

On partitioning interval graphs into proper interval subgraphs and related problems

Frédéric Gardi

Bouygues e-lab, 40 rue Washington, 75008 Paris, France

Abstract

In this paper, we establish that any interval graph (resp. circular-arc graph) with n vertices admits a partition into at most $\lceil \log_3 n \rceil$ (resp. $\lceil \log_3 n \rceil + 1$) proper interval subgraphs, for $n > 1$. The proof is constructive and provides an efficient algorithm to compute such a partition. On the other hand, this bound is shown to be asymptotically sharp for an infinite family of interval graphs. In addition, some results are derived for related problems.

Key words: interval graphs, circular-arc graphs, proper interval graphs, graph partitioning problems, efficient graph algorithms

1991 MSC: 05C70, 05C35, 68R10

1 Introduction

A graph $G = (V, E)$ is an *interval graph* if to each vertex $v \in V$ can be associated an open interval I_v of the real line, such that two distinct vertices $u, v \in V$ are adjacent if and only if $I_u \cap I_v \neq \emptyset$. The family $\{I_v\}_{v \in V}$ is an *interval representation* of G (see Fig. 1). The *left and right endpoints* of the interval I_v are respectively denoted $l(I_v)$ and $r(I_v)$. The class of interval graphs coincide with the intersection of the classes of chordal graphs and of complements of comparability graphs (cf. [14]). A graph is chordal if it contains no induced cycle of length greater than or equal to four; chordal graphs are also known as the intersection graphs of subtrees in a tree (cf. [14]). Comparability graphs are the transitively orientable graphs, they correspond to graphs of partial orders.

Email address: fgardi@bouygues.com (Frédéric Gardi).

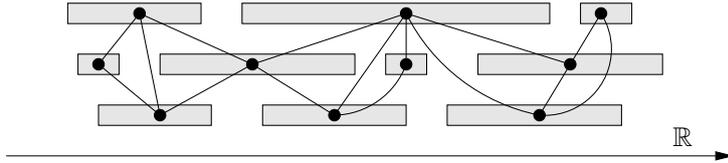


Fig. 1. An interval graph and its representation.

Circular-arc graphs are the intersection graphs of arcs on a circle (see Fig. 2). A circular-arc graph $G = (V, E)$ admits a *circular-arc representation* $\{A_v\}_{v \in V}$ in which each arc A_v is defined by its counterclockwise endpoint and its clockwise endpoint. Note that a circular-arc representation of a graph G which fails to cover some point p on the circle is topologically the same as an interval representation of G [14]. Thus, every interval graph is a circular-arc graph.

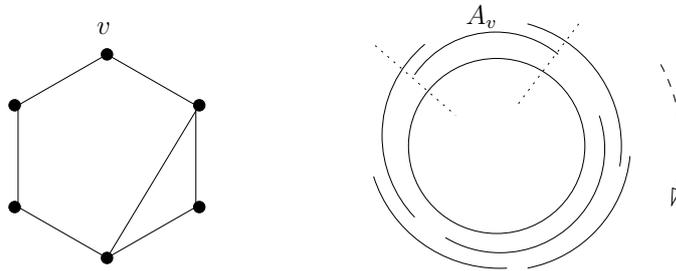


Fig. 2. A circular-arc graph and its representation around the circle.

A graph G is a *proper interval graph* if there is an interval representation of G in which no interval properly contains another (see Fig. 3). *Unit interval graphs* are the graphs having an interval representation in which all the intervals have the same length. *Claw-free interval graphs* are the interval graphs without induced copy of the claw $K_{1,3}$ (the tree composed of one central vertex and three leaves). Roberts (cf. [14, pp. 187–188], see also [11] for a short constructive proof) has shown that the class of proper interval graphs coincide with the classes of unit interval graphs and claw-free interval graphs.

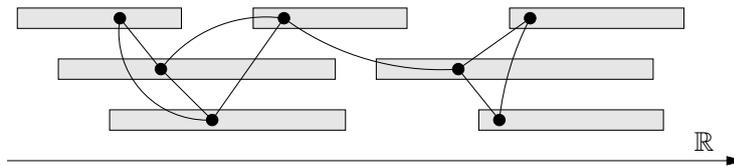


Fig. 3. A proper interval graph and its representation.

1.1 Main results and motivations

Interval graphs and circular-arc graphs have been intensively studied for several decades by both mathematicians and computer scientists. These two

classes of graphs are particularly known for providing numerous models in diverse areas like genetics, psychology, sociology, archaeology, or scheduling. For more details on these graphs and their applications, the reader can consult the books by Roberts [18,19], Golubic [14], or Fishburn [7].

In this paper, the problem of *partitioning interval graphs into proper interval subgraphs* is investigated. Bounds on the size of a minimum partition of an interval graph into proper interval subgraphs have been given in a previous author's paper [8]. Here sharper upper bounds are established, leading to the following theorem. Note that floor and ceiling functions are crucial in the statement of the theorem, which may be seemingly counterintuitive.

Theorem 1.1 *Any interval graph (resp. circular-arc graph) with n vertices and m edges admits a partition into at most $\lceil \log_3 n \rceil$ (resp. $\lceil \log_3 n \rceil + 1$) proper interval subgraphs, for $n > 1$. Moreover, such a partition is computed in $O(n \log n + m)$ time and linear space. On the other hand, for each n , an interval graph exists which admits no partition into less than $\lfloor \log_3(2n + 1) \rfloor$ proper interval subgraphs.*

In addition, some bounds are derived for the related problem of finding the largest proper interval subgraph in an interval graph. Some complexity issues are also discussed.

This theorem could find applications in the design of approximation algorithms for hard optimization problems restricted to interval graphs or circular-arc graphs, since many problems untractable for these graphs become efficiently solvable for proper interval graphs. For example, given a graph G and a positive integer k , the mutual exclusion scheduling problem [1] consists in determining a minimum coloring of G such that each color appears at most k times. The mutual exclusion scheduling problem is NP-hard for interval graphs even if k is a constant greater than or equal to four [2], whereas it is solvable in linear time and space for proper interval graphs [13]. In this case, the above theorem yields immediately an efficient $\lceil \log_3 n \rceil$ -approximation algorithm for the mutual exclusion scheduling problem for interval graphs (partition the interval graph into $\lceil \log_3 n \rceil$ proper interval subgraphs and solve optimally the mutual exclusion scheduling problem for each subgraph). An application of these results in the area of workforce scheduling, which has actually inspired this research, is described in previous author's papers [8,12,13].

This theorem could also be interpreted in order-theoretic terms as follows: any interval order with n elements admits a partition into at most $\lceil \log_3 n \rceil$ subposets isomorphic to unit interval orders (also known as semiorders in mathematical psychology). The interested reader is referred to [7] for more details on interval orders and semiorders.

