Discrete Applied Mathematics 245 (2018) 168–176

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Recognition and characterization of unit interval graphs with integer endpoints

G. Durán [a,b,c,d], F. Fernández Slezak [a,b,*], L.N. Grippo [e], F.de S. Oliveira [f], J.L. Szwarcfiter [f,g]

[a] *CONICET, Argentina*
[b] *Instituto de Cálculo, FCEN, Universidad de Buenos Aires, Argentina*
[c] *Depto. de Matemática, FCEN, Universidad de Buenos Aires, Argentina*
[d] *Depto. de Ingeniería Industrial, FCFM, Universidad de Chile, Chile*
[e] *Instituto de Ciencias, UNGS, Los Polvorines, Buenos Aires, Argentina*
[f] *Instituto de Matemática e Estatística, UERJ, Rio de Janeiro, Brazil*
[g] *IM, COPPE and NCE, UFRJ, Rio de Janeiro, Brazil*

ARTICLE INFO

Article history:
Received 31 December 2015
Received in revised form 11 March 2017
Accepted 19 April 2017
Available online 1 June 2017

Keywords:
Unit interval graphs
Forbidden induced subgraphs
Proper interval graphs

ABSTRACT

We study those unit interval graphs having a model with intervals of integer endpoints and prescribed length. We present a structural result for this graph subclass which leads to a quadratic-time recognition algorithm, giving as positive certificate a model of minimum total length and as negative certificate a forbidden induced subgraph. We also present a quadratic-time algorithm to build, given a unit interval graph, a unit interval model with integer endpoints for which the interval length is as minimum as possible.

1. Introduction

A graph is an *interval graph* if there exists a bijection between a family of open intervals in the real line, and its vertex set, such that two vertices are adjacent if and only if their corresponding intervals intersect. Such a family of intervals is called an *interval model* of the graph. Let $G = (V, E)$ be a graph, $H = (V', E')$ is said to be an *induced subgraph* of G if $V' \subseteq V$ and $E' = \{uv \in E : u, v \in V'\}$. Given a collection of graphs $\mathcal{H}$, G is defined to be $\mathcal{H}$*-free* if for any graph $H \in \mathcal{H}$, G does not contain an induced H. If $\mathcal{H}$ is a set with a single element H, we just use H-free for short. Interval graphs were characterized by forbidden induced subgraphs in a celebrated paper of Boland and Lekkerkerker [6]. Graphs in this class can be recognized in linear-time (see e.g. [5]). A *unit interval graph* is an interval graph having a model with all its intervals of the same length. Such an interval model is called a *unit interval model*. Roberts proves that unit interval graphs are exactly those claw-free interval graphs [9]. Unit interval graphs can be recognized in linear-time. For instance, a direct and simple linear-time recognition algorithm for unit interval graphs based on BFS search can be found in [1]; in case the input graph is a unit interval, the algorithm outputs a unit interval model in which every endpoint is an integer ranging from 0 to n^2.

For definitions and concepts not defined here, see [11]. Two adjacent vertices are *true twins*, or simply *twins*, if they have the same closed neighborhood, where two vertices are *neighbors* if they are adjacent. Given an edge vx and a vertex y of G, if x, y are twin vertices in $G - \{v\}$ but not in G, that is, $N_G[x] \Delta N_G[y] = \{v\}$, we will say that v *distinguishes* x from y. Roberts [9]

* Corresponding author at: Instituto de Cálculo, FCEN, Universidad de Buenos Aires, Argentina.
E-mail addresses: gduran@dm.uba.ar (G. Durán), ffslezak@dm.uba.ar (F. Fernández Slezak), lgrippo@ungs.edu.ar (L.N. Grippo), fabiano.oliveira@ime.uerj.br (F.de S. Oliveira), jayme@nce.ufrj.br (J.L. Szwarcfiter).

http://dx.doi.org/10.1016/j.dam.2017.04.013
0166-218X/

shows that the ordering of the left endpoints of any unit interval model of a connected unit interval graph without twins is unique up to reversal, called a *canonical ordering*. Throughout this paper, the results given will concern finite, undirected and connected graphs. Every connected component can be analyzed separately for disconnected graphs; thus, we will only consider the connected graphs. Extending the results for unconnected graphs is trivial. Given an interval I, denote by $\ell(I)$ and $\mathrm{r}(I)$ the left endpoint and the right endpoint of I, respectively. For a twin-free connected graph, let $\mathcal{M} = \{I_1, \dots, I_n\}$ be a unit interval model indexed in the canonical order, where I_1 has the leftmost left endpoint and I_n has the rightmost one, we define the *total length* of $\mathcal{M}$ as $|\mathcal{M}| = \mathrm{r}(I_n) - \ell(I_1)$. We also define $\mathrm{Left}(\mathcal{M}) = \ell(I_1)$ and $\mathrm{Right}(\mathcal{M}) = \mathrm{r}(I_n)$. For a positive integer k, an open (resp. closed) interval with integer endpoints and length k is a (k)-*interval* (resp. $[k]$-*interval*). Given a graph G and a positive integer k, we say that G is a (k)-*interval graph* (resp. a $[k]$-*interval graph*) if G admits a model $\mathcal{M}$ of (k)-intervals (resp. $[k]$-intervals). This unit interval model is called (k)-*interval model* (resp. $[k]$-*interval model*). Notice that every unit interval graph has a (k)-interval model for some k. Therefore, given a unit interval model $\mathcal{M}$, it is interesting to find another unit interval model $\mathcal{M}'$ such that k and $\left|\mathcal{M}'\right|$ are minimum. This problem was studied in different contexts. Continuing with some ideas developed by Pirlot in [8], Mitas presents a linear-time algorithm that constructs a (k)-interval model with a minimum k for a given unit interval model [7]. This algorithm was developed in the context of the study of interval semiorders representable with intervals of a given length k. Moreover, those semiorders representable with intervals of length k were characterized by forbidden suborders for every integer k. Recently, Soulignac [10] reported a flaw in Mitas' algorithm [7], showed examples in which the integer k, found by the algorithm, is not minimum and the flaw is fixed but the complexity becomes quadratic-time. The problem of finding a unit interval model with minimum k and minimum total length is also studied in the context of gene clusters as intersection of powers of paths [2].

