors can be used; otherwise the total costs are at least $20000K^3t^6$. For the following observations, it is helpful to analyse the interval structure (see also Figure 3) in the NP-completeness proof. Since all intervals $(A_i, X_{ij}]$ and $(B_j, X_{ij}]$ are overlapping and since the number of these intervals is $2t^2$, we have to use $2t^2$ different colors for these intervals. The number of intervals at each time step $i \in (0, 10] \cup (12, 14]$ is equal to $2t^2$. Therefore, a set of intervals colored with one color must form a consecutive chain of intervals without gaps during the intervals $(0, 10]$ and $(12, 14]$.

This implies that one interval $(A_i, X_{ij}]$ must be colored with the same color as the sequence

$$(0, \frac{1}{2000K^2t^4}] \ldots (1 - \frac{1}{2000K^2t^4}, 1]\,(1, 2]\,(2, A_i]$$

or the sequence $(0, A_i]$. Furthermore, the interval $(B_j, X_{ij}]$ must be colored with the same color as the sequence

$$(0, \frac{1}{2000K^2t^4}] \ldots (\frac{100Kt^2}{2000K^2t^4}, 3]\,(3, B_j]$$

or the sequence $(0, B_j]$. Moreover, an interval $(A_i, X_{ij}]$ or an interval $(B_j, X_{ij}]$ must be colored with the same color as one of the following sequences:

(1) $(X_{ij}, 10 + a_i + b_j]\,(11 - c_k, 13]\,(13, 13 + \frac{1}{2000K^2t^4}] \ldots (14 - \frac{1}{2000K^2t^4}, 14]$,
(2) $(X_{ij}, 10 + a_i + b_j]\,(12, 14 - \frac{100Kt^2}{2000K^2t^4}] \ldots (14 - \frac{1}{2000K^2t^4}, 14]$,
(3) $(X_{ij}, 14]$.

From these observations, we conclude that the cheapest coloring is as follows:

(1) each color with costs 1 colors $4000K^2t^4 + 5$ intervals,
(2) each color with costs $10Kt^2$ colors $200Kt^2 + 5$ intervals and
(3) each color with costs $100K^2t^4$ colors 3 intervals.

Then, the total costs are equal to

$$t \cdot (4000K^2t^4 + 5) \cdot 1 + (t^2 - t) \cdot (200Kt^2 + 5) \cdot 10Kt^2 + t^2 \cdot 3 \cdot 100K^2t^4 = costs(K).$$

If the total costs do not exceed $costs(K)$, then we obtain the following sequences of intervals: Each color with costs 1 colors a sequence:

$$\begin{array}{l}(0, \frac{1}{2000K^2t^4}]\,(\frac{1}{2000K^2t^4}, \frac{2}{2000K^2t^4}]\,\ldots\,(1 - \frac{1}{2000K^2t^4}, 1]\,(1, 2] \\ (2, A_i]\,(A_i, X_{ij}]\,(X_{ij}, 10 + a_i + b_j](11 - c_k, 13]\,(13, 13 + \frac{1}{2000K^2t^4}] \\ (13 + \frac{1}{2000K^2t^4}, 13 + \frac{2}{2000K^2t^4}]\,\ldots\,(14 - \frac{1}{2000K^2t^4}, 14]\end{array}$$

of length $4000K^2t^4 + 5$, where each i and each k occur exactly once. Each color with costs $10Kt^2$ colors a sequence:

$$\begin{array}{l}(0, \frac{1}{2000K^2t^4}]\,(\frac{1}{2000K^2t^4}, \frac{2}{2000K^2t^4}]\,\ldots\,(\frac{100Kt^2-1}{2000K^2t^4}, \frac{100Kt^2}{2000K^2t^4}]\,(\frac{100Kt^2}{2000K^2t^4}, 3]\,(3, B_j] \\ (B_j, X_{ij}]\,(X_{ij}, 10 + a_i + b_j]\,(12, 14 - \frac{100Kt^2}{2000K^2t^4}]\,\ldots\,(14 - \frac{1}{2000K^2t^4}, 14]\end{array}$$

of length $200Kt^2+5$, where each j occurs exactly $t-1$ times. This implies that each j occurs exacly once among the intervals $(X_{ij}, 10+a_i+b_j]$ colored with colors with costs 1. The remaining intervals can be colored with t^2 colors using $t^2 - t$ sequences $(0, A_i]\,(A_i, X_{ij}]\,(X_{ij}, 14]$ and t sequences $(0, B_j]\,(B_j, X_{ij}]\,(X_{ij}, 14]$ of length three.

Next, we consider other assignments f of colors to intervals and calculate the **coloring costs**.

Case A: $f((X_{ij}, 14]) \in \{1, \ldots, t^2\}$. In this case, at least one chain of length $2000K^2t^4 + 2$ or one chain of length $100Kt^2 + 2$ is colored with one color $i \in \{t^2 + 1, \ldots, 2t^2\}$. This sequences gives total costs of at least

$$(100Kt^2 + 2) \cdot 100K^2t^4 > 10000K^3t^6 > costs(K).$$

Therefore, the t^2 intervals $(X_{ij}, 10 + a_i + b_j]$ are colored with the first t^2 colors and the intervals $(X_{ij}, 14]$ with the last t^2 colors.

Case B: $f((11 - c_k, 13]) \notin \{1, \ldots, t\}$ for a fixed k. In this case, at least one chain of length $2000K^2t^4 + 1$ is colored with a color with costs $\geq 10Kt^2$. This chain creates total costs of at least

$$(2000K^2t^4 + 1) \cdot 10Kt^2 \geq 20000K^3t^6 > costs(K).$$

This implies, that $f((11 - c_k, 13]) \in \{1, \ldots, t\}$ for each $1 \leq k \leq t$.

Case C.1: $\{f((A_i, X_{ij}]), f((A_i, X_{ij'}])\} \subset \{1, \ldots, t^2\}$ with fixed i and $j \neq j'$. In this case, at least one of these intervals is colored with the same color as the interval $(0, A_i]$. Then, at least one chain of length $2000K^2t^4+3$ or length $100Kt^2+3$ is colored with a color $i \in \{t^2 + 1, \ldots, 2t^2\}$. This chain creates total costs of at least

$$(100Kt^2 + 3) \cdot 100K^2t^4 \geq 10000K^3t^6 > costs(K).$$

Case C.2: $f((A_i, X_{ij}]) \notin \{1, \ldots, t\}$ for fixed i and each $1 \leq j \leq t$. This case gives costs of at least

$$(2000K^2t^4 + 3) \cdot 10Kt^2 \geq 20000K^3t^6 > costs(K).$$

Therefore, exactly one interval of the form $(A_i, X_{ij}]$ for each $1 \leq i \leq t$ is colored with a color with costs 1. Let $\rho(i) \in \{1, \ldots, t\}$ be the index j with $f((A_i, X_{ij}]) \in \{1, \ldots, t\}$. Using case C.1, the other intervals $(A_i, X_{ij'}]$ with $j' \neq \rho(i)$ are colored with colors with costs $100K^2t^4$. Using case A, the intervals $(X_{ij}, 10 + a_i + b_j]$ for

$j \neq \rho(i)$ are colored with colors with costs $10Kt^2$. This implies that exactly $t-1$ intervals of the form $(B_j, X_{ij}]$ for a fixed i are colored with colors with costs $10Kt^2$.

Case D: $\{f((A_i, X_{ij}]), f((A_{i'}, X_{i'j}])\} \subset \{1, \dots, t\}$ with fixed j and $i \neq i'$. In this case, we have at most $t-2$ intervals of the form $(X_{i''j}, 10+a_{i''}+b_j]$ colored with colors with costs $10Kt^2$. Since the chains fit seamlessly one after another during the interval $(0, 10]$, at most $t-2$ intervals of the form $(B_j, X_{i''j}]$ for fixed j are colored with colors with costs $10Kt^2$. Therefore, at least one chain of length $100Kt^2+3$ is colored with costs $100K^2t^4$ and we obtain costs of at least $10000K^3t^6 > costs(K)$.

This analysis shows that

(1) $f((11-c_k, 13]) \in \{1, \dots, t\}$ for each $1 \leq k \leq t$,
(2) $f((X_{i\rho(i)}, 10+a_i+b_{\rho(i)}]) \in \{1, \dots, t\}$ for each $1 \leq i \leq t$ and
(3) ρ forms a permutation of $\{1, \dots, t\}$.

