Optimal Linear Arrangement of Interval Graphs

Johanne Cohen[1], Fedor Fomin[2], Pinar Heggernes[2], Dieter Kratsch[3], and Gregory Kucherov[4]

[1] LORIA, 54506 Vandoeuvre-lès-Nancy Cedex, France. Johanne.Cohen@loria.fr
[2] Department of Informatics, University of Bergen, 5020 Bergen, Norway.
(Fedor.Fomin|Pinar.Heggernes)@ii.uib.no
[3] LITA, Université de Metz, 57045 Metz Cedex 01, France. kratsch@sciences.univ-metz.fr
[4] LIFL/CNRS, 59655 Villeneuve d'Ascq, France. Gregory.Kucherov@lifl.fr

Abstract. We study the optimal linear arrangement (OLA) problem on interval graphs. Several linear layout problems that are NP-hard on general graphs are solvable in polynomial time on interval graphs. We prove that, quite surprisingly, optimal linear arrangement of interval graphs is NP-hard. The same result holds for permutation graphs. We present a lower bound and a simple and fast 2-approximation algorithm based on any interval model of the input graph.

1 Introduction

A *linear layout* (or simply *layout*) of a given graph $G = (V, E)$ is a linear ordering of its vertices. Assuming that the vertices of G are numbered from 1 to n, a layout is a permutation $L(1), L(2), \ldots, L(n)$. The *weight* of a layout L on G is $\mathcal{W}(G, L) = \sum_{(u,v)\in E} |L(u) - L(v)|$. An *optimal linear arrangement* (OLA) of G is a layout with the minimum weight, i.e., $\mathrm{argmin}_L \mathcal{W}(G, L)$. We denote $\mathcal{W}(G) = \min_L \mathcal{W}(G, L)$ and call it the *minimum weight* on G.

Computing the optimal linear arrangement (the OLA problem) is NP-hard [10], and it remains NP-hard for bipartite graphs [5]. The problem is solvable in polynomial time for trees [6, 3, 18], and for some other restricted graph classes such as grids or hypercubes [4]. There is an approximation algorithm for general graphs with performance ratio $O(\log n)$ [17].

A well-known vertex ordering problem related to OLA is the Bandwidth Minimization problem. The bandwidth of a layout L on G is $b(G, L) = \max_{(u,v)\in E} |L(u) - L(v)|$. The *bandwidth* of G is the minimum bandwidth of any layout of G, i.e., $bw(G) = \min_L b(G, L)$. The bandwidth minimization problem is also NP-hard on general graphs [9]. It remains NP-hard even on the restricted class of trees [16]. Furthermore, for general graphs, bandwidth cannot be approximated by a polynomial time algorithm within a constant factor [20], but it can be approximated in polynomial time with a factor of $O(\log^{9/2} n)$ [8].

It is well known that many NP hard-problems are solvable in polynomial time on interval graphs. In 1985, Johnson wrote in his NP-completeness column: "Indeed, I know of no NP-completeness results for interval graphs, although there are still some possibilities in Table 1, in addition to such naturals as BANDWIDTH and SUBGRAPH ISOMORPHISM" [12]. Interestingly, a bit later, it appeared that the bandwidth minimization problem is solvable in polynomial time for interval graphs. For an interval graph with n vertices given by an interval model, Kleitman and Vohra's algorithm solves the decision problem "Is $bw(G) \leq k$?" in $O(nk)$ time, and it can be used to produce a layout with the minimum bandwidth in $O(n^2 \log n)$ time [13]. Furthermore, Sprague has shown how to implement Kleitman and Vohra's algorithm to answer the decision problem in $O(n \log n)$ time, and thus produce a minimum bandwidth layout in $O(n \log^2 n)$ time [19]. We refer the reader to [4] for a survey of known results on the OLA, bandwidth and other related layout problems.

To our knowledge, optimal linear arrangement of interval graphs has not been studied so far. In this paper, we show that, in contrast to bandwidth minimization, the OLA problem is NP-hard on interval graphs. We also show that the problem can be approximated within a constant factor of 2 by a simple algorithm.

Besides its theoretical interest, the class of interval graphs is widely acknowledged as an important graph class, due to a number of applications. Interval graphs are extensively used in bioinformatics, typically to model the genome physical mapping problem, which is the problem of reconstructing the relative positions of DNA fragments, called *clones*, out of information of their pairwise overlaps (see e.g. [21]). However, interval graphs appear also in other situations in bioinformatics, such as for gene structure prediction for example [1]. In [7], interval graphs are used to model temporal relations in protein-protein interactions. In that paper, an optimal linear arrangement of an interval graph models an "optimal" molecular pathway, and the problem of efficiently computing this arrangement is explicitly raised. This provides a direct motivation for the present study.

This paper is organized as follows. In Section 2, graph notations are introduced. We obtain a lower bound for the minimum weight of a linear arrangement for general graphs in terms of the degrees of the vertices. In Section 3, we prove that the OLA problem is NP-complete for interval graphs. In Section 4, using the lower bound we show that both the left endpoint ordering and the right endpoint ordering of an interval graph are 2-approximations for the Optimal Linear Arrangement problem. In Section 5, we first show that the NP-completeness result holds also for permutation graphs, and then discuss approximation algorithms for OLA of the more general class of cocomparability graphs.

2 Preliminaries

We consider only finite, undirected and simple graphs. For $G = (V, E)$, we will denote $|V|$ as n and $|E|$ as m. We sometimes refer to the vertex set of G as $V(G)$ and the edge set as $E(G)$. We let $N(v)$ denote the set of vertices adjacent to v. The *degree* of a vertex v in graph G, $d_G(v)$, is the number of vertices adjacent to v in G. $\Delta(G)$ denotes the maximum degree of a vertex in graph G. The subgraph of $G = (V, E)$ induced by $V' \subseteq V$ will be referred to as $G[V']$. The complement of a graph G is denoted by $\overline{G}$ and has the same vertex set as G, and $(x, y) \in E(\overline{G})$ if and only if $(x, y) \notin E(G)$.