In this work, we give a new approach to solve the following problems:

1. recognize if a graph is a (k)-interval graph for a given k and exhibit a (k)-interval model of minimum total length,
2. find the minimum k in which a graph is a (k)-interval graph.

More precisely, given a graph G and a nonnegative integer k, we develop an algorithm that determines (in quadratic-time) if G is a (k)-interval graph and if so, shows a (k)-interval model of minimum total length. Otherwise, it shows a forbidden induced subgraph contained in G (a negative certificate). In addition, we derive from this algorithm another one that, given a unit interval graph G and a unit interval model of it, finds the minimum k in which G is a (k)-interval graph. This algorithm also constructs a (k)-interval model. Furthermore, we present a characterization by forbidden induced subgraphs for the class of (k)-interval graphs for every integer k. Notice that recognizing a unit interval graph and obtaining a unit interval model is linear [1]; therefore, we shall assume that a unit interval model of an input graph is given whenever the input graph is a unit interval graph.

This paper is organized as follows. In Section 2, we present a characterization for the class of (k)-interval graphs for every integer k by forbidden induced subgraphs and a quadratic-time algorithm that decides for a given integer k whether a unit interval graph has a (k)-interval model or not. Finally, in Section 3, we present an algorithm that finds in quadratic-time a (k)-interval model of minimum k and minimum total length.

An extended abstract of this work was published in [3].

2. Structural characterization

In this section, we present a family of forbidden induced subgraphs for a (k)-interval graph. Our main result is a characterization of this class and an algorithm to exhibit a model of minimum total length.

Before presenting the main result of this section, we need to introduce new definitions. Given an interval $I = (a, b)$, we say that I is *left shifted* if it is replaced by the interval $(a-1, b-1)$, that is, if I is shifted one unit to the left. Given a (k)-interval model $\mathcal{M}$ of a (k)-interval graph G and $I \in \mathcal{M}$, we define $\mathcal{L}_k(I)$ as the minimum submultiset $\mathcal{S}$ of $\mathcal{M}$ such that left shifting all intervals in $S \cup I$ yields a new (k)-interval model of G and $\mathcal{L}_k(A)$ for some multiset A of (k)-intervals as $\cup_{I \in A}\mathcal{L}_k(I)$, where $\mathcal{L}_k(\emptyset) = \emptyset$. Let $\mathrm{RO}(I)$ be the multiset of the intervals of a (k)-interval model which overlap I one unit on the right endpoint and the emptyset otherwise, and $\mathrm{L}(I)$ be the multiset of the intervals of a (k)-interval model whose right endpoint coincides with the left endpoint of I and the emptyset otherwise. Fig. 1 shows $\mathrm{RO}(I)$ and $\mathrm{L}(I)$ of a certain interval $I = (k, 2k)$. If I is left shifted, $\mathrm{RO}(I)$ and $\mathrm{L}(I)$ must also be left shifted so that $\mathcal{M}$ is still a (k)-interval model of G. But, as $\mathrm{RO}(I)$ and $\mathrm{L}(I)$ will be left shifted, $\mathcal{L}_k(L(I))$ and $\mathcal{L}_k(RO(I))$ must also be left shifted; therefore, a recursive way of defining $\mathcal{L}_k : \mathcal{M} \to \mathcal{P}(\mathcal{M})$ is as follows:

$$\mathcal{L}_k(I) = \mathcal{L}_k(\mathrm{RO}(I)) \cup \mathcal{L}_k(\mathrm{L}(I)) \cup \mathrm{RO}(I) \cup \mathrm{L}(I),$$

where $\mathcal{P}(\mathcal{M})$ is the family of all submultisets $S \subseteq \mathcal{M}$. When the length k is clear in the context, we shall omit the subindex and use $\mathcal{L} = \mathcal{L}_k$.

Consider the (k)-interval model $\widetilde{\mathcal{M}}$ obtained from $\mathcal{M}$ left shifting all intervals in $\mathcal{L}(I) \cup \{I\}$ for some $I \in \mathcal{M}$. Notice that $\widetilde{\mathcal{M}}$ is a (k)-interval model of the same graph as $\mathcal{M}$. In fact, $\mathcal{L}(I)$ is the minimum multiset of intervals that has to be left shifted when translating I one unit to the left so as to preserve adjacencies. Suppose $\mathcal{M}$ is a (k)-interval model with two coinciding intervals $I = I'$. The interval I belongs to $\mathcal{L}(I')$ if and only if there exists a sequence of intervals $I_1, I_2, \dots, I_N$ in $\mathcal{L}(I') \cup \{I'\}$ such that $I_1 = I'$ and I_{i+1} equals either $L(I_i)$ or $RO(I_i)$ for each $1 \le i \le N-1$ and either $L(I_N) = I$ or $RO(I_N) = I$. Suppose that $L(I_N) = I$, the other case is analogous as noted in Remark 1. Given two integers a and b, $[\![a, b]\!]$ stands for the set formed by

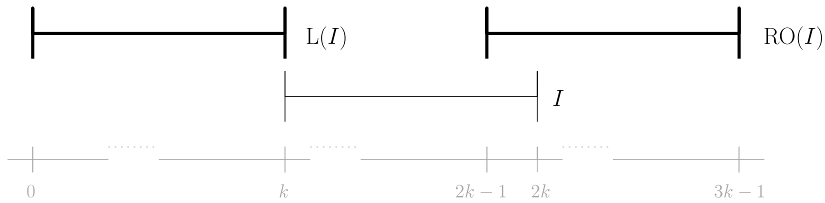

Fig. 1. Configurations that do not preserve adjacencies if I is the only interval left shifted.

