Optimal Linear Arrangement of Interval Graphs

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

[1]LORIA/CNRS, Nancy, France.

[2]Department of Informatics, University of Bergen, Norway.

[3]LITA, Université de Metz, France.

[4]LIFL/CNRS, Lille, France

MFCS, August,31 2006

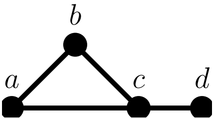
Outline

Outline

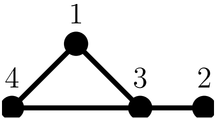
Definition

Definition

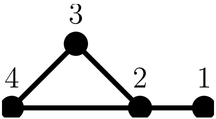
A *linear layout* (or simply *layout*) of a given graph $G = (V, E)$ is a linear ordering of its vertices.



graph G



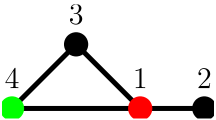
linear layout L_1



linear layout L_2

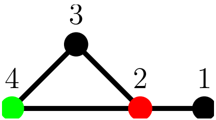
Definition

The *weight* of L on G is $\mathcal{W}(G, L) = \sum_{(u,v)\in E} |L(u) - L(v)|$.



linear layout L_1

$\mathcal{W}(G, L_1) = 1 + 3 + 2 + 1$



linear layout L_2

$\mathcal{W}(G, L_2) = 1 + 1 + 2 + 1$

Definition

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.

Previous work

- Computing an OLA is NP-hard (for general graphs [Garey, Johnson 1979], for bipartite graphs [Even, Shiloach 1975]).
- But the problem is polynomial for trees [Goldberg, Klipker 1976], for grids, for hypercubes [Diaz and al 2002].
- There exists an approximation algorithm with performance ratio $O(\log(n))$ [Rao, Richa 1998].

Definition

A graph $G = (V, E)$ is an *interval graph* if there is an 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.

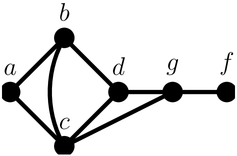


FIG.: An interval graph G

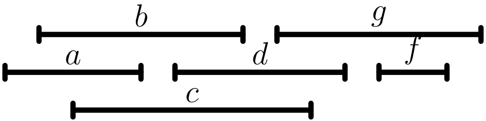


FIG.: Its interval representation

Linear arrangements and interval graphs

- The bandwidth ($b(G, L) = \max_{(u,v)\in E} |L(u) - L(v)|$) minimization problem is polynomial for interval graphs [Kleitman, Vohra1990].
- Interval graphs are used in bioinformatic (reconstruction of relative positions of DNA fragments, gene structure prediction, model of temporal relations in protein-protein interactions).
- An optimal linear arrangement of an interval graph models an "optimal" molecular pathway [Farach-Colton and al 2004].

Our results

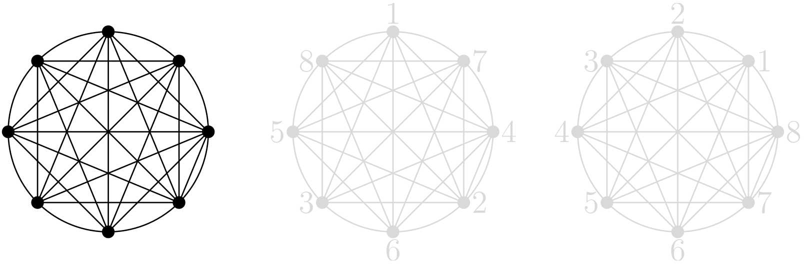
- a proof of NP-hardness of the optimal linear arrangement (OLA) problem on interval graphs.
- a 2-approximation algorithm for interval graphs.
- a 8-approximation algorithm for cocomparability graphs.

Outline

Lemma

Let K_n be the complete graph on n vertices. Then

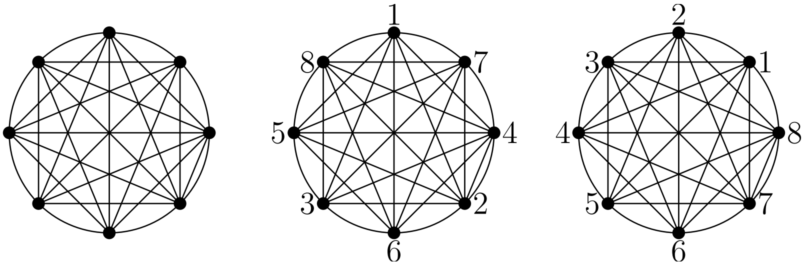
$$\mathcal{W}(K_n) = \frac{(n-1)n(n+1)}{6}.$$



Lemma

Let K_n be the complete graph on n vertices. Then

$$\mathcal{W}(K_n) = \frac{(n-1)n(n+1)}{6}.$$



OLA problem on star graphs.