All the results presented in this paper appear in the author's thesis [9], written

in French, and have been announced in [10].

1.2 Preliminaries

All the graphs considered throughout the paper are undirected, unless explicit notice is made to the contrary. The number of vertices and the number of edges of the graph $G = (V, E)$ are respectively denoted by n and m throughout the paper. A *complete set* or *clique* is a subset of pairwise adjacent vertices. The clique C is *maximum* if no other clique of the graph has a size strictly greater than the one of C ; the clique number $\omega(G)$ denotes the size of a maximum clique in G . On the other hand, an *independent set* or *stable* is a subset of pairwise non-adjacent vertices and the *stability* $\alpha(G)$ denotes the size of a maximum stable in G . A q -*coloring* of the graph G corresponds to a partition of G into q stables and the chromatic number corresponds to the size of a minimum coloring of G .

The graph $K_{1,t}$, called *t-star*, is a tree composed of one central vertex and t leaves. $K_{1,3}$ is also called *claw*. A graph is $K_{1,t}$ -free if it contains no graph $K_{1,t}$ as induced subgraph. An interval representation I_1, \dots, I_n whose intervals are ordered according to nondecreasing left (resp. right) endpoints is a $<_l$ -ordered (resp. $<_r$ -ordered) *interval representation*. Besides, we write $I_i \prec I_j$ if $r(I_i) \leq l(I_j)$ (the relation \prec induces a partial order on the intervals). For the sake of brevity, we denote by $l(C_j)$ (resp. $r(C_j)$) the largest left endpoint (resp. smallest right endpoint) of an interval in a clique C_j ; $l(C_j)$ (resp. $r(C_j)$) is called the left endpoint (resp. right endpoint) of the clique C_j . In this way, we say that an interval contains the clique C_j if it contains the portion $]l(C_j), r(C_j)[$ of the line.

A graph $G = (V, E)$ is a *split graph* if its vertices can be partitioned into two sets S and C such that S induce a stable and C a clique. By analogy with bipartite graphs, split graphs are denoted by $G = (S, C, E)$. A graph $G = (V, E)$ is a *threshold graph* if to each vertex $v \in V$ can be associated a positive integer a_v such that $X \subseteq V$ is a stable if and only if $\sum_{x \in X} a_x \leq t$ with t being an integer (called the threshold). The vertices of a threshold graph can be partitioned into a clique $C = C_1 \cup \dots \cup C_r$ and a stable $S = S_1 \cup \dots \cup S_r$ (with all C_i, S_i not empty except S_r) such that a vertex of S_i is adjacent to a vertex of $C_{i'}$ if and only if $i' > i$ for any $i, i' \in \{1, \dots, r\}$. Thus, every threshold graph is a split graph. Threshold graphs form also a subclass of interval graphs [14], as shown on Fig. 4.

All interval graphs are perfect, which is not true for circular-arc graphs (see [3,14]). Interval graphs, circular-arc graphs, proper interval graphs, and thresh-

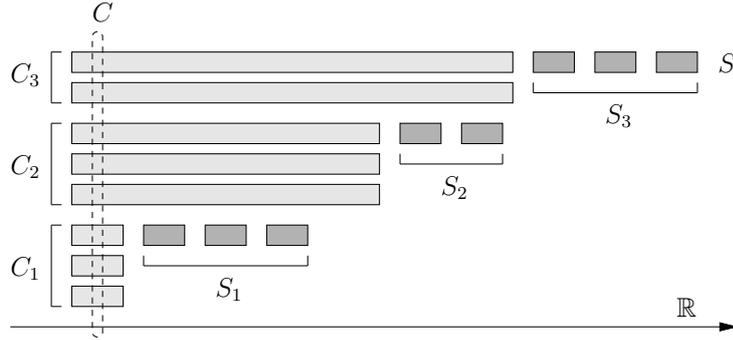


Fig. 4. An interval representation of a threshold graph.

old graphs are recognized in linear time and space (see [3,5,6,14,16,17]).

All the graph-theoretical terms which are not defined here can be found in [3,14].

2 Proof of the theorem

In this section, the different parts of the theorem announced in the introduction are established. We start by proving the lower bound, previously stated in [8] (minor flaws appearing in the original proof are corrected). Let H_k be a k -partite graph whose interval representation is built by defining recursively the stables S_1, \dots, S_k as follows. The stable S_1 consists of only one open interval of length 3^{k-1} . For $i = 2, \dots, k$, the stable S_i is obtained by copying the stable S_{i-1} and subdividing each one of its intervals into three open intervals of equal length (see Fig. 5 for an example of construction). The set of stables S_1, \dots, S_k resulting of this construction induces a k -partite graph with $n = \sum_{i=1}^k 3^{i-1} = (3^k - 1)/2$ vertices.

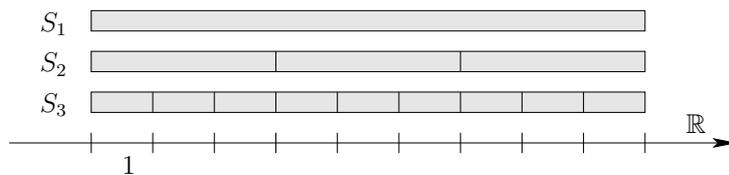


Fig. 5. An interval representation of the graph H_3 .

Lemma 2.1 *For every $k \geq 1$, the k -partite graph H_k with $n = (3^k - 1)/2$ vertices admits no partition into less than $k = \log_3(2n + 1)$ proper interval subgraphs.*

Proof. Since each stable induces trivially a proper interval subgraph, H_k admits immediately a partition into k proper interval subgraphs. Using in-

duction, we show that this partition is of minimum cardinality. Denote by $\xi(H_k)$ the size of a minimum partition of H_k into proper interval subgraphs. One can easily observe that $\xi(H_1) = 1$ and $\xi(H_2) = 2$. Now, assume that $\xi(H_{i-1}) = i - 1$. To demonstrate that $\xi(H_i) = i$, suppose the contrary and consider a partition of H_i into $i - 1$ sets $\mathcal{I}_1, \dots, \mathcal{I}_{i-1}$ of intervals, inducing each one a proper interval subgraph.

Consider without loss of generality that the sole interval $I^* \in S_1$ belongs to the set \mathcal{I}_1 . Then, we claim that the intervals of the set $\mathcal{I}_1 \setminus \{I^*\}$ induce at most two disjoint cliques. Indeed, the contrary implies the existence of $K_{1,3}$ as induced subgraph in \mathcal{I} (with I^* as central vertex and one interval of each disjoint clique as leaves). This claim implies that at least one interval of S_2 as well as all the intervals coming from its subdivision in S_3, \dots, S_i do not belong to the set \mathcal{I}_1 . Now, these intervals induce a copy of the graph H_{i-1} , which requires $i - 1$ proper interval subgraphs to be partitioned according to the induction hypothesis. Consequently, a contradiction is observed, since the only $i - 2$ sets $\mathcal{I}_2, \dots, \mathcal{I}_{i-1}$ are available to realize this partition. This allows to conclude that $\xi(H_i) = i$ for all $i > 2$, which completes the proof by induction.