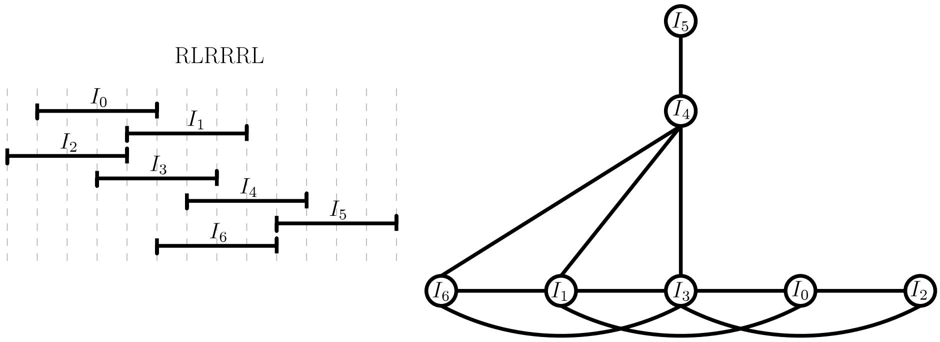

Fig. 2. The (4)-interval model of the string RLRRRL and its corresponding (4)-interval graph.

the integers from a to b. Let i and j be, respectively, the number of "right overlap" intervals in the sequence and the number of "left" intervals in the sequence, that is, $i = |\{l \in [\![1, N-1]\!] : I_{l+1} = RO(I_l)\}|$ and $j = |\{l \in [\![1, N-1]\!] : I_{l+1} = L(I_l)\}|$.

Let $l_i = \ell(I_i)$, we have that

$$l_1 + (k-1)i - jk = l_N = l_1 + k,$$

and, therefore,

$$k = \frac{i}{i-j-1}.$$

Then, as k is an integer, for some $m \geq 1,\ i = km$ and $j = m(k-1) - 1$. In the next lemma, we show that if $m = 1$ then the sequence in which $I \in \mathcal{L}(I')$ is minimum.

To simplify the notation, we will denote these particular (k)-interval models defined by a sequence $I_1, \dots, I_N$ as a string of length N having mk R's and $m(k-1)-1$ L's in the following way: for all $1 \leq i \leq N-1$, an L in the $(i+1)$th position is to denote that $I_{i+1} = L(I_i)$ and an R in that position is to denote that $I_{i+1} = RO(I_i)$. Fig. 2 shows a particular string of (4)-intervals and its corresponding unit interval graph, taking $k = 4$ and $m = 1$.

Lemma 1. *Let G be the unit interval graph corresponding to a (k)-model represented by a string $\mathcal{S}$ of $i = mk$ R's and $j = m(k-1) - 1$ L's with $m > 1$. Then, there exists a contiguous substring $\mathcal{S}'$ of $\mathcal{S}$ consisting of k R's and $k-2$ L's.*

Proof. Let $\mathcal{S}$ be a string of mk R's and $m(k-1)-1$ L's with $m > 1$. We will prove that there exists a contiguous substring $\mathcal{S}'$ of $\mathcal{S}$ consisting of k R's and $k-2$ L's.

Let x_1 be the number of R's before the first L, x_i be the number of R's between the $(i-1)$th and the ith L, $2 \leq i \leq m(k-1)-1$, and $x_{m(k-1)}$ be the number of R's after the last L. Clearly, $\sum_{i=1}^{m(k-1)} x_i = km$. In order to exist a substring $\mathcal{S}'$ of k R's and $k-2$ L's, there should exist i satisfying simultaneously

$$x_i + x_{i+1} + \cdots + x_{i+k-3} + x_{i+k-2} \geq k \tag{1}$$

$$x_{i+1} + \cdots + x_{i+k-3} \leq k. \tag{2}$$

Suppose there is no such $\mathcal{S}'$, then for all i, $x_i + x_{i+1} + \cdots + x_{i+k-3} + x_{i+k-2} \leq k-1$ or $x_{i+1} + \cdots + x_{i+k-3} \geq k+1$. It is easy to see that there exists some i satisfying (1). For this i, since $\sum_{i=1}^{m(k-1)} x_i = km$, the inequality (2) does not hold therefore.

Then, the next sequence $x_{i+1}, x_{i+2}, \dots, x_{i+k-1}$ also satisfies inequality (1). By analogous arguments, this sequence contradicts inequality (2). Repeating this for the next sequences as well as for the previous sequences, we have m disjoint substrings with at least $k+1$ R's each, which results in a contradiction as there are only mk R's in $\mathcal{S}$.

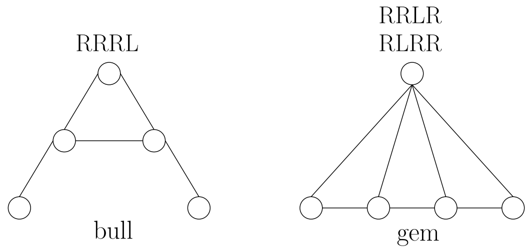

Fig. 3. The graphs bull and gem.

Remark 1. The k R, $k-2$ L strings construct, as detailed above, (k)-intervals in which if adding a last L to the string, this last interval coincides with the first one. Notice that the (k)-interval graph represented by a string after adding the extra L, that is, a k R, $k-1$ L string, is isomorphic to the intersection graph of the representation of any circular rotation of the string: it only changes the interval in which the construction starts.

To illustrate Remark 1 see Fig. 2 in which there is the string RLRRRL and by adding an L at the end of it, the interval I_0 is repeated. A circular rotation of this new string, such as RRRLLRL is represented by the same (k)-intervals starting from interval I_2 instead of I_0.

The circular rotation of the k R, $k-1$ L strings in which there is an R at the end coincides with those strings coming from the analysis if $RO(I_N)=I$.