A layout L of a graph $G = (V, E)$ can be seen as an ordering $(v_1, v_2, \dots, v_n)$ of V, meaning that $L(v_j) = j$, for $1 \leq j \leq n$. We extend this notation to subsets of vertices. Let $V_1, \dots, V_i$ be a partition of V. If a layout L of G has the form $(V_1, \dots V_i)$, then it implies that

– $\forall j, \forall \ell, 1 \leq j < \ell \leq i$, $\forall u \in V_j$, $\forall w \in V_\ell$, $L(u) < L(w)$
– $\forall \ell, 1 \leq \ell \leq i$, the order of L inside V_ℓ is an arbitrary order of V_ℓ.

A graph $G = (V, E)$ is an *interval graph* if there is a one-to-one correspondence between V and a set of intervals of the real line such that, for all $u, v \in V$, $(u, v) \in E$ if and only if the intervals corresponding to u and v have a nonempty intersection. Such a set of intervals $\mathcal{I}$ is called an *interval model* for G. We assume that an interval model is given by a left endpoint and a right endpoint for each interval, namely, $l(v)$ and $r(v)$ for all $v \in V$. Furthermore, we assume that we are also given a sorted list of the endpoints, and that the endpoints are distinct.

First, we study OLA of simple topologies, like stars and complete graphs. A *star*, denoted by S_α, is a tree such that one vertex, called the center, is adjacent to α leaves. A *complete graph*, denoted by K_n, is a graph on n vertices such that all vertices are pairwise adjacent. The following lemmas give the weight of the optimal linear arrangement for these particular topologies.

Lemma 1. *Let K_n be the complete graph on n vertices. Then $\mathcal{W}(K_n) = \frac{(n-1)n(n+1)}{6}$.*

Proof. Straightforward, as all layouts yield the same weight.

Lemma 2. *Let S_α be the star with a center vertex c and α leaves. Then every layout L of S_α satisfies the following:*

– $\lfloor\frac{\alpha}{2}\rfloor(\lfloor\frac{\alpha}{2}\rfloor + 1) \leq \mathcal{W}(S_\alpha, L) \leq \lfloor\frac{\alpha}{2}\rfloor(\alpha + 1)$ *and* $\mathcal{W}(S_\alpha) = \frac{\alpha}{2}(\frac{\alpha}{2} + 1)$*, if α is even,*
– $(\lfloor\frac{\alpha}{2}\rfloor + 1)^2 \leq \mathcal{W}(S_\alpha, L) \leq (\lfloor\frac{\alpha}{2}\rfloor + 1)\alpha$ *and* $\mathcal{W}(S_\alpha) = (\lfloor\frac{\alpha}{2}\rfloor + 1)^2$*, if α is odd,*

and a permutation L is an optimal linear arrangement if and only if L places c at the middle position ($L(c) = \lfloor\frac{\alpha}{2}\rfloor + 1$).

Proof. Assume that $L(c) = k$. Then $\mathcal{W}(S_\alpha, L) = \sum_{i=1}^{k} i + \sum_{i=1}^{a-k} i = (\frac{\alpha^2}{2} + \frac{3\alpha}{2}) + (k^2 - (\alpha + 2)k)$. For the case where α is even, $\mathcal{W}(S_\alpha, L)$ reaches its minimum for $k = \frac{\alpha}{2} + 1$. In this case, $\mathcal{W}(S_\alpha) = \frac{\alpha}{2}(\frac{\alpha}{2} + 1)$. Moreover $\mathcal{W}(S_\alpha, L)$ reaches its maximum for $k = 1$ or for $k = \alpha + 1$. The same arguments can be applied for the case where α is odd.

These results will be needed to prove the NP-completeness of the OLA problem on interval graphs and to give a 2-approximation algorithm for it. The following lower bound for optimal linear arrangement of any graph is obvious, and it will be useful when analyzing the performance ratio of some algorithms.

Lemma 3. *Let $G = (V, E)$ be a graph, $E = E_1 \cup E_2$ and $E_1 \cap E_2 = \emptyset$. Then $\mathcal{W}(G) \geq \mathcal{W}(G_1) + \mathcal{W}(G_2)$, where $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$.*

Corollary 1. *Let $G = (V, E)$, $V = V_1 \cup \cdots \cup V_n$, and $E = E_1 \cup \cdots \cup E_n$, where $E_1, \cdots, E_n$ are pairwise disjoint. Then $\mathcal{W}(G) \geq \mathcal{W}(G_1) + \ldots + \mathcal{W}(G_n)$, where $G_i = (V_i, E_i)$, $1 \leq i \leq n$.*

All these results will be useful to compute the lower and upper bounds of the weight $\mathcal{W}(G, L)$ of a layout L of G. For example, consider a graph G composed of two disjoint complete graphs K_α and K_b and an additional vertex c adjacent to all other vertices of the graph. The set of edges of this graph can be easily partitioned into three sets. From Corollary 1, by construction we have $\mathcal{W}(G) \geq \mathcal{W}(K_b) + \mathcal{W}(K_\alpha) + \mathcal{W}(S_{\alpha+b})$. Moreover, the following layout L of G is considered: $V(K_\alpha), c, V(K_b)$. Layout L has weight $\mathcal{W}(K_b) + \mathcal{W}(K_\alpha) + \mathcal{W}(S_{\alpha+b})$. The previous inequality implies that L is an optimal linear arrangement.

3 The complexity of the OLA problem on interval graphs

The goal of this section is to prove the following theorem.

