



# A one-to-one correspondence between potential solutions of the cluster deletion problem and the minimum sum coloring problem, and its application to $P_4$ -sparse graphs <sup>☆</sup>

Flavia Bonomo <sup>a</sup>, Guillermo Duran <sup>b,c</sup>, Amedeo Napoli <sup>d</sup>,  
Mario Valencia-Pabon <sup>e,\*</sup>,<sup>1</sup>

<sup>a</sup> CONICET and Dep. de Computación, FCEN, Universidad de Buenos Aires, Argentina

<sup>b</sup> CONICET and Dep. de Matemática and Instituto de Cálculo, FCEN, Universidad de Buenos Aires, Argentina

<sup>c</sup> Dep. de Ingeniería Industrial, FCFM, Universidad de Chile, Santiago, Chile

<sup>d</sup> LORIA, CNRS, Inria Nancy Grand Est, Université de Lorraine, Vandoeuvre-les-Nancy, France

<sup>e</sup> Université Paris-13, Sorbonne Paris Cité LIPN, CNRS UMR7030, Villetaneuse, France

## ARTICLE INFO

### Article history:

Received 11 May 2014

Accepted 8 February 2015

Available online 13 February 2015

Communicated by R. Uehara

### Keywords:

Cluster deletion

Sum coloring

$P_4$ -sparse

Graph algorithms

## ABSTRACT

In this note we show a one-to-one correspondence between potentially optimal solutions to the cluster deletion problem in a graph  $G$  and potentially optimal solutions for the minimum sum coloring problem in  $\bar{G}$  (i.e. the complement graph of  $G$ ). We apply this correspondence to polynomially solve the cluster deletion problem in a subclass of  $P_4$ -sparse graphs that strictly includes  $P_4$ -reducible graphs.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

A *cluster graph* is a graph in which every connected component is a clique (i.e. a complete subgraph). Cluster graphs have been used in a variety of applications whenever clustering of objects is studied or when consistent data is sought among noisy or error-prone data [1]. The *cluster deletion* problem asks for the minimum num-

ber of edges that can be removed from an input graph to make the resulting graph a cluster graph. There exist several results for the cluster deletion problem (see for example [3,14,12] and references therein). The cluster deletion problem is known to be NP-complete [14]. Recently, Gao et al. [6] have shown that the greedy algorithm that finds iteratively maximum cliques, gives an optimal solution for the class of graphs known as *cographs*. It implies that the cluster deletion problem is polynomial-time solvable on cographs.

A *vertex coloring* of a graph is an assignment of positive integers to the vertices of the graph such that adjacent vertices receive different integers. The *sum* of a vertex coloring of a graph is the sum of the integers assigned to the vertices. The *minimum sum coloring* problem asks for the smallest sum that can be achieved by any vertex coloring of an input graph. The minimum sum coloring problem is motivated by applications in scheduling [2,7] and VLSI design [15]. In [13] it is shown that the problem is NP-hard

<sup>☆</sup> Partially supported by UBACyT Grant 20020100100980, CONICET PIP 112-200901-00178 and 112-201201-00450CO and ANPCyT PICT 2012-1324 (Argentina), Fondo Nacional de Desarrollo Científico y Tecnológico Grant 1140787 (Chile), and MathAmSud Project 13MATH-07 (Argentina–Brazil–Chile–France).

\* Corresponding author.

E-mail addresses: fbonomo@dc.uba.ar (F. Bonomo), gduran@dm.uba.ar (G. Duran), amedeo.napoli@loria.fr (A. Napoli), valencia@lipn.univ-paris13.fr (M. Valencia-Pabon).

<sup>1</sup> Currently visiting the INRIA Nancy, Grand Est, 2013–2014.

in general, but polynomial time solvable for trees. The dynamic programming algorithm for trees can be extended to partial  $k$ -trees, block graphs and cographs [8]. Recently, Bonomo and Valencia-Pabon [4,5] have shown that the minimum sum coloring problem can be solved in polynomial time on a wide subclass of  $P_4$ -sparse graphs.

A graph is  $P_4$ -sparse if every 5-vertex subset contains at most one  $P_4$ . The family of  $P_4$ -sparse graphs generalize the family of cographs (i.e.  $P_4$ -free graphs) and they can be recognized in linear time [10].

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  $S = \{s_1, \dots, s_k\}$  ( $k \geq 2$ ) is an independent set;  $C = \{c_1, \dots, c_k\}$  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 [10]. 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 1.1.** (See [11].) *If  $G$  is a non-trivial  $P_4$ -sparse graph, then either  $G$  or  $\bar{G}$  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  $G_x = (V_x, E_x)$ . A 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  $G[S]$ ,  $G[C]$ , and  $G[R]$  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 [11].

