On the Minimum Sum Coloring of P_4-sparse graphs

Flavia Bonomo[1], Mario Valencia-Pabon[2]

[1] IMAS-CONICET and Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina.
e-mail: `fbonomo@dc.uba.ar`.
[2] LIPN, Université Paris-Nord, 99 Av. J.-B. Clément, 93430 Villetaneuse, France.
e-mail: `valencia@lipn.univ-paris13.fr`

Abstract. In this paper, we study the Minimum Sum Coloring (MSC) problem on P_4-sparse graphs. In the MSC problem, we aim to assign natural numbers to vertices of a graph such that adjacent vertices get different numbers, and the sum of the numbers assigned to the vertices is minimum. Based in the concept of maximal sequence associated with an optimal solution of the MSC problem of any graph, we show that there is a large sub-family of P_4-sparse graphs for which the MSC problem can be solved in polynomial time. Moreover, we give a parameterized algorithm and a 2-approximation algorithm for the MSC problem on general P_4-sparse graphs.

Key words. graph coloring, minimum sum coloring, P_4-sparse graphs.

1. Introduction

In this paper, we study the *Minimum Sum Coloring (MSC)* problem for the family of P_4-sparse graphs. A *vertex coloring* of a graph $G = (V, E)$ is an assignment of colors to the vertices in V such that adjacent vertices receive different colors. We assume that the colors are positive integers. A vertex k-coloring of a graph G is a coloring such that the color of each vertex in V is taken from the set $\{1, 2, \ldots, k\}$. Given a vertex coloring of a graph G, the *sum* of the coloring is the sum of the colors assigned to the vertices. The *chromatic sum* $\Sigma(G)$ of G is the smallest sum that can be achieved by any proper coloring of G. In the *Minimum Sum Coloring* (MSC) problem we have to find a coloring of G with sum $\Sigma(G)$. The minimum number of colors needed in a minimum sum coloring of G is called the *strength* of G and is denoted by $s(G)$. Clearly, for any graph G we have $s(G) \geq \chi(G)$, where $\chi(G)$ denotes the chromatic number of G.

The MSC problem was introduced by Kubicka [12]. The problem is motivated by applications in scheduling [1,2,6] and VLSI design [14,16]. The computational complexity of determining the vertex chromatic sum of a simple graph has been extensively studied since then. In [13] it is shown that the problem is NP-hard in general, but polynomial time solvable for trees. The dynamic programming algorithm for trees can be extended to partial k-trees and block graphs [11]. Furthermore, the MSC problem is NP-hard even when restricted to some classes of graphs for which finding the chromatic number is easy,

---

such as bipartite or interval graphs [2, 16]. A number of approximability results for various classes of graphs were obtained in the last ten years [1, 5, 6, 4].

Jansen has shown in [11] that a more general optimization problem where each color has an integer cost, but this cost is not necessarily equal to the color itself, *the Optimal Cost Chromatic Partition (OCCP) problem*, can be solved in polynomial time for cographs (i.e. P_4-free graphs) and block graphs (i.e. diamond-free chordal graphs), but it remains NP-hard for permutation graphs. Salavatipour has shown in [15] that the OCCP problem can be solved in polynomial time for the family of P_4-reducible graphs, a superclass containing the family of cographs. P_4-reducible graphs were introduced by Jamison and Olariu [8] as a generalization of cographs: a graph is P_4-reducible if every vertex belongs to at most one P_4. A generalization of P_4-reducible graphs are the P_4-sparse graphs introduced in [7]. A graph is P_4-sparse if every 5-vertex subset contains at most one P_4. The family of P_4-sparse graphs can be recognized in linear time [9], and is a subclass of perfect graphs [7].

If G_1 and G_2 are two vertex disjoint graphs, then their *union* $G_1 \cup G_2$ is the graph with vertex set $V(G_1 \cup G_2) = V(G_1) \cup V(G_2)$ and edge set $E(G_1 \cup G_2) = E(G_1) \cup E(G_2)$. Similarly, their *join* $G_1 \vee G_2$ is the graph with $V(G_1 \vee G_2) = V(G_1) \cup V(G_2)$ and $E(G_1 \vee G_2) = E(G_1) \cup E(G_2) \cup \{(x, y) : x \in V(G_1), y \in V(G_2)\}$.

A *spider* is a graph whose vertex set can be partitioned into S, C and R, where $S = \{s_1, \ldots, s_k\}$ ($k \geq 2$) is an independent set; $C = \{c_1, \ldots, c_k\}$ is a complete set; s_i is adjacent to c_j if and only if $i = j$ (a *thin spider*), or s_i is adjacent to c_j if and only if $i \neq j$ (a *thick spider*); R is allowed to be empty and if it is not, then all the vertices in R are adjacent to all the vertices in C and non-adjacent to all the vertices in S. Clearly, the complement of a thin spider is a thick spider, and vice-versa. The triple (S, C, R) is called the *spider partition*, and can be found in linear time [9]. The sets S, C and R are called the *legs*, *body* and *head* of the spider, respectively. The *size* of the spider will be $|C|$. P_4-sparse graphs have a nice decomposition theorem as follows.