Theorem 1. *The problem of deciding, for an interval graph $G = (E, V)$ and a constant K, whether $\mathcal{W}(G) \leq K$ is NP-complete.*

The proof will be by reduction from the 3-PARTITION problem [10]:

3-PARTITION
Instance: A finite set A of $3m$ integers $\{a_1, \ldots, a_{3m}\}$, a bound $B \in Z^+$ such that $\sum_{i=1}^{3m} a_i = mB$.
Question: Can A be partitioned into m disjoint sets $A_1, A_2, \ldots, A_m$ such that, for all $1 \leq i \leq m$, $\sum_{a \in A_i} a = B$?

3-PARTITION is known to be NP-complete in the strong sense [10]. Note that we do not require here that each A_i is composed of exactly three elements.

The structure of our proof will be as follows. We first construct a graph $\mathcal{H}(B, m)$ depending on two natural numbers B and m, and we describe the structure of its optimal linear arrangement. In the second part, we describe a polynomial-time reduction from 3-PARTITION, i.e., we encode numbers $\{a_1, \ldots, a_{3m}\}$ by adding some additional edges to graph $\mathcal{H}(B, m)$, and show that an optimal linear arrangement of this extended graph corresponds precisely to a 3-partition of $\{a_1, \ldots, a_{3m}\}$.

For simplicity of notation in our proofs, in this section we will let $\mathcal{K}(n) = \mathcal{W}(K_n)$ and $\mathcal{S}(\alpha) = \mathcal{W}(S_\alpha)$, where K_n is the complete graph on n vertices, and S_α is the star with α leaves.

3.1 Construction of $\mathcal{H}(B, m)$ and its optimal linear arrangement

Let m and B be two integers. We assume that m is even. The set of vertices of $\mathcal{H}(B, m)$ will be the union of several disjoint sets

$$V(\mathcal{H}(B, m)) = R_1 \cup X \cup V \cup Y \cup Z \cup R_2.$$

The number of vertices in each set is defined as follows.

– Each of R_1 and R_2 has $3m^3(B + 1)$ vertices,
– X is the union of disjoint sets $X_1, \ldots, X_{m/2}$, where each X_i has $2(B + 1)$ vertices; similarly, Z is the union of disjoint sets $Z_1, \ldots, Z_{m/2}$, where each Z_i has $2(B + 1)$ vertices,
– V has $(m + 1)$ vertices,
– Y has mB vertices.

The set of edges of $\mathcal{H}(B, m)$ is defined as follows.

– Vertices of $R_1 \cup X$ form a clique, i.e., they are all pairwise adjacent; vertices in R_1 have no other neighbors,
– vertices of $R_2 \cup Z$ form a clique; vertices in R_2 have no other neighbors,

– vertices $V = \{v_1, \dots, v_{m+1}\}$ form a clique,
– for each $1 \le i \le m/2$, v_i is adjacent to all vertices of $X_i \cup \ldots \cup X_{m/2}$,
– for each $1 \le i \le m/2$, v_{m+2-i} is adjacent to all vertices of $Z_i \cup \ldots \cup Z_{m/2}$,
– each vertex of Y is adjacent to all vertices of V,
– $\mathcal{H}(B, m)$ has no edges other than those defined above.

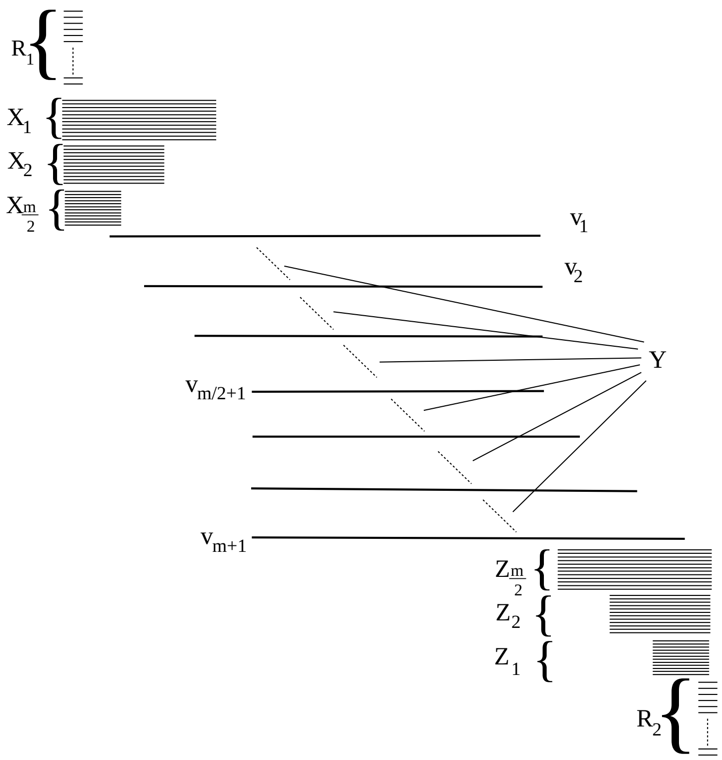

Fig. 1. Interval representation of graph $\mathcal{H}(B, m)$

An interval representation of graph $\mathcal{H}(B, m)$ is given in Figure 1. From this figure, it is clear that $\mathcal{H}(B, m)$ is an interval graph. From Lemma 1, a lower bound on $\mathcal{W}(\mathcal{H}(B, m))$ can be now established as follows.

Lemma 4. $\mathcal{W}(\mathcal{H}(B, m)) \ge 2\mathcal{K}(3m^3(B+1)+m(B+1))+2\sum_{i=1}^{m/2} \mathcal{S}(2(m-i+1)(B+1)-m)+\mathcal{S}(mB)+\mathcal{K}(m+1)$.

