REPORTS
IN
INFORMATICS

ISSN 0333-3590

A new representation of proper interval graphs with an application to clique-width

Pinar Heggernes Daniel Meister
Charis Papadopoulos

REPORT NO 354 June 2007

Department of Informatics
UNIVERSITY OF BERGEN
Bergen, Norway

This report has URL http://www.ii.uib.no/publikasjoner/texrap/pdf/2007-354.pdf

Reports in Informatics from Department of Informatics, University of Bergen, Norway, is available at http://www.ii.uib.no/publikasjoner/texrap/.

Requests for paper copies of this report can be sent to:

Department of Informatics, University of Bergen, Høyteknologisenteret,
P.O. Box 7800, N-5020 Bergen, Norway

A new representation of proper interval graphs with an application to clique-width*

Pinar Heggernes† Daniel Meister† Charis Papadopoulos†

Abstract

We introduce a new representation of proper interval graphs that can be computed in linear time, and stored in $\mathcal{O}(n)$ space. This representation is a two dimensional vertex partition and it is particularly interesting with respect to clique-width. Based on this, we prove new non-trivial upper bounds on the clique-width of proper interval graphs.

1 Introduction

Proper interval graphs are the intersection graphs of intervals of the real line where no interval properly contains another. The latter property is equivalent to all intervals being of the same (unit) size [18], hence these graphs are also referred to as unit interval graphs. Among applications of proper interval graphs is the Physical Mapping of DNA, or genome reconstruction, where fragments of a chromosome and the overlap information between pairs of these fragments are used to get information on the arrangement of genes on the chromosome [9, 19], and in some biological frameworks the fragments are always of the same length [13]. This graph class has been subject to extensive theoretical study, and there are several representations and many characterisations of proper interval graphs. In this paper we give a new representation of them that can be seen as a generalisation of previous representations, and we show how some problems can be solved very easily on proper interval graphs by finding simple patterns in this representation.

An important characterisation of proper interval graphs is through proper interval orderings [15]. If a proper interval ordering is given together with the leftmost neighbour of each vertex according to the ordering then this is an $\mathcal{O}(n)$-space representation of proper interval graphs that can be computed in linear time [3, 7]. Also other characterisations of this graph class exist through vertex orderings, with properties that neighbourhoods of each vertex, or the cliques of the graph, appearing consecutively, or the ordering and its reverse being a perfect elimination ordering [10]. These can all be seen as equivalent to proper interval orderings. During the last decade, many linear-time recognition algorithms for proper interval graphs have been developed. More recent ones generate vertex orderings that happen to be one of the above kind if and only if the input graph is a proper interval graph [2, 16, 12, 17]. Most of these algorithms are elegantly based on special breadth-first search (BFS) strategies. Other recognition algorithms also apply BFS strategies but with a different approach: for every connected component, find a vertex of special kind and run BFS starting with this vertex. A graph is then a proper interval graph if and only if the BFS-levels are cliques and the neighbourhoods between consecutive levels satisfy the so called chain property [3, 14]. On the representation side, these latter algorithms compute an ordered

*This work is supported by the Research Council of Norway through grant 166429/V30.

†Department of Informatics, University of Bergen, N-5020 Bergen, Norway. Emails: pinar@ii.uib.no, danielm@ii.uib.no, charis@ii.uib.no

vertex partition and verify neighbourhood properties. Similar to the linear orderings mentioned above, these partitions can be turned into graph representation by adding adjacency information.

For the representation of proper interval graphs that we introduce in this paper, we define a 2-dimensional structure similar to a matrix. The elements, called *bubbles*, are sets of vertices, and they define a partition of the vertex set of graph. Two vertices are adjacent only if they belong to bubbles appearing in the same column or in consecutive columns. This new representation is called a *bubble model*, and the exact definition is given in Section 3. The two types of representations mentioned above, orderings and vertex partitions, are captured by the bubble model, which means that those representations can be "embedded" into our representation, hence it generalises previous representations. Besides this, the bubble model provides structural information that can particularly be useful for the design of simple algorithms. To illustrate this we discuss the listing of maximal cliques and maximal independent sets in Section 5. We show that finding such objects reduces to finding simple patterns in the bubble model. Before this we explain that our new representation of proper interval graphs can be computed in linear time from an adjacency list representation of the graph, and in $\mathcal{O}(n)$ time from a proper interval ordering representation of the graph. It requires $\mathcal{O}(n)$ space through a compact storage scheme that we will explain in Section 4.

In Section 6, we address a more challenging algorithmic task, and we give two new non-trivial upper bounds for the clique-width of proper interval graphs. Clique-width is a graph parameter similar to treewidth and pathwidth, that can be used for measuring the complexity of problems. In particular, all problems that can be expressed in a certain kind of monadic second order logic can be solved in linear time on graph classes whose clique-width is bounded by a constant. Naturally, computing the clique-width in general is an NP-hard problem, even when restricted to complements of bipartite graphs [8]. Courcelle and Olariu showed that the clique-width of a graph cannot be more than $2^{t+1} + 1$ for t the treewidth of the graph [6]. Corneil and Rotics improved this bound slightly and showed that a dramatic improvement is not possible [4]. Fellows et al. showed that the clique-width of a graph is bounded by its pathwidth plus 2 [8], which automatically gives that the clique-width of a proper interval graph is at most the size of its largest clique plus 1. By now, grids are the only class known with unbounded clique-width for which the clique-width can be computed in polynomial time [11]. Golumbic and Rotics showed that proper interval graphs have unbounded clique-width [11], hence upper bounds are particularly of interest for this graph class. In this paper, we give two new upper bounds on the clique-width of proper interval graphs that are not based on pathwidth or maximum clique size. We show that there are graphs on which our results give a better upper bound than the size of the maximum clique. In particular, we show that the clique-width of a proper interval graph is bounded by the size of its largest independent set, which also enables us to give a tight bound on the clique-width of co-chain graphs. Our new representation is of essential importance for proving this bound. Furthermore, we are able to efficiently construct a clique-width expression that corresponds to the computed bound. This is a non-trivial task in general, since such bounds may also follow from purely combinatorial arguments, and not necessarily have constructive proofs.

2 Preliminaries

We consider undirected finite graphs with no loops or multiple edges. For a graph $G = (V, E)$, we denote its vertex and edge set by $V(G) = V$ and $E(G) = E$, respectively, with $n = |V|$ and $m = |E|$. For a vertex subset $S \subseteq V$, the *subgraph of G induced by S* is denoted by $G[S]$. Moreover, we denote by $G - S$ the graph $G[V \setminus S]$ and by $G - v$ the graph $G[V \setminus \{v\}]$.

The *neighbourhood* of a vertex x of G is $N_G(x) = \{v \mid xv \in E\}$. The *closed neighbourhood* of x is $N_G[x] = N_G(x) \cup \{x\}$. The *degree* of a vertex x in a graph G is $d_G(x) = |N_G(x)|$. For $S \subseteq V$,

$N_G(S) = \bigcup_{x \in S} N_G(x) \setminus S$. Two vertices x, y of G are called *true twins* if $N_G[x] = N_G[y]$.

A graph is *connected* if there is a path between any pair of vertices. A *connected component* of a disconnected graph is a maximal connected subgraph of it. A *clique* is a set of pairwise adjacent vertices, while an *independent set* is a set of pairwise non-adjacent vertices. The *clique number* of G, $\omega(G)$, is the size of a largest clique in G, and the *independent set number* of G, $\alpha(G)$, is the size of a largest independent set.

The notion of *clique-width* of graphs was first introduced by Courcelle, Engelfriet, and Rozenberg in [5]. The *clique-width* of a graph G, denoted by $\mathrm{cwd}(G)$, is defined as the minimum number of labels needed to construct G, using the following operations:

(i) Creation of a new vertex v with label i, denoted by $i(v)$;

(ii) Disjoint union, denoted by $\oplus$;

(iii) Changing all labels i to j, denoted by $\rho_{i \to j}$;

(iv) Adding edges between all vertices with label i and all vertices with label j, $i \neq j$, denoted by $\eta_{i,j} = \eta_{j,i}$.

An expression built by using the above four operations is called a *clique-width expression*. If k labels are used in a clique-width expression then it is called a k-expression. We say that a k-expression t *defines* a graph G if G is isomorphic to the graph obtained by using the operations in t in the order given by t.