**2. Maximal sequences and optimal solutions**

The following approach was used by Bonomo and Valencia-Pabon [4,5] in order to deal with the minimum sum coloring (MSC) problem on  $P_4$ -sparse graphs. A  $k$ -coloring of a graph  $G = (V, E)$  is a partition of the vertex set  $V$  into  $k$  independent sets  $S_1, \dots, S_k$ , 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  $i = 1, \dots, k$  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 2.1.** 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  $\sum_{1 \leq i \leq t} p[i] \geq \sum_{1 \leq i \leq t} q[i]$ .

Let  $p$  be a sequence. We denote by  $\tilde{p}$  the sequence that results from  $p$  when we order it in a non-increasing way. Clearly,  $\tilde{p} \succeq p$ .

The following lemma is the key to study the MSC problem on graphs.

**Lemma 2.2.** (See Lemma 3 in [5].) *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  $\sum_{1 \leq i \leq n} i \cdot p[i] \leq \sum_{1 \leq i \leq n} i \cdot q[i]$ .*

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  $\succeq$ .

A similar approach has been used by Gao et al. [6] in order to deal with the cluster deletion problem on cographs. In fact, notice that an optimal solution of the cluster deletion problem in a graph  $G$  is a partition of the vertex set  $V$  into cliques  $M_1, \dots, M_t$ . So, for any partition of  $V$  into  $t$  cliques, we can associate a sequence  $p$  such that  $p[i] = |M_i|$  for  $i = 1, \dots, t$  and  $p[i] = 0$  for  $i > t$ . Notice that if  $M = (|M_1|, \dots, |M_t|)$  is an integer sequence associate to a partition into cliques of the set  $V$ , such that  $|M_1| \geq |M_2| \geq \dots \geq |M_t|$ , then  $M$  is an integer partition of the integer  $|V|$ . Gao et al. [6] show the following result.

**Theorem 2.3.** (See Theorem 4 in [6].) *Let  $p$  and  $q$  be two integer partitions of some positive integer  $n$ , with  $p \neq q$ . If  $p \succeq q$ , then  $|p| \leq |q|$ , and  $\sum_{i=1}^{|p|} \binom{|p|}{i} > \sum_{i=1}^{|q|} \binom{|q|}{i}$ .*

By Theorem 2.3, we can also restrict us to study maximal sequences w.r.t. the partial order  $\succeq$  in order to solve the cluster deletion problem on graphs. Notice also that maximal sequences for both problems (MSC and cluster deletion) are non-increasing sequences.

Clearly, there is a one-to-one correspondence between the maximal sequences corresponding to partitions of the vertex set of  $G$  into cliques and the maximal sequences corresponding to partitions of the vertex set of  $\bar{G}$  into independent sets. Thus, by Theorem 2.3 and Lemma 2.2, there is a one-to-one correspondence between the potentially optimal solutions for the cluster deletion problem of a graph  $G$  and the potentially optimal solutions for the minimum sum coloring of the complement graph  $\bar{G}$ .

Nevertheless, when a graph  $G$  has more than one maximal sequence corresponding to partitions of its vertex set into cliques, the optimal solution for the cluster deletion problem of  $G$  and the optimal solution for the minimum

sum coloring of its complement  $\bar{G}$  may not correspond to the same sequence. We will provide an example of this in the next section.

This suggests that the computational complexity of the cluster deletion problem on a graph class  $\mathcal{G}$  and the computational complexity of the minimum sum coloring problem on the class  $\text{co-}\mathcal{G}$  (the class of complement graphs of graphs in  $\mathcal{G}$ ) can be different. But, to the best of our knowledge, there are not known examples of this behaviour.

### 3. Maximal sequences in $P_4$ -sparse graphs

The following results can be obtained directly from the definitions of union and join of graphs, and from the definition of a spider graph (see Section 1).

**Lemma 3.1.** *Let  $G_1$  and  $G_2$  be graphs. Then,  $\overline{G_1 \cup G_2} = \overline{G_1} \vee \overline{G_2}$  and  $\overline{G_1 \vee G_2} = \overline{G_1} \cup \overline{G_2}$ .*

**Lemma 3.2.** *Let  $SP$  be a spider and  $(S, C, R)$  its spider partition. If  $SP$  is a thin (resp. thick) spider then,  $\overline{SP}$  is a thick (resp. thin) spider with partition  $(C, S, R)$ .*

By Lemmas 3.1 and 3.2 we have that if  $G$  is a  $P_4$ -sparse graph, then the complement graph  $\bar{G}$  is also a  $P_4$ -sparse graph. In fact, by using the tree decomposition of these graphs, we can deduce that union-nodes (resp. join-nodes) in  $G$  correspond to a join-nodes (resp. union-nodes) in  $\bar{G}$ , and that thin spiders (resp. thick spiders) in  $G$  correspond to thick spiders (resp. thin spiders) in  $\bar{G}$ .