Proof. Using Corollary 1, we can estimate the lower bound as follows: $\mathcal{W}(\mathcal{H}(B, m)) \ge \mathcal{K}(|R_1|+|X|)+\sum_{i=1}^{m/2} \mathcal{S}(|X_i|+\ldots+|X_{m/2}|+|Y|)+\sum_{i=1}^{m/2} \mathcal{S}(|Z_i|+\ldots+|Z_{m/2}|+|Y|)+\mathcal{K}(|V|)+\mathcal{S}(|Y|)+\mathcal{K}(|Z|+|R_2|)$. Here terms $\mathcal{K}(|R_1|+|X|)$ and $\mathcal{K}(|Z|+|R_2|)$ correspond to complete graphs formed respectively by vertex sets $R_1 \cup X$ and $Z \cup R_2$. Each term $\mathcal{S}(|X_i|+\ldots+|X_{m/2}|+|Y|)$, $1 \le i \le m/2$, corresponds to the star with center v_i and leaves $X_i \cup \cdots \cup X_{m/2} \cup Y$. Similarly, term $\mathcal{S}(|Z_i|+\ldots+|Z_{m/2}|+|Y|)$, $1 \le i \le m/2$, corresponds to the star with center v_{m+2-i} and leaves $Z_i \cup \cdots \cup Z_{m/2} \cup Y$. Finally term $\mathcal{S}(|Y|)$ corresponds to the star with center $v_{m/2+1}$ and leaves Y, and $\mathcal{K}(|V|)$ corresponds to the clique V. By substituting the cardinalities of the sets, we obtain the bound of Lemma 4.

We now show the following upper bound on $\mathcal{W}(\mathcal{H}(B, m))$.

Lemma 5. $\mathcal{W}(\mathcal{H}(B, m)) \le 2\mathcal{K}(3m^3(B+1)+m(B+1))+2\sum_{i=1}^{m/2} \mathcal{S}(2(m-i+1)(B+1))+\mathcal{S}(m(B+1)) - (B+1)\mathcal{K}(m+1)$.

Proof. Consider the following layout of $\mathcal{H}(B,m)$:

$$R_1, X_1, \cdots, X_{m/2}, v_1, Y_1, v_2, Y_2, \ldots, Y_m, v_{m+1}, Z_{m/2}, \cdots, Z_1, R_2, \tag{1}$$

where $Y_1 \cup \cdots \cup Y_m = Y$, and for each $1 \le i \le m$, $|Y_i| = B$. Observe that the order of vertices inside R_1, X_i, Y_i, Z_i, $1 \le i \le \frac{m}{2}$, and R_2 is irrelevant.

Since vertices in $R_1 \cup X$ and $Z \cup R_2$ are consecutive in the layout, the contribution of cliques $R_1 \cup X$ and $Z \cup R_2$ is respectively $\mathcal{K}(|R_1| + |X|) = \mathcal{K}(3m^3(B+1) + m(B+1))$ and $\mathcal{K}(|Z| + |R_2|) = \mathcal{K}(3m^3(B+1) + m(B+1))$.

Now consider vertices $v_1, \ldots, v_{m/2}$. Each vertex $v_i,\ 1 \le i \le m/2$, has $2(m-i+1)(B+1)$ neighbors in graph $\mathcal{H}(B,m)$: $2(m/2-i+1)(B+1)$ neighbors belonging to $X_i, \ldots, X_{m/2}$, m neighbors $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_{m+1}$, and mB neighbors in Y. Observe now that these $2(m-i+1)(B+1)$ neighbors of v_i appear in (1) at consecutive positions before and after v_i and moreover, v_i appears exactly in the middle of those vertices. This implies that the contribution of each star centered at v_i $1 \le i \le m/2$ is $\mathcal{S}(2(m-i+1)(B+1))$ and the overall contribution is $\sum_{i=1}^{m/2} \mathcal{S}(2(m-i+1)(B+1))$.

Symmetrically, the contribution of the stars centered at $v_{m/2+1}, \ldots, v_{m+1}$ is also $\sum_{i=1}^{m/2} \mathcal{S}(2(m-i+1)(B+1))$. By the same argument, the star with center $v_{m/2}+1$ and leaves $\{v_1, \ldots, v_{m/2}, v_{m/2+2}, \ldots, v_{m+1}\}$ contributes with $\mathcal{S}(m(B+1))$.

Observe that each edge with both endpoints in $\{v_1, \ldots, v_{m+1}\}$ has been counted twice. We therefore have to subtract $(B+1)\mathcal{K}(m+1)$ to take this into account.

By summing up all the terms, we obtain the lemma.

To proceed, we need to estimate from above the difference between the upper (Lemma 5) and lower (Lemma 4) bounds. By straightforward arithmetics, one can establish that for any x and $y \le x$, we have $\mathcal{S}(x) - \mathcal{S}(x-y) \le xy$. Using this, the difference between the upper and lower bounds is

$$\begin{aligned}
&2\sum_{i=1}^{m/2}[\mathcal{S}(2(m-i+1)(B+1)) - \mathcal{S}(2(m-i+1)(B+1)-m)] + [\mathcal{S}(m(B+1)) - \mathcal{S}(mB)] - \\
&\qquad (B+2)\mathcal{K}(m+1) \le 2\sum_{i=1}^{m/2} 2(m-i+1)(B+1)m + m^2(B+1) - (B+2)\mathcal{K}(m+1) \le \\
&\qquad 4m(B+1)\sum_{i=1}^{m/2}(m-i+1) + m^2(B+1) - (B+2)m(m+1)(m+2)/6 < 3m^3(B+1)
\end{aligned} \tag{2}$$

The next step is to prove that layout (1) of Lemma 5 is actually an optimal linear arrangement. Let L^* by an optimal linear arrangement of $\mathcal{H}(B,m)$. We first show that L^* maps vertices of $R_1 \cup X$ to consecutive positions.