Define $\delta(i) = k$ whenever interval $(X_{i\rho(i)}, 10+a_i+b_{\rho(i)}]$ is colored with the same color with costs 1 as interval $(11-c_k, 13]$. Using the fact $\sum_{i=1}^{t}(a_i+b_i+c_i) = t$, the intervals $(X_{i\rho(i)}, 10+a_i+b_{\rho(i)}]$ and $(11-c_{\delta(i)}, 13]$ fit seamlessly after another. This implies that $10+a_i+b_{\rho(i)} = 11-c_{\delta(i)}$ and that $a_i+b_{\rho(i)}+c_{\delta(i)} = 1$. Therefore, I_1 is a yes-instance of N3DM.

Conversely, given a solution for N3DM, a coloring of all intervals of I_2 with total costs $costs(K)$ can be constructed using the description above.

The second part of claim 1 follows by the following observation. Consider a coloring f of all intervals in I_2.

Case 1: We have more than $2t^2$ colors in f. Then, the total costs are greater than $k_{2t^2+1} = 20000K^3t^6$.

Case 2: We have at most $2t^2$ colors in f. Then, at least one of the four cases A-D above must occur (otherwise we would get a yes instance of N3DM) and we obtain costs of at least $10000K^3t^6$.

Therefore, in both cases an optimum coloring f generates costs greater than $10000K^3t^6$.

The **second part** of the proof is the specification of the parameter K to achieve our non-approximability result. Let H be an approximation algorithm for the OCCP problem that computes a coloring with costs $H(I_2) \leq c|V|^{0.5-\epsilon}OPT(I_2)$ where c is a constant and $OPT(I_2)$ are the minimum costs of a solution of instance I_2. We construct for an instance I_1 of the N3DM problem an instance I_2 of the OCCP problem as described above. We obtain an interval graph with $8t^2 + 4000K^2t^5 + 200Kt^4 - 200Kt^3 \leq 4208K^2t^5$ vertices.

If there exists a solution of the N3DM problem, the optimum solution of the OCCP instance I_2 has costs at most $4355K^2t^6$. In this case, our approximation algorithm H produces the value $H(I_2) \leq 4355K^2t^6c|V|^{0.5-\epsilon}$. Since the number of vertices in I_2 is at most $4208K^2t^5$, we have $c|V|^{0.5-\epsilon} \leq (4208)^{0.5}cK^{1-2\epsilon}t^{2.5}$. This implies that $H(I_2) \leq (4208)^{0.5}4355cK^{3-2\epsilon}t^{8.5}$.

If there exists no solution of the N3DM problem, then $OPT(I_2) > 10000K^3t^6$ and, therefore, algorithm H generates a solution with costs greater than $10000K^3t^6$. We consider the following inequality:

$$10000K^3t^6 > (4208)^{0.5}4355K^{3-2\epsilon}ct^{8.5}.$$

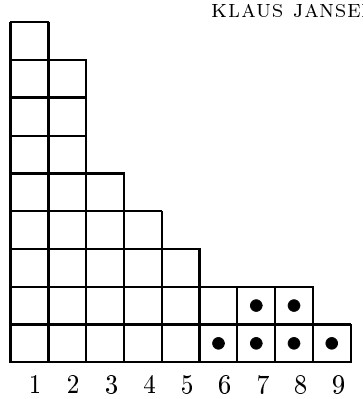

FIGURE 4. An optimum solution for a graph with 36 vertices that uses $b_{opt} = 9$ colors and $a_{opt} = 6$ colors.

This inequality is satisfied if and only if

$$K^{2\epsilon} > \frac{(4208)^{0.5}4355}{10000} c t^{2.5}.$$

We define

$$K = \left\lceil \sqrt[2\epsilon]{\frac{(4208)^{0.5}4355}{10000} t^{2.5}} \right\rceil + 1.$$

Since c and ϵ are constant, K is a polynomial in t and, therefore, I_2 with this parameter K can be constructed in polynomial time. If there exists no solution of the N3DM problem, then H generates a solution with costs of at least

$$10000K^3t^6 > (4208)^{0.5}4355K^{3-2\epsilon}ct^{8.5} \geq 4355K^2t^6c|V|^{0.5-\epsilon}.$$

Therefore, by using the polynomial time approximation algorithm H we could decide the existence of a solution for the N3DM problem, which would imply $P = NP$. □

3.2. Approximability result. Next, we propose an approximation algorithm A with ratio $O(|V|^{0.5})$ for the OCCP problem restricted to interval graphs or to unimodular graphs. The key idea is to analyse the structure of the optimum solution and to solve a special coloring problem.

Suppose that the optimum solution consists of $b_{opt} \geq \omega(G)$ colors. In Figure 4, we have an optimum solution for a graph with 36 vertices such that $b_{opt} = 9$. Each vertex of the graph is represented by a square and each column of squares corresponds to an independent set of G. Furthermore, we assume that the colors $a_{opt}, a_{opt}+1, \ldots, b_{opt}$ cover at least $\lceil\sqrt{|V|}\rceil$ vertices and that the colors $a_{opt}+1, \ldots, b_{opt}$ cover less than $\lceil\sqrt{|V|}\rceil$ vertices of G. This implies that $a_{opt} \in \{1, \ldots, b_{opt}\}$ (see also Figure 4 with $a_{opt} = 6$). Let $n_{a_{opt},b_{opt}-a_{opt}}$ be the number of vertices colored with colors $1, \ldots, a_{opt}$ and let $\bar{n}_{a_{opt},b_{opt}-a_{opt}}$ be the number of vertices colored with the other colors $a_{opt}+1, \ldots, b_{opt}$. Therefore, $\bar{n}_{a_{opt},b_{opt}-a_{opt}}$ is bounded by $\sqrt{|V|}$.

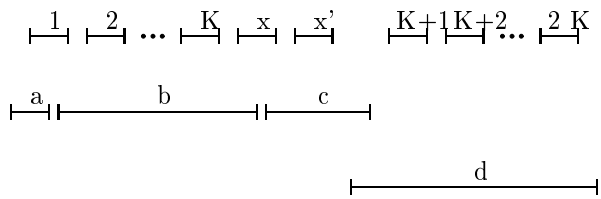

FIGURE 5. An interval graph with maximum clique size $\omega(G) = 2$ and coloring costs $k_1 = 1, k_2 = K$ and $k_3 = K^2$

Using these assumptions, we obtain the following lower bounds for the minimum costs $OPT(I)$ of a coloring:

$$\begin{array}{ll} (1) & OPT(I) \geq \lceil \sqrt{|V|} \rceil \cdot k_{a_{opt}}, \\ (2) & OPT(I) \geq k_{b_{opt}}. \end{array}$$

The first inequality is satisfied, since $\lceil \sqrt{|V|} \rceil$ vertices are colored with the colors $a_{opt}, a_{opt}+1, \ldots, b_{opt}$ and since $k_{a_{opt}} \leq k_{a_{opt}+1} \leq \ldots \leq k_{b_{opt}}$. The second inequality follows from the fact that color b_{opt} occurs at least once in the optimum coloring.

For our approximation algorithm we have to solve the following graph theoretical problem (called *maximum* $(a, b-a)$*-colorable subgraph problem*).

Maximum $(a, b-a)$-colorable subgraph:
Given: A graph $G = (V, E)$, and numbers $a, b \in \mathbb{N}$ with $a \leq b$ and $b \geq \chi(G)$.
Question: Compute a partition $(V', V \setminus V')$ of V such that V' has maximum cardinality and can be colored with a colors and $V \setminus V'$ can be colored with $b - a$ colors.

In Figure 5, an example is given to illustrate that maximum $(a, b-a)$-colorable subgraphs are useful for the approximation. If we take a greedy 2-coloring, then we could get the partition $\{a, b, c, K+1, \ldots, 2K\}$, $\{1, \ldots, K, x, x', d\}$ with costs $(K+3) + K(K+3) = \Omega(K^2)$. If we take first a maximum independent set $\{1, \ldots, 2K, x, x'\}$, then we have to use three colors with costs $2K + 2 + 3K + K^2 = \Omega(K^2)$. On the other hand, an optimum solution generated by a maximum $(1,1)$-colorable subgraph consists of two sets $\{1, \ldots, 2K, c\}$, $\{a, b, x, x', d\}$ with costs $2K + 1 + 5K = 7K + 1$. Since the number of vertices is equal to $2K + 6$, the first two heuristics generate a coloring with costs $\Omega(|V|)OPT(I)$.