Finally, the equality $n = (3^k - 1)/2$ yields $\xi(H_k) = \log_3(2n + 1)$. \square

Proposition 2.2 *For every $n \geq 1$, an interval graph with n vertices exists which admits no partition into less than $\lfloor \log_3(2n + 1) \rfloor$ proper interval subgraphs. For every $t > 1$, a $K_{1,t}$ -free interval graph with at least $\lfloor (3t - 4)/2 \rfloor$ vertices exists which admits no partition into less than $\lfloor \log_3(3t - 3) \rfloor$ proper interval subgraphs.*

Proof. Define the graph H_k with k the largest integer such that $n \geq (3^k - 1)/2$. Then, add to this graph $n - (3^k - 1)/2$ isolated vertices. According to the previous lemma, this graph admits no partition into less than $\lfloor \log_3(2n + 1) \rfloor$ proper interval subgraphs.

To prove the second assertion, observe that the graph H_k is $K_{1,t}$ -free for $t = 3^{k-1} + 1$. Having expressed n and $\xi(H_k)$ in function of t , we obtain that H_k contains at least $\lfloor (3t - 4)/2 \rfloor$ vertices and admits no partition into less than $\lfloor \log_3(3t - 3) \rfloor$ proper interval subgraphs, for any value of $k \geq 1$. \square

Now, new results are given which improve the upper bound established in [8]. The next lemma is central in the proof of the logarithmic upper bound. This lemma requires the following classical linear-time and space algorithm for partitioning interval graphs into cliques (see for example [15]).

Algorithm CANONICAL-PARTITION-INTO-CLIQUEs;
Input: an interval graph $G = (V, E)$;

Output: a canonical partition of G into cliques;
Begin;
 compute a $<_l$ -ordered interval representation I_1, \dots, I_n of G ;
 $i \leftarrow 1, j \leftarrow 0$;
while $i \leq n$ **do**
 $C_j \leftarrow \{I_i\}, r_{\text{clique}} \leftarrow r(I_i), i \leftarrow i + 1$;
 while $i \leq n$ and $l(I_i) \leq r_{\text{clique}}$ **do**
 $C_j \leftarrow C_j \cup \{I_i\}$;
 if $r(I_i) < r_{\text{clique}}$ **then** $r_{\text{clique}} \leftarrow r(I_i)$;
 $i \leftarrow i + 1$;
 $j \leftarrow j + 1$;
return $\{C_0, \dots, C_{j-1}\}$;
End;

Computing an ordered interval representation of G is done in linear time and space [16] (see also [3,14]). Then, the computation of cliques takes $O(n)$ time and space. These cliques form a partition of the interval graph G into cliques. By picking the interval having the smallest right endpoint in each clique C_j , we obtain a maximum stable of G , which implies that this partition is minimum [15]. Thus, such a partition into cliques shall be called *canonical* throughout the paper.

Lemma 2.3 *Any $K_{1,t}$ -free interval graph admits a partition into*

$$\left\lceil \log_3 \frac{3t-3}{2} \right\rceil$$

proper interval subgraphs, for $t > 1$. Moreover, such a partition is computed in $O(n \log t + m)$ time and linear space.

Proof. Here is described the algorithm computing such a partition. Recall that an interval contains a clique C_j if it contains the portion $]l(C_j), r(C_j)[$ of the line, where $l(C_j)$ (resp. $r(C_j)$) is the largest left endpoint (resp. smallest right endpoint) of an interval in C_j . Synthetically, having computed a canonical partition into cliques, the algorithm colors these cliques in a logarithmic fashion with the set of colors $\{0, \dots, i^*\}$; the output is the partition of G induced by these $i^* + 1$ colors.

Algorithm COLOR-CLIQUE-LOGARITHMIC;

Input: a $K_{1,t}$ -free interval graph $G = (V, E)$ with $t > 1$;

Output: a partition of G into $\lceil \log_3((3t-3)/2) \rceil$ proper interval subgraphs;

Begin;

 compute a $<_l$ -ordered interval representation I_1, \dots, I_n of G ;
 $\{C_0, \dots, C_{q-1}\} \leftarrow \text{CANONICAL-PARTITION-INTO-CLIQUE}(G)$;
 $i^* \leftarrow \lceil \log_3((t-1)/2) \rceil, C_0 \leftarrow \dots \leftarrow C_{i^*} \leftarrow \emptyset, i \leftarrow i^*$;
 while $i \geq 0$ **do**

```

for  $j$  from 0 to  $q - 1$  by step of  $3^i$  do
    let  $\mathcal{U}$  be the set of all unmarked intervals containing the clique  $C_j$ ;
    mark and add to  $\mathcal{C}_i$  all intervals in  $\mathcal{U}$ ;
     $i \leftarrow i - 1$ ;
return  $\mathcal{C}_0, \dots, \mathcal{C}_{i^*}$ ;
End;

```

First, the complexity of the algorithm is analyzed. Computing an ordered interval representation of G is done in linear time and space [16] (see also [3,14]). Then, a canonical partition of G into cliques is obtained in $O(n)$ time and space [15]. Searching the unmarked intervals containing the cliques C_j takes $O(n)$ time and space too. Indeed, the left endpoints of these intervals are necessarily in the portion $]r(C_{j'}), r(C_j)]$ of the line, where $C_{j'}$ is the clique examined previously. Consequently, only one sweep of the intervals in the order $<_l$ suffices to realize the search. Since the while loop is repeated $i^* + 1$ times, that is $\lceil \log_3((t-1)/2) \rceil + 1$ times, the complete algorithm runs in $O(n \log t + m)$ time and linear space.

Then, the correctness of the algorithm is demonstrated. At each step j of the inner loop, the unmarked intervals which are included in the set \mathcal{C}_i form a clique (because all these intervals contain the clique C_j). Thus, each set \mathcal{C}_i ($0 \leq i \leq i^* - 1$) is a union of cliques. Now, consider three cliques C_u, C_v, C_w included in \mathcal{C}_i , with $u < v < w$. We claim that at least one clique C_k with $u < k < w$ is included in $\mathcal{C}_{i'}$ with $i' > i$ (in other words, three consecutive cliques included in \mathcal{C}_i are always separated by a clique previously included in $\mathcal{C}_{i'}$ with $i' > i$). Indeed, the cliques included in \mathcal{C}_i correspond to the set of unmarked intervals containing C_j for $j = \beta \cdot 3^i$ with $\beta \geq 0$. But according to the algorithm, the cliques C_j such that β is a multiple of three have been previously included in a set $\mathcal{C}_{i'}$ with $i' > i$.