Lemma 6. *Let L^* be an optimal linear arrangement of $\mathcal{H}(B,m)$. Then the set $\{L^*(w) | w \in R_1 \cup X\}$ contains $|R_1| + |X|$ consecutive integers.*

Proof. Assume for contradiction that some vertex from $V \cup Y \cup R_2$ appears at a position p which is between the smallest and the largest positions of $\{L^*(w) | w \in R_1 \cup X\}$. Then the contribution of each edge of $\{(w_1,w_2) | w_1, w_2 \in R_1 \cup X, L^*(w_1) < p, L^*(w_2) > p\}$ is increased by at least one. The total increase is then at least $\min_{1 \le L \le |R_1|+|X|-1}(L \cdot (|R_1| + |X| - L)) = |R_1| + |X| - 1 = 3m^3(B+1) + m(B+1) - 1$. Observe now that this quantity is larger than the maximal possible difference (2) between the upper and the lower bound on $\mathcal{W}(\mathcal{H}(B,m))$, which gives the desired contradiction.

Lemma 7. *Let L^* be an optimal linear arrangement of $\mathcal{H}(B,m)$. Then the set $\{L^*(w) | w \in Z \cup R_2\}$ contains $|Z| + |R_2|$ consecutive integers.*

Proof. By symmetry, the proof is similar to that of Lemma 6.

Thus, Lemmas 6 and 7 imply that any optimal linear arrangement maps vertices of $R_1 \cup X$ and $Z \cup R_2$ into sets of consecutive positions. By an argument similar to that of Lemma 6, we further deduce that vertices of $R_1 \cup X$ appear in the beginning of an optimal layout, and vertices of $Z \cup R_2$ appear in the end of this layout, while the other vertices ($V \cup Y$) appear between them. Indeed, if it is not the case, edges "crossing" $R_1 \cup X$ (or $Z \cup R_2$) would give an increase in the weight that would be larger than the maximal possible difference (2) between the upper and the lower bound.

To further specify an optimal linear arrangement of $\mathcal{H}(B, m)$, we have to clarify the layout of $V \cup Y$. The following lemma completes this part of the proof.

Lemma 8. *Any optimal linear arrangement of $\mathcal{H}(B, m)$ has the form*

$$R_1 \cup X, v_1, Y_1, v_2, Y_2, \ldots, Y_m, v_{m+1}, Z \cup R_2, \tag{3}$$

where $Y_1 \cup \cdots \cup Y_m = Y$ and for each $1 \leq i \leq m$, $|Y_i| = B$.

Proof. It is easy to see that v_1 appears immediately after $R_1 \cup X$, as otherwise it can be moved down to that position which only decreases the resulting weight. By symmetry, v_{m+1} appears immediately before $Z \cup R_2$. From similar considerations, we can deduce that the ordering of vertices in V is the "natural" ordering $v_1, v_2, \ldots, v_{m+1}$ (otherwise by permuting the vertices we would decrease the total weight).

It remains only to show that between each v_i and v_{i+1} there are exactly B vertices of Y. If this is the case, then observe (see the proof of Lemma 5) that each star centered at v_i has exactly the same number of neighbors to the left of $L^*(v_i)$ as to the right of $L^*(v_i)$, and all these neighbors appear at consecutive positions. Thus, each star centered at v_i is optimally arranged and reaches the absolute lower bound of the contributed weight. Any other arrangement of $v_1, \ldots, v_{m+1}$ would break the parity at least for one of these stars, and therefore, by the remark after Lemma 2, would necessarily increase the weight contributed by this star. This completes the proof. □

3.2 NP-completeness proof

Using the construction of graph $\mathcal{H}(B, m)$ from the previous section, we now prove Theorem 1 by reduction from the 3-PARTITION.

Consider an instance of 3-PARTITION, $(\{a_1, \ldots, a_{3m}\}, B)$, where $\sum_{i=1}^{3m} a_i = mB$. We transform it into the graph $\mathcal{H}(B, m)$ extended by additional edges over vertices in Y. Consider a partition $Y = Y_1 \cup \cdots \cup Y_{3m}$, where $Y_i \cap Y_j = \emptyset$ for $i \neq j$, and $|Y_i| = a_i$ for all $i,\ 1 \leq i \leq 3m$. We turn each Y_i into a clique by adding a set of edges E_i over all pairs of vertices of Y_i. Consider an extended graph $G = \mathcal{H}(B, m) \cup (\cup_{i=1}^{3m}(Y_i, E_i))$. Again, from Figure 1, it is clear that G is an interval graph. Let $K = \mathcal{W}(\mathcal{H}(B, m)) + \sum_{i=1}^{3m} \mathcal{K}(a_i)$. Obviously the whole transformation can be carried out in polynomial time.

Theorem 2. *There exists a 3-partition of $\{a_1, \ldots, a_{3m}\}$ if and only if $\mathcal{W}(G) = K$.*

Proof. **Only if part:** Assume that $A = \{a_1, \ldots, a_{3m}\}$ can be partitioned into m disjoint subsets $A_1, \cdots, A_m$, each summing up to B. Let $A_i = \{a_1^i, \ldots, a_{|A_i|}^i\} \subseteq A$. We construct a layout L^* defined by

$$R_1 \cup X, v_1, Y_1^1, \ldots, Y_{|A_i|}^1, v_2, \ldots, Y_1^m, \ldots, Y_{|A_m|}^m, v_{m+1}, Z \cup R_2, \tag{4}$$

where $Y_j^i \in \{Y_1, \ldots, Y_{3m}\}$ is the subset corresponding to a_j^i ($|Y_j^i| = a_j^i$). Observe that in (4), there are exactly B vertices of Y between every v_i and v_{i+1} and that all edges between vertices of Y are edges of cliques with vertices mapped by L^* to consecutive positions. Therefore, using Lemma 8, the weight of L^* is $\mathcal{W}(G, L^*) = \mathcal{W}(\mathcal{H}(B, m)) + \sum_{i=1}^{3m} \mathcal{K}(a_i) = K$. By Corollary 1, this is the smallest possible weight, i.e., $\mathcal{W}(G) = K$.