Theorem 1. [7, 10] *If G is a non-trivial P_4-sparse graph, then either G or $\overline{G}$ is not connected, or G is a spider.*

To each P_4-sparse graph G one can associate a corresponding decomposition rooted tree T in the following way. Each non-leaf node in the tree is labeled with either "$\cup$" (union-nodes), or "$\vee$" (join-nodes) or "SP" (spider-partition-nodes), and each leaf is labeled with a vertex of G. Each non-leaf node has two or more children. Let T_x be the subtree of T rooted at node x and let V_x be the set of vertices corresponding to the leaves in T_x. Then, each node x of the tree corresponds to the graph $G_x = (V_x, E_x)$. An union-node (join-node) corresponds to the disjoint union (join) of the P_4-sparse graphs associated with its children. A spider-partition-node corresponds to the spider with spider-partition (S, C, R) where S, C, and R are its children. Finally, the P_4-sparse graph associated with the root of the tree is just G, the P_4-sparse graph represented by this decomposition tree. The decomposition tree associated to a P_4-sparse graph can be computed in linear time [10].

The paper is organized as follows. In Section 2 we give some preliminaries concerning maximal sequences associated with optimal solutions of the MSC problem on any graph. In Section 3, we study the maximal sequences associated with optimal solutions of the MSC problem on P_4-sparse graphs. As a consequence, we obtain a parameterized algorithm for this problem and we show that there is a large sub-family of P_4-sparse graphs for which the MSC problem can be solved in polynomial time. In Section 4 we analyze the strength

of P_4-sparse graphs. Finally, in Section 5 we present a simple 2-approximation algorithm for the MSC problem on P_4-sparse graphs.

As far as we know, the results that we have obtained in this work are the first ones concerning the MSC problem on the family of P_4-sparse graphs. The computational complexity of the MSC problem on P_4-sparse graphs remains an open problem.

An extended abstract of this paper was presented at LAGOS'09 and appears in [3].

2. Maximal sequences and optimal solutions of the MSC problem

A k-coloring of a graph $G = (V, E)$ is a partition of the vertex set V into k independent sets $S_1, \ldots, S_k$, where each vertex in S_i is colored with color i, for $1 \leq i \leq k$. So, for any such k-partition of V into independent sets, we can associate a non-negative *sequence* p such that $p[i] = |S_i|$ for $i = 1, \ldots, k$ and $p[i] = 0$ for $i > k$. In the sequel, we deal with finite-support non-negative integer sequences only. Let $|p| = \max\{i : p[i] > 0\}$.

Definition 1. *Let p and q be two integer sequences. We say that p dominates q, denoted by $p \succeq q$, if for all $t \geq 1$ it holds that $\sum_{1 \leq i \leq t} p[i] \geq \sum_{1 \leq i \leq t} q[i]$.*

Definition 2. *Let p be a sequence. We denote by $\widetilde{p}$ the sequence that results from p when we order it in a non-increasing way.*

The following two lemmas are direct consequences of Definition 1.

Lemma 1. *The dominance relation $\succeq$ is a partial order.*

Lemma 2. *Let p be a sequence. Then, $\widetilde{p} \succeq p$.*

The following lemma will be very useful in order to study the MSC problem on graphs.

Lemma 3. *Let p and q be two sequences and let $n = \max\{|p|, |q|\}$. If $p \succeq q$ and $\sum_{1 \leq i \leq n} p[i] = \sum_{1 \leq i \leq n} q[i]$, then it holds that $\sum_{1 \leq i \leq n} i \cdot p[i] \leq \sum_{1 \leq i \leq n} i \cdot q[i]$.*

Proof. Let $N = \sum_{1 \leq i \leq n} p[i] = \sum_{1 \leq i \leq n} q[i]$. Let P and Q be two sequences obtained from p and q such that $|P| = |Q| = N$, and defined by $P[j] = \min\{k : \sum_{1 \leq i \leq k} p[i] \geq j\}$ (resp. $Q[j] = \min\{k : \sum_{1 \leq i \leq k} q[i] \geq j\}$) for $j = 1, \ldots, N$. By hypothesis, $p \succeq q$, and so, $P[j] \leq Q[j]$ for all $1 \leq j \leq N$. Therefore, $\sum_{1 \leq i \leq n} i \cdot p[i] = \sum_{1 \leq j \leq N} P[j] \leq \sum_{1 \leq j \leq N} Q[j] = \sum_{1 \leq i \leq n} i \cdot q[i]$. $\square$

Notice that if the sequences represent partitions of the vertex set of a graph into independent sets, where the value of the i-th element of the sequence represents the size of the i-th independent set in the partition, then for the sum-coloring problem on graphs we can restrict us to study maximal sequences w.r.t. the partial order $\succeq$. Notice also that maximal sequences are non-increasing sequences. We will call *maximal partition* to a partition of the vertex set of a graph into independent sets associated to a maximal sequence. In the following, we define some operations between sequences.

Definition 3. *Let p and q be two sequences. The join of p and q, denoted by $p \star q$, is the sequence that results by sorting the concatenation of the sequences p and q in non-increasing order.*

Definition 4. *Let p and q be two sequences. The sum of p and q, denoted by $p + q$, is the sequence whose i-th value is equal to $p[i]+q[i]$, for $i \geq 1$. Notice that $|p+q| = \max\{|p|, |q|\}$.*

Definition 5. *Let p and q be two sequences. We say that p and q are non-comparable, denoted by $p||q$, if $p \not\succeq q$ and $q \not\succeq p$.*

The following two lemmas will be useful in order to study the MSC problem on P_4-sparse graphs.