From this claim, we deduce that each set of $\mathcal{C}_0, \dots, \mathcal{C}_{i^*-1}$ induces a proper interval graph. Assume that an interval I_a and three intervals $I_b \prec I_c \prec I_d$ induce a copy of $K_{1,3}$ in the set \mathcal{C}_i (with $0 \leq i \leq i^* - 1$). Clearly, the three intervals I_b, I_c, I_d belong to three different cliques in \mathcal{C}_i and the interval I_a must contain these three cliques. Now, according to the previous claim, I_a also contains one clique included in $\mathcal{C}_{i'}$ with $i' > i$. Since the latter was formed before the set \mathcal{C}_i according to the algorithm, it should contain the interval I_a , which is in contradiction with the hypothesis.

Finally, we show that the existence of an induced copy of $K_{1,3}$ in the set \mathcal{C}_{i^*} implies the existence of an induced copy of $K_{1,t}$ in the graph G , which is a contradiction.

Let $\mathcal{C}_{i^*} = \{C_1^{i^*}, \dots, C_r^{i^*}\}$ be the set of cliques which are included into the set \mathcal{C}_{i^*} , in the order of their extraction by the algorithm (see Fig. 6). If $r \leq 2$, then

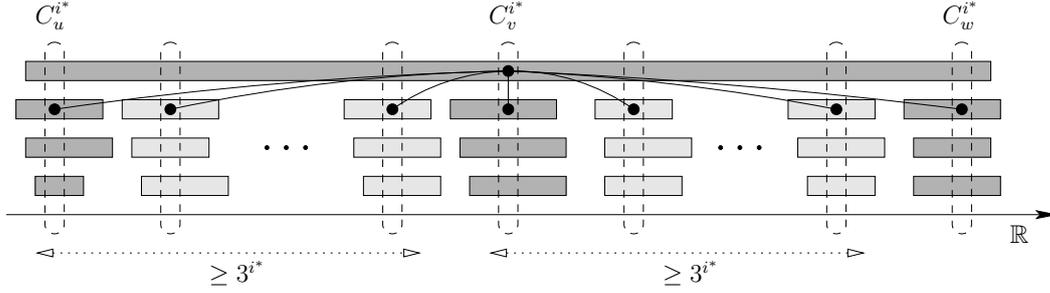


Fig. 6. The proof of $K_{1,3} \Rightarrow K_{1,t}$ for the set \mathcal{C}_{i^*} .

\mathcal{C}_{i^*} is trivially claw-free. Otherwise, suppose that \mathcal{C}_{i^*} contains an induced copy of $K_{1,3}$ with I_a as central vertex and $I_b \prec I_c \prec I_d$ its three leaves. Clearly, these leaves belong to disjoint cliques and we can set $I_b \in C_u^{i^*}$, $I_c \in C_v^{i^*}$ and $I_d \in C_w^{i^*}$ with $1 \leq u < v < w \leq r$. Since the cliques of \mathcal{C}_{i^*} have been extracted from the left to the right by the algorithm CANONICAL-PARTITION-INTO-CLIQUEs, the interval I_a belongs necessarily to the clique $C_u^{i^*}$.

Now, select in each clique of the canonical partition between $C_u^{i^*}$ and $C_w^{i^*}$ the interval having the smallest right endpoint and add it to the set S , initially empty. We claim that S induces a stable of size at least $t' = 2 \cdot 3^{i^*} + 1$. Indeed, two intervals of S cannot overlap according to the algorithm CANONICAL-PARTITION-INTO-CLIQUEs. At least 3^{i^*} cliques appear in the canonical partition between $C_u^{i^*}$ (included) and $C_v^{i^*}$ (excluded), and still at least 3^{i^*} between $C_v^{i^*}$ (included) and $C_w^{i^*}$ (excluded). Consequently, the set S contains at least t' elements and the claim is demonstrated.

Since the interval $I_a \in C_u^{i^*}$ overlaps the interval $I_d \in C_w^{i^*}$, this one overlaps all the intervals of S too, except maybe the last one (according to the $<_l$ order) which belongs to the clique $C_w^{i^*}$. Thus, this last interval is replaced by I_d which cannot overlap the penultimate interval of S (otherwise, I_d would not belong to $C_w^{i^*}$). The resulting stable S has a size $t' = 2 \cdot 3^{i^*} + 1 \geq t$, for all $t > 1$. Then, we obtain that at least t pairwise disjoint intervals are intersected by I_a , contradicting the initial assumption that G is $K_{1,t}$ -free. Consequently, the set \mathcal{C}_{i^*} induces a proper interval graph too, which completes the proof. \square

Since a graph with n vertices is trivially $K_{1,n}$ -free, Lemma 2.3 provides immediately the following general upper bound.

Proposition 2.4 *Any interval graph admits a partition into at most*

$$g_0(n) = \left\lceil \log_3 \frac{3n-3}{2} \right\rceil$$

proper interval subgraphs, for $n > 1$. Moreover, this partition is computed in $O(n \log n + m)$ time and linear space.

According to Propositions 2.2 and 2.4, the gap between the lower bound and the upper bound is at most one (resp. two) for interval graphs (resp. circular-arc graphs). Nevertheless, a direct application of the algorithm COLOR-CLIQUE-LOGARITHMIC may lead to unsatisfactory results for some instances. An example of such instances is the star $K_{1,n-1}$: whereas the minimum partition into proper interval subgraphs is trivially of size two, the algorithm returns a partition of cardinal $\lceil \log_3((3n-3)/2) \rceil$. In fact, all graphs having a chromatic number in $O(1)$ and having an induced copy of $K_{1,t}$ with $t = \Omega(n)$ are pathological for the algorithm COLOR-CLIQUE-LOGARITHMIC. The following corollary helps to further refine the upper bound. Remind that $\alpha(G)$ denotes the size of a maximum stable in a graph G .