If part: Let $\mathcal{W}(G) = K$, i.e., there exists a layout L^* such that $\mathcal{W}(G, L^*) = K$. Decompose G as the edge-disjoint union of graph $\mathcal{H}(B, m)$ and cliques $(Y_1, E_1), \ldots, (Y_{3m}, E_{3m})$. For any layout L of G, $\mathcal{W}(\mathcal{H}(B, m), L) \geq \mathcal{W}(\mathcal{H}(B, m))$ and $\mathcal{W}((Y_i, E_i), L) \geq \mathcal{K}(a_i)$ for all $i,\ 1 \leq i \leq 3m$. On the other hand, by Corollary 1, $\mathcal{W}(G) \geq \mathcal{W}(\mathcal{H}(B, m)) + \sum_{i=1}^{3m} \mathcal{K}(a_i)$. Therefore, if a layout L^* verifies $\mathcal{W}(G, L^*) = K$, this implies that *(i)* $\mathcal{W}(\mathcal{H}(B, m), L^*) = \mathcal{W}(\mathcal{H}(B, m))$ and *(ii)* $\mathcal{W}((Y_i, E_i), L^*) = \mathcal{K}(a_i)$, for all $i,\ 1 \leq i \leq 3m$.

Condition *(i)* implies that layout L^* verifies Lemma 8, and, in particular, splits vertices of Y by vertices $v_1, \ldots, v_{m+1}$ into m groups, each of cardinality B. Condition *(ii)* ensures that each subset Y_i is mapped by L^* into consecutive positions and therefore falls inside one such group. This means that numbers $\{a_1, \ldots, a_{3m}\}$ (cardinalities of $\{Y_1, \ldots, Y_{3m}\}$) are split into m disjoint subsets each of which sums up to B. This completes the proof of Theorem 2. $\square$

Since the optimal linear arrangement problem for interval graphs is NP-complete, the next section describes a 2-approximation algorithm for interval graphs.

4 A 2-approximation algorithm for OLA of interval graphs

Before describing an approximation algorithm, we study two layouts of an interval graph G, defined by any fixed interval model. Let $\mathcal{I}$ be an interval model of G. The layout of G consisting of vertices ordered by the left endpoints of their corresponding intervals is called the *left endpoint ordering* (*leo*) of G with respect to the interval model $\mathcal{I}$. Similarly, the layout of G consisting of vertices ordered by the right endpoints of their corresponding intervals is called the *right endpoint ordering* (*reo*) of G with respect to $\mathcal{I}$.

It has been shown in [14] that leo and reo are good approximations for the bandwidth of interval graphs: $b(G, leo) \leq 2 \cdot bw(G)$ and $b(G, reo) \leq 2 \cdot bw(G)$. This is based on the fact that:

- in a left endpoint ordering, *leo*, for every pair of adjacent vertices $leo(u) < leo(w)$, each vertex between u and w is adjacent to u, and
- in a right endpoint ordering *reo*, for every pair of adjacent vertices $reo(u) < reo(w)$ each vertex between u and w is adjacent to w.

This can be used to show that left endpoint and right endpoint orderings are 2-approximations for the OLA problem on interval graphs.

Theorem 3. *Let $G = (V, E)$ be an interval graph, and let $\mathcal{I}$ be an interval model of G. Then, $\mathcal{W}(G, leo) \leq 2\mathcal{W}(G)$, and $\mathcal{W}(G, reo) \leq 2\mathcal{W}(G)$.*

Proof. We focus on the ordering reo. For any integer i, $1 \leq i \leq V(G)$, we define graph G_i such that

- $V(G_i) = \{u \mid u \in V(G) \wedge reo(u) \leq i\}$, and
- $E(G_i) = \{e = (u, v) \in E(G) \mid u \in V(G_i) \wedge v \in V(G_i)\}$.

We prove this theorem by induction on the number of vertices. The induction hypothesis is that $\mathcal{W}(G_i, reo) \leq 2\mathcal{W}(G_i)$ for any integer i , $1 \leq i \leq V(G)$.

The basis of the induction is the situation where G_1 contains only one vertex ($i = 1$). The induction hypothesis holds here because $\mathcal{W}(G_1, reo) = 0$ and $\mathcal{W}(G_1) = 0$. Then $\mathcal{W}(G_1, reo) \leq 2\mathcal{W}(G_1)$.

For the induction step, we assume that the induction hypothesis for i holds. Now, we will prove that the induction hypothesis holds for $i + 1$. Let u be the vertex such that $reo(u) = i + 1$.

First we give a lower bound for $\mathcal{W}(G_{i+1})$. We can notice that sets $E(G_i)$ and $\{e = (v, u) \mid v \in V(G_i) \wedge e \in E(G_{i+1})\}$ form a partition of set $E(G_{i+1})$. By Lemma 3, $\mathcal{W}(G_{i+1}) \geq \mathcal{W}(G_i) + \mathcal{W}(S_{d_{G_{i+1}}(u)})$,

Secondly, we give an upper bound for $\mathcal{W}(G_{i+1}, reo)$ by considering the partition $E(G_i)$ and $\{e = (v, u) \mid v \in V(G_i) \wedge e \in E(G_{i+1})\}$ of set $E(G_{i+1})$.

For the edge set $E(G_i)$, we have $\sum_{e=(u,v)\in E(G_i)} |reo(u) - reo(v)| = \mathcal{W}(G_i, reo)$.