For $k \geq 2$, let $\mathcal{F}_k$ be a family of the (k)-interval graphs with the following (k)-representations defined recursively: (i) Place the first interval I_0 of length k with its endpoints on integer numbers, this interval is called *the generator vertex*. (ii) Let I be the last so far interval placed, choose one of the following two options: **Left-choice:** Build a new interval J of length k with its right endpoint coinciding with the left endpoint of the previous one, that is, J satisfies that $\mathrm{r}(J)=\ell(I)$. This choice can be done if and only if the new interval is not fully to the left of I_0, or **Right-choice:** Build a new interval J of length k such that its left endpoint intersects in one unit the right endpoint of the previous one, that is, J satisfies left$(J)=$ right$(I)-1$. (iii) Repeat step (ii) until exactly $k-2$ Left-choices and k Right-choices are carried out.

In Fig. 2, a graph of $\mathcal{F}_4$ is exhibited; its construction as described above is after choosing the string RLRRRL when repeating step (ii) for 2 Left-choices and 4 Right-choices ($k=4$). In Fig. 3, the graphs in $\mathcal{F}_3$ are shown, with the corresponding strings of R's and L's that construct them. All graphs in $\mathcal{F}_4$ are depicted in Fig. 4 removed of isomorphic graphs. Some of the graphs in $\mathcal{F}_{k+1}$ are shown in Fig. 5.

The following theorem gives a structural characterization of (k)-interval graphs:

Theorem 2. *Let $G=(V,E)$ be a connected unit interval graph without twins and $n=|V|$. The following statements are equivalent:*

1. *G is a (k)-interval graph,*
2. *G is an induced subgraph of $(P_l)^{k-1}$, the $(k-1)$th power of the path P_l with $n \leq l \leq d(k-1)+1$, where d is the diameter of G,*
3. *G is $\mathcal{F}_{k+1}$–free.*
4. *G is a $[k-1]$–interval graph.*

Unit interval graphs $\mathcal{F}_4$–free can be represented in a (3)-interval model. The proof of Theorem 2 will require some definitions and lemmas. For a graph G, we define $q(G)$ as the number of maximal cliques of it.

Lemma 3. *Let G be a unit interval graph without twins, $C_1, C_2, \ldots, C_q$ its maximal cliques in the canonical order, $v \in C_1$ a simplicial vertex of G and $\widetilde{G}=G-\{v\}$. Then,*

1. *$q(\widetilde{G})=q(G)-1$. Moreover, the maximal cliques of $\widetilde{G}$ are $C_2, C_3, \ldots, C_q$.*
2. *if $\widetilde{G}$ has twin vertices $x, y \in V(G)$ such that $\ell(I_x) \leq \ell(I_y)$, then $x, y \in C_2$, $x \in C_1$, and x, y are the only twins in $\widetilde{G}$.*

Proof. Such a vertex v exists because given a unit interval graph without twins, the order of the left (and right) endpoints of the intervals is invariant no matter the chosen representation, except for the reverse [9].

(1) Since $C_2, C_3, \ldots, C_q$ do not contain the vertex v which was removed from G, they are maximal cliques in $\widetilde{G}$. Arguing towards a contradiction, suppose that $C_1-\{v\}$ is a maximal clique of $\widetilde{G}$. As $\widetilde{G}$ is a unit interval graph, there is a simplicial vertex u in the first maximal clique $C_1-\{v\}$, then $u \neq v$, $u \in C_1$, $u \notin C_2, C_3, \ldots, C_q$. Therefore, u and v are twins in G, a contradiction. The contradiction arose from supposing that $C_1-\{v\}$ is a maximal clique of $\widetilde{G}$.

(2) Suppose there are two pairs of twins in $\widetilde{G}$, namely x, y and $\bar{x}, \bar{y}$. As G is a graph without twins, vertex v must distinguish in G each vertex from its twin in $\widetilde{G}$. We can assume, without loss of generality, that x and $\bar{x}$ are adjacent to v and that y and $\bar{y}$ are

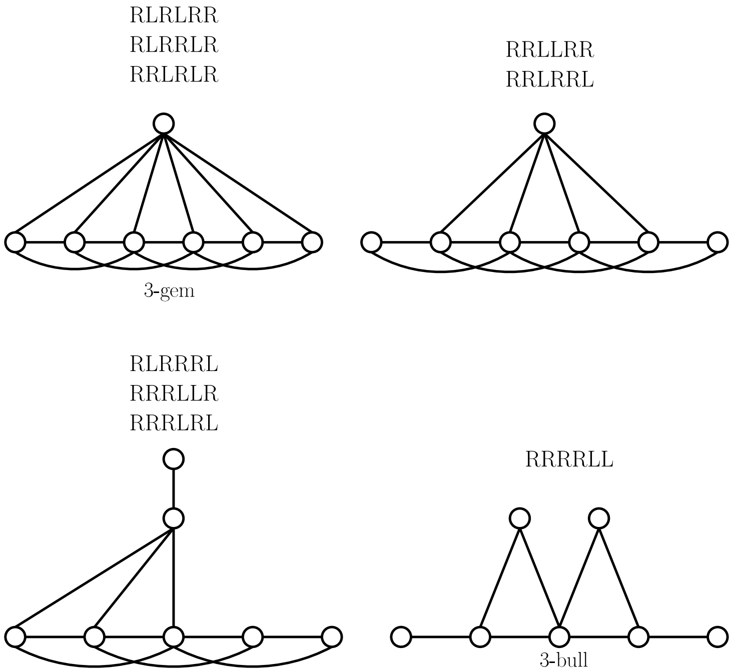

Fig. 4. Forbidden subgraphs for (3)-interval graphs and their corresponding strings of R's and L's.