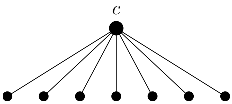
Lemma

Let S_α be the star with a center vertex c and α leaves. Then,

1. *a permutation L is an optimal linear arrangement if and only if L places c at the middle position.*

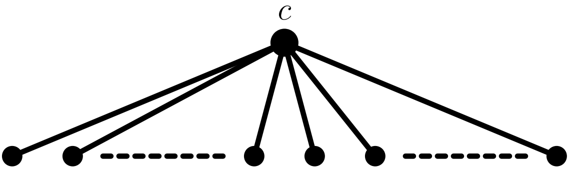
$$\mathcal{W}(S_\alpha) = \left\lceil \frac{\alpha}{2} \right\rceil \left(\left\lfloor \frac{\alpha}{2} \right\rfloor + 1 \right)$$

2. $\mathcal{W}(S_\alpha) \le \mathcal{W}(S_\alpha, L) \le 2\mathcal{W}(S_\alpha)$ *for any layout L*

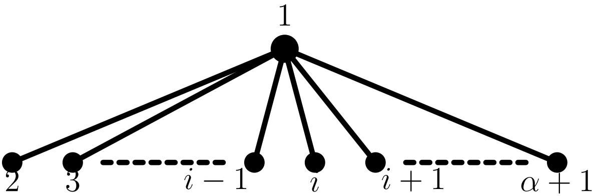


Proof of Lemma

Let S_α be a star with a center vertex c and α leaves (α even).



Let S_α be a star with a center vertex c and α leaves (α even).

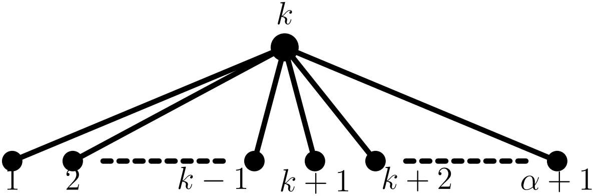


Case where $L(c) = 1$:

- $\mathcal{W}(S_\alpha, L) = \sum_{i=1}^{\alpha}(i + 1 - 1) = \sum_{i=1}^{\alpha} i = \frac{\alpha(\alpha+1)}{2}$

Proof of Lemma

Let S_α be a star with a center vertex c and α leaves (α even).



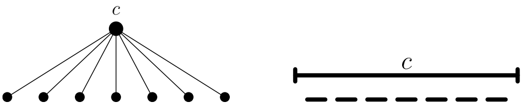
Case where $L(c) = 1$: the worst case ($\mathcal{W}(S_\alpha, L) = \frac{\alpha(\alpha+1)}{2}$).
Case where $L(c) = k$:

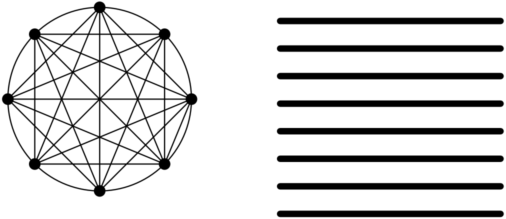
- $\mathcal{W}(S_\alpha, L) = \sum_{i=1}^{k-1} i + \sum_{i=1}^{\alpha+1-k} i$.

Case where $L(c) = \alpha/2 + 1$: the better case ($\mathcal{W}(S_\alpha, L) = \frac{\alpha}{2}(\frac{\alpha}{2} + 1)$).

Remark

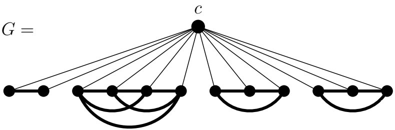
The stars and the complete graphs are interval graphs.





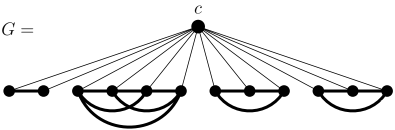
Partition of edge set.