For the edge set $\{e = (v, u) \mid v \in V(G_i) \wedge e \in E(G_{i+1})\}$, since vertex u and its neighborhood in G_{i+1} are consecutive in the layout reo, the linear arrangement reo gives $d_{G_{i+1}}(u) + 1$ consecutive numbers. We can compute an upper bound of $\sum_{v\in N_{G_{i+1}}(u)} |reo(u) - reo(v)|$ because according to the linear arrangement reo, we are in the situation of the worst case for the star. So, we have

$$\sum_{v\in N_{G_{i+1}}(u)} |reo(u) - reo(v)| \leq 2\mathcal{W}(S_{d_{G_{i+1}}(u)})$$

This yields an upper bound for $\mathcal{W}(G_{i+1}, reo)$. We get $\mathcal{W}(G_{i+1}, reo) \leq \mathcal{W}(G_i, reo) + 2\mathcal{W}(S_{d_{G_{i+1}}(u)})$.
Since we have $\mathcal{W}(G_i, reo) \leq 2\mathcal{W}(G_i)$ by induction hypothesis, we have

$$\mathcal{W}(G_{i+1}, reo) \leq 2\mathcal{W}(G_i) + 2\mathcal{W}(S_{d_{G_{i+1}}(u)}) \leq 2\mathcal{W}(G_{i+1})$$

So, the Theorem holds.

Theorem 3 shows that left endpoint and right endpoint orderings are 2-approximation algorithms for this problem. This is the best possible bound for these orderings. In fact, a star S_α with an even number α of leaves has an interval representation such that $\mathcal{W}(S_\alpha, reo) = \frac{\alpha(\alpha+1)}{2}$ and $\mathcal{W}(S_\alpha) = \frac{\alpha}{2}(\frac{\alpha}{2} + 1)$. So the ratio $\frac{\mathcal{W}(S_\alpha, reo)}{\mathcal{W}(S_\alpha)}$ equals to $2 - \frac{1}{\alpha+2}$.

In the next section, we focus on close relatives of interval graphs – permutation graphs – and on their generalization – cocomparability graphs.

5 OLA of permutation and cocomparability graphs

Cocomparability, interval, and permutation graphs are well-known classes of perfect graphs. All of them have geometric intersection models. Many references, including [2, 11], contain comprehensive overviews of the many known structural and algorithmic properties of (co)comparability, interval, and permutation graphs.

Permutation graphs are intersection graphs of straight line segments between two parallel lines. Vertices of the graph are associated to segments and two vertices are adjacent iff corresponding segments intersect.

Our first remark here is that graph $\mathcal{H}(B, m)$ considered in Section 3 is a permutation graph. Figure 2 shows a permutation representation for $\mathcal{H}(B, m)$.

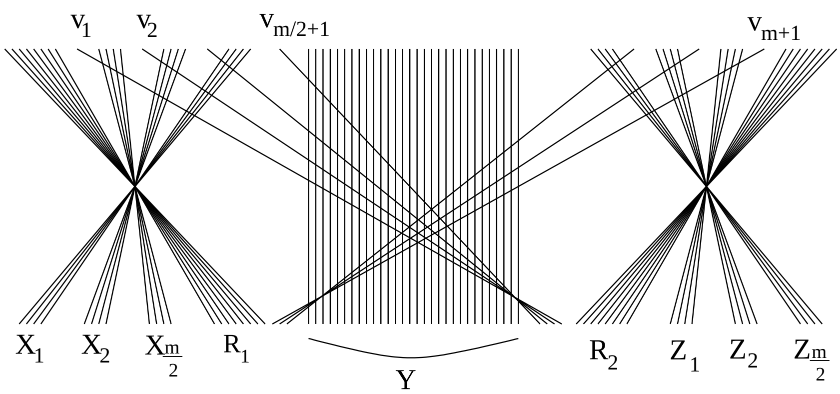

Fig. 2. Permutation representation of graph $\mathcal{H}(B, m)$

This immediately implies

Lemma 9. *The problem of deciding, for a permutation graph $G = (E, V)$ and a constant K, whether $\mathcal{W}(G) \leq K$ is NP-complete.*

Let us now turn to cocomparability graphs that are generalizations of both interval and permutation graphs. A graph G is cocomparability if its complement $\overline{G}$ is a comparability graph, i.e., the comparability graph of a poset $P = (V, \prec)$ is the graph with vertex set V for which vertices x and y are adjacent if and only if either $x \prec y$ or $y \prec x$ in P.

The following property of cocomparability graphs is well known (see e.g. [2]), and it is crucial for our arguments.

Proposition 1. *A graph $G = (V, E)$ is a cocomparability graph if and only if it has a* cocomparability ordering, *i.e., an ordering $(v_1, v_2, \ldots, v_n)$ of its vertices such that $(v_i, v_k) \in E$ and $i < j < k$ imply either $(v_i, v_j) \in E$ or $(v_j, v_k) \in E$.*

Since every interval graph is a cocomparability graph, the OLA problem remains NP-complete on cocomparability graphs. Now, we focus on the approximation problem. First, the following lower bound for the weight of an optimal linear arrangement of any graph will be useful when analyzing the performance ratio of some algorithms and orderings respectively.

Lemma 10. *For every graph $G = (V, E)$,*

$$\mathcal{W}(G) \geq \frac{m}{2} + \frac{1}{8} \sum_{v \in V} d^2(v).$$

Proof. Let v be a vertex of G. Then to minimize the sum over all edges incident to v in a layout, half of the neighbors of v must be placed immediately to the left of v and half of the neighbors of v must be placed immediately to the right of v. Thus the sum over all edges incident to v is at least $1 + 1 + 2 + 2 + \cdots + \frac{d(v)}{2} + \frac{d(v)}{2}$ if $d(v)$ is even, and $1 + 1 + 2 + 2 + \cdots + \frac{d(v)-1}{2} + \frac{d(v)-1}{2} + \frac{d(v)+1}{2}$ if $d(v)$ is odd.