Lemma 4. *Let p, p' and q be sequences. If $\widetilde{p} \succeq \widetilde{p'}$ then $p \star q \succeq p' \star q$.*

Proof. Consider the sequence $p' \star q$. By definition of join, $\widetilde{p'}$ is a subsequence of $p' \star q$. Let s be the sequence that results from $p' \star q$ by replacing each element $\widetilde{p'}[i]$ by $\widetilde{p}[i]$. As by hypothesis, $\widetilde{p} \succeq \widetilde{p'}$, then we have that $s \succeq p' \star q$. But now, note that $p \star q = \widetilde{s}$ and thus, $p \star q \succeq s \succeq p' \star q$. $\square$

Lemma 5. *Let p, p', and q be sequences. Then, $p||p'$ if and only if $p + q||p' + q$.*

Proof. Note that $p||p'$ if and only if there exist two different positive integers j_1 and j_2 such that $\sum_{i=1}^{j_1} p[i] > \sum_{i=1}^{j_1} p'[i]$ and $\sum_{i=1}^{j_2} p[i] < \sum_{i=1}^{j_1} p'[i]$. This happens if and only if $\sum_{i=1}^{j_1}(p[i] + q[i]) > \sum_{i=1}^{j_1}(p'[i] + q[i])$ and $\sum_{i=1}^{j_2}(p[i] + q[i]) < \sum_{i=1}^{j_1}(p'[i] + q[i])$, which is equivalent to $p + q||p' + q$. $\square$

The following result can be proved similarly.

Lemma 6. *Let p, p', and q be sequences. Then, $p \succeq p'$ if and only if $p + q \succeq p' + q$.* $\square$

3. Maximal sequences of P_4-sparse graphs

In the sequel, sequences of a graph will represent partitions of its vertex set into independent sets. The following two lemmas show that if we are looking for maximal sequences of a graph that is either the union or the join of two vertex disjoint graphs G_1, G_2, then it is sufficient to consider maximal sequences of the graphs G_1 and G_2.

Lemma 7. *Let G_1, G_2 be two vertex disjoint graphs, and let $G = G_1 \cup G_2$. Then, every maximal sequence p of G can be expressed as $p = p_1 + p_2$, where p_1 (resp. p_2) is a maximal sequence of G_1 (resp. G_2).*

Proof. Let p be a maximal sequence of G and let $S_1, \ldots, S_k$ be a partition associated with p. Let $S_{i_1}, \ldots, S_{i_t}$ be the sets in the partition having nonempty intersection with $V(G_1)$. We claim that $\{i_1, \ldots, i_t\} = \{1, \ldots, t\}$. Otherwise, there is some value i such that $S_i \cap V(G_1) = \emptyset$ and $S_{i+1} \cap V(G_1) \neq \emptyset$. Since no vertex of G_1 is adjacent to a vertex of G_2, vertices in $S_{i+1} \cap V(G_1)$ can migrate to S_i, obtaining a sequence that strictly dominates p, a contradiction. The same happens for G_2, so p can be expressed as $p = p_1 + p_2$, where p_1 and p_2 are sequences of G_1 and G_2, respectively. By Lemma 6, they must be maximal for the corresponding graphs. $\square$

Lemma 8. *Let G_1, G_2 be two vertex disjoint graphs, and let $G = G_1 \vee G_2$. Then, every maximal sequence p of G can be expressed as $p = p_1 \star p_2$, where p_1 (resp. p_2) is a maximal sequence of G_1 (resp. G_2).*

Proof. Let p be a maximal sequence of G and let $S_1, \ldots, S_k$ be a partition associated with p. Since every vertex of G_1 is adjacent to all the vertices of G_2 in G, each S_i, $1 \leq i \leq k$, is entirely contained either in G_1 or in G_2. Besides, as p is maximal, it is non-increasing. So p can be expressed as $p = p_1 \star p_2$, where p_1 and p_2 are sequences of G_1 and G_2, respectively. By Lemma 4, they must be maximal for the corresponding graphs. $\square$

A similar result holds in general for homogeneous sets. Let G be a graph. A set $H \subseteq V(G)$ of vertices is called *homogeneous* if, for each vertex $w \in V(G) \setminus H$, either w is adjacent to all the vertices in H or to none of them. For any subset of vertices $X \subseteq V(G)$, denote by $G[X]$ the subgraph of G induced by X.

Lemma 9. *Let G be a graph and H an homogeneous set of G. Let $S_1, \ldots, S_k$ be a maximal partition of G, and let $S_{i_1}, \ldots, S_{i_t}$ be the sets in the partition having nonempty intersection with H. Then, $S_{i_1} \cap H, \ldots, S_{i_t} \cap H$ is a maximal partition of $G[H]$.*