$G =$



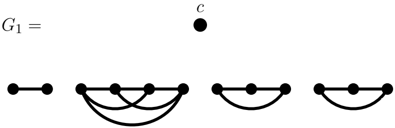
Partition of edge set.

$G =$

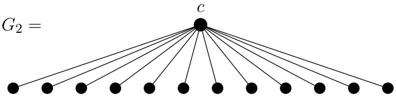


can be split into 2 graphs : G_1 and G_2

$G_1 =$

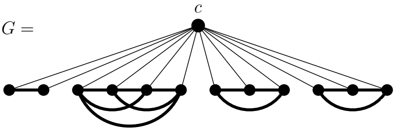


$G_2 =$



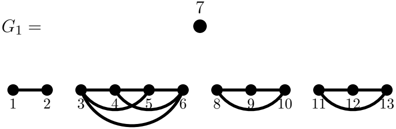
Partition of edge set.

$G =$



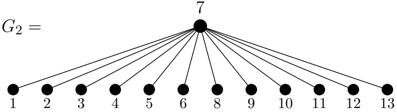
$$\mathcal{W}(G, L) = \mathcal{W}(G_1, L) + \mathcal{W}(G_2, L)$$

$G_1 =$



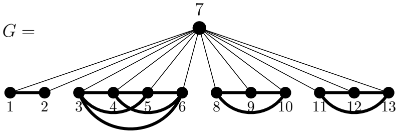
$\mathcal{W}(G_1, L)$

$G_2 =$

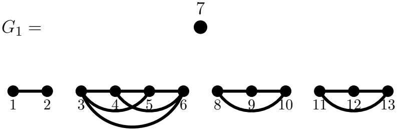


$\mathcal{W}(G_2, L)$

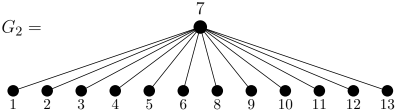
Partition of edge set.



$$\mathcal{W}(G) \geq \mathcal{W}(G_1) + \mathcal{W}(G_2)$$



$\mathcal{W}(G_1)$



$\mathcal{W}(G_2)$

Some results :

Lemma

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* $G_2 = (V, E_2)$*.*

Some results :

Lemma

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 $G_2 = (V, E_2)$.

More generally,

Corollary

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 $G_i = (V_i, E_i)$, $1 \leq i \leq n$.

Outline

The NP-completeness result

Theorem

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

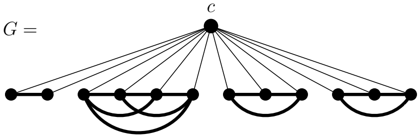
- This problem belongs to NP.
- The proof is by reduction from the 3-PARTITION problem :
  Instance : A finite set A of $3m$ integers $\{a_1, \ldots, a_{3m}\}$, a bound $B \in Z^+$ such that $\sum_{i=1}^{3m} a_i = mB$.
  Question : Can A be partitioned into m disjoint sets $A_1, A_2, \ldots, A_m$ such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

Sketch of proof :

Let f be the following reduction from *PARTITION* problem to *OLA* problem

$$f(A,B) = (G,k) \text{ with } \begin{cases} G \leftarrow \cup_{i=1}^{|A|} K_{a_i} \cup S_{2B} \text{ with a center } v \\ k \leftarrow \mathcal{W}(S_{2B}) + \sum_{i=1}^{|A|} \mathcal{W}(K_{a_i}) \end{cases}$$

$A = \{2,4,3,3\} \quad \rightarrow$
$B = 6$



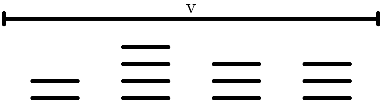
Sketch of proof :

Let f be the following reduction from *PARTITION* problem to *OLA* problem

$$f(A,B) = (G,k) \text{ with } \begin{cases} G \leftarrow \cup_{i=1}^{|A|} K_{a_i} \cup S_{2B} \text{ with a center } v \\ k \leftarrow \mathcal{W}(S_{2B}) + \sum_{i=1}^{|A|} \mathcal{W}(K_{a_i}) \end{cases}$$

G is an interval graph.