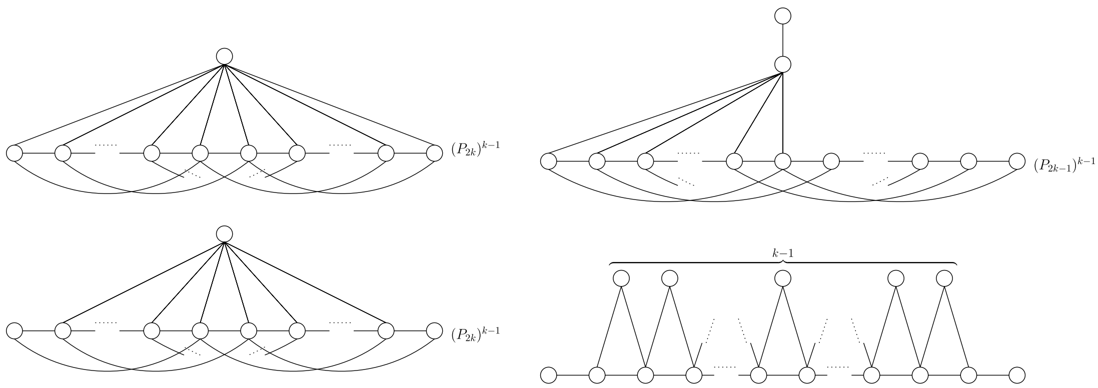

Fig. 5. Some of the forbidden subgraphs for (k)-interval graphs.

not. As $v \in C_1$ is simplicial, we have that $x, \bar{x} \in C_1$, while $y, \bar{y} \notin C_1$. If x and $\bar{x}$ are the same vertex, then $N_{\widetilde{G}}[y] = N_{\widetilde{G}}[x] = N_{\widetilde{G}}[\bar{y}]$, and then y and $\bar{y}$ would be twins in G as they are not adjacent to v. Therefore, x and $\bar{x}$ are distinct vertices. As the maximal cliques are indexed in the canonical order, there exist $i, \bar{i} \in [\![1, q]\!]$ such that $x \in C_1, C_2, \ldots, C_i$ and $x \notin C_{i+1}, \bar{x} \in C_1, C_2, \ldots, C_{\bar{i}}$ and $\bar{x} \notin C_{\bar{i}+1}$. If $i = \bar{i}$ then $N_G[x] = N_G[\bar{x}]$ and so they would be twins in G, which cannot happen. Without loss of generality, suppose that $\bar{i} < i$. Since x and y are twins in $\widetilde{G}$, that is, y is adjacent to $N[x] \setminus \{v\}$, then $y \in C_2, C_3, \ldots, C_i$ and y $\notin C_1, C_{i+1}$. Similarly, $\bar{y} \in C_2, C_3, \ldots, C_{\bar{i}}$ and $\bar{y} \notin C_1, C_{\bar{i}+1}$, as shown in Fig. 6.

As $\bar{y} \in C_l \; \forall l = 2, \ldots, \bar{i}$ and $x \in C_l \; \forall l = 1, \ldots, i$ with $\bar{i} < i$, the model is non-proper ($I_{\bar{y}} \subsetneq I_x$), a contradiction.

Vertices x and y of Lemma 3, that is two vertices that become twins after removing the simplicial vertex v, will be called *temporary twins*.

Recall that the first interval I_0 in the construction of a member of $\mathcal{F}_k$ is called the *generator vertex*. The following lemma gives a recursive method to build $\mathcal{F}_{k+1}$ from $\mathcal{F}_k$:

Lemma 4. *Let $G = (V, E)$ be a graph in $\mathcal{F}_k$ with $V = \{v_1, v_2, \ldots, v_n\}$ indexed in the canonical ordering. Let $x = v_i$ be the generator vertex of G, and H be the twin-free graph obtained by adding to G a twin x' to x and a simplicial vertex s adjacent to $v_1, \ldots, v_{i-1}, v_i = x$, thus removing the twin condition between x and x'. Then, $H \in \mathcal{F}_{k+1}$.*

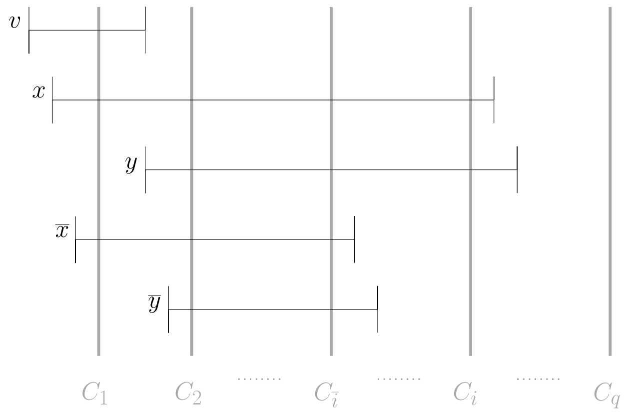

Fig. 6. A representation of G having two pairs of twins x, y and $\bar{x}, \bar{y}$. The vertex sets $C_1, C_2, \dots, C_q$ are the maximal cliques in the canonical order.

Proof. The vertex s is in fact a simplicial vertex of H as the generator vertex $x = v_i$, by construction, is always in the first clique, ordering the cliques in the canonical order. Then, as the vertices are in the canonical order $v_1, \dots, v_{i-1}, v_i$ are all in the first clique.

The graph G is a member of $\mathcal{F}_k$ with a particular string $\mathcal{S}$ of k *Right* choices and $k-2$ *Left* choices resulting in a (k)-model with the intervals $I_x, I_1, I_2, \dots, I_{2k-2}$, in the order they were built. If this same string $\mathcal{S}$ is taken to construct a member of the $(k+1)$-interval family $\mathcal{F}_{k+1}$, interval $(0, k+1)$ being the generator, then the last interval I_{2k-2} would be placed as follows:

$$\ell_{2k-2} = \ell_x + k \cdot k - (k-2)(k+1)$$
$$\ell_{2k-2} = k+2$$

Adding x' a twin to x such that $I'_x = (1, k+2)$ and the simplicial $I_s = (-k, 1)$, this $(k+1)$–model results in a member of the $\mathcal{F}_{k+1}$ with the sequence of choices $R\mathcal{S}L$, being the generator the interval I_s. Notice that the number of *Right* choices is $k+1$ and the number of *Left* choices is $k-2+1 = k-1$.