A graph $G = (V, E)$ is called *proper interval graph* if every vertex of G can be assigned an interval of the real line such that no interval is properly contained in another (equivalently, every interval has a unit length [18]), and two vertices are adjacent if and only if their corresponding intervals overlap. An *ordering* σ on an arbitrary graph $G = (V, E)$ is a permutation of V. We write $u \prec_\sigma v$ if u appears before v in the ordering. Ordering σ is called a *proper interval ordering* if for every triple u, v, w of vertices of G where $u \prec_\sigma v \prec_\sigma w$, the following condition is satisfied: $uw \in E \ \Rightarrow \ uv \in E$ and $vw \in E$. We call this condition the *umbrella* property.

Theorem 1 ([15]). *A graph G is a proper interval graph if and only if G has a proper interval ordering.*

A *proper interval ordering representation* of G is obtained by storing both the vertices in the order σ and for each vertex its leftmost neighbour according to σ. It is not difficult to see that if a vertex ordering σ is a proper interval ordering, so is its reversal. If a connected proper interval graph G has no twin vertices then G admits exactly two proper interval orderings [7]. Thus a proper interval ordering of a connected graph with no true twins is unique up to reversal.

3 A representation of proper interval graphs – bubble models

We introduce a new representation of proper interval graphs that places vertices in a 2-dimensional structure. Let A be a finite and non-empty set. A 2*-dimensional bubbles structure* $\mathcal{B}$ for A is a 2-dimensional arrangement of *bubbles*, $\langle B_{i,j} \rangle_{1 \leq j \leq k, 1 \leq i \leq r_j}$, and every bubble $B_{i,j}$ contains a subset of A where every object of A appears in exactly one bubble. Some bubbles may be empty. To give an intuition, bubbles are put into a matrix-like setting, and bubble $B_{i,j}$ appears in row i and column j. For every $j \in \{1, \ldots, k\}$, bubbles $B_{1,j}, \ldots, B_{r_j,j}$ are grouped to the j-th column of $\mathcal{B}$. Column j starts with bubble $B_{1,j}$ and ends with bubble $B_{r_j,j}$.

Definition 1. *Let A be a finite and non-empty set. Let $\mathcal{B} = \langle B_{i,j} \rangle_{1 \leq j \leq k, 1 \leq i \leq r_j}$ be a 2-dimensional bubbles structure for A. The* graph defined by $\mathcal{B}$*, denoted as $G(\mathcal{B})$, is defined as follows:*

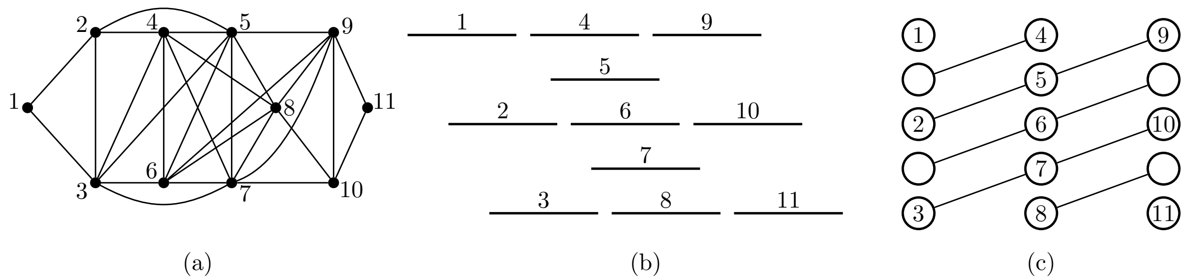

Figure 1: (a) A proper interval graph G, (b) a proper interval model for G and (c) a bubble model for G.

(1) $G(\mathcal{B})$ has a vertex for every element in A, and

(2) uv is an edge of $G(\mathcal{B})$ if and only if there are indices i, j, i', j' such that $a_u \in B_{i,j}$ and $a_v \in B_{i',j'}$, where a_u and a_v are the elements of A corresponding to u and v, respectively, and $|j - j'| \le 1$ and one of the two conditions holds: either $j = j'$ or $(i - i') \cdot (j - j') < 0$.

In particular, adjacent vertices of the graph defined by a 2-dimensional bubbles structure are contained in the same column or in neighbouring columns. It follows directly that vertices that appear in the same column form a clique. Furthermore, vertices in the same bubble are true twins, since they are adjacent and they have the same neighbours in the neighbouring columns. An alternative definition for adjacency of vertices in neighbouring columns is: $u \in B_{i,j}$ and $v \in B_{i',j+1}$ are adjacent if and only if $i > i'$. We can say that u is in a "lower row" than v.

Definition 2. *A* bubble model *for a graph $G = (V, E)$ is a 2-dimensional bubbles structure $\mathcal{B}$ for V such that $G = G(\mathcal{B})$.*

Figure 1 shows an example of a proper interval graph, that is represented by a proper interval model, and a bubble model for the same graph. The line segments between bubbles in neighbouring columns emphasise the neighbourhood property. For example, vertex 10 is adjacent to 7 and 8 but not to 4, 5, 6.

A first property of bubble models might be that the columns give a partition of the vertex set of the defined graph. And if the defined graph is connected, there is no column containing no vertex. Let $G = (V, E)$ be a graph, and let $\mathcal{B} = \langle B_{i,j} \rangle_{1 \le j \le k, 1 \le i \le r_j}$ be a bubble model for G. $\mathcal{B}$ defines a partial order on the vertices of G: we say that vertex u is *to the left* of v, if $u \in B_{i,j}$ and $v \in B_{i',j'}$ and $j < j'$ or $j = j'$ and $i < i'$. Informally, u is to the left of v if the column of u is before the column of v or if the two vertices are in the same column but the bubble containing u has a lower row index. Vertices in the same bubble are not comparable. We call this partial order the *partial order defined by $\mathcal{B}$*. A *linear extension* of a partial order is an ordering that obeys the partial order. A linear extension of the partial order defined by a bubble model just adds orders on vertices contained in the same bubble.

Theorem 2. *A graph is a proper interval graph if and only if it has a bubble model.*

Proof We show two implications. Let $G = (V, E)$ be a graph, and let $\mathcal{B} = \langle B_{i,j} \rangle_{1 \le i \le k, 1 \le j \le r_i}$ be a bubble model for G. Let P be the partial order on V defined by $\mathcal{B}$. Let σ be a vertex ordering for G defining a linear extension of P. We show that σ is a proper interval ordering for G, which means we have to verify the umbrella property for every triple of vertices. Let u and w be adjacent vertices of G, $u \prec_\sigma w$, and let i, j, i'', j'' be such that $u \in B_{i,j}$ and $w \in B_{i'',j''}$. If $j = j''$, the definition of σ shows that $i \le i''$, and if $j \ne j''$, Definition 1 shows that $j'' = j + 1$ and $i'' < i$. Let v be a vertex of G such that $u \prec_\sigma v \prec_\sigma w$. We need to show that v is a neighbour of u and w in G. Let i', j' such that $v \in B_{i',j'}$. With arguments analogous to the ones above, it follows that $j \le j' \le j''$ and $i \le i'$ (if $j = j'$) or $i' \le i''$

(if $j' = j''$). Clearly, v is adjacent to u (if $j = j'$) or w (if $j' = j''$). If $j = j' = j''$ then u, v and w are in the same column and pairwise adjacent. If $j = j' = j'' - 1$ then $i'' + 1 \leq i \leq i'$, i.e., $i'' < i'$, and v and w are adjacent in G. If $j + 1 = j' = j''$ then $i' \leq i'' \leq i - 1$, i.e., $i' < i$, and v and u are adjacent in G. Hence, u, v and w are pairwise adjacent in G. It follows that σ satisfies the umbrella property, i.e., σ is a proper interval ordering for G, thus G is a proper interval graph due to Theorem 1. Since σ was an arbitrary linear extension of the partial order defined by $\mathcal{B}$, we conclude that G is a proper interval graph.