Proof. Let p be the sequence of G associated to $S_1, \ldots, S_k$, that is, $p[i] = |S_i|$ for $i = 1, \ldots, k$, $p[i] = 0$ for $i > k$. Let q be the sequence of $G[H]$ associated to $S_{i_1} \cap H, \ldots, S_{i_t} \cap H$, that is, $q[j] = |S_{i_j} \cap H|$ for $j = 1, \ldots, t$, $q[j] = 0$ for $j > t$. Let q' be a maximal sequence for $G[H]$ such that $q' \succeq q$. Since $\sum_{j=1}^{|q'|} q'[j] = \sum_{j=1}^{|q|} q[j] = |H|$, we have $|q'| \leq |q| = t$. Let $S_1', \ldots, S_t'$ be a partition of H associated with q', where maybe some of the sets are empty. Notice that every vertex in $(S_{i_1} \cup \ldots \cup S_{i_t}) \setminus H$ has at least a non-neighbor in H, since $S_{i_1}, \ldots, S_{i_t}$ are the sets having nonempty intersection with H and they are independent sets. Since H is an homogeneous set of G, every vertex in $(S_{i_1} \cup \ldots \cup S_{i_t}) \setminus H$ has no neighbors in H. So we can consider the partition of $V(G)$ obtained from $S_1, \ldots, S_k$ by replacing S_{i_j} by $(S_{i_j} \setminus H) \cup S_j'$, for $j = 1, \ldots, t$, that is a partition of $V(G)$ into independent sets. Let p' be the sequence associated to this new partition. Then $p'[i] = p[i]$ for $i \notin \{i_1, \ldots, i_t\}$, while $p'[i_j] = p[i_j] - q[j] + q'[j]$ for $j = 1, \ldots, t$. It is easy to see that $p' \succeq p$ because $q' \succeq q$, and that the domination is strict for p' and p if it is strict for q' and q. Since p is maximal for G, it follows that $q' = q$ and q is maximal for $G[H]$. $\square$

We describe next the maximal sequences of spiders.

Lemma 10. *Let $G = (S, C, R)$ be a spider such that $R \neq \emptyset$, and let p be a maximal sequence for G. Then there exists a partition $S_1, \ldots, S_{|p|}$ associated with p in which there are only three kinds of sets: sets entirely contained in R, sets entirely contained in C, and sets intersecting both R and S; moreover, sets entirely contained in C are the last $|C|$ sets, and only S_1 intersects both R and S, with $S \subseteq S_1$.*

Proof. Let $S_1, \ldots, S_{|p|}$ be a partition associated with p. It is clear that no set of the partition can contain both vertices from C and R, since all the vertices in C are adjacent to all the vertices in R. If there is a set S_i containing vertices of S and no vertex of R, then either $S_i \subseteq S$ or S_i contains exactly one vertex of C, because C is a complete set. Since $R \neq \emptyset$, there is some set S_j containing vertices from R, thus $S_j \subseteq R \cup S$. Then the vertices in $S_i \cap S$ can migrate to S_j, possibly swapping S_i and S_j if $i < j$, thus obtaining a partition associated with a sequence p' such that $p' \succeq p$. Since p is maximal, $p' = p$. Therefore, there exists a partition associated with p in which every set is either entirely contained in R, or entirely contained in C, or intersects both R and S. Sets entirely contained in C have only one element, and since p is maximal, thus non-increasing, we may assume that these are the last $|C|$ sets. From now on, we will assume that $S_1, \ldots, S_{|p|}$ is such a partition. In particular, $S_1 \subseteq S \cup R$. Suppose that there is a set S_i, $i > 1$, containing vertices of S. Then the vertices in $S_i \cap S$ can migrate to S_1, obtaining a sequence that strictly dominates p, a contradiction. $\square$

Lemma 11. *Let $G = (S, C, R)$ be a spider such that $R \neq \emptyset$. Then, the number of maximal sequences of G is equal to the number of maximal sequences of $G[R]$. Moreover, for each maximal sequence q of $G[R]$ there exists only one maximal sequence q' of G with $|q'| =$*

$|q| + |C|$ and where $q'[1] = q[1] + |C|$, $q'[i] = q[i]$ for $2 \leq i \leq |q|$ (if $|q| \geq 2$), and $q'[i] = 1$ for $|q| + 1 \leq i \leq |q| + |C|$.

Proof. Let p be a maximal sequence of G, and $S_1, \ldots, S_{|p|}$ a maximal partition associated with p. Since $R \neq \emptyset$, by Lemma 10, we may assume that $S_1, \ldots, S_{|p|}$ is such that sets $S_{|p|-|C|+1}, \ldots, S_{|p|}$ are entirely contained in C, S_1 intersects both R and S and $S_2, \ldots, S_{|p|-|C|}$ (when $2 \leq |p| - |C|$) are entirely contained in R. By Lemma 9, $p[1] - |C|, p[2], \ldots, p[|p| - |C|]$ (or simply $p[1] - |C|$ when $|p| = |C| + 1$) is a maximal sequence for $G[R]$. Conversely, for each maximal sequence q of $G[R]$ associated with partition $T_1, \ldots, T_{|q|}$, define sequence q' of G with $|q'| = |q| + |C|$ and where $q'[1] = q[1] + |C|$, $q'[i] = q[i]$ for $2 \leq i \leq |q|$ (if $|q| \geq 2$), and $q'[i] = 1$ for $|q| + 1 \leq i \leq |q| + |C|$, associated with the partition $T_1 \cup S, \ldots, T_{|q|}, \{c_1\}, \ldots, \{c_k\}$ if $|q| \geq 2$, $T_1 \cup S, \{c_1\}, \ldots, \{c_k\}$ otherwise. Let q_1 and q_2 be maximal sequences of $G[R]$, and let q'_1 and q'_2 be their respective maximal sequences of G constructed as below. It is easy to see that if $q_1 || q_2$ then $q'_1 || q'_2$, so the lemma holds. □