Let H be an optimum algorithm to solve the maximum $(a, b-a)$-colorable subgraph problem. Later, we prove that there are polynomial optimum algorithms for interval graphs and also for unimodular graphs. A call of this algorithm with parameters a and b is denoted by $H(a, b-a)$. Note, that the maximum $(a, b-a)$-colorable subgraph problem is harder as the maximum q-colorable subgraph problem. This implies that the decision problem corresponding to the maximum $(a, b-a)$-colorable subgraph problem is NP-complete for e.g. split graphs, undirected path graphs and their complements and for k-trees with unbounded k.

We denote by $\alpha_{a,b-a}(G)$ the maximum cardinality of such a subset V' and with $\bar{\alpha}_{a,b-a}(G)$ the number of vertices in $V \setminus V'$. Clearly, $\alpha_{a_{opt},b_{opt}-a_{opt}}(G) \geq n_{a_{opt},b_{opt}-a_{opt}}$ and $\bar{\alpha}_{a_{opt},b_{opt}-a_{opt}}(G) \leq \bar{n}_{a_{opt},b_{opt}-a_{opt}}$. Given a solution with sets V' (and $V \setminus V'$), a coloring with at most a (and $b-a$) colors can be computed with an optimum coloring algorithm for several classes of graphs (e.g. interval or unimodular graphs). Since the colors $a_{opt}+1, \ldots, b_{opt}$ cover less than $\sqrt{|V|}$ vertices

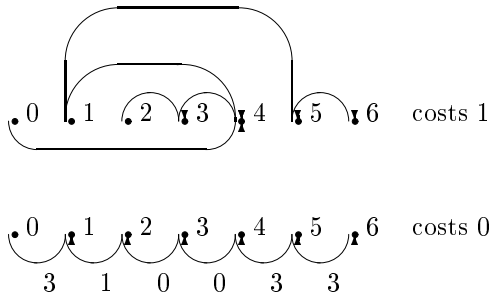

FIGURE 6. The digraph with orginal and dummy arcs for the intervals $(3,4]$, $(5,6]$, $(0,4]$, $(1,5]$, $(2,3]$ and $(1,4]$ and parameter $b=4$

in the optimum solution and since $\bar{\alpha}_{a_{opt},b_{opt}-a_{opt}}(G) \leq \bar{n}_{a_{opt},b_{opt}-a_{opt}}$, the value $\bar{\alpha}_{a_{opt},b_{opt}-a_{opt}}(G)$ is bounded by $\sqrt{|V|}$. Let $C_H(a_{opt}, b_{opt}-a_{opt})$ be the costs of a coloring computed by a call of the algorithm $H(a_{opt}, b_{opt}-a_{opt})$ and a corresponding coloring algorithm. Then, we can bound the costs $C_H(a_{opt}, b_{opt}-a_{opt})$ using the lower bounds (1) and (2) as follows:

$$\begin{aligned} C_H(a_{opt}, b_{opt}-a_{opt}) &\leq \alpha_{a_{opt},b_{opt}-a_{opt}}(G) \cdot k_{a_{opt}} + \bar{\alpha}_{a_{opt},b_{opt}-a_{opt}}(G) \cdot k_{b_{opt}} \\ &\leq |V| \cdot k_{a_{opt}} + \sqrt{|V|} \cdot k_{b_{opt}} \\ &\leq \sqrt{|V|}\sqrt{|V|} k_{a_{opt}} + \sqrt{|V|} \cdot k_{b_{opt}} \\ &\leq 2\sqrt{|V|} OPT(I). \end{aligned}$$

Now, we prove that the maximum $(a, b-a)$-colorable subgraph problem can be solved in polynomial time for interval and also unimodular graphs. Later, we show how we can find the parameters a_{opt} and b_{opt} efficiently.

THEOREM 3.2. *The maximum $(a, b-a)$-colorable subgraph problem for interval graphs is solvable in polynomial time using a mincost flow algorithm.*

PROOF. We show that the maximum $(a, b-a)$-colorable subgraph problem corresponds to a mincost flow problem. Let G be an interval graph with corresponding intervals $(a_i, b_i]$ such that $a_i, b_i \in \{0, \ldots, n\}$ for $1 \leq i \leq n$. The maximum cliquesize $\omega(G)$ is given by

$$max_{1 \leq i \leq n} |\{j | j \in \{1, \ldots, n\}, i \in (a_j, b_j]\}|.$$

We construct for G an acyclic digraph $D = (\{0, \ldots, n\}, A)$ as follows. For each vertex $j \in V$ with corresponding interval $(a_j, b_j]$ we choose an arc from a_j to b_j in D with capacity 1 and cost 1. Furthermore, for $i \in \{1, \ldots, n\}$ we include a dummy arc from $i-1$ to i. These arcs get the capacities $b - |\{j | j \in V, i \in (a_j, b_j]\}|$ and costs 0. The goal is to compute a minimum cost flow from 0 to n with flow value $b-a$.

In Figure 6, we have given the digraph generated for the intervals $(3,4]$, $(5,6]$, $(0,4]$, $(1,5]$, $(2,3]$ and $(1,4]$ and parameter $b=4$. The original arcs with costs 1 and capacities 1 are given in the upper part and the dummy arcs with costs 0 and capacities between 0 and b are given in the lower part of Figure 6. A maximum $(2,2)$-colorable subgraph of the interval graph with $|V'| = 4$ corresponds to the following paths

(1) $(0,4], (4,5](5,6]$ with two original arcs $(0,4], (5,6]$,

(2) $(0,1](1,2],(2,3](3,4],(4,5](5,6]$ with two original arcs $(2,3],(3,4]$.

We prove the following **claim**: The maximum cardinality of a set $V' \subset V$ of a maximum $(a, b-a)$ colorable subgraph is equal to $|V|$ minus the costs of a cost minimum flow from 0 to n with flow value $b-a$.

Suppose that f is a minimum cost flow from 0 to n with flow value $b-a$ and costs $costs(f)$. Since the capacity of each arc is integer, we may assume that f has an integer flow through each arc. Therefore, the flow f corresponds to $b-a$ paths from 0 to n with unit flow and each of these paths goes through original and dummy arcs. The costs $costs(f)$ of this flow are equal to the number of original arcs covered by the $b-a$ paths. Since the sum of the capacities of the arcs at each time step $i \in (0,n]$ is exactly b, it remains a set of uncovered original arcs and dummy arcs with capacity sum equal to a for each time step $i \in (0,n]$. Therefore, the intervals corresponding to the uncovered original arcs can be colored with a colors. Let V' be the set of the intervals corresponding to the uncovered arcs. Then, the cardinality $|V'|$ is equal to $|V| - costs(f)$ and $(V', V \setminus V')$ is a $(a, b-a)$ colorable subgraph of G. Next, we prove that V' is a maximum set with this property.

On the contrary, suppose that V' is not a maximum set. Let $\bar{V}'$ be a subset of V with $|\bar{V}'| > |V'|$ such that

(1) $\bar{V}'$ can be colored with a colors and
(2) $V \setminus \bar{V}'$ can be colored with $b-a$ colors.

Using the dummy arcs, these sets can be assigned

(1) a paths from 0 to n in D that go through the original arcs corresponding to the intervals in $\bar{V}'$ and
(2) $b-a$ paths from 0 to n in D that go through the original arcs corresponding to the intervals in $V \setminus \bar{V}'$.

These $b-a$ paths correspond to a flow $\bar{f}$ with flow value $b-a$ and costs $|V \setminus \bar{V}'|$. Since $|V \setminus \bar{V}'| < |V \setminus V'|$, the costs $costs(\bar{f}) < costs(f)$ and we have a contradiction.

Conversely, suppose that $V' \subset V$ is a set with maximum cardinality in an $(a, b-a)$-colorable subgraph. Then, V' is colorable with a colors and $V \setminus V'$ is colorable with $b-a$ colors. As above, we obtain a flow f with flow value $b-a$ in D and $costs(f) = |V \setminus V'|$. Suppose that f is not a minimum cost flow. Then, there exists another flow $\bar{f}$ with $costs(\bar{f}) < costs(f)$. Again as above, $\bar{f}$ corresponds to a subset $\bar{V} \subset V$ of original arcs (or intervals) that lie in one of the $b-a$ paths of $\bar{f}$. We have $|\bar{V}| = costs(\bar{f})$ and $|V \setminus \bar{V}| = |V| - costs(\bar{f})$. Then, the intervals in $V \setminus \bar{V}$ can be colored with a colors and $|V \setminus \bar{V}| > |V'|$, since $costs(\bar{f}) < costs(f)$. This gives a contradiction and, therefore, f is a cost minimum flow with flow value $b-a$. □

Notice that the simpler maximum a-colorable subgraph problem is solvable in linear time $O(|V| + |E|)$ [**25**] for interval graphs.