For the converse, we give an algorithm that computes a bubble model by iteratively adding vertices. Let $G = (V, E)$ be a proper interval graph, and let $\sigma = \langle x_1, \ldots, x_n \rangle$ be a proper interval ordering for G. Let $G_i =_{\text{def}} G[\{x_1, \ldots, x_i\}]$ for $i \in \{1, \ldots, n\}$. Based on σ, we define 2-dimensional bubbles structures, $\mathcal{B}_1, \ldots, \mathcal{B}_n$, for which we will prove that they are bubble models for $G_1, \ldots, G_n$, respectively. A crucial property of the constructed bubbles structures will be that they respect ordering σ. Bubbles structure $\mathcal{B}_1$ contains only one bubble, that contains vertex x_1. Obviously, $\mathcal{B}_1$ is a bubble model for G_1. Since $\mathcal{B}_1$ contains exactly one column, it satisfies the property that every vertex in a column that is not the first column has a non-neighbour in the previous column. We call this property the "non-neighbour" property. We show by induction that there is a bubble model $\mathcal{B}_i$ for G_i for every $i \in \{1, \ldots, n\}$ that has the following properties:

- $\mathcal{B}_i$ respects the order defined by $\langle x_1, \ldots, x_i \rangle$
- $\mathcal{B}_i$ has the non-neighbour property.

It is obvious that $\mathcal{B}_1$ has both properties. Now, assume that $\mathcal{B}_i$ has been defined and proved to be a bubble model for G_i, $i \in \{1, \ldots, n-1\}$, that has the two properties. For constructing $\mathcal{B}_{i+1}$, we distinguish three cases with respect to x_{i+1}.

(a) x_{i+1} is a true twin in G_{i+1}. Then, by the definition of proper interval orderings, x_{i+1} and x_i are true twins in G_{i+1}. To obtain $\mathcal{B}_{i+1}$, add vertex x_{i+1} to the bubble of $\mathcal{B}_i$ containing x_i. Since $\mathcal{B}_i$ is a bubble model for G_i and x_i and x_{i+1} have the same closed neighbourhood in the graph defined by $\mathcal{B}_{i+1}$, $\mathcal{B}_{i+1}$ is a bubble model for G_{i+1} that has the two properties.

The two other cases assume that x_{i+1} is not a true twin vertex in G_{i+1}. This particularly means that, if x_i and x_{i+1} are adjacent, x_i has a neighbour in G_{i+1} that is not a neighbour of x_{i+1} in G_{i+1}. This distinction vertex can be in the last or second last column of $\mathcal{B}_i$.

(b) x_{i+1} is not a true twin vertex in G_{i+1} and it is adjacent to every vertex in the last column of $\mathcal{B}_i$. Let $\mathcal{B}_i$ have k columns. By the discussion above, $k \geq 2$. Let a be largest possible such that the bubble at position $(a, k-1)$ in $\mathcal{B}_i$ contains a non-neighbour of x_{i+1}. Note that a exists, since x_i has a neighbour in column $k - 1$ that is not a neighbour of x_{i+1}. Since $\mathcal{B}_i$ respects the order defined by $\langle x_1, \ldots, x_i \rangle$ and according to the umbrella property of σ, all vertices in bubbles below row a in column $k - 1$ are neighbours of x_{i+1} and x_i. Let column k of $\mathcal{B}_i$ contain r bubbles. Certainly, x_i is contained in the bubble at position (r, k). Definition 1 then shows that $a > r$. To define the bubble model for G_{i+1}, we distinguish two cases.

(b1) Assume that the non-empty bubble at position $(a, k - 1)$ contains only non-neighbours of x_{i+1}. Let $\mathcal{B}_{i+1}$ be obtained from $\mathcal{B}_i$ by adding $a-1-r$ many empty bubbles to column k and then adding a bubble that contains x_{i+1}. This means that x_{i+1} is contained in the bubble at position (a, k) in $\mathcal{B}_{i+1}$. By the considerations above, $\mathcal{B}_{i+1}$ is indeed a bubble model for G_{i+1}, since the neighbours of x_{i+1} appear consecutively in the ordering $\langle x_1, \ldots, x_{i+1} \rangle$. Furthermore, $\mathcal{B}_{i+1}$ has the two properties.

(b2) Assume that the non-empty bubble at position $(a, k-1)$ contains a neighbour of x_{i+1}. Then, the vertices in this bubble are not true twins in G_{i+1}, which means that they have to be in different bubbles. We modify $\mathcal{B}_i$ by *splitting* the bubble at position $(a, k-1)$. We first add an "empty row" in the following way: shift all bubbles below row a by one row further down and add empty bubbles in row $a+1$ of every column (unless a column has no bubble below row a). Let the result be $\mathcal{B}'_i$. It is easy to check that $\mathcal{B}'_i$ is still a bubble model for G_i. Now, obtain $\mathcal{B}''_i$ from $\mathcal{B}'_i$ by moving the neighbours of x_{i+1} in the bubble at position $(a, k-1)$ to the bubble at position $(a+1, k-1)$. Since the bubble at position $(a+1, k-2)$ is empty (if it exists), $\mathcal{B}''_i$ is also a bubble model for G_i, and the neighbours of x_{i+1} in column $k-1$ are exactly the vertices in the bubbles below row a. By the umbrella property, $\mathcal{B}''_i$ respects the order defined by $\langle x_1, \ldots, x_i \rangle$, so that we can conclude that $\mathcal{B}''_i$ has the two properties. Now, $\mathcal{B}''_i$ also satisfies the condition of case (b1), which we apply to obtain $\mathcal{B}_{i+1}$.

(c) x_{i+1} is not a true twin vertex in G_{i+1} and it is not adjacent to a vertex in the last column of $\mathcal{B}_i$. This case is analogous to case (b) with the difference that a new column is added. The rest is similar and obtained by just transferring the arguments for column $k-1$ to column k.

The final bubble model for G then is $\mathcal{B}_n$. ∎

The proof of Theorem 2 also shows that a vertex ordering for a proper interval graph is a proper interval ordering for the graph if and only if it is a linear extension of the partial order defined by a bubble model of the graph. As a second result, we obtain that a bubble model of a proper interval graph can be computed in $\mathcal{O}(n^2)$ time. In the next section, we will study a special type of bubble models, that have desirable properties which are useful for the design of algorithms.

4 Restricting bubble models for algorithmic purposes

Regarding representations of graph classes in general, for characterisation purposes and efficient design of algorithms it is common to define and use a restricted version of the representation. Observe that a proper interval graph can have many different bubble models. In this section we define a bubble models that are minimal with respect to a vertex being placed in the first possible bubble. We will see that a proper interval graph has a unique representation (up to reversal) through such bubble models.

For the definition of the restricted bubble model, we need the following relation on bubble models. Let $\mathcal{B}$ and $\mathcal{B}'$ be bubble models for a graph G with bubbles $B_{i,j}$ and $B'_{i,j}$, respectively. We say that a vertex x *improves its position in* $\mathcal{B}'$ *with respect to* $\mathcal{B}$, if $x \in B_{i,j}$ and $x \in B'_{i',j'}$ and $j' < j$ (x appears in an earlier column in $\mathcal{B}'$ with respect to $\mathcal{B}$) or $j = j'$ and $i' < i$ (x appears in the same column in $\mathcal{B}$ and $\mathcal{B}'$ but "improves" its row number).

Definition 3. *Let $\mathcal{B}$ be a bubble model for a graph $G = (V, E)$, and let P be the partial order on V defined by $\mathcal{B}$. We call $\mathcal{B}$ a* first-fit bubble model *(*ff-bubble model*), for G if and only if no column ends with an empty bubble and there is no bubble model for G that respects P and contains a vertex that improves its position with respect to $\mathcal{B}$.*

It is clear that a graph has a first-fit bubble model if and only if it has a bubble model. A first observation is that true twins appear in the same bubble of an ff-bubble model. Thus a connected proper interval graph has exactly two ff-bubble models by the corresponding uniqueness of its proper interval ordering. The following result provides a characterisation of ff-bubble models through patterns in the model.