The following operations between sequences were defined in [4,5]: Let  $p$  and  $q$  be two sequences.

- The *join* of  $p$  and  $q$ , denoted by  $p \star q$ , is the sequence that results by ordering in a non-increasing way the concatenation of sequences  $p$  and  $q$ .
- The *sum* of  $p$  and  $q$ , denoted by  $p + q$ , is the sequence such that its  $i$ -th value is equal to  $p[i] + q[i]$ , for  $i \geq 1$ . Notice that  $|p + q| = \max\{|p|, |q|\}$ .
- $p$  and  $q$  are *non-comparable*, denoted by  $p \parallel q$ , if  $p \not\leq q$  and  $q \not\leq p$ .

Moreover, in [4,5] the following two lemmas have been obtained.

**Lemma 3.3.** (See Lemma 4 in [5].) *Let  $p, p'$  and  $q$  be sequences. If  $\tilde{p} \geq \tilde{p}'$  then  $p \star q \geq p' \star q$ .*

**Lemma 3.4.** (See Lemma 6 in [5].) *Let  $p, p'$  and  $q$  be sequences. Then,  $p \geq p'$  if and only if  $p + q \geq p' + q$ .*

In the sequel, maximal sequences of a graph  $G$  will correspond to potential solutions of the cluster deletion problem on  $G$ .

**Lemma 3.5.** *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  $p = p_1 + p_2$ , where  $p_i$  is a maximal sequence of  $G_i$ , for  $i = 1, 2$ .*

**Proof.** The results follows by induction on  $|G|$ , by Lemma 3.1 on  $\bar{G}$ , and by Lemma 7 in [5] concerning maximal sequences for the minimum sum coloring of the union of two graphs.  $\square$

**Lemma 3.6.** *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  $p = p_1 \star p_2$ , where  $p_i$  is a maximal sequence of  $G_i$ , for  $i = 1, 2$ .*

**Proof.** The results follows by induction on  $|G|$ , by Lemma 3.1 on  $\bar{G}$ , and by Lemma 8 in [5] concerning maximal sequences for the minimum sum coloring of the join of two graphs.  $\square$

**Lemma 3.7.** *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  $|q'| = |q| + |S|$  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  $|q| + 1 \leq i \leq |q| + |S|$ .*

**Proof.** Notice that the spider  $G$  is a  $P_4$ -sparse graph and so,  $G[R]$  (i.e. the subgraph of  $G$  induced by  $R$ ) is also a  $P_4$ -sparse graph. Then, the result follows by induction on  $|G|$ , by Lemma 3.2 on  $\bar{G}$ , and by Lemma 11 in [5] concerning maximal sequences for the minimum sum coloring of spiders.  $\square$

**Lemma 3.8.** *Let  $G = (S, C, R)$  be a thick 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.** The results follows by Lemma 3.2 and by Lemma 12 in [5] concerning maximal sequences for the minimum sum coloring of thin spiders without head.  $\square$

**Lemma 3.9.** *Let  $G = (S, C, R)$  be a thin 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.** The results follows by Lemma 3.2 and by Lemma 13 in [5] concerning maximal sequences for the minimum sum coloring of thick spiders without head.  $\square$

Notice also that the trivial graph has only one maximal sequence  $p$ , with  $|p| = 1$ , where  $p[1] = 1$ . Therefore, we have the following two theorems whose proofs are similar to the ones of Theorem 2 and Theorem 3 in [5], respectively.

**Theorem 3.10.** *Let  $G$  be a  $P_4$ -sparse graph such that in its modular decomposition there are no thin spiders  $(S, C, R)$  with  $|C| \geq 3$  and  $R = \emptyset$ . Then,*

1.  $G$  has a unique maximal sequence and an optimal solution for the cluster deletion problem of  $G$  can be computed from its modular decomposition in polynomial time.

2. In such an optimal solution, each  $M_i$  is a maximum clique of  $G \setminus \bigcup_{1 \leq j < i} M_j$ .

**Theorem 3.11.** Let  $G$  be a  $P_4$ -sparse graph on  $n$  vertices. Let  $t$  be the number of thin spiders  $(S, C, R)$  with  $|C| \geq 3$  and  $R = \emptyset$  in the modular decomposition of  $G$ . Then, the number of maximal sequences of  $G$  is at most  $2^t$ , and an optimal solution for the cluster deletion problem of  $G$  can be computed in  $2^t P(n)$  time, where  $P(n)$  is a polynomial in  $n$ .