Lemma 12. *Let $G = (S, C, R)$ be a thin spider such that $R = \emptyset$. Then, G has only one maximal sequence p, with $|p| = |C|$, where $p[1] = |C|$, $p[2] = 2$, and $p[i] = 1$ for $3 \leq i \leq |C|$.*

Proof. Let $S = \{s_1, \ldots, s_k\}$ and $C = \{c_1, \ldots, c_k\}$, with $k \geq 2$. Let $S_1, \ldots, S_t$ be a partition of the vertex set of G into independent sets, with $t \geq 1$, such that its associated sequence p is maximal. By hypothesis, we have that $R = \emptyset$. Note first that each vertex $c_i \in C$ must belong to a different independent set S_j and so, $t \geq k$. Now, by definition of a thin spider, each vertex s_i is adjacent to vertex c_j if and only if $i = j$. We claim that there is S_i such that $S_i = S$ or there are S_i and S_j, with $i \neq j$, such that $S_i = (S \setminus \{s_n\}) \cup \{c_n\}$ and $S_j = \{s_n, c_m\}$, for some $n, m \in \{1, \ldots, k\}$, with $m \neq n$. Assume that it is not true. Suppose first that there are three sets S_i, S_j, S_l, with $i < j < l$, such that each one of them contains at least one vertex of S. Let $s_q \in S$ be a vertex in S_l. Then vertex $c_q \in C$ belongs to at most one of S_i or S_j but not to both. Thus, vertex s_q must migrate to one of S_i or S_j that contains no vertex c_q, which gives a sequence that strictly dominates p, a contradiction. Therefore, vertices in S belong either to only one set S_i or to two different sets S_i and S_j, with $i < j$. If S_i contains no vertex of C, all the vertices in $S \cap S_j$ must migrate to S_i, which gives a sequence that strictly dominates p, a contradiction. Else, S_i contains exactly one vertex c_n of C. In that case, by similar arguments, only vertex s_n could be in S_j. Since p is maximal, S_j contains also a vertex in C. (Otherwise we can merge two sets, obtaining a sequence that strictly dominates p, a contradiction.) As p is maximal, then p is such that: (*i*) $p[1] = k$ and $p[i] = 1$ for $2 \leq i \leq k + 1$, that is, $S_1 = S$ and $S_j = \{c_{j-1}\}$ for $2 \leq j \leq k + 1$; or (*ii*) $p[1] = k$, $p[2] = 2$, and $p[i] = 1$ for $3 \leq i \leq k$, that is, sequence p is associated with the partition $S_1 = (S \setminus \{s_1\}) \cup \{c_1\}$, $S_2 = \{s_1, c_2\}$, and $S_l = \{c_l\}$ for $3 \leq l \leq k$. Clearly, the sequence of Case (*ii*) dominates the one of Case (*i*), and it is the only maximal sequence for G. □

Lemma 13. *Let $G = (S, C, R)$ be a thick spider such that $|C| \geq 3$ and $R = \emptyset$. Then, G has only two maximal sequences p_1 and p_2, with $|p_1| = |C|$ and $|p_2| = |C| + 1$, where $p_1[i] = 2$ for $1 \leq i \leq |C|$, and $p_2[1] = |C|$ and $p_2[i] = 1$ for $2 \leq i \leq |C| + 1$.*

Proof. Let $S = \{s_1, \ldots, s_k\}$ and $C = \{c_1, \ldots, c_k\}$, with $k \geq 3$. By hypothesis, we have that $R = \emptyset$. By definition of a thick spider, each vertex s_i is adjacent to vertex c_j if and only if $i \neq j$.

The sequence p_1 with $|p_1| = k$, and such that $p_1[i] = 2$ for $1 \le i \le k$, can be obtained by defining $S_i = \{c_i, s_i\}$ for $1 \le i \le k$. The sequence p_2 with $|p_2| = k + 1$, and such that $p_2[1] = k$ and $p_2[i] = 1$ for $2 \le i \le k + 1$, can be obtained by defining $S_1 = S$ and $S_i = \{c_{i-1}\}$ for $2 \le i \le k+1$. Moreover, we have that $p_1 || p_2$. In fact, let $j_1 = 1$ and $j_2 = k$. Then, $p_1[1] = 2 < k = p_2[1]$ and $\sum_{i=1}^{j_2} p_1[i] = 2k > 2k - 1 = \sum_{i=1}^{j_2} p_2[i]$.

We will show now that these are the only two maximal sequences for G. Let $S_1, \ldots, S_t$ be a partition of the vertex set of G into independent sets, with $t \ge 1$, such that its associated sequence p is maximal. First notice that there is at most one set entirely contained in S, because two such sets could be merged obtaining a sequence that strictly dominates p, a contradiction.

Suppose first that there is some set S_i containing more than one vertex of S. Since no two vertices of S have a common non-neighbor in C, then $S_i \subseteq S$ and it is the only set entirely contained in S. Every other set is either contained in C or has one vertex of C and one of S. Since p is non-increasing, we may assume $i = 1$. If some set S_i with $i > 1$ contains a vertex of S, it can migrate to S_1, leading to a sequence that strictly dominates p, a contradiction. So $S_1 = S$ and $p = p_2$.