Theorem 3. *Let $G = (V, E)$ be a graph, and let $\mathcal{B} = \langle B_{i,j} \rangle_{1 \le j \le k, 1 \le i \le r_j}$ be a bubble model for G. $\mathcal{B}$ is an ff-bubble model for G if and only if $\mathcal{B}$ does not satisfy any of the following conditions:*

(1) a column of $\mathcal{B}$ ends with an empty bubble

(2) there are $j \in \{2, \ldots, k\}$ and $i \in \{1, \ldots, r_j\}$ such that $B_{1,j-1}, \ldots, B_{i,j-1}$ all are empty and $B_{i,j}$ is non-empty

(3) there are $j \in \{1, \ldots, k\}$ and $i \in \{2, \ldots, r_j\}$ such that $B_{i,j-1}$ and $B_{i-1,j+1}$ are empty and $B_{i,j}$ is non-empty.

Proof We first show that $\mathcal{B}$ cannot be an ff-bubble model for G, if it satisfies any of the conditions (1–3). If $\mathcal{B}$ satisfies condition (1), then $\mathcal{B}$ is clearly not an ff-bubble model for G according to Definition 3. Let P be the partial order on V defined by $\mathcal{B}$. For the two other conditions, we explicitly construct a bubble model for G that respects P and contains a vertex that improves its position. Assume that $\mathcal{B}$ does not satisfy condition (1).

(a) Assume that $\mathcal{B}$ satisfies condition (3). Let $j \in \{1, \ldots, k\}$ and $i \in \{2, \ldots, r_j\}$ such that $B_{i,j-1}$ and $B_{i-1,j+1}$ are empty and $B_{i,j}$ is non-empty. We obtain $\mathcal{B}'$ from $\mathcal{B}$ by the following operation: move the vertices in $B_{i,j}$ to $B_{i-1,j}$ and if $i = r_j$, delete $B_{i,j}$. It is obvious that $\mathcal{B}'$ is a bubble model for G, since $\mathcal{B}'$ defines the same neighbourhood for the vertices in $B_{i,j}$ as $\mathcal{B}$. Furthermore, $\mathcal{B}'$ clearly respects P; to be precise, the partial order defined by $\mathcal{B}'$ is equal to P. Since $\mathcal{B}$ does not satisfy condition (1), $\mathcal{B}'$ does not satisfy condition (1) either. The vertices in $B_{i,j}$ have improved their position in $\mathcal{B}'$ with respect to $\mathcal{B}$, so that $\mathcal{B}$ is not an ff-bubble model according to Definition 3. Note that it is crucial for the conclusion that $B_{i,j}$ is a non-empty bubble.

(b) Assume that $\mathcal{B}$ satisfies condition (2). Let $j \in \{2, \ldots, k\}$ and $i \in \{1, \ldots, r_j\}$ such that $B_{1,j-1}, \ldots, B_{i,j-1}$ are empty and $B_{i,j}$ is non-empty. Choose j largest possible and then i smallest possible. In particular, $B_{1,j}, \ldots, B_{i-1,j}$ are empty and $B_{1,j+1}, \ldots, B_{i-1,j+1}$ are empty. If $i \ge 2$ then $\mathcal{B}$ satisfies condition (3) with the bubbles $B_{i,j-1}, B_{i,j}, B_{i-1,j+1}$, and $\mathcal{B}$ is not an ff-bubble model for G due to case (a). Let $i = 1$. We construct $\mathcal{B}'$ from $\mathcal{B}$ in the following way: let $r =_{\text{def}} \max\{r_{j-2} - 1, r_{j-1}, r_j\}$ (where we assume $r_0 = 0$); add a bubble at position $(r + 1, j - 1)$ containing the vertices in $B_{i,j}$ (if necessary, i.e., if $r > r_{j-1}$, add empty bubbles to column $j - 1$) and make $B_{i,j}$ into an empty bubble. By assumption and choice of j and i, the vertices in $B_{i,j}$ are adjacent in G exactly to the vertices in columns $j - 1$ and j. By the definition of r, these vertices still have the same neighbourhoods in $\mathcal{B}'$. Thus, $\mathcal{B}'$ is a bubble model for G and respects P. If column j of $\mathcal{B}'$ contains a non-empty bubble, no column of $\mathcal{B}'$ ends with an empty bubble, and we conclude analogous to case (a) that $\mathcal{B}$ is not an ff-bubble model. If column j of $\mathcal{B}'$ does not contain a non-empty bubble, $B_{i,j}$ is the only non-empty bubble in column j of $\mathcal{B}$. Modify $\mathcal{B}'$ by adding r empty bubbles to the top of columns $j + 1, \ldots, k$ and then delete column j. The result is a bubble model for G with the required properties.

We conclude that a bubble model satisfying one of the three conditions (1–3) is not an ff-bubble model for G.

For the converse, assume that $\mathcal{B}$ does not contain a column that ends with an empty bubble. Otherwise, $\mathcal{B}$ satisfies condition (1). Furthermore, assume that $\mathcal{B}$ does not satisfy condition (2). Let $\mathcal{B}' = \langle B'_{i,j} \rangle_{1 \le j \le k', 1 \le i \le r'_j}$ be a bubble model for G respecting P (the partial order defined by $\mathcal{B}$) and containing a vertex that improves its position with respect to $\mathcal{B}$. By assumption about satisfaction of condition (2), $\mathcal{B}$ and $\mathcal{B}'$ have the same number of columns, i.e., $k = k'$, and the vertex sets of corresponding columns are equal (otherwise the leftmost column of $\mathcal{B}$ containing a vertex that improves its position in $\mathcal{B}'$ with respect to $\mathcal{B}$ by jumping to another column contains a pattern satisfying condition (2)). Let

$j \in \{1, \ldots, k\}$ and $i \in \{2, \ldots, r_j\}$ be such that $B_{i,j}$ contains a vertex x that improves its position in $\mathcal{B}'$ with respect to $\mathcal{B}$. We show that $\mathcal{B}$ satisfies condition (3). Consider the following marking procedure:

> mark $B_{i,j}$. If $j+1 \leq k$ and $i-1 \geq 1$ and $B_{i-1,j+1}$ is non-empty, mark $B_{i-1,j+1}$ and repeat this marking step with $B_{i-1,j+1}$ in place of $B_{i,j}$.

Consider the last marked bubble $B_{a,b}$. Suppose that $a = 1$. Note that every vertex in the marked bubbles must improve its position in $\mathcal{B}'$ with respect to $\mathcal{B}$, since the vertices in $B_{i,j}$ improve their positions. However, the vertices in $B_{1,b}$ can improve their positions only by moving to the previous column, and this contradicts our assumptions. Hence, $a > 1$, and $B_{a-1,b+1}$ is empty. If $B_{a,b-1}$ is empty, $\mathcal{B}$ satisfies condition (3). Let $B_{a,b-1}$ be non-empty. Since the vertices in $B_{a,b}$ improve their positions in $\mathcal{B}'$ with respect to $\mathcal{B}$, the vertices in $B_{a,b-1}$ must improve their positions in $\mathcal{B}'$ with respect to $\mathcal{B}$, as well. Otherwise, the vertices in $B_{a,b}$ and $B_{a,b-1}$ would be adjacent in the graph defined by $\mathcal{B}'$, which contradicts $\mathcal{B}'$ being a bubble model for G. Start at $B_{a,b-1}$ and execute the marking procedure; let $B_{a',b'}$ be the bubble on which the procedure stops. With similar arguments as above, $a' > 1$ and $b' \leq k$ and $B_{a',b'}$ is non-empty and $B_{a'-1,b'+1}$ is empty. If $B_{a',b'-1}$ is empty, $\mathcal{B}$ satisfies condition (3). If $B_{a',b'-1}$ is non-empty, we repeat the construction. Since the start row always decreases, there must be an iteration that finds a pattern in $\mathcal{B}$ that satisfies condition (3). If not, the vertices in $B_{i,j}$ could not improve their positions in $\mathcal{B}'$ with respect to $\mathcal{B}$. Thus, if $\mathcal{B}$ is not an ff-bubble model for G then $\mathcal{B}$ satisfies one of the conditions (1–3), which concludes the proof. ∎

The proof of Theorem 3 implicitly gives an algorithm for obtaining an ff-bubble model from a given bubble model: find a forbidden pattern and apply the appropriate operation from the proof. This algorithm is certainly polynomial in the size of the input bubble model. Note that the size of an arbitrary bubble model is not related to the size of the represented graph, since empty bubbles destroy every relationship.