Another idea is to formulate this problem as a zero-one linear program. Suppose we have an instance I of the maximum $(a, b-a)$-colorable subgraph problem containing a perfect graph $G = (V, E)$ with $V = \{1, \ldots, n\}$ and positive integers a, b. The integer program contains binary decision variables $x_i \in \{0, 1\}$ indicating

whether $i \notin V'$ or $i \in V'$. The objective function and the constraints are:

(0)	$max \sum_{i=1}^{n} x_i$	such that
(1)	$\sum_{i \in C} x_i \leq a$	for each clique C of G,
(2)	$\sum_{i \in C} x_i \geq \lvert C \rvert - (b-a)$	for each clique C of G,
(3)	$x_i \geq 0$	for each $1 \leq i \leq n$,
(4)	$x_i \leq 1$	for each $1 \leq i \leq n$.

The objective function (0) specifies the number of vertices in V' that we want to maximize. An inequality in (1) specifies that for each clique at most a variables are equal to one, and an inequality in (2) declares that for a clique C at most $b-a$ variables x_i are equal to zero. Since G is perfect, the constraints (1) guarantee that V' can be colored with at most a colors and the constraints (2) that $V \setminus V'$ can be colored with at most $b-a$ colors.

If G is an unimodular graph, G is perfect and the clique matrix A is totally unimodular (each square submatrix of A has determinant equal to 0, +1 or −1). The constraints (1) − (4) can be written in the following way:

$$\begin{bmatrix} -I \\ I \\ A \\ -A \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \leq \begin{bmatrix} 0 \\ 1 \\ a \\ -|C| + (b-a) \end{bmatrix}$$

Since A is totally unimodular, the matrix above is also totally unimodular. This implies that the polyhedron given by the constraints (1)−(4) for unimodular graphs has only integral extrema. Therefore, we get the following result:

THEOREM 3.3. *The maximum $(a, b-a)$-colorable subgraph problem for unimodular graphs is solvable in polynomial time.*

For the approximation algorithm for the OCCP problem the values a and $b-a$ can be bounded by $\omega(G)$. If $a_{opt} \geq \omega(G)$, the optimum costs $OPT(I)$ are greater than $\lceil \sqrt{|V|} \rceil k_{\omega(G)}$. In this case, we get an approximate solution with $a = \omega(G)$ and $b = a$ using

$$C_H(\omega(G), 0) \leq \alpha_{\omega(G),0}(G) k_{\omega(G)} \leq \sqrt{|V|}\sqrt{|V|} k_{\omega(G)} \leq \sqrt{|V|} OPT(I).$$

If $b_{opt} > \omega(G) + a_{opt}$ and $a_{opt} \leq \omega(G)$, then the optimum costs $OPT(I) \geq k_{\omega(G)+a_{opt}}$. In this case, we get an approximate solution with $a = a_{opt}$ and $b = a + \omega(G)$ using

$$C_H(a_{opt}, \omega(G)) \leq |V| k_{a_{opt}} + \sqrt{|V|} k_{\omega(G)+a_{opt}} \leq 2\sqrt{|V|} OPT(I).$$

These arguments imply that at most $O(\omega(G)^2)$ calls of the maximum $(a, b-a)$ - colorable subgraph are sufficient for our approximation algorithm. In the next part of this section we improve this bound. We show that at most $O(\log \omega(G))$ calls of the maximum $(a, b-a)$ colorable subgraph algorithm H are needed.

For each $a \in \{1, \ldots, \omega(G)-1\}$, let x be the smallest integer with $x \in \{1, \ldots, \omega(G)\}$ (if possible) such that $\bar{\alpha}_{a,x}(G) < \lceil \sqrt{|V|} \rceil$. We notice that $\bar{\alpha}_{a,\omega(G)}(G)$ can be greater than $\lceil \sqrt{|V|} \rceil$ and that

$$\bar{\alpha}_{a,1}(G) \geq \bar{\alpha}_{a,2}(G) \geq \ldots \geq \bar{\alpha}_{a,\omega(G)}(G).$$

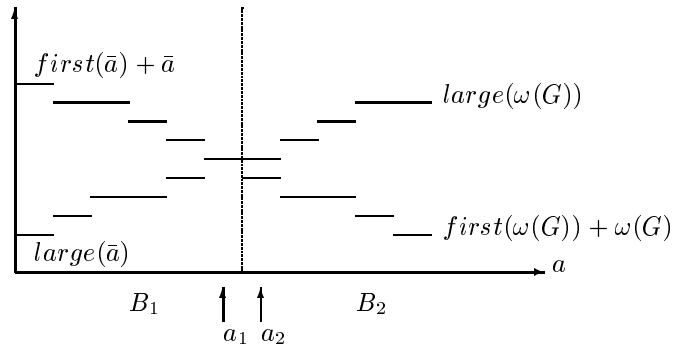

FIGURE 7. The mappings $\varphi : a \to first(a) + a$ and $\psi : a \to large(a)$

For $a \in \{1, \ldots, \omega(G) - 1\}$ we define

$$first(a) = \begin{cases} x & \text{if } \bar{\alpha}_{a,\omega(G)}(G) < \lceil \sqrt{|V|} \rceil \\ \infty & \text{otherwise} \end{cases}$$

Since G can be colored with $\omega(G)$ colors, we define $first(\omega(G)) = 0$.

LEMMA 3.4. *If $a < a'$ and $first(a), first(a') \neq \infty$ then $first(a') + a' \leq first(a) + a$.*

PROOF. Consider a solution with $\bar{\alpha}_{a,first(a)}(G) < \lceil \sqrt{|V|} \rceil$. In this solution, less than $\lceil \sqrt{|V|} \rceil$ vertices are colored with the colors $a + 1, \ldots, first(a) + a$. We take this solution as a solution for a'. We get a solution with less than $\lceil \sqrt{|V|} \rceil$ vertices colored with the colors $a' + 1, \ldots, first(a) + a$. The number of used colors in the second part is equal to $first(a) + a - a'$. Since the solution of the maximum $(a', first(a'))$ colorable subgraph problem has the same number of colors or fewer colors in the second component, it holds $first(a') \leq first(a) + a - a'$. This proves the Lemma. □

The smallest $\bar{a}$ with $first(\bar{a}) < \infty$ can be found using binary search with calls $H(a, \omega(G))$. Therefore, $\bar{a}$ can be found with $O(\log \omega(G))$ calls of the maximum a-colorable subgraph algorithm. We notice that $first(a') < \infty$ for each $a' \in \{\bar{a}, \ldots, \omega(G)\}$. This implies that the mapping $\varphi : a \to first(a) + a$ is non-increasing.

For each $a \in \{1, \ldots, \omega(G)\}$, let x be the smallest number in $\{a + 1, \ldots, |V|\}$ (if existent) with $k_x > \sqrt{|V|} k_a$. Notice that $k_{|V|}$ can be smaller than $\sqrt{|V|} k_a$. We define

$$large(a) = \begin{cases} x & \text{if } k_{|V|} > \sqrt{|V|} k_a \\ \infty & \text{otherwise} \end{cases}$$

For $\omega(G) = |V|$, we define $large(|V|) = \infty$.

LEMMA 3.5. *If $a < a'$ then $large(a) \leq large(a')$.*

PROOF. First, we consider the case with $large(a), large(a') \neq \infty$. Suppose that $large(a) > large(a')$. Since $k_a \leq k_{a'}$ and $k_{large(a')} \leq k_{large(a)}$, we obtain

$$k_{large(a)} \geq k_{large(a')} > \sqrt{|V|} k_{a'} \geq \sqrt{|V|} k_a.$$

Thus, $k_{large(a')}$ is also larger than $\sqrt{|V|}k_a$. Since $large(a') < large(a)$, we obtain a contradiction.