$A = \{2,4,3,3\} \quad \rightarrow$
$B = 6$

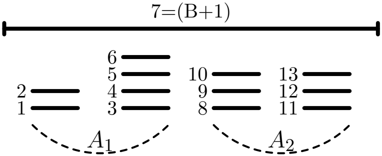


Sketch of proof :

Let f be the following reduction from *PARTITION* problem to *OLA* problem

$$f(A,B) = (G,k) \text{ with } \begin{cases} G \leftarrow \cup_{i=1}^{|A|} K_{a_i} \cup S_{2B} \text{ with a center } v \\ k \leftarrow \mathcal{W}(S_{2B}) + \sum_{i=1}^{|A|} \mathcal{W}(K_{a_i}) \end{cases}$$

$A = \{2,4,3,3\} \quad \rightarrow$
$B = 6$
$A_1 = \{2,4\}$
$A_2 = \{3,3\}$



Sketch of proof :

Let f be the following reduction from *PARTITION* problem to *OLA* problem

$$f(A,B) = (G,k) \text{ with } \begin{cases} G \leftarrow \cup_{i=1}^{|A|} K_{a_i} \cup S_{2B} \text{ with a center } v \\ k \leftarrow \mathcal{W}(S_{2B}) + \sum_{i=1}^{|A|} \mathcal{W}(K_{a_i}) \end{cases}$$

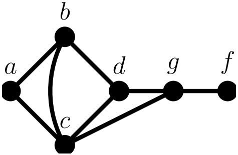
A can be partitioned into 2 disjoint sets A_1, A_2 such that $\sum_{a \in A_1} a = \sum_{a \in A_2} a = B$ $\Leftrightarrow$ there exists a linear layout L with $\mathcal{W}(G,L) \leq k$

A polynomial-time reduction from $3-PARTITION$ problem to *OLA* problem is based on the same idea.

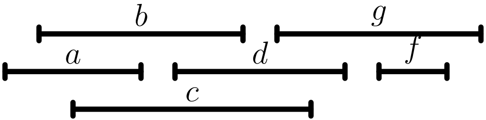
Outline

Definition

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 $\mathcal{I}$.



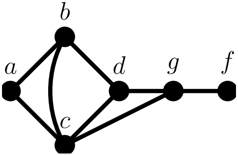
an interval graph G



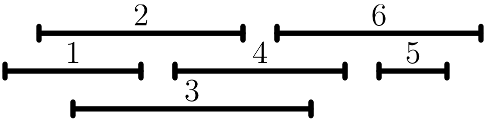
its interval representation

Definition

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 $\mathcal{I}$.



an interval graph G

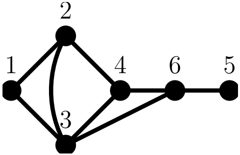


its interval representation

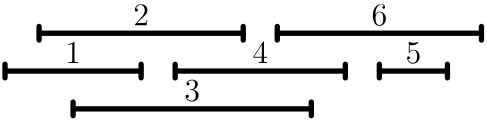
Right endpoint orderings

Definition

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 $\mathcal{I}$.



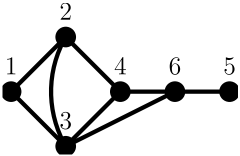
an interval graph G

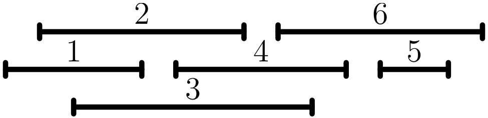


its interval representation

Remark

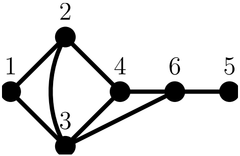
In a right endpoint ordering *reo*, for every pair of adjacent vertices $reo(u) < reo(w)$, each vertex between u and w is adjacent to w.

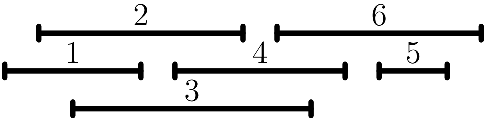




Remark

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.



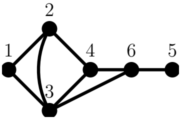


Theorem

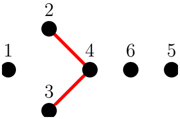
Let $G = (V, E)$ be an interval graph, and let $\mathcal{I}$ be an interval model of G. Then, $\mathcal{W}(G, reo) \leq 2\mathcal{W}(G)$.