Before adding $\mathcal{S}$, it is easy to prove that x and x' are twins since there is no interval I whose L(I) or RO(I) is I_x. In case this happens, it would contradict that in $\mathcal{F}_{k+1}$ we have the minimum string of R's and L's in which we can create a copy of an interval, by virtue of Lemma 1. By construction, I_s intersects I_x and all intervals with smaller left endpoint. Therefore, a string in $\mathcal{F}_{k+1}$ has been found to represent H. Since the process can be reversed to obtain a member of $\mathcal{F}_k$ from one of $\mathcal{F}_{k+1}$, family $\mathcal{F}_{k+1}$ can be built from $\mathcal{F}_k$.

Let us now prove Theorem 2:

Proof. (2) ⇒ (1) Let $P_l = v_1, v_2, \dots, v_l$, being G an induced subgraph of $(P_l)^{k-1}$. Represent each v_i as a (k)-interval with its left endpoint on the integer i. In this model, remove every interval associated with the vertices that has to be eliminated to construct G.

(1) ⇒ (2) Reverse construction as (2) ⇒ (1). It remains to prove that the bounds of l are correct. Since vertices from $(P_l)^{k-1}$ are removed, then $n \leq l$. To prove the upper bound, we first claim the following:

We claim that for a graph with a unit interval model $\mathcal{M} = \{I_1, I_2, \dots, I_n\}$, intervals index as in the canonical order associated to $V = \{1, 2, \dots, n\}$, there exists an induced path that starts at the vertex 1 and ends at n which realizes the diameter of G.

Suppose the claim does not hold. Let $P = u_1, u_2, \dots, u_l$ be an induced path that realizes d. Therefore, at least one of its endvertices is neither 1 nor n. Assume, without loss of generality, that $u_1 \neq 1$. The proof continues depending on the existence of the edge $(1, u_1)$.

Case $(1, u_1) \in E$: Vertex 1 can also be adjacent only to u_2, since otherwise the path P would not be induced, as noted in Fig. 7. If $(1, u_2) \in E$, the induced path $1, u_2, u_3, \dots, u_l$ has the same length as P and therefore it is a path that realizes the diameter, with first vertex 1. Else, the path $1, u_1, u_2, \dots, u_l$ is an induced path longer than P.

Case $(1, u_1) \notin E$: Let $P' = 1, v_2, v_3, \dots, v_j, u_1$, with $j \geq 2$, be an induced path between vertices 1 and u_1. If there are no edges between P and P', then there exists an induced path longer than P, by joining $P'P$. The only possible edge between these induced paths is (v_j, u_2), since any other would contradict the fact that P and P' are induced paths. The induced path $1, v_2, \dots, v_j, u_2, u_3, \dots, u_l$ is longer than P, which results in a contradiction.

Consider an induced path $P = 1, v_2, \dots, v_d, n$ which satisfies the claim. The model of P of maximum total length is that in which the intersecting intervals are placed overlapping each other by one unit. The left endpoint of all the other intervals

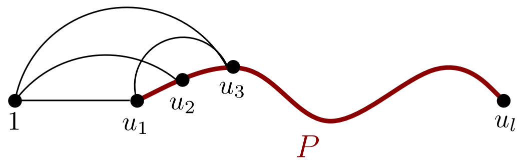

Fig. 7. If there is an edge between 1 and u_3 then all vertices between them must induce a complete subgraph.

must fit between ℓ_1 and r_n as they are labeled in the canonical order. The maximum number of possible intervals without twins between I_1 and I_n is $d \cdot (k-1)+1$.
(1) $\Leftrightarrow$ (3) The proof is by induction on $n = |V|$:

If $n = 1$, an isolated vertex can be represented in a (k)-interval model and does not have any graph in $\mathcal{F}_{k+1}$ as an induced subgraph.

Suppose the claim holds for $i < n$. Let G be a unit interval graph of n vertices without twins. Let us index the vertices in the canonical ordering $\{v_1, v_2, \ldots, v_n\}$. Remove the first vertex v_1, which corresponds to a simplicial vertex. In the remaining graph, by virtue of Lemma 3, there is at most one pair of temporary twins, which must be consecutive, v_j and v_{j+1}; let $\widetilde{G} = G - \{v_1, v_j\}$.

If G is $\mathcal{F}_{k+1}$–free, we will show that G can be represented in a (k)-interval model. Since $\left|V(\widetilde{G})\right| < n$ and $\widetilde{G}$ is free of $\mathcal{F}_{k+1}$ as induced subgraphs, by inductive hypothesis, there exists a (k)-interval model $\mathcal{M} = \{I_2, \ldots, I_{j-1}, I_{j+1}, \ldots I_n\}$ of $\widetilde{G}$, where interval I_i is associated to vertex v_i. Notice that if temporary twins did not appear after removing v_1, the model $\mathcal{M}$ would also include the interval I_j. The proof is done as the removed vertices $\{v_1, v_j\}$ are represented with (k)-intervals in $\mathcal{M}$.

First, in case there is a temporary twin, place I_j initially coinciding with I_{j+1}. Since I_1 must intersect only the I_j, we shall slide I_j one unit to the left. In order to make sure the adjacencies are preserved, we proceed as follows.

After placing I_j coinciding with I_{j+1}, slide all intervals in $\mathcal{L}(I_j) \cup \{I_j\}$ one unit to the left. Now, place I_1 such that $r(I_1) = \ell(I_{j+1})$. Therefore, I_1 intersects I_j but not I_{j+1}, completing the construction. Such a strategy will work if and only if $I_{j+1} \notin \mathcal{L}(I_j)$.

In order to hold that $I_{j+1} \in \mathcal{L}(I_j)$, by virtue of Lemma 1, there must be a string of k R's and $k-2$ L's having I_j as the generator. Then, we would have a graph in $\mathcal{F}_k$ as an induced subgraph of G. After adding the I_{j+1} and I_1, we would be in the hypothesis of Lemma 4 in which we build recursively $\mathcal{F}_{k+1}$ from $\mathcal{F}_k$. So, G is (k)-interval if and only if G is $\mathcal{F}_{k+1}$-free.