Next, we consider $large(a) = \infty$. In this case, $k_{|V|} \leq \sqrt{|V|}k_a \leq \sqrt{|V|}k_{a'}$ (using $a < a'$) and, therefore, $large(a') = \infty$. Finally, if $large(a') = \infty$ then the statement follows directly. □

This Lemma implies that the mapping $\psi : a \to large(a)$ is non-decreasing. Next, we define two regions $B_1 = \{a \in \{\bar{a}, \dots, \omega(G)\} | large(a) \leq first(a) + a\}$ and $B_2 = \{a \in \{\bar{a}, \dots, \omega(G)\} | large(a) > first(a) + a\}$. We define $a_1 = max(B_1)$ and $a_2 = min(B_2)$ (if the corresponding sets are non-empty). In Figure 7, we have illustrated the mappings $\psi : a \to large(a)$ and $\varphi : a \to first(a) + a$. Since the mappings φ and ψ are non-increasing and non-decreasing, for each pair $a \in B_1$, $a' \in B_2$ we have $a < a'$. There are two special cases:

(1): $first(\bar{a}) + \bar{a} < large(\bar{a})$. In this case, $B_1 = \emptyset$ and $B_2 = \{\bar{a}, \dots, \omega(G)\}$.

(2): $first(\omega(G)) + \omega(G) \geq large(\omega(G))$. Since $first(\omega(G)) = 0$, we get $large(\omega(G)) \leq \omega(G)$. Using the definition of the mapping $large$, we know that $large(\omega(G)) \geq \omega(G) + 1$. Therefore, this case is not possible. Furthermore, this implies that the case $B_2 = \emptyset$ and $B_1 = \{\bar{a}, \dots, \omega(G)\}$ is also not possible.

Consider an optimum solution with parameters (a_{opt}, b_{opt}) where $a_{opt} \leq \omega(G)$. We prove that is sufficient to compute at most two solutions using a $(a_1, first(a_1))$ and a $(a_2, first(a_2))$ maximum colorable subgraph.

LEMMA 3.6. *Let $(V', V \backslash V')$ be an optimum solution with parameters (a_{opt}, b_{opt}) such that $a_{opt} \leq \omega(G)$. If $B_1 \neq \emptyset$ and $B_2 \neq \emptyset$ then*

$$min(C_H(a_1, first(a_1)), C_H(a_2, first(a_2))) \leq 2\sqrt{|V|}OPT(I).$$

If $B_1 = \emptyset$, then $a_2 = \bar{a}$ and

$$C_H(\bar{a}, first(\bar{a})) \leq 2\sqrt{|V|}OPT(I).$$

PROOF. We analyse the case with $B_1 \neq \emptyset$ and $B_2 \neq \emptyset$; the other case with $B_1 = \emptyset$ follows then directly. Clearly, we have $a_2 = a_1 + 1$. We have to consider three cases.

Case 1: $large(a_{opt}) \leq first(a_{opt}) + a_{opt}$ and $a_{opt} \geq \bar{a}$. In this case, the optimum solution lies in the first region B_1. Notice that $first(a_{opt}) + a_{opt} \leq b_{opt}$. Since the mapping $\varphi : a \to first(a) + a$ is non-increasing and $a_1 \geq a_{opt}$, we have $first(a_1) + a_1 \leq first(a_{opt}) + a_{opt}$. Moreover, it holds that $large(a_1) \leq first(a_1) + a_1$ and that $k_{large(a_1)} > \sqrt{|V|}k_{a_1}$.

The costs $C_H(a_1, first(a_1)) \leq |V|k_{a_1} + \sqrt{|V|}k_{first(a_1)+a_1}$. Using $\sqrt{|V|}k_{a_1} < k_{large(a_1)}$ and $large(a_1) \leq first(a_1) + a_1 \leq first(a_{opt}) + a_{opt} \leq b_{opt}$, we obtain

$$C_H(a_1, first(a_1)) \leq 2\sqrt{|V|}k_{b_{opt}} \leq 2\sqrt{|V|}OPT(I).$$

Case 2: $large(a_{opt}) > first(a_{opt}) + a_{opt}$ and $a_{opt} \geq \bar{a}$. In this case, the optimum solution lies in region B_2. Using the non-decreasing mapping $\psi : a \to large(a)$ and $a_2 \leq a_{opt}$, we have $large(a_2) \leq large(a_{opt})$. If $large(a_2) \neq \infty$, it holds that $k_{large(a_2)} > \sqrt{|V|}k_{a_2}$. Using the definition of $large(a_2)$ and the fact that $first(a_2) + a_2 < large(a_2)$, we have $k_{first(a_2)+a_2} \leq \sqrt{|V|}k_{a_2}$. The costs $C_H(a_2, first(a_2)) \leq |V|k_{a_2} + \sqrt{|V|}k_{first(a_2)+a_2} \leq 2|V|k_{a_2}$. Since $k_{a_2} \leq k_{a_{opt}}$ and

$\sqrt{|V|}k_{a_{opt}} \leq OPT(I)$, we get

$$C_H(a_2, first(a_2)) \leq 2|V|k_{a_2} \leq 2\sqrt{|V|}\sqrt{|V|}k_{a_{opt}} \leq 2\sqrt{|V|}OPT(I).$$

If $large(a_2) = \infty$, then we have $k_{|V|} \leq \sqrt{|V|}k_{a_2}$. In this case,

$$\begin{aligned} C_H(a_2, first(a_2)) &\leq |V|k_{a_2} + \sqrt{|V|}k_{first(a_2)+a_2} \leq |V|k_{a_2} + \sqrt{|V|}k_{|V|} \leq 2|V|k_{a_2} \\ &\leq 2\sqrt{|V|}OPT(I). \end{aligned}$$

Case 3: $a_{opt} < \bar{a}$. In this case, $first(a_{opt}) = \infty$ and $\bar{\alpha}_{a_{opt},\omega(G)} \geq \lceil\sqrt{|V|}\rceil$. Using the fact that (a_{opt}, b_{opt}) are the parameters of the optimum solution, we have $\bar{n}_{a_{opt},b_{opt}-a_{opt}} < \lceil\sqrt{|V|}\rceil$ and, therefore, $\bar{\alpha}_{a_{opt},\omega(G)} < \lceil\sqrt{|V|}\rceil$. This gives a contradiction and shows that $a_{opt} \geq \bar{a}$. □

The next Lemma implies how the values a_1 and a_2 can be computed using binary search and calls $H(a, large(a) - a)$ and $H(a, large(a) - a - 1)$.

LEMMA 3.7. *Let $a \in \{\bar{a}, \ldots, \omega(G)\}$. By using calls $H(a, large(a) - a)$ and $H(a, large(a) - a - 1)$, we can decide whether $a \in B_1$ or $a \in B_2$.*

PROOF. If $large(a) = \infty$ then using $a \geq \bar{a}$ we have $first(a) + a < large(a)$ and, therefore, $a \in B_2$. Suppose that $large(a) < \infty$. To decide whether $a \in B_1$ or $a \in B_2$, consider the cardinalities $\bar{\alpha}_{a,large(a)-a}$ and $\bar{\alpha}_{a,large(a)-a-1}$ computed by the corresponding calls.

Case 1: $\bar{\alpha}_{a,large(a)-a} < \lceil\sqrt{|V|}\rceil$. In this case, $first(a) \leq large(a) - a$.

Case 1.1: $\bar{\alpha}_{a,large(a)-a-1} < \lceil\sqrt{|V|}\rceil$. Again, we get $first(a) \leq large(a) - a - 1 < large(a) - a$. Therefore, we obtain $a \in B_2$.

Case 1.2: $\bar{\alpha}_{a,large(a)-a-1} \geq \lceil\sqrt{|V|}\rceil$. Using the definition of $first(a)$, we get $first(a) = large(a) - a$ and, therefore, $a \in B_1$.

Case 2: $\bar{\alpha}_{a,large(a)-a} \geq \lceil\sqrt{|V|}\rceil$. In this case, we obtain $first(a) > large(a) - a$ and, therefore, $a \in B_1$. □

Now, we are ready for our approximation algorithm.

Algorithm B:

(1): compute $large(a)$ for each $a \in \{1, \ldots, \omega(G)\}$ (using preprocessing in $O(\omega(G) + |V|)$ time),

(2): compute a first solution with the call $H(\omega(G), 0)$ (this is an arbitrary coloring for the case $a_{opt} > \omega(G)$),

(3): find the smallest $\bar{a}$ with $first(\bar{a}) < \infty$ using binary search with calls $H(a, \omega(G))$ (these are maximum a - colorable subgraphs),

(4): compute a_1 (if it exists) and a_2 by using a binary search with calls $H(a, large(a) - a)$ and $H(a, large(a) - a - 1)$ (these are maximum (a, x)-colorable subgraphs),

(5): compute $first(a_1)$ (if a_1 exists) and $first(a_2)$ using binary search with calls $H(a_i, x)$ (these are maximum (a_i, x)-colorable subgraphs),

(6): choose the cheapest among the solutions with costs $C_H(a_1, first(a_1))$ (if a_1 exists), $C_H(a_2, first(a_2))$ and $C_H(\omega(G), 0)$.