Suppose now that no set contains more than one vertex of S. Then each set is either composed of one vertex of C, or of one vertex of S, or of a vertex $s_i \in S$ and its only non-adjacent vertex $c_i \in C$. Clearly, p_1 dominates every such a sequence, so $p = p_1$. $\square$

Notice also that the trivial graph has only one maximal sequence p, with $|p| = 1$, where $p[1] = 1$. Therefore, we have the following theorems.

Theorem 2. *Let G be a P_4-sparse graph such that in its modular decomposition there are no thick spiders (S, C, R) with $|C| \ge 3$ and $R = \emptyset$. Then,*

1. *$s(G) = \chi(G)$, G has a unique maximal sequence, and $\Sigma(G)$ and an optimal coloring of G can be computed from its modular decomposition in polynomial time.*
2. *In such an optimal coloring, each set S_i is a maximum independent set of $G \setminus \bigcup_{1 \le j < i} S_j$ which verifies $\chi(G \setminus \bigcup_{1 \le j \le i} S_j) = \chi(G \setminus \bigcup_{1 \le j < i} S_j) - 1$.*

Proof. Let G be a P_4-sparse graph such that in its modular decomposition there are no thick spiders (S, C, R) with $|C| \ge 3$ and $R = \emptyset$.

1. We will prove by induction that G admits a unique maximal sequence p and that $|p| = \chi(G)$. This implies $s(G) = \chi(G)$. By Theorem 1, G is either trivial, or the union or join of two smaller P_4-sparse graphs G_1 and G_2, or G is a spider (S, C, R) and $G[R]$ is P_4-sparse. If G is trivial, the property holds. Suppose G is the union or join of G_1 and G_2. By inductive hypothesis, for $i = 1, 2$, G_i has a unique maximal sequence p_i, and $|p_i| = \chi(G_i)$. If $G = G_1 \cup G_2$ then, by Lemma 7, G has a unique maximal sequence $p = p_1 + p_2$. Therefore, $|p| = \max\{|p_1|, |p_2|\} = \max\{\chi(G_1), \chi(G_2)\} = \chi(G)$. If $G = G_1 \vee G_2$ then, by Lemma 8, G has a unique maximal sequence $p = p_1 \star p_2$. Therefore, $|p| = |p_1| + |p_2| = \chi(G_1) + \chi(G_2) = \chi(G)$. Finally, if G is a spider (S, C, R), then either G is a thin spider with $R = \emptyset$ or $R \ne \emptyset$. In the first case the property follows by Lemma 12. In the second case, by inductive hypothesis, $G[R]$ has a unique maximal sequence q, and $|q| = \chi(G[R])$. By Lemma 11, there exists only one maximal sequence p of G, and $|p| = |q| + |C| = \chi(G[R]) + |C| = \chi(G)$.

   Now, let n be the number of vertices in G. Let T be the decomposition rooted tree associated with G. It is well known that T can be computed in linear time [10]. We will show that the unique maximal sequence p of G and a partition associated with p

can be computed from T in polynomial time. In order to compute an optimal coloring with $s(G)$ colors and sum $\Sigma(G)$ for this case, we proceed from the leaves to the root in T as follows. If x is a leaf in T then its associated partition is $S_1 = \{x\}$ having as maximal sequence p, with $|p| = 1$ and $p[1] = 1$. If node $x \in T$ is a union-node (resp. join-node) then, by Lemma 7 (resp. Lemma 8), the unique maximal sequence and its corresponding optimal partition of the vertex set of G_x into independent sets can be computed from the unique maximal sequences and their corresponding optimal partitions of the children of x. If node $x \in T$ is a spider-partition node representing the spider $\sigma = (S, C, R)$ then, the unique maximal sequence and its corresponding optimal partition of the vertex set of G_x into independent sets can be computed either as in Lemma 12 (if σ is a thin spider with $R = \emptyset$) or from the unique maximal sequence and their corresponding optimal partitions of the child $G_x[R]$ of x as shown in Lemma 11, if $R \neq \emptyset$. Finally, notice that each node $x \in T$ needs $O(n)$ time to compute its optimal partition. As there are at most $2n - 1$ nodes in T, then the complexity time of the algorithm is bounded by $O(n^2)$.

2. It follows by induction from Theorem 1, and using Lemma 7 (resp. Lemma 8) if G is a disjoint union (resp. join) of P_4-sparse graphs, and Lemma 12 (resp. Lemma 11) if G is a thin spider (S, C, R) with $R = \emptyset$ (resp. G is a spider (S, C, R) with $R \neq \emptyset$).