Corollary 2.5 *Let G be an interval graph. If $1 \leq \alpha(G) < t$, then G admits a partition into at most $g_0(t)$ proper interval subgraphs. If $t \leq \alpha(G) < n - 1$, then G admits a partition into at most $g_0(n - t) + 1$ proper interval subgraphs. Moreover, in both cases, the partition is computed in $O(n \log t + m)$ time and linear space.*

Proof. The first assertion follows immediately from Lemma 2.3, since any graph is trivially $K_{1,t}$ -free for $t = \alpha(G) + 1$. The second assertion is proved as follows. First, extract a maximum stable from G in linear time and space [15]. Since $\alpha(G) \geq t$, at most $n - t$ intervals remain in G (with $n - t > 1$). Then, partition these remaining intervals into at most $g_0(n - t)$ proper interval subgraphs according to Lemma 2.3. Since a stable induces trivially a proper interval subgraph, the desired bound is obtained. \square

Remark 2.6 *Here is another formulation of the previous corollary. Let G be an interval graph satisfying one of the two following conditions, with $t = O(1)$: (i) G contains no induced copy of $K_{1,t}$ (or no stable of size greater than t); (ii) G contains an induced copy of $K_{1,n-t}$ (or a stable of size greater than $n - t$). Then, G admits a partition into $O(1)$ proper interval subgraphs, which is computed in linear time and space.*

By joining the two assertions of the previous corollary, we obtain that any interval graph admits a partition into at most $g_1(n, t) = \max\{g_0(t), g_0(n - t) + 1\}$, for any value of $t \in \{2, \dots, n - 2\}$. Now, we can search a minimum of the function $g_1(n, t)$ along the dimension t . Since the functions $g_0(t)$ and $g_0(n - t) + 1$ are respectively monotonically increasing and monotonically decreasing, a lower bound of the minimum is given by the solution of the equation $\tilde{g}_0(t) = \tilde{g}_0(n - t) + 1$ for $2 \leq t \leq n - 2$ with $\tilde{g}_0(t) = \log_3((3t - 3)/2)$. A straightforward calculation yields $\tilde{t} = (3n - 2)/4$ as solution to this equation. Then, a minimum of $g_1(n, t)$ is obtained by taking the minimum among the two values $g_0(\lceil \tilde{t} \rceil)$ and $g_0(n - \lfloor \tilde{t} \rfloor) + 1$, which can be bounded as follows:

$$g_0(\lceil \tilde{t} \rceil) \leq \lceil \log_3((9n - 6)/8) \rceil$$

$$g_0(n - \lfloor \tilde{t} \rfloor) + 1 \leq \lceil \log_3((9n + 18)/8) \rceil$$

Finally, the best bound is $g_1(n) = \lceil \log_3((9n - 6)/8) \rceil$, reached for the value $t = \lceil (3n - 2)/4 \rceil$ and $n \geq 6$ (but the bound can be extended until $n > 1$ by verifying that $g_1(2) = g_1(3) = 1$ and $g_1(4) = g_1(5) = 2$). In this way, the coefficient $3/2$ in $g_0(n)$ is reduced to $9/8$ in $g_1(n)$, improving yet the upper bound of Proposition 2.4.

The paradigm used in Corollary 2.5 can be generalized by the following recursive algorithm.

Algorithm COLOR-CLIQUES-RECURSIVE;
Input: a $<_l$ -ordered interval representation \mathcal{I} of the graph G ;
Output: a partition of G into proper interval subgraphs;
Begin;
 if $|\mathcal{I}| \leq 3$ **then return** \mathcal{I} ;
 else
 choose a value of t in $\{2, \dots, n - 2\}$;
 compute a maximum stable S in \mathcal{I} ;
 if $|S| < t$ **then return** COLOR-CLIQUES-LOGARITHMIC(\mathcal{I});
 else
 $\mathcal{I} \leftarrow \mathcal{I} \setminus S$;
 return $S \cup$ COLOR-CLIQUES-RECURSIVE(\mathcal{I});
 End;

Set $t = \lceil \alpha n \rceil$, with $0 \leq \alpha \leq 1$. Now, assume that the algorithm COLOR-CLIQUES-RECURSIVE stops after i recursive calls with $i \geq 0$. At call $i + 1$, the cardinality of the set \mathcal{I} is lower than $(1 - \alpha)^i n$ and i sets have been added to the partition. Then, the cardinality of the output partition is lower than

$$i + \left\lceil \log_3 \frac{3 \lceil \alpha(1 - \alpha)^i n \rceil - 3}{2} \right\rceil$$

when the algorithm ends by calling COLOR-CLIQUES-LOGARITHMIC, which can be bounded by

$$\left\lceil \log_3 \frac{3^{i+1}}{2} \alpha(1 - \alpha)^i n \right\rceil$$

Now, we can determine the optimal value of α , and thus of t , for which the above expression is minimum over i . By differentiating the expression (without ceiling) with respect to i , we obtain

$$\frac{3}{2} \alpha (3 - 3\alpha)^i (\ln 3 + \ln(1 - \alpha))$$

whose zeros are 0 and $2/3$. Since $t = 0$ causes a contradiction (such a value prevents the algorithm COLOR-CLIQUES-LOGARITHMIC to be called), the optimum value for t is $\lceil 2n/3 \rceil$, leading to the upper bound $\lceil \log_3 n \rceil$. On the other

hand, if the algorithm terminates due to the condition $|\mathcal{I}| \leq 3$, then we have $(1 - \alpha)^i n \leq 3$, implying that the number of previous recursive calls cannot exceed $\lceil \log_\beta(n/3) \rceil$ with $\beta = 1/(1 - \alpha)$. In this case, the cardinality of the partition is lower than $\lceil \log_\beta(\beta n/3) \rceil$, leading to the upper bound $\lceil \log_3 n \rceil$ for $\alpha = 2/3$.

Finally, the complexity of the algorithm COLOR-CLIQUE-RECURSIVE is analyzed for $t = \lceil 2n/3 \rceil$. According to the previous discussion, the number of recursive calls is bounded by $O(\log n)$. Thus, carefully implemented, the entire algorithm takes $O(n \log n)$ time and linear space. Indeed, computing a maximum stable is done in $O(n)$ time since an ordered interval representation is given in input, and the algorithm COLOR-CLIQUE-LOGARITHMIC runs in $O(n \log n)$ time according to the previous proposition.

Proposition 2.7 *Any interval graph admits a partition into at most $\lceil \log_3 n \rceil$ proper interval subgraphs, for $n > 1$. Moreover, this partition is computed in $O(n \log n + m)$ time and linear space.*

As corollary, we also obtain an upper bound for the class of circular-arc graphs.