In case there is no pair of temporary twins after removing v_1, we add it in $\mathcal{M}$ as follows. Let $j \in [\![1, n]\!]$ be such that $I_{j+1} \in C_2 \setminus C_1$ and $I_j \in C_1$ where C_1 and C_2 denote for the first two maximal cliques. Then, place I_1 such that $I_1 = L(I_{j+1})$.
(4) $\Rightarrow$ (1) Let $\mathcal{M}$ be a $[k-1]$-interval model of G. Let $\widetilde{\mathcal{M}}$ be the model obtained from $\mathcal{M}$ by opening all intervals. Adjacencies of each vertex either remains the same or decreases only in the case the interval shares just an endpoint with another interval. Augment the right endpoint of all intervals by one unit. This transforms $\mathcal{M}$ into a (k)-interval model $\widetilde{\mathcal{M}}$ where $\mathrm{r}(\widetilde{I}) = \mathrm{r}(I)+1$ and $\ell(\widetilde{I}) = \ell(I)\, \forall \widetilde{I} \in \widetilde{\mathcal{M}}$ and $I \in \mathcal{M}$.

It does not change its adjacencies since for $I, J \in \mathcal{M}$ and $\widetilde{I}, \widetilde{J} \in \widetilde{\mathcal{M}}$, I precedes $J \Leftrightarrow \mathrm{r}(I) < \ell(J) \Leftrightarrow \mathrm{r}(\widetilde{I}) - 1 < \ell(\widetilde{J}) \Leftrightarrow \mathrm{r}(\widetilde{I}) \leq \ell(J) \Leftrightarrow \widetilde{I}$ precedes $\widetilde{J}$.
(1) $\Rightarrow$ (4) Symmetrical to the proof of (4) $\Rightarrow$ (1), closing the intervals and decreasing their right endpoint by one unit.

Theorem 2 plays a central role in this characterization. The proof of (3) $\Rightarrow$ (1) is constructive; i.e, it describes how to construct a (k)-interval model (in quadratic-time) for a given $\mathcal{F}_{k+1}$–free unit interval graph. We apply this construction in the algorithm to solve the associated optimization problem of finding a (k)-interval model of minimum k and minimum total length, described in the next section.

3. Minimum representation and complexity

In this section, we prove that the (k)-interval model constructed in the previous section minimizes the total length, and we present a quadratic-time algorithm to find the minimum k such that a unit interval graph is a (k)-interval graph.

3.1. Model of minimum total length

Given a (k)-interval graph G, a vertex v of G, and a positive integer T, we define the parameter MAXL(v) as the rightmost left endpoint of the corresponding interval I_v of v among all possible (k)-interval models $\mathcal{M}$ of G such that Right$(\mathcal{M}) = T$. We assume hereafter that T is some positive constant. Using the uniqueness of the ordering of left endpoints of a unit interval model with no twins, it is easy to prove that, for a fixed integer T, a (k)-interval model of minimum total length will be that which has its first interval as right as possible.

Lemma 5. *Let $G = (V, E)$ be a connected (k)-interval graph and $\mathcal{M}$ a (k)–interval model. If $\ell(I_v) = MAXL(v)$ for every vertex $v \in V$ and its corresponding interval $I_v \in \mathcal{M}$, then $\mathcal{M}$ is a minimum length model of G.*

Next, we will summarize the steps of the main algorithm of our work. This algorithm, given a unit interval model of a graph G and a fixed k as an input, finds in quadratic-time a (k)-interval model of minimum total length if G is (k)-interval or a forbidden subgraph for a graph being (k)-interval otherwise.

The (k)-interval model described in Theorem 2 consists in disassembling the unit interval model by removing a simplicial interval and one twin of the possible pair of twins appearing after the removal. Then, the model can be reconstructed inductively by adding the eliminated twin and the simplicial vertex to the (k)-interval model given by induction, keeping the model as (k)-interval. This construction can be computed in $\mathcal{O}(n^2)$–time. In [1], an $\mathcal{O}(n+m)$-time algorithm is described to recognize unit interval graphs, removing all twins of a graph can be done in linear time [4]. In addition, it is proved that the model constructed by Theorem 2 satisfies the hypothesis of Lemma 5 and thus it is a (k)-interval model of minimum total length.

Using the uniqueness of the ordering of unit interval graphs, it is easy to prove the following:

Lemma 6. *Let G be a (k)-interval graph without twins and $\mathcal{M}$ a (k)–model. If $r(I_v) \leq \ell(I_u)$, then $MAXL(v) \leq MAXL(u) - k$.*

Proposition 7. *The model obtained from Theorem (2), (1) $\Leftrightarrow$ (3), satisfies Lemma 5.*

Proof. By induction on $|V|$:

The basis step is trivial. Inductive step: Consider G a (k)-interval graph and $\widetilde{G} = G - \{v_1, v_j\}$ with the model $\mathcal{M}$ as described in Theorem 2. By inductive hypothesis, $\widetilde{G}$ satisfies Lemma 5. In order to obtain a (k)-interval model for G, I_1 and I_j are added to $\mathcal{M}$ as follows:

- I_j satisfies that $\ell_j = \ell_{j+1} - 1$ and, by the inductive hypothesis, $\ell_{j+1} = \mathrm{MAXL}(v_{j+1})$. As G does not contain twins, $\mathrm{MAXL}(v_j) \leq \mathrm{MAXL}(v_{j+1}) - 1 = \ell_j$. Therefore, $\ell_j = \mathrm{MAXL}(v_j)$.
- $I_1 = \mathrm{L}(I_{j+1})$, then $\ell_1 = \ell_{j+1} - k = \mathrm{MAXL}(I_{j+1}) - k$, using the inductive hypothesis. By Lemma 6, $\mathrm{MAXL}(v_1) \leq \mathrm{MAXL}(v_{j+1}) - k = \ell_1$. Therefore, $\ell_1 = \mathrm{MAXL}(v_1)$.