$\square$

Theorem 3. *Let G be a P_4-sparse graph on n vertices. Let t be the number of thick spiders (S, C, R) with $|C| \geq 3$ and $R = \emptyset$ in the modular decomposition of G. Then, $s(G) \leq \chi(G) + t$, the number of maximal sequences of G is at most 2^t, and an optimal coloring of G can be computed in $2^t P(n)$ time, where $P(n)$ is a polynomial in n.*

Proof. The statement holds for $t = 0$ by Theorem 2. Suppose $t \geq 1$, and let $\sigma^1, \ldots, \sigma^t$ be the thick spiders in the decomposition tree T of G such that $\sigma_j = (S^j, C^j, \emptyset)$ and $|C^j| \geq 3$, for $j = 1, \ldots, t$. By Lemma 13, each σ^j has exactly two maximal sequences. Clearly, there are 2^t ways of choosing maximal sequences (and their corresponding partitions) for the t thick spiders σ^j. Now, given a fixed choice for the thick spiders σ^j and by using the algorithm in the proof of item (1) of Theorem 2, we can compute in $O(n^2)$ time a maximal sequence and its corresponding partition into independent sets for G. This shows that G has at most 2^t maximal sequences and that an optimal coloring with $s(G)$ colors and sum $\Sigma(G)$ can be computed in $O(2^t n^2)$ time. Finally, note that for each thick spider σ^j, one of its maximal sequences has length $\chi(\sigma^j) + 1$ and thus, by induction, it can be proved that the number of colors used in an optimal solution for G is upper bounded by $\chi(G) + t$. $\square$

The algorithm described in Theorem 3 allows us to find also the minimum sum that can be attained by a coloring of G with at most r colors, for some given value $r \geq \chi(G)$, and the corresponding coloring.

4. The strength of P_4-sparse graphs

As we have shown in the last section, the difference between the strength of a P_4-sparse graph and its chromatic number is upper bounded by the number of thick spiders of size at least three and empty head.

We will exhibit now some examples were this bound is tight, where it is not, and were it can be seen that Lemma 9 does not hold if we replace "maximal sequence" by "sequence achieving the minimum sum".

Let G_1 be the join of t disjoint thick spiders of size 3 with empty head. The possible maximal sequences for each of them are $[3, 1, 1, 1]$ and $[2, 2, 2]$. Since the graph is symmetric, suppose that x of the spiders are partitioned $[3, 1, 1, 1]$ and $t - x$ are partitioned $[2, 2, 2]$, where $0 \leq x \leq t$. By Lemma 8, the corresponding maximal sequence of G_1 will be

$$[\underbrace{3, \ldots, 3}_{x}, \underbrace{2, \ldots, 2}_{3(t-x)}, \underbrace{1, \ldots, 1}_{3x}]$$

So, the sum $\Sigma(x)$ of the obtained coloring is a function of x, more precisely, $\Sigma(x) = \sum_{i=1}^{3t+x} i + \sum_{i=1}^{3t-2x} i + \sum_{i=1}^{x} i = 3x^2 - 3tx + 9t^2 + 3t$. Since it is a convex quadratic function, the minimum is attained only by the value of x where $\Sigma'(x) = 6x - 3t = 0$, that is, for $x = \frac{t}{2}$. Since we are looking for the minimum attained by an integer value, it will be $x = \frac{t}{2}$ if t is even, and either $x = \lfloor \frac{t}{2} \rfloor$ or $x = \lceil \frac{t}{2} \rceil$ if t is odd. So, $s(G_1) = \chi(G_1) + \lfloor \frac{t}{2} \rfloor$.

Let now $G_2 = G_1 \vee K_{3t}$, that is, the join of G_1 and a complete graph on $3t$ vertices, and let x denote again the number of spiders with partitions $[3, 1, 1, 1]$. By Lemma 8, the corresponding maximal sequence of G_2 will be

$$[\underbrace{3, \ldots, 3}_{x}, \underbrace{2, \ldots, 2}_{3(t-x)}, \underbrace{1, \ldots, 1}_{3x+3t}]$$

Now, $\Sigma(x) = \sum_{i=1}^{6t+x} i + \sum_{i=1}^{3t-2x} i + \sum_{i=1}^{x} i = 3x^2 + \frac{45}{2}t^2 + \frac{9}{2}t$ and the minimum is attained only by the value of x where $\Sigma'(x) = 6x = 0$, that is, for $x = 0$. So, $s(G_2) = \chi(G_2)$.

Finally, let $G_3 = G_1 \vee \overline{3tK_2}$, where $\overline{3tK_2}$ is the join of $3t$ disjoint independent sets of size 2, and let x as above. By Lemma 8, the corresponding maximal sequence of G_3 will be

$$[\underbrace{3, \ldots, 3}_{x}, \underbrace{2, \ldots, 2}_{3(t-x)+3t}, \underbrace{1, \ldots, 1}_{3x}]$$

Now, $\Sigma(x) = \sum_{i=1}^{6t+x} i + \sum_{i=1}^{6t-2x} i + \sum_{i=1}^{x} i = 3x^2 - 6tx + 36t^2 + 6t$ and the minimum is attained only by the value of x where $\Sigma'(x) = 6x - 6t = 0$, that is, for $x = t$. So, $s(G_3) = \chi(G_3) + t$, and in this case the bound of Theorem 3 is tight.

5. A 2-approximation algorithm for the MSC problem on P_4-sparse graphs

Let G be a P_4-sparse graph on n vertices. Let $\sigma^1, \sigma^2, \ldots, \sigma^t$ be the thick spiders in the decomposition tree of G, such that $\sigma^j = (S^j, C^j, \emptyset)$ and $|C^j| \geq 3$, for $j = 1, \ldots, t$. We assume that t is of order $\Omega(n)$, otherwise, by Theorem 3, an optimal solution for the MSC problem on G can be computed in polynomial time.