Using the calculations above, we obtain the following result.

THEOREM 3.8. *The approximation algorithm B above computes a coloring of the OCCP problem restricted to interval graphs (and unimodular graphs) with approximation ratio $O(|V|^{0.5})$.*

The time complexity of this algorithm for interval graphs is $O(\log \omega(G))$ calls of a minimum cost flow algorithm. For unimodular graphs, we need at most $O(\log \omega(G)))$ calls of a linear programming algorithm. It is an interesting question to improve the time complexities of the maximum $(a, b-a)$ colorable subgraph algorithms.

4. Permutation and comparability graphs

In this section we show that the OCCP problem restricted to permutation graphs (and therefore also for comparability graphs) is hard to approximate with ratio $O(|V|^{1-\epsilon})$. In our proof, we use the point model, where the permutation $\Pi = (i_1, \ldots, i_n)$ is interpreted as a set of points $\{p_j = (j, \Pi^{-1}(j)) | 1 \leq j \leq n\}$. An independent set $U = \{a_1, \ldots, a_u\}$ corresponds to the point set $S_U = \{(a_1, \Pi^{-1}(a_1)), \ldots, (a_u, \Pi^{-1}(a_u))\}$ in increasing order ($a_1 < \ldots < a_u$ and $\Pi^{-1}(a_1) < \ldots < \Pi^{-1}(a_u)$). Conversely, a clique C corresponds to a point set S_C in decreasing order.

For our non-approximability result, we use a reduction from 3-SAT.

3-SAT:

Given: A set of unnegated variables $X = \{x_1, \ldots, x_m\}$ and negated variables $\bar{X} = \{\bar{x}_1, \ldots, \bar{x}_m\}$, a collection of clauses $c_1, \ldots, c_n$ over $X \cup \bar{X}$ such that each clause c_i contains three literals $y_{i,1}, y_{i,2}, y_{i,3} \in X \cup \bar{X}$.

Question: Does there exist a truth assignment for the variables such that each clause is satisfied?

First, we show the NP-completeness of the OCCP problem using an integer parameter K. Later, we specify the parameter K to achieve our non - approximability result. The proof uses a technique of Wagner [**24**] who proved that the cochromatic number problem is NP-complete for permutation graphs.

THEOREM 4.1. *The OCCP problem for permutation graphs is NP-complete if there are at least three different cost values.*

PROOF. The theorem is proved by a reduction from 3-SAT. Let I_1 be an instance of 3-SAT containing a collection of n clauses $c_1, \ldots, c_n$ over a set of unnegated and negated variables $\{x_1, \bar{x}_1, \ldots, x_m, \bar{x}_m\}$. In the first step, we construct a new instance I_1' with $r = (m+1)n$ clauses such that $c'_{in+j} = c_j$ for $0 \leq i \leq m$ and $1 \leq j \leq n$. In other words, I_1' contains $m+1$ identical copies of each clause in I_1. The literals of clause c_i' are denoted by $y'_{i,1}, y'_{i,2}$ and $y'_{i,3}$. Clearly, I_1 is a yes-instance if and only if I_1' is a yes-instance of 3-SAT. An instance I_2 of the OCCP problem is constructed as follows. The permutation graph G is given by the corresponding point set

$$P = \bigcup_{i=1}^{m} (A_i \cup B_i) \cup \bigcup_{i=1}^{r} (C_i \cup D_i \cup \{a_{i1}, a_{i2}, a_{i3}\})$$

where

- A_i, B_i consists each of $10Kr$ points situated on the increasing line marked by A_i, B_i in Figure 8.
- C_i, D_i consists each of two points as given in Figure 8.
- a_{ik} is situated in the crossing of the decreasing line marked by h_i and the increasing line marked by f_ℓ if $y'_{i,k} = \bar{x}_\ell$ or g_ℓ if $y'_{i,k} = x_\ell$.

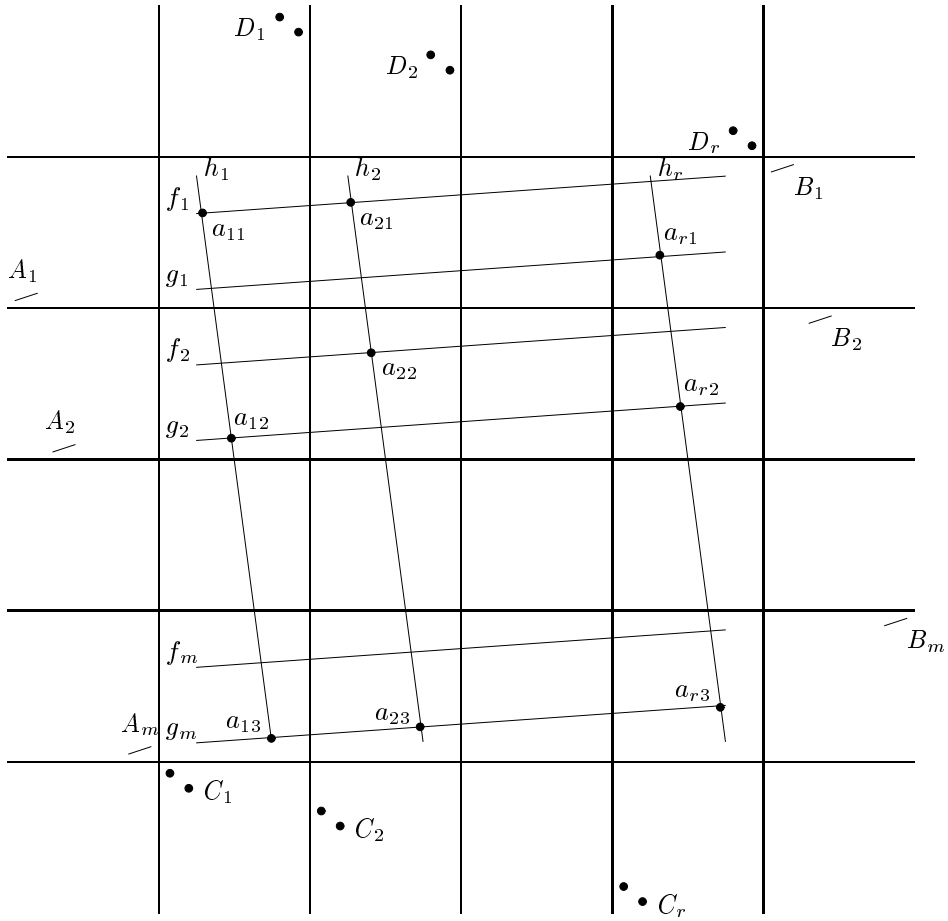

FIGURE 8. The point set P constructed for the 3-SAT instance $(\bar{x}_1 \vee x_2 \vee x_m) \wedge (\bar{x}_1 \vee \bar{x}_2 \vee x_m) \wedge \ldots \wedge (x_1 \vee x_2 \vee x_m)$

Furthermore, there are m colors with costs 1, $2r$ colors with costs $10Km$ and all other colors have costs $100K^2rm$.

As illustrated in Figure 8, the point set $\{a_{ik} | 1 \leq i \leq r, 1 \leq k \leq 3\}$ can be divided into r vertical stripes and m horizontal stripes. Since the instance I_1' consists of $m+1$ copies of instance I_1, we can subdivide the $r = (m+1)n$ vertical stripes into $m+1$ sections of n consecutive stripes.

We prove the following **claim**: I_1' is a yes-instance of 3-SAT if and only if the minimum total costs of coloring the permutation graph in instance I_2 do not exceed

$$costs(K) = (20Krm + 3r) + 6r \cdot 10Km = 80Krm + 3r \leq 83Krm.$$

Suppose that α is a truth assignment for the variables which makes the clauses $c_1', \ldots, c_r'$ true. The first m colors cover the points of $A_j \cup B_j$ and all points situated on f_j if $\alpha(x_j) = false$ or all points situated on g_j if $\alpha(x_j) = true$ for $1 \leq j \leq m$. Then, at least one of the points a_{i1}, a_{i2} or a_{i3} is colored with one of these m colors for each clause c_i'. The next $2r$ colors with costs $10Km$ are needed to color the

points of $C_j \cup D_j$ and the remaining uncolored points a_{i1}, a_{i2}, a_{i3} (at most two points for each clause c'_i). This gives a coloring of the corresponding permutation graph with costs bounded by $costs(K)$.