3.2. *Finding the minimum k*

We present an algorithm to find the minimum k for which a unit interval graph can be represented with a (k)-interval model, and that outputs a (k)-interval model of minimum total length.

Recall that all intervals in $\mathcal{L}_k(I)$ have length k and, if I is moved to the left, must be moved along with I the same number of units.

Given G a (k)-interval graph without twins, G is also a $(k+1)$–interval graph. Considering the algorithm detailed in Theorem 2, (1) $\Leftrightarrow$ (3), in the constructions of the (k) and the $(k+1)$–interval model, at every iteration the sets $\mathcal{L}_k(I)$ and $\mathcal{L}_{k+1}(I)$ will coincide. For a (k)-interval graph, in the algorithm proposed at every iteration we left shift the set $\mathcal{L}(I)$, where I is the last twin we have modeled. As $\mathcal{L}_k(I) = \mathcal{L}_{k+1}(I)$, we have the same construction for a (k) as for a $(k+1)$–interval model. Thus, the relative positions between the temporary twins is the same for both models.

Therefore, we can find the minimum k, for a unit interval graph without twins, with the following simple sketch of algorithm. If it is a connected (1)–interval graph, then it is an isolated vertex since for two or more vertices it would have twins. Thus, we start with length $k=2$, place the last temporary twins I_j, I_{j+1} in the elimination of the induction in Theorem 2, (1) $\Leftrightarrow$ (3), slide one unit to the left the twin I_j and all intervals in $\mathcal{L}(I_j)$ as long as $I_{j+1} \notin \mathcal{L}(I_j)$. Place the interval associated to the simplicial vertex v such that $I_v = L(I_{j+1})$. Continue with the next temporary twins.

Everytime $I_{j+1} \in \mathcal{L}(I_j)$, we are running into a member of $\mathcal{F}_{k+1}$ and, therefore, the graph is not (k)-interval. In that case, in order to enlarge in one unit the intervals already placed, do a sweep along the intervals in the order they were built. In the sweep place every temporary twins I_j, I_{j+1} preserving their relative position and the simplicial vertex v such that $I_v = L(I_{j+1})$, all of length $k+1$. The complexity of this sweep is $\mathcal{O}(n)$. Increase k in one unit and continue with the next temporary twins until the model is finished.

To sum up, the minimum k and a (k)-interval model of minimum total length can easily be computed in $\mathcal{O}(n^2)$–time.

4. Conclusions

We have presented several characterizations of (k)-interval graphs and $[k-1]$-interval graphs for every $k \in \mathbb{N}$. We generate the family $\mathcal{F}_{k+1}$ of forbidden induced subgraphs by particular sequences of intervals of length $k+1$. Furthermore, we construct a positive certificate for their recognition giving a (k)-interval model of minimum total length, for the minimum k possible or for a fixed k. This can be computed in $\mathcal{O}(n^2)$-time with a simple algorithm.

We mention here some directions for future investigations. First, characterize the unit interval graphs that admit a $[k)$-or a $(k]$-interval model, that is, models in which the endpoints of the intervals are closed-open or open-closed, respectively. Second, characterize thoroughly the families $\mathcal{F}_k$ removing isomorphic graphs. Third, characterize interval graphs that admit two different lengths with integer endpoints.

Acknowledgments

The authors thank the anonymous referees for their generous observations, questions and comments that helped to improve the paper. The first author was partially supported by Complex Engineering Systems Institute, Chile (CONICYT – PIA – FB0816; ICM P-05-004-F). The first, second and third authors were partially supported by ANPCyT PICT-2012-1324 and UBACyT Grant 20020130100808BA (Argentina). The third author was partially supported by PIO CONICET UNGS–144-20140100011-CO. The fourth author was partially supported by FAPERJ. The fifth author was partially supported by CNPq and CAPES.

References

[1] D.G. Corneil, H. Kim, S. Natarajan, S. Olariu, A.P. Sprague, Simple linear time recognition of unit interval graphs, Inf. Process. Lett. 55 (2) (1995) 99–104.
[2] V. Costa, S. Dantas, D. Sankoff, X. Xu, Gene clusters as intersections of powers of paths, J. Braz. Comput. Soc. 18 (2) (2012) 129–136.
[3] G. Durán, F. Fernández Slezak, L. Grippo, F. Oliveira, J.L. Szwarcfiter, On unit interval graphs with integer endpoints, Electron. Notes Discrete Math. 50 (2015) 445–450.
[4] M. Habib, C. Paul, L. Viennot, A synthesis on partition refinement: A useful routine for strings, graphs, Boolean matrices and automata, in: 15th Annual STACS, Paris, France, 1998, pp. 25–38.
[5] D. Kratsch, R.M. McConnell, K. Mehlhorn, J.P. Spinrad, Certifying algorithms for recognizing interval graphs and permutation graphs, SIAM J. Comput. 36 (2) (2006) 326–353.
[6] C. Lekkerkerker, D. Boland, Representation of finite graphs by a set of intervals on the real line, Fund. Math. 51 (1962) 45–64.
[7] J. Mitas, Minimal Representation of Semiorders with Intervals of Same Length Orders, Algorithms and Applications, in: LNCS, vol. 831, Springer-Verlag, 1994, pp. 162–175.
[8] M. Pirlot, Minimal representation of a semiorder, Theory and Decision 28 (1990) 109–141.
[9] F.S. Roberts, Indifference graphs, in: Proof Techniques in Graph Theory, in: Proc. Second Ann Arbor Graph Theory Conf. Ann Arbor, Mich. Academic Press, New York, 1969, pp. 139–146.
[10] F. Soulignac, Bounded, Minimal, and Short Representations of Unit Interval and Unit Circular-arc Graphs, 2014. http://arxiv.org/abs/12076960.
[11] D.B. West, Introduction To Graph Theory, second ed., Prentice Hall, 2001.