Consider the following algorithm to color G: first, for each thick spider σ^j, we choose as its maximal sequence p^j the sequence p_1 of Lemma 13, that is, $p^j[i] = 2$ for $1 \leq i \leq |C^j|$, and its corresponding maximal partition $S_1^j, \ldots, S_{|C^j|}^j$, where $S_i^j = \{s_i^j, c_i^j\}$, being s_i^j and c_i^j non-adjacent vertices in S^j and C^j, respectively, for $i = 1, \ldots, |C^j|$. Next, we apply the algorithm in the proof of item (1) of Theorem 2 in order to compute in $O(n^2)$ time a

partition into independent sets for G. Let ϕ be the coloring of the vertices of G obtained by the previous algorithm. Clearly, ϕ uses $\chi(G)$ colors. Let $\Sigma_\phi(G)$ be the sum of the colors of the vertices of G induced by the coloring ϕ. We claim the following.

Claim 4 $\Sigma_\phi(G) \leq 2\Sigma(G)$.

Proof. Let H be the induced subgraph of G obtained by removing from G all the vertices of each independent set S^j of the thick spider σ^j, for $1 \leq j \leq t$. Let Φ be an optimal coloring of H with sum $\Sigma(H)$. We extend the coloring Φ of H to a coloring Φ' of G, where each vertex $s_i^j \in S^j$ is assigned the color $\Phi'(c_i^j)$ of vertex $c_i^j \in C^j$, for $1 \leq i \leq |C^j|$, being $(S^j, C^j, \emptyset)$ the thick spider σ^j, for $1 \leq j \leq t$. Let $\Sigma_{\Phi'}(G)$ be the sum induced by the coloring Φ' on G. Clearly, $\Sigma_{\Phi'}(G) \leq 2\Sigma(H)$. On the other hand, since H is a subgraph of G, $\Sigma(H) \leq \Sigma(G)$ and, therefore, $\Sigma_{\Phi'}(G) = 2\Sigma(G)$. □

Hence, we have the following result.

Theorem 5. *There is a 2-approximation algorithm for the MSC problem on P_4-sparse graphs.*

Finally, notice that all our results concerning the MSC problem on P_4-sparse graphs can be easily adapted to the OCCP problem on this family of graphs.

Acknowledgments

References

1. A. Bar-Noy, M. Bellare, M. M. Halldórsson, H. Shachnai, and T. Tamir, *On chromatic sums and distributed resource allocation*, Information and Computation, 140 (1998), pp. 183–202.

2. A. Bar-Noy and G. Kortsarz, *Minimum color sum of bipartite graphs*, Journal of Algorithms, 28 (1998), pp. 339–365.

3. F. Bonomo and M. Valencia-Pabon, *Minimum sum coloring of P_4-sparse graphs*, in Proc. V Latin-American Algorithms, Graphs and Optimization Symposium (LAGOS), vol. 35 of Electronic Notes in Discrete Mathematics, Elsevier, 2009, pp. 293–298.

4. U. Feige, L. Lovász, and P. Tetali, *Approximating min sum set cover*, Algorithmica, 40 (2004), pp. 219–234.

5. K. Giaro, R. Janczewski, M. Kubale, and M. Malafiejski, *A 27/26-approximation algorithm for the chromatic sum coloring of bipartite graphs*, in Proc. 5th International Workshop on Approximation Algorithms for Combinatorial Optimization (APPROX), vol. 2462 of Lecture Notes in Computer Science, Springer-Verlag, 2002, pp. 135–145.

6. M. M. HALLDÓRSSON, G. KORTSARZ, , AND H. SHACHNAI, *Sum coloring interval and k-claw free graphs with application to scheduling dependent jobs*, Algorithmica, 37 (2003), pp. 187–209.

7. C. T. HOÀNG, *Perfect graphs*, PhD thesis, School of Computer Science, McGill University, 1985.

8. B. JAMISON AND S. OLARIU, *P_4-reducible graphs – a class of uniquely tree-representable graphs*, Discrete Mathematics, 51 (1984), pp. 35–39.

9. B. JAMISON AND S. OLARIU, *Recognizing P_4-sparse graphs in linear time*, SIAM Journal on Computing, 21 (1992), pp. 381–406.

10. B. JAMISON AND S. OLARIU, *A tree representation for P_4-sparse graphs*, Discrete Applied Mathematics, 35 (1992), pp. 115–129.

11. K. JANSEN, *Complexity results for the optimum cost chromatic partition problem*, in Proc. 24th International Colloquium on Automata, Languages and Programming (ICALP), vol. 1256 of Lecture Notes in Computer Science, Springer-Verlag, 1997, pp. 727–737.

12. E. KUBICKA, *The Chromatic Sum of a Graph*, PhD thesis, Western Michigan University, 1989.

13. E. KUBICKA AND A. J. SCHWENK, *An introduction to chromatic sums*, in Proc. 17th ACM Annual Computer Science Conference, 1989, pp. 39–45.

14. S. NICOLOSO, M. SARRAFZADEH, AND X. SONG, *On the sum coloring problem on interval graphs*, Algorithmica, 23 (1999), pp. 109–126.

15. M. SALAVATIPOUR, *On sum coloring of graphs*, Master's thesis, Department of Computer Science, University of Toronto, 2000.

16. T. SZKALICZKI, *Routing with minimum wire length in the dogleg-free Manhattan model is NP-complete*, SIAM Journal on Computing, 29 (1999), pp. 274–287.