Corollary 2.8 *Any circular-arc graph admits a partition into at most $\lceil \log_3 n \rceil + 1$ proper interval subgraphs, for $n > 1$. Moreover, this partition is computed in $O(n \log n + m)$ time and linear space.*

Proof. Let $G = (V, E)$ be a circular-arc graph. Compute a circular-arc representation of G in linear time and space [17]. Choose a point p on the circle and compute the set C of vertices corresponding to arcs containing p . To obtain the desired upper bound, observe that the set C induces a clique (which induces trivially a proper interval graph) and that the subgraph induced by $V \setminus C$ is an interval graph. \square

To conclude, an interval graph is built for which the upper bound $\lceil \log_3 n \rceil$ provided by the algorithm COLOR-CLIQUE-RECURSIVE is asymptotically reached. Define $\lceil 2n/3 \rceil - 1$ intervals with unit length inducing a stable and then cover them by a sole interval I . This corresponds to an interval representation of the graph $K_{1, \lceil 2n/3 \rceil - 1}$. Now, complete the representation by arranging the remaining intervals, with unit length too, in such a way that the maximum stable in the graph remains of size $\lceil 2n/3 \rceil - 1$. Whereas this interval graph admits trivially a partition into two proper interval subgraphs (the interval I on one hand, the unit intervals on the other hand), the algorithm COLOR-CLIQUE-RECURSIVE returns a partition of size $\lceil \log_3((3 \lceil 2n/3 \rceil - 3)/2) \rceil \leq \lceil \log_3 n \rceil$, with equality when n tends to infinity.

3 Complexity issues

Here are derived two complexity results about the problem of partitioning graphs into proper interval subgraphs.

Corollary 3.1 *The problem of partitioning into proper interval subgraphs is solvable in linear time and space for $K_{1,7}$ -free interval graphs.*

Proof. Let G be a $K_{1,7}$ -free interval graph. First, test in linear time and space if G is a proper interval graph [5,6]. If not, then use the algorithm COLOR-CLIQUE-LOGARITHMIC with $t = 7$ to partition the graph into two proper interval subgraphs in linear time and space (according to Lemma 2.3). \square

Corollary 3.2 *The problem of partitioning into proper interval subgraphs admits an approximation algorithm for interval graphs (resp. circular-arc graphs) providing a worst-case ratio $\lceil \log_3 n \rceil / 2$ (resp. $(\lceil \log_3 n \rceil + 1) / 2$) in $O(n \log n + m)$ time and linear space.*

Proof. Let G be an interval graph. If G is not a proper interval graph (whose recognition is done in linear time and space [5,6]), then the algorithm COLOR-CLIQUE-RECURSIVE returns a partition of G into $\lceil \log_3 n \rceil$ proper interval subgraphs in $O(n \log n + m)$ time and linear space (according to Proposition 2.7). Since the size of an optimal partition is at least two in this case, the resulting worst-case approximation ratio is $\lceil \log_3 n \rceil / 2$. For circular-arc graphs, the same idea can be employed, having first extracted a clique as done in the proof of Corollary 2.8. \square

Note that the construction given at the end of the previous subsection provides interval graphs for which the worst-case approximation ratio is reached. At our knowledge, the complexity of partitioning interval graphs (or more generally arbitrary graphs) into proper interval subgraphs remains unknown.

4 Related problem

In this section, a related problem is addressed: finding a largest proper interval subgraph in an interval graph. First, an upper bound is given on the size of such a subgraph.

Proposition 4.1 *Any split graph contains a proper interval subgraph of size at least $\lfloor n/2 \rfloor + 1$, for $n \geq 1$. Moreover, for every $n \geq 1$, a threshold graph exists whose largest proper interval subgraph is of size at most $\lceil n/2 \rceil + 1$.*

Proof. Let G be a split graph with S the stable set and C the clique set. If $|C| \geq \lfloor n/2 \rfloor$, then the subgraph induced by the set C and any vertex of the set S admits a unit interval representation and the bound is established. Otherwise, the stable S , inducing trivially a proper interval subgraph, has a size strictly greater than $\lfloor n/2 \rfloor$.

Here is a threshold graph for which this bound is sharp. Consider a stable S with $\lfloor n/2 \rfloor + 1$ vertices such that each vertex of S is connected to $\lfloor n/2 \rfloor - 1$ vertices inducing a clique C . If a proper interval subgraph contains at least three vertices of the stable S , then this one has necessarily a size smaller than $\lfloor n/2 \rfloor + 1$ (because such a subgraph cannot contain one vertex in C without inducing the claw $K_{1,3}$). On the other hand, if a proper interval subgraph contains at most two vertices of S , then this one has a size smaller than $\lfloor n/2 \rfloor + 1$ (because this subgraph contains at best all the vertices of C plus the two vertices of S). \square

As for partitioning into proper interval subgraphs, a lower bound can be derived by exhibiting stables or cliques. The Perfect Graph Theorem of Lovász (cf. [14, pp. 53–58]) provides a first lower bound for the size of a largest proper interval subgraph in perfect graphs (and a fortiori in interval graphs). Lovász (cf. [14, pp. 53–58]) has established that if a graph G is perfect, then $\omega(G)\alpha(G) \geq n$. This implies that if $\omega(G) \leq \lfloor \sqrt{n} \rfloor$, then $\alpha(G) \geq \lfloor \sqrt{n} \rfloor$. Consequently, the largest set among a maximum clique and a maximum stable in a perfect graph induces a proper interval subgraph of size at least $\lfloor \sqrt{n} \rfloor$. But Proposition 2.7 provides better lower bounds for interval graphs and circular-arc graphs (which are not all perfect).

Corollary 4.2 *Any interval graph (resp. circular-arc graph) contains a proper interval subgraph of size at least $\lceil n/\lceil \log_3 n \rceil \rceil$ (resp. $\lceil n/(\lceil \log_3 n \rceil + 1) \rceil$), for $n > 1$. Moreover, such a subgraph is computed in $O(n \log n + m)$ time and linear space.*

Proof. According to Proposition 2.7 and Corollary 2.8, any interval graph (resp. circular-arc graph) admits a partition into at most $\lceil \log_3 n \rceil$ (resp. $\lceil \log_3 n \rceil + 1$). By applying the pigeonhole principle, we obtain that at least one of these subgraphs contains more than $\lceil n/\lceil \log_3 n \rceil \rceil$ (resp. $\lceil n/(\lceil \log_3 n \rceil + 1) \rceil$) vertices. \square

Note that $\lceil n/\lceil \log_3 n \rceil \rceil \geq \lfloor \sqrt{n} \rfloor$ for all $n > 1$. This lower bound can be refined by using the following corollary, which is inspired by Corollary 2.5.