The next result presents an $\mathcal{O}(n)$ time algorithm for computing an ff-bubble model of a proper interval graph. Since an ff-bubble model can have $\Theta(n^2)$ bubbles, our algorithm does not output the model in the full representation. Observe that at most n bubbles of a bubble model are non-empty. This leads to the following space efficient representation. A *compact representation* of a bubble model only lists the non-empty bubbles which are partitioned into columns and ordered, and additionally stores the row number of each bubble. To reconstruct the actual bubble model from a given compact representation, the number of empty bubbles preceding a non-empty bubble B has to be computed, and this number is determined by the row number of B and the bubble preceding B. A compact representation of a bubble model requires then only linear space in the number of vertices, which is comparable to previous representations of proper interval graphs.

Theorem 4. *Let G be a proper interval graph, given in a proper interval ordering representation. An ff-bubble model for G can be computed in $\mathcal{O}(n)$ time, that is output in compact representation.*

Proof We give an algorithm for computing an ff-bubble model, that runs in two phases. The first phase will compute the vertex partition into columns and find dependencies between non-empty bubbles in different columns, and the second phase assigns the proper row to each non-empty bubble. Let σ be the given proper interval ordering.

For the first phase, we apply the algorithm for computing a bubble model from the proof of Theorem 2, where non-empty bubbles are not added. The result $\mathcal{B}$ is a 2-dimensional bubbles structure, that is obtained from the output of the algorithm by deleting all empty bubbles. Note that $\mathcal{B}$ also has the two properties considered in the proof of Theorem 2 ($\mathcal{B}$ respects the order defined by σ, and it has the non-neighbour property).

In the second phase of the algorithm, we compute the row number for every bubble of $\mathcal{B}$, which gives the compact representation of an ff-bubble model. We define a directed graph, which will be acyclic, for which we can compute a topological ordering. Following this topological ordering, we obtain the row numbers by easy calculations. Let H have a vertex for every bubble of $\mathcal{B}$ and the following arcs:

> let x be a vertex of H, then there are arcs pointing to x from the vertices corresponding to these bubbles: downmost bubble containing a non-neighbour in the preceding column, bubble preceding x in the same column, downmost bubble containing a neighbour in the next column. If a described bubble does not exist, the corresponding arc does not exist.

In particular, the vertices corresponding to the bubbles of a column form an induced directed path in H. We show some properties of H.

Claim. *(1) H is acyclic and (2) H contains exactly one source vertex.*

Proof (1) Suppose the contrary and let H contain a (directed) cycle $C = (y_1, \ldots, y_l)$. We show that H then also contains a cycle of a restricted form. Let C contain vertices from the columns $b', \ldots, b''$. Since vertices from different columns are connected only if the columns are consecutive, C really contains a vertex from every of the columns $b', \ldots, b''$. Suppose there are numbers a' and a'' where $a' < a''$ such that $y_{a'}$ and $y_{a''}$ are in the same column b and not all of the vertices $y_{a'+1}, \ldots, y_{a''-1}$ and $y_{a''+1}, \ldots, y_{a'-1}$ are in column b (here, we mean two subpaths in C, so that indices have to be adjusted). Informally, C leaves column b on the path from $y_{a'}$ to $y_{a''}$ and on the path from $y_{a''}$ to $y_{a'}$. Without loss of generality, we can assume that there is a path from $y_{a'}$ to $y_{a''}$ in H containing only vertices from column b (otherwise, re-index the vertices of C). Furthermore, we can assume that C does not contain another vertex from C on the $y_{a'}, y_{a''}$-path in column b. Obtain C' from C by replacing the path $(y_{a'}, \ldots, y_{a''})$ by the path from $y_{a'}$ to $y_{a''}$ containing only vertices from column b. Note that this path is unique. Repeated application of this argument shows the existence of a cycle C^* in H where the vertices from a column appear consecutively. And this implies that C^* contains vertices from only two columns (as we have already noticed the vertices of a single column induce a path). Let $C^* = (z_1, \ldots, z_q)$, and let p be such that $z_1, \ldots, z_p$ are in one column, say b^*, and $z_{p+1}, \ldots, z_q$ are in the other column, $b^* + 1$. By construction, z_q corresponds to the downmost bubble in column $b^* + 1$ containing a neighbour of the vertices in the bubble corresponding to z_1, and z_p corresponds to the downmost bubble in column b^* containing a non-neighbour of the vertices in the bubble corresponding to z_{p+1}. Let u_1, u_2, u_3, u_4 be vertices from the bubbles corresponding to z_1, z_p, z_{p+1}, z_q. Note that u_1 and u_2 as well as u_3 and u_4 might be equal. It follows that u_1 and u_4 are adjacent and u_2 and u_3 are non-adjacent in G. Now, from the definition of a proper interval ordering, the vertices from u_1 to u_4 in σ form a clique, and this contradicts u_2 and u_3 being non-adjacent. Hence, H is acyclic.

(2) By construction, every vertex of H not corresponding to the first bubble in a column of $\mathcal{B}$ is endpoint of an arc from the vertex corresponding to the previous bubble in the same column. Hence, source vertices in H can correspond only to first bubbles in columns. A vertex is in the first bubble of a column, that is not the first column, if there is a non-neighbour in the previous column. Thus, every vertex of H corresponding to the first bubble of a column that is not the first column is endpoint of an arc. So, at most one vertex of H is not endpoint of an arc, and since every acyclic graph has a source vertex, H contains exactly one source vertex. This source vertex corresponds to the first bubble of the first column of $\mathcal{B}$. □

Let τ be a topological ordering for H. Assign 1 to the unique source vertex of H and assign numbers to the other vertices of H that are computed as follows: let u be a vertex of H and let v_1, v_2, v_3 be the predecessors of u, where v_1 corresponds to a bubble in the column preceding the column for u. Let a_1, a_2, a_3 be the numbers already assigned to v_1, v_2, v_3, respectively. Then, $\max\{a_1 - 1, a_2, a_3\} + 1$ is assigned to u. Since the number assignment can follow τ, every vertex can be assigned a number. Let $\mathcal{B}^*$ be the bubbles structure obtained from $\mathcal{B}$ by placing the bubbles in the rows determined by the assigned numbers and filling the space between these bubbles with empty bubbles.

Claim. *(1) $\mathcal{B}^*$ is a bubble model for G and (2) $\mathcal{B}^*$ is an ff-bubble model for G.*

Proof (1) We prove the statement in two steps: we first consider adjacent vertices, and then we consider non-adjacent vertices. But note first that $\mathcal{B}^*$ respects the order defined by σ, since the row number of a bubble in a column is greater than the row number of the preceding bubble. Furthermore, the construction of $\mathcal{B}$ and the properties of proper interval orderings guarantee that the vertices in a column of $\mathcal{B}^*$ are pairwise adjacent in G. So, we have to consider only vertices in different columns. Let u and v be adjacent vertices in G, and let u be in column b of $\mathcal{B}^*$ and v in column $b+1$. Let u and v be in bubbles B and B' in rows a and a', respectively. Let a'' be the row of the downmost bubble in column $b+1$ containing a neighbour of v. By construction, the vertex in H corresponding to B is endpoint of an arc from the vertex corresponding to the bubble at position $(a'', b+1)$ in $\mathcal{B}^*$. Hence, $a > a''$, and since $a'' \geq a'$, we obtain $a > a'$, which means that u and v are adjacent in the graph defined by $\mathcal{B}^*$. Now, let u and v be non-adjacent vertices. As seen above, u and v are not contained in the same column of $\mathcal{B}^*$. If u and v are not contained in consecutive columns, u and v are non-adjacent in the graph defined by $\mathcal{B}^*$. So, let u and v be contained in consecutive columns, in the bubbles at positions (a, b) and $(a', b+1)$, respectively. Let a'' be the row number of the downmost bubble in column b containing a non-neighbour of v. By construction, $a \geq a''$. And by definition of proper interval orderings and properties above, $a'' \geq a'$, so that we conclude $a \geq a'$. Hence, u and v are non-adjacent in the graph defined by $\mathcal{B}^*$. Thus, $\mathcal{B}^*$ is a bubble model for G.