Notice that a cograph is a  $P_4$ -sparse graph without spiders. Moreover,  $P_4$ -sparse graphs having only spiders whose spider partition  $(S, C, R)$  is of size equal to 2 (i.e.  $|S| = |C| = 2$ ) are known as  $P_4$ -reducible graphs. The class of  $P_4$ -reducible graphs was introduced by Jamison and Olariu [9] as a generalization of cographs: a graph is  $P_4$ -reducible if every vertex belongs to at most one  $P_4$ . Finally, notice that thin spiders with size 2 are isomorphic to thick spiders with size 2. Therefore, by Theorem 3.10, we have the following corollary which generalizes the result by Gao et al. [6].

**Corollary 3.12.** Let  $G$  be a  $P_4$ -reducible graph. Then,

1.  $G$  has a unique maximal sequence and an optimal solution for the cluster deletion problem of  $G$  can be computed from its modular decomposition in polynomial time.
2. In such an optimal solution, each  $M_i$  is a maximum clique of  $G \setminus \bigcup_{1 \leq j < i} M_j$ .

Finally, we provide an example of a  $P_4$ -sparse graph  $G$  having more than one maximal sequence and such that the optimal solution for the cluster deletion problem of  $G$  and the optimal solution for the minimum sum coloring of its complement  $\bar{G}$  do not correspond to the same sequence.

Let  $N_3$  be a thin spider of size 3 with no head (also known as *net*), and let  $G = (N_3 \cup N_3) \vee (K_1 \cup K_1)$ . By the theorems above,  $G$  has three different maximal sequences, namely  $p_1 = [3, 3, 2, 2, 2, 2]$ ,  $p_2 = [4, 3, 2, 2, 1, 1, 1]$ , and

$p_3 = [4, 4, 1, 1, 1, 1, 1]$ . The internal edges in the partition into cliques induced by each maximal sequence are, respectively, 10, 11 and 12. So, the number of removed edges are, respectively, 26, 25 and 24, being  $p_3$  the sequence realizing the optimal solution. As for  $\bar{G}$ , the sums of the colorings induced by each maximal sequence are, respectively, 45, 42 and 45, being in this case  $p_2$  the sequence realizing the optimal solution.

## References

- [1] N. Bansal, A. Blum, S. Chawla, Correlation clustering, *Mach. Learn.* 56 (1–3) (2004) 89–113.
- [2] A. Bar-Noy, M. Bellare, M.M. Halldórsson, H. Shachnai, T. Tamir, On chromatic sums and distributed resource allocation, *Inf. Comput.* 140 (1998) 183–202.
- [3] S. Böcker, P. Damaschke, Even faster parametrized cluster deletion and cluster editing, *Inf. Process. Lett.* 111 (2011) 717–721.
- [4] F. Bonomo, M. Valencia-Pabon, Minimum sum coloring of  $P_4$ -sparse graphs, *Electron. Notes Discrete Math.* 35 (2009) 293–298.
- [5] F. Bonomo, M. Valencia-Pabon, On the minimum sum coloring of  $P_4$ -sparse graphs, *Graphs Comb.* 30 (2) (2014) 303–314.
- [6] Y. Gao, D.R. Hare, J. Nastos, The cluster deletion problem for cographs, *Discrete Math.* 313 (2013) 2763–2771.
- [7] M.M. Halldórsson, G. Kortsarz, H. Shachnai, Sum coloring interval and  $k$ -claw free graphs with application to scheduling dependent jobs, *Algorithmica* 37 (2003) 187–209.
- [8] K. Jansen, Complexity results for the optimum cost chromatic partition problem, in: *Proc. 24th ICALP*, in: *Lect. Notes Comput. Sci.*, vol. 1256, 1997, pp. 727–737.
- [9] B. Jamison, S. Olariu,  $P_4$ -reducible graphs – a class of uniquely tree-representable graphs, *Discrete Math.* 51 (1984) 35–39.
- [10] B. Jamison, S. Olariu, Recognizing  $P_4$ -sparse graphs in linear time, *SIAM J. Comput.* 21 (1992) 381–406.
- [11] B. Jamison, S. Olariu, A tree representation for  $P_4$ -sparse graphs, *Discrete Appl. Math.* 35 (1992) 115–129.
- [12] C. Komusiewicz, J. Uhlmann, Cluster editing with locally bounded modifications, *Discrete Appl. Math.* 160 (15) (2012) 2259–2270.
- [13] E. Kubicka, A.J. Schwenk, An introduction to chromatic sums, in: *Proc. 17th ACM Annual Computer Science Conference*, 1989, pp. 39–45.
- [14] R. Shamir, R. Sharan, D. Tsur, Cluster graph modification problems, *Discrete Appl. Math.* 144 (1–2) (2004) 173–182.
- [15] T. Szkaliczki, Routing with minimum wire length in the dogleg-free Manhattan model is NP-complete, *SIAM J. Comput.* 29 (1999) 274–287.