Corollary 4.3 *Let G be an interval graph. If $1 \leq \alpha(G) < t$, then G contains a proper interval subgraph of size at least $\lceil n/\lceil \log_3((3t-3)/2) \rceil \rceil$. If $t \leq \alpha(G) \leq n$, then G contains a proper interval subgraph of size at least t . Moreover, such a subgraph is computed in $O(n \log t + m)$ time and linear space.*

Proof. The proof of the first assertion is done by applying Lemma 2.3 and the pigeonhole principle, as in the proof of the previous corollary. The second assertion relies on the fact that any stable induces a proper interval graph. \square

Remark 4.4 *Here is another formulation of the previous corollary. Let G be an interval graph satisfying one of the two following conditions, with $t = O(1)$: (i) G contains no induced copy of $K_{1,t}$ (or no stable of size greater than t); (ii) G contains an induced copy of $K_{1,n-t}$ (or a stable of size greater than $n-t$). Then, G contains a proper interval subgraph of size $\Omega(n)$, which is computed in linear time and space.*

According to the previous corollary, we obtain that any interval graph with n vertices contains a proper interval subgraph of size at least $\min\{\lceil n/g_0(t) \rceil, t\}$ for any value of $t \in \{2, \dots, n-2\}$, with $g_0(t) = \lceil \log_3((3t-3)/2) \rceil$. Now, we search the value t for which this expression reaches a maximum. The function $\lceil n/g_0(t) \rceil$ is monotonically decreasing, while the function t is monotonically increasing. Thereby, a lower bound of the maximum sought is $\lceil n/g_0(\lceil \tilde{t} \rceil) \rceil$ with \tilde{t} solution of the equation $t = n/\tilde{g}_0(t)$ with $\tilde{g}_0(t) = \log_3(3t/2)$. The value of \tilde{t} is equal to $\tilde{n}/W_0(\beta\tilde{n})$ with $\tilde{n} = n \ln 3$, $\beta = 3/2$ and W_0 the principal branch of the Lambert W function [4].

The Lambert W function satisfies $W(z)e^{W(z)} = z$ for any complex number z . This function is not injective and have an infinity of branches, with only one in positive reals. This latter, called principal branch W_0 , is known in enumerative combinatorics to be used in the counting of some trees. The asymptotic behavior of the branch W_0 is given by

$$W_0(z) \sim \ln z - \ln \ln z + \sum_{k=0}^{\infty} \sum_{m=1}^{\infty} c_{k,m} \frac{(\ln \ln z)^m}{(\ln z)^{k+m}}$$

where $c_{k,m}$ represents constants not depending from z . When $z \geq e/\alpha$, we have that $\ln(\alpha z) - \ln \ln(\alpha z) \leq W_0(\alpha z) \leq \ln(\alpha z)$. For more details on the Lambert W function and its principal branch W_0 , the reader is referred to [4].

Consequently, we have the following upper bound for \tilde{t} :

$$t^* = \frac{n}{\log_3(\beta'n) - \log_3 \ln(\beta'n)}$$

with $\beta' = (3 \ln 3)/2$, and then the following lower bound for $\lceil n/g_0(\lceil \tilde{t} \rceil) \rceil$:

$$\lceil n/\lceil \log_3(\beta t^*) \rceil \rceil$$

with $\beta = 3/2$. Such a bound can be surrounded as follows for $n > 3$:

$$\left\lceil \frac{n}{\lceil \log_3 n \rceil} \right\rceil \leq \left\lceil \frac{n}{\lceil \log_3(\beta t^*) \rceil} \right\rceil \leq \left\lceil \frac{n}{\lceil \log_3(\beta' n) - \log_3 \log_3(\beta' n) \rceil} \right\rceil$$

This confirms that the bound $\lceil n/\lceil \log_3(\beta t^*) \rceil \rceil$ outperforms the bound given in Corollary 4.2.

Proposition 4.5 *Any interval graph contains a proper interval subgraph of size at least*

$$\left\lceil \frac{n}{\lceil \log_3 \left(\frac{\beta n}{\log_3(\beta' n) - \log_3 \ln(\beta' n)} \right) \rceil} \right\rceil$$

with $\beta = 3/2$ and $\beta' = (3 \ln 3)/2$, for $n > 3$. Moreover, such a subgraph is computed in $O(n \log(\frac{n}{\log n - \log \log n}) + m)$ time and linear space.

Note that a similar lower bound can be derived for circular-arc graphs.

We conclude this section by discussing the complexity of the problem. Finding the largest proper interval subgraph in an arbitrary graph generalizes the recognition problem for proper interval graphs, which is solvable in linear time and space [5,6]. The complexity of this problem has not been elucidated, even if restricted to interval graphs. Nevertheless, the problem is shown to be linear-time and space solvable when restricted to the subclass of threshold graphs.

Proposition 4.6 *The problem of finding a largest proper interval subgraph is solvable in linear-time and space for threshold graphs.*

Proof. Remind that a threshold graph admits a partition into one clique $C = C_1 \cup \dots \cup C_r$ and one stable $S = S_1 \cup \dots \cup S_r$ (with $r \leq n$ and all C_i, S_i non empty except S_r) such that a vertex of S_i is adjacent to a vertex of $C_{i'}$ if $i' > i$ for all $i, i' \in \{1, \dots, r\}$. Such a representation is computed in linear time and space, having partitioned the vertices of the graph according to their degree [14, pp. 223–227].

Any proper interval subgraph of maximum cardinality in a threshold graph must contain at least one vertex of C . Indeed, if a proper interval subgraph contains only vertices in S , then the vertices of the set C_1 can be added to this one in order to increase its size (by definition, C_1 is not empty and connected to no set among S_1, \dots, S_r). Now, denote by i the largest index such that a vertex C_i belongs to a maximum proper interval subgraph ($1 \leq i \leq r$). We

claim that such a subgraph contains: (i) all the vertices in $C_1 \cup \dots \cup C_i$, (ii) all the vertices in $S_i \cup \dots \cup S_r$, (iii) one vertex in $S_1 \cup \dots \cup S_{i-1}$ if $i > 1$. The assertions (i) and (ii) follow from the fact that any vertex of C_i induces a clique with the vertices in $C_1 \cup \dots \cup C_i$ while it is adjacent to no vertex in S_i, \dots, S_r . To prove the assertion (iii), one can observe that the subgraph induced by a vertex of C_i , a vertex of C_1 and two vertices in $S_1 \cup \dots \cup S_{i-1}$ is isomorphic to the claw $K_{1,3}$. Then, the size of a maximum proper interval subgraph containing at least one vertex in C_i is given by $t_i = \sum_{j=1}^i |C_j| + \sum_{j=i}^r |S_j|$ (+1 if $i > 1$).

Consequently, finding a largest proper interval subgraph in a threshold graph is reduced to computing the index i (with $1 \leq i \leq r$) for which the value of t_i is maximum. Carefully implemented, the computation of t_i for all $i = 1, \dots, r$ can be done incrementally in linear time and space. \square

Remark 4.7 *According to the previous discussion, the number of maximal proper interval subgraphs in a threshold graph is exactly r , which is lower than the number n of vertices.*

5 Conclusion

The table below summarizes the results established throughout the paper. These ones concern the two following problems: (a) partitioning an interval graph into proper interval subgraphs, (b) finding the largest proper interval subgraph in an interval graph.