(2) We prove the statement by applying the characterisation of Theorem 3. By construction, $\mathcal{B}^*$ does not satisfy condition (1) of Theorem 3, and the non-neighbour property for $\mathcal{B}$ guarantees that $\mathcal{B}^*$ does not satisfy condition (2). So, it remain to prove that $\mathcal{B}^*$ does not satisfy condition (3). Suppose the contrary and let B, B', B'', B''' be bubbles at positions $(a, b), (a, b-1), (a-1, b), (a-1, b+1)$ in $\mathcal{B}^*$ such that B is non-empty and B' and B''' are empty. Then, B'' must be non-empty. This means that the vertices in B and B'' are true twins in G (according to statement (1), $\mathcal{B}^*$ is a bubble model for G), which contradicts the construction of $\mathcal{B}$: true twin vertices of G are true twins in every induced subgraph containing these vertices, so that the algorithm puts these vertices into the same bubble. Since no later added vertex implies a splitting for these vertices, they remain in the same bubble. This is a contradiction, which shows that $\mathcal{B}^*$ does not satisfy condition (3), i.e., $\mathcal{B}^*$ is an ff-bubble model. □

It remains to discuss the running time of the algorithm. The first algorithm phase can be executed in $\mathcal{O}(n)$ time. The computed bubbles structures are implemented as chained lists: the columns are chained lists, and the vertices in a bubble are stored as chained list, an element for every vertex. The list for a bubble keeps the vertices in order defined by σ. For case (a) (algorithm in the proof of Theorem 2), x_i and x_{i+1} are true twins if and only if they have the same leftmost neighbour. Adding a vertex to an already existing bubble is a simple list operation. For cases (b–c), adding a new bubble to a column and adding a new column are also simple list operations. Distinguishing between the two cases is done by looking at the leftmost neighbour. The leftmost neighbour also determines the bubble that has to be split. Visiting the vertices in the order defined by σ starting with the non-neighbours finds the splitting point in time proportional to the number of non-neighbours in the bubble. This is due to the umbrella property of proper interval orderings and the vertex order in the chained list of the bubble. However, non-neighbours of x_{i+1} are non-neighbours of every further vertex, which means that the bubble containing the visited non-neighbours will never be visited again. Hence, $\mathcal{B}$ can be obtained in $\mathcal{O}(n)$ time.

For the second algorithm step, the crucial part is to construct graph H. We have to find (at most) three bubbles for every bubble. The downmost bubble in the previous column containing a non-neighbour is the bubble containing x_{d-1} where x_d is the leftmost neighbour. The preceding bubble in the same column is directly taken from the list. And the downmost bubble in the next column containing a neighbour is obtained by the inverse leftmost neighbour. Hence, H can be constructed in $\mathcal{O}(n)$ time. Note that H has at most $3n$ arcs, and since a topological ordering of an acyclic graph can be computed in linear time, a topological ordering for H is obtained in $\mathcal{O}(n)$ time. The assignment of the row numbers then is straightforward in $\mathcal{O}(n)$ time. We conclude that a first-fit model for G can be computed in $\mathcal{O}(n)$ time given a proper interval ordering with leftmost neighbour assignment. ∎

5 Maximal independent sets and cliques in proper interval graphs

In this section, we give two applications of ff-bubble models. We show that all maximal independent sets and cliques can be listed by scanning the bubble model from left to right and making only local decisions. The main advantage is that the characterisation of maximal independent sets and cliques relies on finding local patterns, which only consider columns, rows and emptiness. We first consider maximal independent sets.

Proposition 5. *Let G be a proper interval graph, and let $\mathcal{B} = \langle B_{i,j} \rangle_{1 \leq j \leq k, 1 \leq i \leq r_j}$ be an ff-bubble model for G. Then, $\{u_1, \ldots, u_l\}$ is a maximal independent set of size l of G if and only if there are integers $a_1, b_1, \ldots, a_l, b_l$ and a permutation π over $\{1, \ldots, l\}$ such that $u_{\pi(i)} \in B_{a_i, b_i}$ for every $i \in \{1, \ldots, l\}$ and the following three conditions hold for every $i \in \{2, \ldots, l\}$:*

(1) $b_1 = 1$

(2) if $b_i \leq b_{i-1} + 1$ *then* $b_i = b_{i-1} + 1$ *and* $a_i \geq a_{i-1}$

(3) if $b_i > b_{i-1} + 1$ *then* $b_i = b_{i-1} + 2$ *and either* $a_i < a_{i-1}$ *or* $B_{a_{i-1}, b_{i-1}+1} \cup \cdots \cup B_{a_i, b_{i-1}+1} = \emptyset$.

Proof Let $U = \{u_1, \ldots, u_l\}$ be a maximal independent set in G. Since every column of $\mathcal{B}$ defines a clique in G, the vertices in U are in different columns. Let π be the permutation over $\{1, \ldots, l\}$ that orders the vertices according to the columns they appear in. Let a_i, b_i be such that $u_{\pi(i)} \in B_{a_i, b_i}$ for every $i \in \{1, \ldots, l\}$. As an intermediate result, we know that $b_1 < \cdots < b_l$. Let $i \in \{2, \ldots, l\}$. If $b_i \leq b_{i-1} + 1$ then $b_i = b_{i-1} + 1$, which directly follows from $b_{i-1} < b_i$, and $a_i \geq a_{i-1}$, which follows from the adjacency definition for vertices in consecutive columns. Let $b_i > b_{i-1} + 1$. If $b_i > b_{i-1} + 2$ then we can choose a vertex in column $b_i + 2$ that is not adjacent to any vertex in column $b_i + 3$. Such a vertex exists due to condition (2) of Theorem 3, and it is not adjacent to any vertex in U. Then, however, U cannot be a maximal independent set in G. Hence, $b_i = b_{i-1} + 2$. If $a_i \geq a_{i-1}$ and one of the bubbles $B_{a_{i-1}, b_{i-1}+1}, \ldots, B_{a_i, b_{i-1}+1}$ is non-empty then a vertex can be picked that can be added to U and still obtain an independent set. Thus we know that $a_i < a_{i-1}$ or all these bubbles are empty. Finally, condition (2) of Theorem 3 implies that U must contain a vertex contained in column 1. The converse directly follows with similar arguments. ∎

Corollary 6. *Let G be a proper interval graph, and let $\mathcal{B}$ be an ff-bubble model for G. The number of columns of $\mathcal{B}$ is equal to $\alpha(G)$.*

Proof According to Proposition 5, the cardinality of independent sets of G is bounded above by the number of columns of $\mathcal{B}$. For equality, let U be a set containing a vertex from the topmost non-empty bubble of every column of $\mathcal{B}$. Condition (2) of Theorem 3 shows that no column of $\mathcal{B}$ contains only empty bubbles and that U indeed is an independent set. Hence, the number of columns of $\mathcal{B}$ is bounded above by the independent-set number of G, which shows the result. ∎

It follows directly that ff-bubble models have the smallest possible number of columns among all bubble models for a proper interval graph.

Maximal cliques can be listed even more easily in ff-bubble models compared to maximal independent sets, as shown in the following result.

Proposition 7. *Let G be a proper interval graph, and let $\mathcal{B} = \langle B_{i,j} \rangle_{1 \le j \le k, 1 \le i \le r_j}$ be an ff-bubble model for G. Then U is a maximal clique of G if and only if there are a, a', b where $a \ge 1$ and $a' > a$ and $1 \le b \le k$ such that the following three conditions hold:*

(1) $U = B_{a+1,b-1} \cup \cdots \cup B_{r_{b-1},b-1} \cup B_{1,b} \cup \cdots \cup B_{a,b}$

(2) $B_{a,b}$ *is non-empty*

(3) either $a = r_b$ *or* $B_{a+1,b} \cup \cdots \cup B_{a'-1,b} = \emptyset$ *and* $B_{a',b} \neq \emptyset$ *and* $B_{a+1,b-1} \cup \cdots \cup B_{a',b-1} \neq \emptyset$.