Thus we obtain

$$\begin{aligned}\mathcal{W}(G) &\geq \frac{1}{2} \sum_{v \in V} \Big(\big(\frac{d(v)}{2} + 1\big)\big(\frac{d(v)}{2}\big) \Big) \geq \sum_{v \in V} \big(\frac{d^2(v)}{8} + \frac{d(v)}{4} \big) \\ &\geq \frac{1}{8} \sum_{v \in V} d^2(v) \ + \frac{m}{2}.\end{aligned}$$

We use the lower bound of the previous section to show that every cocomparability ordering of a cocomparability graph has weight at most $8 \cdot \mathcal{W}(G)$.

Theorem 4. *Let $G = (V, E)$ be a cocomparability graph and let L be a cocomparability ordering of G. Then, $\mathcal{W}(G, L) \leq 8 \cdot \mathcal{W}(G)$.*

Proof. By the definition of L, if u and w are adjacent in G then all vertices between u and w in L are either adjacent to u or adjacent to w. Therefore

$$|L(u) - L(w)| \leq |N(u) \cup N(w)| \leq d(u) + d(v),$$

and by Lemma 10,

$$\begin{aligned}\mathcal{W}(G, L) &= \sum_{e=(u,v)} |L(u) - L(v)| \\ &\leq \sum_{e=(u,v)} (d(u) + d(v)) \\ &\leq \sum_{v \in V} d^2(v) \\ &\leq 8 \cdot \mathcal{W}(G).\end{aligned}$$

Since a cocomparability ordering can be found in polynomial time $O(n^{2.376})$ [15], Theorem 4 immediately implies an 8-approximation polynomial-time algorithm for OLA on cocomparability graphs.

6 Conclusion and open problems

In this paper, we resolved the complexity of the OLA problem for interval, permutation and consequently for cocomparability, graphs. We have given simple approximation algorithms for those classes. There are several other linear layout problems, like CUTWIDTH, whose complexity is not resolved for the class of interval graphs [4].

References

1. T. BIEDL, B. BREJOVA, E. DEMAINE, A. HAMEL, A. LOPEZ-ORTIZ, T. VINAR, Finding Hidden Independent Sets in Interval Graphs, *Theoretical Computer Science* 310 (1-3), Jan 2004, 287–307.
2. A. BRANDSTÄDT, V.B. LE, J. SPINRAD Graph Classes: A Survey, *SIAM Monographs on Discrete Math. Appl., Vol.* 3, SIAM, Philadelphia (1999)
3. F.R.K. CHUNG, Labelings of graphs, *Selected Topics in graph theory*, 151-168, Academic Press, San Diego, 1988.
4. J. DÍAZ, J. PETIT, M.J. SERNA, A survey of graph layout problems, *ACM Computing Surveys*, vol.34, No.3, Sept 2002, 313-356.
5. S. EVEN, Y. SHILOACH, NP-Completeness of Several Arrangement Problems, *Technical Report #43*, Computer Science Dept., The Technion, Haifa, Israel, 1975.
6. M.A. GOLDBERG, I.A. KLIPKER, Minimal placing of trees on a line, *Technical report, Physico-Technical Institute of Low Temperatures, Academy of Sciences of Ukranian SSR, USSR*, 1976.
7. M. FARACH-COLTON, Y. HUANG, J.L.L. WOOLFORD, Discovering temporal relations in molecular pathways using protein-protein interactions, *Proceedings of the 8th Annual International Conference on Computational Molecular Biology (RECOMB)*, 2004, San Diego, California, USA, March 27-31, 2004, ACM Press, 150–156.
8. U. FEIGE, Approximating the bandwidth via volume respecting embeddings, *J. Comput. System Sci.* **60** (2000) 510-539.
9. M. R. GAREY, R. L. GRAHAM, D. S. JOHNSON, AND D. E. KNUTH, Complexity results for bandwidth minimization, *SIAM J. Appl. Math.*, **34** (1978), 477–495.
10. M.R. GAREY, D.S. JOHNSON, Computers and Intractability - A Guide to the Theory and Practice of NP-Completeness, W.H. Freeman, San Francisco, 1979
11. M.C. GOLUMBIC, Algorithmic Graph Theory and Perfect Graphs, *Academic Press, New York* (1980)
12. D. JOHNSON, The NP-completeness column: an ongoing guide, *J. Algorithms* **6** (1985), 434–451.
13. D.J. KLEITMAN, R.V. VOHRA, Computing the bandwidth of interval graphs, *SIAM J. Discrete Math.* **3** (1990) 373-375.
14. D. KRATSCH, L.K. STEWART, Approximating bandwidth by mixing layouts of interval graphs, *SIAM Journal on Discrete Mathematics* **15** (2002) 435-449.
15. R. MCCONNELL, J. SPINRAD, Modular decomposition and transitive orientation, *Discrete Math.* **201** (1999), 189–241.
16. B. MONIEN, The bandwidth minimization problem for caterpillars with hair length 3 is NP-complete. *SIAM J. Algebraic Discrete Methods* **7** (1986), 505–512.
17. S. RAO, A. RICHA, New approximation techniques for some ordering problems, *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'98)* 211–218, ACM, New York, 1998.
18. Y. SHILOACH, A minimum linear arrangement algorithm for undirected trees, *SIAM Journal on Computing* **8** (1979) 15-32.
19. A.P. SPRAGUE, An $O(n \log n)$ algorithm for bandwidth of interval graphs, *SIAM J. Discrete Math.* **7** (1994) 213-220
20. W. UNGER, The complexity of the approximation of the bandwidth problem, *Proceedings of the Thirty-ninth Annual IEEE Symposium on Foundations of Computer Science* (Palo Alto, CA, 1998).
21. M. WATERMAN, Introduction to computational biology: Maps, sequences and genomes, Chapman & Hall, 1995.