Problem	Lower bound	Upper bound	Complexity
(a)	$\lfloor \log_3(2n + 1) \rfloor$	$\lceil \log_3 n \rceil$	open
(b)	$\left\lceil \frac{n}{\log_3 \left(\frac{\beta n}{\log_3(\beta' n) - \log_3 \ln(\beta' n)} \right)} \right\rceil$	$\lceil n/2 \rceil + 1$	open

All these lower and upper bounds have been obtained by constructive proofs.

5.1 Open questions

The lower and upper bounds established in Theorem 1.1 differ by *at most one* for interval graphs. The first values for which lower and upper bounds are not equal are $t = 8, 9$ and $n = 10, 11, 12$ (see the table below).

Interval graph	Lower bound	Upper bound
$K_{1,3}$ -free	1	1
$K_{1,4}$ -free	2	2
$K_{1,5}$ -free	2	2
$K_{1,6}$ -free	2	2
$K_{1,7}$ -free	2	2
$K_{1,8}$ -free	2	3
$K_{1,9}$ -free	2	3
$K_{1,10}$ -free	3	3
$K_{1,t}$ -free	$\lfloor \log_3(3t - 3) \rfloor$	$\lceil \log_3((3t - 3)/2) \rceil$
arbitrary	$\lfloor \log_3(2n + 1) \rfloor$	$\lceil \log_3 n \rceil$

For still sharpening bounds of Theorem 1.1, answering to the following question seems to be decisive: do $K_{1,8}$ -free (or $K_{1,9}$ -free) interval graphs exist admitting no partition into less than three proper interval subgraphs? Or does any $K_{1,8}$ -free (or $K_{1,9}$ -free) interval graph admit a partition into two proper interval subgraphs? The reader is invited to look at Fig. 7 where is illustrated an interval graph containing no induced copy of $K_{1,8}$ but one of $K_{1,7}$. This graph, which is not a proper interval graph, admits a partition into three proper interval subgraphs. But does it admit a partition into two proper interval subgraphs? The answer is yes: the partition is obtained by separating clear intervals from dark ones on Fig. 7.

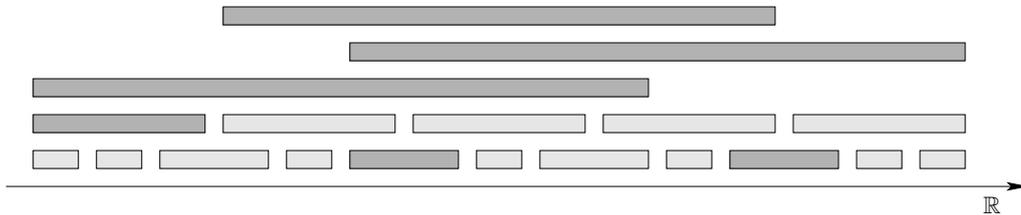


Fig. 7. An example of challenging $K_{1,8}$ -free interval graph.

More generally, the characterization of interval graphs admitting a partition into k proper interval subgraphs remains an open question. We have shown that any interval graph containing H_k as induced subgraph admits no partition into less than k proper interval subgraphs. But is the converse true?

References

- [1] B.S. Baker, E.G. Coffman, Jr. (1996). Mutual exclusion scheduling. *Theoretical Computer Science* 162, pp. 225–243.

- [2] H.L. Bodlaender, K. Jansen (1995). Restrictions of graph partition problems. Part I. *Theoretical Computer Science* 148, pp. 93–109.
- [3] A. Brandstädt, V.B. Le, J.P. Spinrad (1999). *Graph Classes: a Survey*. SIAM Monographs on Discrete Mathematics and Applications 3, SIAM Publications, Philadelphia, PA.
- [4] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, D.E. Knuth (1996). On the Lambert W function. *Advances in Computational Mathematics* 5, pp. 329–359.
- [5] D.G. Corneil, H. Kim, S. Natarajan, S. Olariu, A. Sprague (1995). Simple linear time recognition of unit interval graphs. *Information Processing Letters* 55, pp. 99–104.
- [6] X. Deng, P. Hell, J. Huang (1996). Linear-time representation algorithms for proper circular-arc graphs and proper interval graphs. *SIAM Journal on Computing* 25(2), pp. 390–403.
- [7] P.C. Fishburn (1985). *Interval Orders and Interval Graphs*. Wiley-Interscience Series in Discrete Mathematics, John Wiley & Sons, New York, NY.
- [8] F. Gardi (2004). On partitioning interval and circular-arc graphs into proper interval subgraphs with applications. In *Proceedings of the 6th Latin American Symposium on Theoretical Informatics* (M. Farach-Colton, ed.), *Lecture Notes in Computer Science* 2976, pp. 129–140. Springer-Verlag, Berlin, Germany.
- [9] F. Gardi (2005). Ordonnancement avec exclusion mutuelle par un graphe d’intervalles ou d’une classe apparentée: complexité et algorithmes. PhD Thesis. Université de la Méditerranée - Aix-Marseille II, Marseille, France.
http://www.lif-sud.univ-mrs.fr/~garddi/downloads/These_Gardi.pdf
- [10] F. Gardi (2006). Mutual exclusion scheduling with interval graphs or related classes: complexity and algorithms. *4OR, a Quarterly Journal of Operations Research* 4(1), pp. 87–90.
- [11] F. Gardi (2007). The Roberts characterization of proper and unit interval graphs. *Discrete Mathematics* 307(22), pp. 2906–2908.
- [12] F. Gardi (2008). Mutual exclusion scheduling with interval graphs or related classes. Part II. *Discrete Applied Mathematics* 156(5), pp. 794–812.
- [13] F. Gardi (2009). Mutual exclusion scheduling with interval graphs or related classes. Part I. *Discrete Applied Mathematics* 157(1), pp. 19–35.
- [14] M.C. Golumbic (1980). *Algorithmic Graph Theory and Perfect Graphs*. Computer Science and Applied Mathematics Series, Academic Press, New York, NY.
- [15] U.I. Gupta, D.T. Lee, J.Y.-T. Leung (1982). Efficient algorithms for interval graphs and circular-arc graphs. *Networks* 12, pp. 459–467.

- [16] M. Habib, R.M. McConnell, C. Paul, L. Viennot (2000). Lex-BSF and partition refinement, with applications to transitive orientation, interval graph recognition and consecutive ones testing. *Theoretical Computer Science* 234, pp. 59–84.
- [17] R.M. McConnell (2003). Linear time recognition of circular-arc graphs. *Algorithmica* 37(2), pp. 93–147.
- [18] F.S. Roberts (1976). *Discrete Mathematical Models with Applications to Social, Biological, and Environmental Problems*. Prentice-Hall, Englewood Cliffs, NJ.
- [19] F.S. Roberts (1978). *Graph Theory and its Application to the Problems of Society*. CBMS-NSF Regional Conference Series in Applied Mathematics 29, SIAM Publications, Philadelphia, PA.