Proof First, let $b \in \{1, \ldots, k\}$ and $a \ge 1$ and $a' > a$ such that the assumptions of the lemma are satisfied. According to Definition 2, $B_{a+1,b-1} \cup \cdots \cup B_{r_{b-1},b-1}$ and $B_{1,b} \cup \cdots \cup B_{a,b}$ are cliques in G. Furthermore, the vertices of both sets are pairwise adjacent, so that the union U of both sets is also a clique in G. If there is a vertex x such that $U \cup \{x\}$ is a clique in G then x must be vertex in column $b-1$ or b; otherwise x cannot be adjacent to vertices in these two columns. If x is vertex in column $b-1$ then it is contained in $B_{1,b-1} \cup \cdots \cup B_{a,b-1}$ and therefore not adjacent to a vertex in $B_{a,b}$, which is non-empty by assumption. If x is vertex in column b then we know that $a < r_b$ and x is contained in $B_{a',b} \cup \cdots \cup B_{r_b,b}$ by assumption. But then there is a vertex of U contained in column $b-1$ which is not adjacent to x. Hence, U is a maximal clique in G. For the converse, let U be a maximal clique in G. Since vertices in non-consecutive columns are non-adjacent and U is non-empty, there are numbers a, b such that $B_{a,b}$ is non-empty and U is contained in $U' =_{\text{def}} B_{a+1,b-1} \cup \cdots \cup B_{r_{b-1},b-1} \cup B_{1,b} \cup \cdots \cup B_{a,b}$. Since U' is a clique, U and U' are equal by maximality of U. Let $a \neq r_b$. Then, U contains a vertex that is not adjacent to all vertices in $B_{r_b,b}$, which means that $b > 1$. Let a' be smallest such that $a' > a$ and $B_{a',b}$ is non-empty, and let x be a vertex in $B_{a',b}$. Clearly, $B_{a+1,b}, \ldots, B_{a'-1,b}$ are empty by the choice of a'. Since x is not in U and U is a maximal clique, U contains a vertex y that is non-adjacent t x. Since y is adjacent to the vertices of $B_{a,b}$, y is contained in $B_{a+1,b-1} \cup \cdots \cup B_{a',b-1}$, and we conclude the proof. ∎

6 Upper bounds on the clique-width of proper interval graphs

In this section we show two different approaches for computing a clique-width expression of a proper interval graph. For computing both of them efficiently we take advantage of an ff-bubble model. It is well-known that for any graph G, $\text{cwd}(G) = \max\{\text{cwd}(G') \mid G' \text{ is a connected component of } G\}$ [6]. Hence for the rest of the section we will consider connected proper interval graphs. First we give an upper bound on the clique-width with respect to the number of columns of an ff-bubble model of a proper interval graph G. Recall that this number is unique by Corollary 6.

Theorem 8. *Let G be a proper interval graph. Then $cwd(G) \le \alpha(G) + 1$. Moreover, given a compact representation of an ff-bubble model for G, an $(\alpha(G) + 1)$-expression defining G can be constructed in $\mathcal{O}(n)$ time.*

Proof Let $\mathcal{B} = \langle B_{i,j} \rangle_{1 \le j \le k, 1 \le i \le r_j}$ be an ff-bubble model for G. First we give an algorithm for constructing a $(k+1)$-expression, which defines G. We use labels $1, \ldots, k+1$, where the first k labels are assigned to the columns of $\mathcal{B}$, and label $k+1$ is used to add a new vertex. Our algorithm visits the bubbles of $\mathcal{B}$ row by row in a top-down manner and within a row from left to right. Vertices in the same bubble are treated sequentially. This defines an ordering on the vertices of G. Let x be the first vertex. The expression for $G[\{x\}]$ is $1(x)$. Now, let y be a vertex of G, and assume that the expression t for the graph on the vertices preceding y has already been defined. Let y be in a bubble in column j. The expression for the graph induced by the vertices not succeeding y then is

$$\rho_{k+1\to j}(\eta_{k+1,j+1}(\eta_{k+1,j}((k+1)(y)\oplus t))).$$

Correctness follows directly from the properties of ff-bubble models, since the neighbours of y are only in the subgraph in columns j and $j+1$. Note that, if $j = k$, the subexpression involving column $j+1$ is obsolete. Corollary 6 then shows that $k = \alpha(G)$, which implies $\text{cwd}(G) \le \alpha(G) + 1$.

For the running time, note that an $(\alpha(G)+1)$-expression for G can be computed in $\mathcal{O}(n)$ time given the vertex ordering. Since vertices in the same bubble appear consecutively in the ordering, it suffices to show that the corresponding ordering of the non-empty bubbles of $\mathcal{B}$ can be computed in $\mathcal{O}(n)$ time. Let r be the largest row number among the non-empty bubbles of $\mathcal{B}$. We start with an array of size r, whose entries point to ordered lists. Every list corresponds to a row of $\mathcal{B}$ containing the non-empty bubbles in the left-right order. The algorithm scans the bubbles of $\mathcal{B}$ in the compact representation column by column from left to right. A bubble is appended to the list corresponding to its row number. It is clear then that the required vertex ordering can be obtained in $\mathcal{O}(n)$ time, which concludes the proof. ∎

As an interesting consequence, we mention that the previous result immediately gives a bound on the clique-width of co-chain graphs, which constitute a subclass of proper interval graphs[1]. Combining the already known results on the clique-width of chain graphs [11] and complements of graphs [6], clique-width of co-chain graphs is at most 6. Hence by the following corollary, we are able to give a better bound. Furthermore, the bound given below is tight, as it is known that cographs are exactly the class of graphs of clique-width at most 2 [6], and there exist co-chain graphs that are not cographs. Observe that every co-chain graph G has $\alpha(G) \le 2$ since G is the complement of a bipartite graph. Therefore the following result holds by combining the fact that a proper interval ordering can be obtained in $\mathcal{O}(n+m)$ time [3, 7], and Theorems 4 and 8.

Corollary 9. *For every co-chain graph G, cwd$(G) \le 3$, and a 3-expression defining it can be constructed in $\mathcal{O}(n+m)$ time.*

Next we proceed with the second approach regarding the clique-width of a proper interval graph. For that purpose we need to introduce the following notion. Let G be a connected proper interval graph let $\mathcal{B}$ be an ff-bubble model for G. A set of non-empty bubbles of a given column j of $\mathcal{B}$ is called a *group* if the vertices of the bubbles have the same neighbourhood in G with respect to the vertices of column $j+1$. By the definition of the bubble models, we know that every vertex of G belongs to exactly one group. Moreover observe that only consecutive non-empty bubbles may belong to the same group. That is, if $B_{i,j}$ and $B_{i',j}$ belong to the same group such that $i < i'$ then every non-empty bubble $B_{i'',j}$ for which $i < i'' < i'$ belongs to the given group. For instance, in Figure 1 every non-empty bubble of the first column defines a group by itself, whereas in the second column there are three groups defined as $\{\{4\},\{5,6\},\{7,8\}\}$, and in the third column altogether the non-empty bubbles define a single group.

The notion of groups gives rise to a new parameter for proper interval graphs that we will call the group number. Let v be a vertex of G and let $L(\mathcal{B}, v)$ be the set of vertices to the left of v or in the same bubble as v (excluding v) in $\mathcal{B}$. Let $n_{\mathcal{B}}(v)$ be the number of groups consisting of vertices from $L(\mathcal{B}, v)$ containing at least one neighbour of v. Notice that for G there are exactly two ff-bubble models, $\mathcal{B}$ and $\mathcal{B}'$. The *group number* of G, denoted by $\varphi(G)$, is defined as follows: $\varphi(G) = \min\{\max_{v\in V} n_{\mathcal{B}}(v)\,,\ \max_{v\in V} n_{\mathcal{B}'}(v)\}$. We show how to compute $\varphi(G)$ in linear time and we prove that it gives an upper bound for the clique-width of G.

[1]We refer to [1, 10] for the definitions of graph classes mentioned in this paragraph.