Conversely, suppose that the minimum costs of coloring all points in I_2 do not exceed $80Krm + 3r$. If we have more than $2r + m$ colors, then the costs are at least $100K^2rm > 80Krm + 3r$. Since the maximum clique size $\omega(G) = 2r + m$ (a clique of this size is given by the points in the sets $D_1, \ldots, D_r$ and arbitrary points $b_1, \ldots, b_m$ with $b_i \in B_i$), we have exactly $2r + m$ colors. Let $U_1, \ldots, U_{2r+m}$ be the independent sets in such a coloring.

The first m independent sets contain at least one point of A_i and one point of B_i, for each $1 \leq i \leq m$; otherwise at least one chain A_j or B_j is colored with a color with costs $10Km$. This would imply a coloring with costs of at least $100K^2rm > 83Krm$. Therefore, we may assume that at least one point in A_i and one in B_i is colored with color i. For $1 \leq i \leq m$, the corresponding independent set U_i can cover only points in the horizontal stripe i; otherwise U_i cannot be independent. Then, we may assume that all points in $A_i \cup B_i$ and a maximal number of points in the horizontal stripe i are colored with color i. Since the point set colored with i must be in increasing order, we cannot take first some points on line f_i and then some points on line g_i. The only possibility of assigning color i to some points on f_i and on g_i is to cover first some points on line g_i and then some points on line f_i.

The colors $m + 1, \ldots, m + 2r$ cover at least one point in C_i and at least one point in D_i; otherwise we need more than $m + 2r$ colors and would obtain costs of at least $100K^2rm$. Therefore, we may assume that the points in $C_i \cup D_i$ are colored with the colors $m+2i$ and $m+2i-1$ for $1 \leq i \leq r$. Furthermore, the corresponding independent sets U_{m+2i} and U_{m+2i-1} can cover only points in the vertical stripe i and at most two of the points a_{i1}, a_{i2} and a_{i3}. Since we have $m + 1$ sections and only m independent sets $U_1, \ldots, U_m$, there exists at least one section i_0 in which no independent set U_i $(1 \leq i \leq m)$ jumps from the lower line g_i to the upper line f_i. The section i_0 consists of the vertical stripes $(i_0 - 1)n + 1, \ldots, i_0 n$. We define the following truth assignment α for the variables $x_1, \ldots, x_m$:

$$\alpha(x_j) = \begin{cases} true & \text{if } U_j \text{ covers the points situated} \\ & \text{in section } i_0 \text{ and on line } g_j \\ false & \text{if } U_j \text{ covers the points situated} \\ & \text{in section } i_0 \text{ and on line } f_j \end{cases}$$

Since the independent sets U_{m+2i-1} and U_{m+2i} cover at most two of the points a_{i1}, a_{i2} and a_{i3}, the sets $U_1, \ldots, U_m$ have to cover at least one of these points for each i with $(i_0 - 1)n + 1 \leq i \leq i_0 n$. Let $a_{(i_0-1)n+j,s_j}$ be one point covered by the sets $U_1, \ldots, U_m$, for $1 \leq j \leq n$. If $a_{(i_0-1)n+j,s_j}$ lies on line f_ℓ, then we have $y'_{(i_0-1)n+j,s_j} = \bar{x}_\ell$ and $\alpha(x_\ell) = false$. If $a_{(i_0-1)n+j,s_j}$ lies on line g_ℓ, then we have $y'_{(i_0-1)n+j,s_j} = x_\ell$ and $\alpha(x_\ell) = true$. In both cases, we get $\alpha(y'_{(i_0-1)n+j,s_j}) = true$ and, therefore, α satisfies all clauses in I'_1. □

If I_1 is a no-instance of 3-SAT, then each coloring of the permutation graph in the I_2 instance produces costs of at least $100K^2rm$.

THEOREM 4.2. *For each $\epsilon < 1$, there exists no polynomial approximation algorithm with ratio $O(|V|^{1-\epsilon})$ for the OCCP problem restricted to permutation graphs, unless $P = NP$.*

PROOF. Let H be an approximation algorithm for the OCCP problem that computes a coloring with costs $H(I_2) \leq c|V|^{1-\epsilon}OPT(I_2)$, where c is a constant and $OPT(I_2)$ are the minimum costs of a solution of instance I_2.

We construct for an instance I_1 of the 3-SAT problem an instance I_2 of the OCCP problem as described in the proof above. We obtain a permutation graph with at most $20Krm + 7r \leq 27Krm$ vertices. If there exists a solution of the 3-SAT problem, the optimum solution of the OCCP instance I_2 has costs of at most $83Krm$. In this case, our approximation algorithm produces the value $H(I_2) \leq 83Krmc|V|^{1-\epsilon}$. Since the number of vertices in I_2 is at most $27Krm$, we have $c|V|^{1-\epsilon} \leq 27K^{1-\epsilon}crm$ and $H(I_2) \leq 27 \cdot 83K^{2-\epsilon}cr^2m^2$.

If there exists no solution of the 3-SAT instance, then $OPT(I_2) > 100K^2rm$ and, therefore, algorithm H generates a solution with costs greater than $100K^2rm$. Again, we consider an inequality

$$100K^2rm > 27 \cdot 83K^{2-\epsilon}cr^2m^2$$

or equivalent

$$K^\epsilon > \frac{27 \cdot 83}{100}crm.$$

We define

$$K = \left\lceil \sqrt[\epsilon]{\frac{27 \cdot 83}{100}crm} \right\rceil + 1.$$

Since c and ϵ are constant, K is a polynomial in r and m and, therefore, the instance I_2 can be constructed in polynomial time. If there exists no solution of the 3-SAT problem, then H generates a solution with costs of at least

$$100K^2rm > 27 \cdot 83K^{2-\epsilon}cr^2m^2 \geq 83Krmc|V|^{1-\epsilon}.$$

Therefore, by using the polynomial time approximation algorithm H, we could decide the existence of a satisfying truth assignment for the 3-SAT problem, which would imply $P = NP$. □

5. Split and chordal graphs

In this section we show that the OCCP problem restricted to split graphs (and therefore also for chordal graphs) is hard to approximate with ratio $O(|V|^{1-\epsilon})$. For the reduction we use the classic Minimum Cover (MC) problem that was proved NP-complete by Karp [**16**].

Minimum cover (MC):
Given: A set $S = \{1, \ldots, n\}$, subsets $S_i \subset S$ for $1 \leq i \leq m$, and a positive integer h.
Question: Do there exist h subsets $S_{i_1}, \ldots, S_{i_h}$ with $\cup_{j=1}^{h} S_{i_j} = \{1, \ldots, n\}$?

THEOREM 5.1. *For each $\epsilon < 1$, there exists no polynomial approximation algorithm with ratio $O(|V|^{1-\epsilon})$ for the OCCP problem restricted to split graphs, unless $P = NP$.*

PROOF. First, we give a reduction from MC with variable parameter $K \in \mathbb{N}$ and, later, we specify the parameter $K \in \mathbb{N}$ to achieve our non-approximability result. Let I_1 be an instance of MC containing a set $S = \{1, \ldots, n\}$, m subsets $S_i \subset S$ and a positive integer h. We construct an instance I_2 of the OCCP problem.

color	vertices
1	$\{s_{i_1}\} \cup \{i + nj \mid 0 \le j \le K-1, 1 \le i \le n, i \in S_{i_1}\}$
.	.
.	.
.	.
h	$\{s_{i_h}\} \cup \{i + nj \mid 0 \le j \le K-1, 1 \le i \le n, i \in S_{i_h}, i \notin \cup_{j=1}^{h-1} S_{i_j}\}$
$h+1$	$\{s_{i_{h+1}}\}$
.	.
.	.
.	.
m	$\{s_{i_m}\}$

TABLE 3. The optimum coloring of G

First, we define a split graph $G = (V, E)$ where $V = \{1, \dots, nK\} \cup \{s_1, \dots, s_m\}$ and where E contains a clique with vertices $s_1, \dots, s_m$ and the edge set

$$\{\{i + nj, s_\ell\} \mid 0 \le j \le K-1, 1 \le i \le n, 1 \le \ell \le m, i \notin S_\ell\}.$$

In other words, $\{s_1, \dots, s_m\}$ forms a clique and $\{1, \dots, nK\}$ forms an independent set in G. Furthermore, there are h colors with costs 1, $m-h$ colors with costs $10K$ and all other colors have costs $100K^2mn$.