Proof of the Theorem :

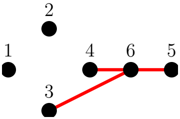
- E is split : $E_i = \{(u, v) \mid reo(u) = i \wedge reo(v) < i\} \cap E(G)$



G and reo



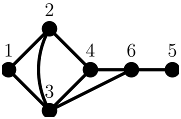
E_4



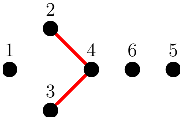
E_6

Proof of the Theorem :

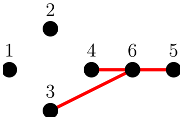
- E is split : $E_i = \{(u, v) \mid reo(u) = i \wedge reo(v) < i\} \cap E(G)$
- $G_i = (V, E_i)$ is composed of a star of $n_i (= \mid E_i \mid)$ leaves



G and reo



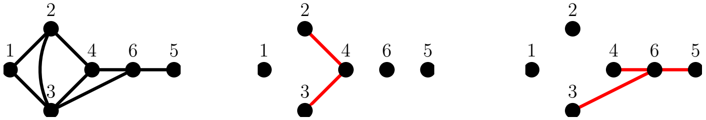
E_4 and graph G_4



E_6 and graph G_6

Proof of the Theorem :

- E is split : $E_i = \{(u,v) \mid reo(u) = i \wedge reo(v) < i\} \cap E(G)$
- $G_i = (V, E_i)$ is composed of a star of $n_i (= \mid E_i \mid)$ leaves

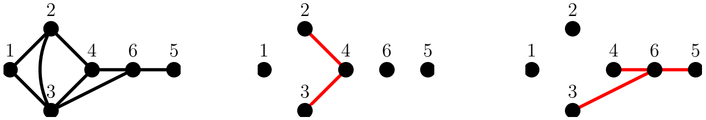


G and reo — E_4 and graph G_4 — E_6 and graph G_6

- For $G_i = (V, E_i)$:$\mathcal{W}(G_i, reo) = \sum_{v \in V(G_i)} |reo(u) - reo(v)|$

Proof of the Theorem :

- E is split : $E_i = \{(u, v) \mid reo(u) = i \wedge reo(v) < i\} \cap E(G)$
- $G_i = (V, E_i)$ is composed of a star of $n_i (=\mid E_i \mid)$ leaves



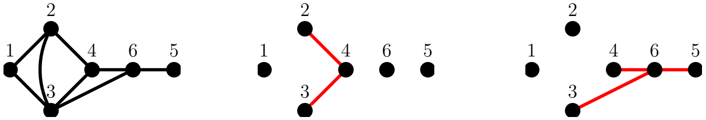
G and reo $\quad$ E_4 and graph G_4 $\quad$ E_6 and graph G_6

- For $G_i = (V, E_i)$:$\mathcal{W}(G_i, reo) = \sum_{v \in V(G_i)} |reo(u) - reo(v)|$ can be the worst case for the star of n_i leaves.

$$2 \times \mathcal{W}(S_{n_i}) \geq \mathcal{W}(G_i, reo) \geq \mathcal{W}(S_{n_i})$$

Proof of the Theorem :

- E is split : $E_i = \{(u,v) \mid reo(u) = i \wedge reo(v) < i\} \cap E(G)$
- $G_i = (V, E_i)$ is composed of a star of $n_i (=\mid E_i \mid)$ leaves



G and *reo* E_4 and graph G_4 E_6 and graph G_6

- For $G_i = (V, E_i)$:$\mathcal{W}(G_i, reo) = \sum_{v \in V(G_i)} |reo(u) - reo(v)|$ can be the worst case for the star of n_i leaves.

$$2 \times \mathcal{W}(S_{n_i}) \geq \mathcal{W}(G_i, reo) \geq \mathcal{W}(S_{n_i})$$

- For G :$\mathcal{W}(G) \geq \sum_{i=1}^{n} \mathcal{W}(G_i) \geq \sum_{i=1}^{n} \mathcal{W}(S_{n_i})$
- So $\mathcal{W}(G, reo) = \sum_{i=1}^{n} \mathcal{W}(G_i, reo)$ and $2\mathcal{W}(G) \geq \mathcal{W}(G, reo)$

Outline

Conclusion

- Optimal linear arrangement of interval graphs is NP-hard.
- There exists a fast 2-approximation algorithm based on any interval model of the input graph.
- In the paper, we extend this result to cocomparability graphs (NP-completeness, 8-approximation polynomial-time algorithm)
- The complexity of several other linear layout problems, like CUTWIDTH is not resolved for the class of interval graphs.