Theorem 10. *Let G be a proper interval graph. Then $cwd(G) \leq \varphi(G) + 2 \leq \omega(G) + 1$. Moreover, a $(\varphi(G) + 2)$-expression defining G can be constructed in $\mathcal{O}(n + m)$ time.*

Proof First we prove that $\varphi(G) + 1 \leq \omega(G)$. Let $\mathcal{B}$ be an ff-bubble model for G. Let us consider $n_{\mathcal{B}}(v)$ for any vertex v. We call *active groups of* v those groups consisting of vertices from $L(\mathcal{B}, v)$ and that contain at least a neighbour of v. We show that the vertices of the active groups of v together with v form a clique in G for each vertex v. Observe that by the definition of the groups every vertex of an active group is adjacent to v. Moreover the vertices of any group induce a clique in G as they belong to the same column of $\mathcal{B}$. Consider the active group F of v that is furtherest away from v in $\mathcal{B}$. Then, for any other group F' which is between F and v we know that every vertex of F is adjacent to every vertex of F', and furthermore, every vertex of F' is adjacent to v by the properties of $\mathcal{B}$. This means that every group between F and v is active and an easy induction on the groups consisting of vertices of $L(\mathcal{B}, v)$ shows that every vertex of an active group of v is adjacent to v and to every vertex of another active group of v. Thus the vertices of the active groups of v together with v form a clique in G which implies that $\varphi(G) + 1 \leq \omega(G)$.

Now we show that there exists a $(\varphi(G) + 2)$-expression defining G. For a vertex v, a group consisting of vertices from $L(\mathcal{B}, v)$ that is not active is called a *dead group of* v. ¿From the discussion above it follows that a dead group of v cannot become an active group of another vertex u such that v is to the left of u. Without loss of generality assume that $\varphi(G) = \max_{v \in V} n_{\mathcal{B}}(v)$. For every vertex v starting from the leftmost vertex in $\mathcal{B}$, we add v in a proper way to an expression t defined by the vertices of $L(\mathcal{B}, v)$. Let $q = n_{\mathcal{B}}(v)$ for a vertex v. Assume that a label is assigned for every group and all the vertices that belong to the same group have the same label. First we change the labels of all the vertices of the dead groups of v to 1 and then we assign a distinct label from $\{2, \ldots, q + 1\}$ for every active group of v. This can be done by using the appropriate ρ operation at most $q + 1$ times. Next we use label $q + 2$ in order to add v and join the vertices of label $i \in \{2, \ldots, q + 1\}$ with the vertex v of label $q + 2$. Thus $(q + 2)v$ and $\eta_{i,q+2}$ define the appropriate operations. Recall that v is adjacent to every vertex of an active group of v. Finally we put v into the group that v belongs to. Notice that v cannot belong to a dead group of v since G is connected. If v belongs to an already existed group consisting of vertices of $L(\mathcal{B}, v)$ then we change the label $q + 2$ to the label of that group. Otherwise, v does not belong to any of the active groups of v and there is no need to change the expression as v belongs to a group that does not appear in t. We keep applying the same operations for every vertex of $\mathcal{B}$ so that every vertex of G is defined by the expression and the join operations imply edges of G. Therefore, $\text{cwd}(G) \leq \varphi(G) + 2 \leq \omega(G) + 1$.

What is left to prove is the running time for the construction of the given expression. By using one of the linear-time algorithms [3, 7] together with Theorem 4, we obtain a compact representation of an ff-bubble model $\mathcal{B}$ for G. In order to compute the groups of $\mathcal{B}$ we use the following observation. Let B and B' be two non-empty bubbles of column j in $\mathcal{B}$ with row number i and i', respectively, such that $i < i'$. Let B'' be the bubble of column $j+1$ that contains a neighbour of a vertex of B' such that its row number i'' is as large as possible. If $i > i''$ then both B, B' belong to the same group since the vertices of B have the same neighbourhood with the vertices of B'. Otherwise, $i \leq i''$, which implies that B' belongs to a different group of B, since the vertices of B' are adjacent to at least one vertex x of column $j + 1$ so that x is non-adjacent to the vertices of B. We visit every column of the compact representation starting from the first row and we apply the previous observation on two consecutive bubbles in order to compute the groups. Notice that for every bubble B' we need to find the bubble B'' of the next column of maximum row number that is adjacent to B'. This can be done in $\mathcal{O}(d_G(v))$ time for a vertex $v \in B'$ by looking one by one the bubbles of the next column and starting from the first row, since v is adjacent to every vertex of those bubbles. Hence computing the groups takes linear time in the size of G. As explained before, for computing the expression we apply a vertex-incremental approach by visiting the bubbles of $\mathcal{B}$ column by column and within a column in a top-down manner. Having the groups, we require $\mathcal{O}(d_G(v))$ time for every vertex v that we visit, since v is adjacent to every vertex of an active group of v. Thus the overall running time for constructing a $(\varphi(G)+2)$-expression is $\mathcal{O}(n+m)$. ∎

By Theorems 8 and 10 we obtain the following result.

Theorem 11. *For a proper interval graph G, $cwd(G) \leq \min\{\alpha(G)+1, \varphi(G)+2\} \leq \omega(G)+1$.*

We point out that there are proper interval graphs G for which $\varphi(G)$ or $\alpha(G)$ is significantly smaller than $\omega(G)$. An easy example can be derived from the graph G shown in Figure 1 by extending the three columns of the bubble model so that $2\,\varphi(G) = \omega(G)$. Hence our bounds are better than the previously known bound on clique-width of proper interval graphs.

References

[1] A. Brandstädt, V. B. Le, and J. P. Spinrad. Graph classes : A survey. SIAM Monographs on Discrete Mathematics and Applications, 1999.

[2] D. G. Corneil. A simple 3-sweep LBFS algorithm for the recognition of unit interval graphs. Disc. Appl. Math. 138, pp. 371–379, 2004.

[3] D. G. Corneil, H. Kim, S. Natarajan, S. Olariu, and A. P. Sprague. Simple linear time recognition of unit interval graphs. Inf. Proc. Let. 55, pp. 99–104, 1995.

[4] D. G. Corneil and U. Rotics. On the relationship between clique-width and treewidth. SIAM J. Comput. 34, pp. 825–847, 2005.

[5] B. Courcelle, J. Engelfriet, and G. Rozenberg. Handle-rewriting hypergraph grammars. J. Comput. System Sci. 46, pp. 218–270, 1993.

[6] B. Courcelle and S. Olariu. Upper bounds to the clique width of graphs. Discrete Applied Mathematics 101, pp. 77–114, 2000.

[7] X. Deng, P. Hell, and J. Huang. Linear-time representation algorithms for proper circular-arc graphs and proper interval graphs. SIAM J. Comput. 25, pp. 390–403, 1996.

[8] M. R. Fellows, F. A. Rosamond, U. Rotics, and S. Szeider. Clique-width Minimization is NP-hard. Proceedings of the thirty-eighth annual ACM Symposium on Theory of Computing, STOC'06, pp. 354–362, 2006.

[9] P. W. Goldberg, M. C. Golumbic, H. Kaplan, and R. Shamir. Four strikes against physical mapping of DNA. J. Comput. Bio. 2, pp. 139–152, 1995.

[10] M. C. Golumbic. Algorithmic Graph Theory and Perfect Graphs. Second edition. Annals of Discrete Mathematics 57. Elsevier, 2004.

[11] M. C. Golumbic and U. Rotics. On the clique-width of some perfect graph classes. Int. J. Found. Comp. Sci. 11, pp. 423–443, 2000.

[12] P. Hell and J. Huang. Certifying LexBFS Recognition Algorithms for Proper Interval Graphs and Proper Interval Bigraphs. SIAM J. Disc. Math. 18, pp. 554–570, 2005.

[13] P. Hell, R. Shamir, and R. Sharan. A fully dynamic algorithm for recognizing and representing proper interval graphs. SIAM J. Comput. 31, pp. 289–305, 2001.

[14] C. M. Herrera de Figueiredo, J. Meidanis, and C. Picinin de Mello. A linear-time algorithm for proper interval graph recognition. Inf. Proc. Lett. 56, pp. 179–184, 1995.

[15] P. J. Looges and S. Olariu. Optimal greedy algorithms for indifference graphs. Computers & Mathematics with Applications 25, pp. 15–25, 1993.

[16] D. Meister. Recognition and computation of minimal triangulations for AT-free claw-free and co-comparability graphs. Disc. Appl. Math. 146, pp. 193–218, 2005.

[17] B. S. Panda and S. K. Das. A linear time recognition algorithm for proper interval graphs. Inf. Proc. Lett. 87, pp. 153–161, 2003.

[18] F. S. Roberts. Indifference graphs. In: F. Harary (Ed.), Proof techniques in graph theory, pp. 139–146, Academic Press, New York, 1969.

[19] J. Spinrad. Efficient Graph Representations. Fields Institute Monographs 19. ACM, 2003.