Claim 1: I_1 is a yes-instance of MC if and only if the minimum total costs of coloring all vertices of I_2 do not exceed $Kn + h + (m-h)10K \le 11Kmn$. If I_1 is a no-instance of MC, then the total costs of coloring all vertices of I_2 are greater than $10K^2$.

Suppose that I_1 is a yes-instance of MC. Then, there exist h subsets $S_{i_1}, \dots, S_{i_h}$ with $\cup_{j=1}^{h} S_{i_j} = \{1, \dots, n\}$. Let $\{i_{h+1}, \dots, i_m\} = \{1, \dots, m\} \setminus \{i_1, \dots, i_h\}$. Then, we can construct an optimum solution as given in Table 3.

Since $\{s_{i_\ell}, i + nj\} \notin E$ for $i \in S_{i_\ell}$ and $0 \le j \le K-1$, each color set forms an independent set. Moreover, we obtain a proper coloring, because the h subsets $S_{i_1}, \dots, S_{i_h}$ cover the set $\{1, \dots, n\}$. The total costs of this coloring are $Kn + h + (m-h)10K$.

Conversely, let f be a coloring of G with total costs bounded by $Kn + h + (m-h)10K$. Then, only the first m cheapest colors can be used; otherwise the total costs are at least $100K^2mn$. Since $\{s_1, \dots, s_m\}$ forms a clique in G, we have exactly m colors in f and all vertices in $\{s_1, \dots, s_m\}$ are colored differently. Let s_{i_j} be the vertex with $f(s_{i_j}) = j$ for $1 \le j \le m$.

Since the costs of f do not exceed $Kn+h+(m-h)10K$, the vertices $1, \dots, nK$ must be colored with colors with costs 1. Then, we can conclude that the corresponding sets $S_{i_1}, \dots, S_{i_h}$ cover $\{1, \dots, n\}$.

If we have no solution of the MC problem, then we have either more than m colors or a m-coloring of G. In the first case, we have total costs greater than $100K^2mn \ge 10K^2$. In the second case, the vertices $s_1, \dots, s_m$ must be colored differently. Furthermore, a vertex $i + nj$ can be colored with the same color as vertex s_{i_ℓ} only if $i \in S_{i_\ell}$. Since there exists no solution of the MC problem, at least K vertices $i + nj$ $(0 \le j \le K-1)$ are colored with a color with costs $10K$. Therefore, the total costs are greater than $10K^2$.

The **second part** of the proof is the specification of the parameter K. Let H be an approximation algorithm for the OCCP problem that computes a coloring with costs $H(I_2) \leq c|V|^{1-\epsilon}OPT(I_2)$. We construct for an instance I_1 of the MC problem an instance I_2 of the OCCP problem as described above. We obtain a split graph with $Kn + m \leq K(n+m)$ vertices.

If there exists a solution of the MC problem, the optimum solution of the OCCP instance I_2 has costs $Kn+h+(m-h)10K \leq 10K(n+m)$. In this case, our approximation algorithm H produces a coloring with $H(I_2) \leq 10K(n+m)c|V|^{1-\epsilon}$ colors. Since the number of vertices in I_2 is at most $K(n+m)$, the value $|V|^{1-\epsilon} \leq K^{1-\epsilon}(n+m)$. It follows that $H(I_2) \leq 10cK^{2-\epsilon}(n+m)^2$.

If there exists no solution of the MC problem, then $OPT(I_2) > 10K^2$ and, therefore, algorithm H generates a coloring with costs greater than $10K^2$. Again, we consider the inequality:

$$10K^2 > 10cK^{2-\epsilon}(n+m)^2$$

or equivalent

$$K^{\epsilon} > c(n+m)^2.$$

We define

$$K = \left\lceil \sqrt[\epsilon]{c(n+m)^2} \right\rceil + 1.$$

Since c and ϵ are constant, K is a polynomial in n and m and, therefore, the instance I_2 with this parameter K can be constructed in polynomial time. Hence, if there exists no solution of the MC problem, then H generates a solution with costs of at least

$$10K^2 > 10cK^{2-\epsilon}(n+m)^2 \geq 10cK(n+m)|V|^{1-\epsilon}.$$

Therefore, by using the polynomial time approximation algorithm H we could decide the existence of a solution for the MC problem, which would imply $P = NP$. □

Acknowledgement. I thank Thomas Erlebach (TU München) for his helpful comments and many fruitful discussions.

References

[1] E.M. Arkin and E.L. Silverberg, Scheduling jobs with fixed start and finish times. *Discrete Applied Mathematics* 18 (1987), 1–8.

[2] C. Berge and M. Las Vergnas, Sur un theoreme du type König pour hypergraphes, *International Conference on Combinatorial Mathematics*, Annals of the New York Academy of Sciences 175 (1970), 32–40.

[3] H.L. Bodlaender, K. Jansen and G. Woeginger, Scheduling with incompatible jobs, *Graph Theoretic Concepts in Computer Science*, LNCS, 657 (1992), 37-49.

[4] M.S. Chang, Y.H. Chen, G.J. Chang and J.H. Yan, Algorithmic aspects of the generalized clique-transversal problem on chordal graphs, *Discrete Applied Mathematics* 66 (1996), 189–203.

[5] W.H. Cunningham, *private communication.*

[6] V.R. Dondeti and H. Emmons, Job scheduling with processors of two types, *Operations Research* 40 (1992) S76–S85.

[7] M. Fischetti, S. Martello and P. Toth, The Fixed Job Schedule Problem with spread time constraints, *Operations Research* 6 (1987), 849–858.

[8] M. Fischetti, S. Martello and P. Toth, The Fixed Job Schedule Problem with working time constraints, *Operations Research* 3 (1989), 395–403.

[9] M. Fischetti, S. Martello and P. Toth, Approximation algorithms for Fixed Job Schedule Problems, *Operations Research* 40 (1992), S96–S108.

[10] A. Frank, On chain and antichain families of a partially ordered set, *Journal Combinatorial Theory* (B) 29 (1980), 176–184.

[11] M.R. Garey and D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completness, Freeman, San Francisco, 1979.

[12] F. Gavril, Algorithms for maximum k-colorings and k-coverings of transitive graphs, *Networks* 17 (1987), 465–470.

[13] M.C. Golumbic: Algorithmic graph theory and perfect graphs, Academic Press, New York, 1980.

[14] I. Heller, On linear systems with integral valued solutions, *Pacific Journal of Mathematics* 7 (1957), 1351–1364.

[15] K. Jansen, P. Scheffler and G.J. Woeginger, Maximum covering with D cliques, *Fundamentals of Computation Theory*, LNCS, 710 (1993) 319-328.

[16] R.M. Karp, Reducibility among combinatorial problems, in: Miller and Thatcher: Complexity of Computer Computations, Plenum Press (1972), 85–104.

[17] A.W.J. Kolen and L.G. Kroon, Licence Class Design: complexity and algorithms, *European Journal of Operational Research* 63 (1992), 432–444.

[18] A.W.J. Kolen and L.G. Kroon, Analysis of shift class design problems, *European Journal of Operational Research* 79 (1994), 417–430.

[19] L.G. Kroon, A. Sen, H. Deng and A. Roy, The optimal cost chromatic partition problem for trees and interval graphs, to appear in: *Workshop on Graph Theoretical Concepts in Computer Science* (1996).

[20] E. Kubicka and A. Schwenk, Introduction to chromatic sums, *Proc. ACM Computer Science Conference*, (1989).

[21] A. Sen, H. Deng and S. Guha, On a graph partition problem with an application to VLSI layout, *Information Processing Letters* 43 (1992), 87–94.

[22] P.D. Seymour, Decomposition of regular matroids, *Journal of Combinatorial Theory* B 28 (1980), 305–359.

[23] K.J. Supowit, Finding a maximum planar subset of a set of nets in a channel, *IEEE Transactions on Computer Aided Design*, CAD 6, 1 (1987) 93–94.

[24] K. Wagner: Monotonic coverings of finite sets, *Journal of Information Processing and Cybernetics - EIK* 20 (1984), 633 – 639.

[25] M. Yannakakis and F. Gavril, The maximum k-colorable subgraph problem for chordal graphs, *Information Processing Letters* 24 (1987), 133–137.

Fachbereich IV - Mathematik, Universität Trier, 54 286 Trier, Germany
E-mail address: jansen@dm3.uni-trier.de