The k-Sparsest Subgraph Problem in (Proper) Interval Graphs

Rémi Watrigant, Marin Bougeret and Rodolphe Giroudeau

LIRMM, Montpellier, France



Séminaire AIGCo, 20/09/2012

Contents



- 2 FPT Algorithm in Interval Graphs
- PTAS in Proper Interval Graphs
- Open Problems and Future Work

k-Sparsest Subgraph Problem (k-SS)

Input: a graph G = (V, E), $k \le |V|$. **Output:** a set $S \subseteq V$ of size exactly k. **Goal:** minimize E(S) (the number of edges induced by S)

k-Sparsest Subgraph Problem (k-SS)Input:a graph $G = (V, E), k \le |V|$.Output:a set $S \subseteq V$ of size exactly k.Goal:minimize E(S) (the number of edges induced by S)

• generalization of independent set

 \Rightarrow k-SS *NP*-hard in general graphs (+ W[1]-hard, inapproximable)

k-Sparsest Subgraph Problem (k-SS)Input:a graph $G = (V, E), k \le |V|$.Output:a set $S \subseteq V$ of size exactly k.Goal:minimize E(S) (the number of edges induced by S)

- generalization of independent set
 ⇒ k-SS NP-hard in general graphs (+ W[1]-hard, inapproximable)
- maximization version (k-Densest Subgraph) NP-hard on chordal graphs
 ⇒ k-SS NP-hard in co-chordal ⊆ perfect graphs

k-Sparsest Subgraph Problem (k-SS)Input:a graph $G = (V, E), k \leq |V|$.Output:a set $S \subseteq V$ of size exactly k.Goal:minimize E(S) (the number of edges induced by S)

- generalization of independent set
 ⇒ k-SS NP-hard in general graphs (+ W[1]-hard, inapproximable)
- maximization version (k-Densest Subgraph) NP-hard on chordal graphs
 ⇒ k-SS NP-hard in co-chordal ⊆ perfect graphs
- k-SS polynomial in split graphs

k-Sparsest Subgraph Problem (k-SS)Input:a graph $G = (V, E), k \leq |V|$.Output:a set $S \subseteq V$ of size exactly k.Goal:minimize E(S) (the number of edges induced by S)

- generalization of independent set
 ⇒ k-SS NP-hard in general graphs (+ W[1]-hard, inapproximable)
- maximization version (k-Densest Subgraph) NP-hard on chordal graphs
 ⇒ k-SS NP-hard in co-chordal ⊆ perfect graphs
- k-SS polynomial in split graphs
- complexity of *k*-DS unknown in (proper) interval graphs. PTAS in interval graphs, 3-approximation in chordal graphs

k-Sparsest Subgraph Problem (k-SS)

Input: a graph G = (V, E), $k \le |V|$. **Output:** a set $S \subseteq V$ of size exactly k. **Goal:** minimize E(S) (the number of edges induced by S)

In this talk:

k-Sparsest Subgraph Problem (k-SS)

Input: a graph G = (V, E), $k \le |V|$. **Output:** a set $S \subseteq V$ of size exactly k. **Goal:** minimize E(S) (the number of edges induced by S)

In this talk:

- FPT algorithm in interval graphs (parameterized by the cost of the solution)
- PTAS in proper interval graphs

k-Sparsest Subgraph Problem (k-SS)

Input: a graph G = (V, E), $k \le |V|$. **Output:** a set $S \subseteq V$ of size exactly k. **Goal:** minimize E(S) (the number of edges induced by S)

In this talk:

- FPT algorithm in interval graphs (parameterized by the cost of the solution)
- PTAS in proper interval graphs

FPT Algorithm

An *FPT* algorithm for a parameterized problem is an algorithm that exactly solves the problem in O(f(k).poly(n)) where n is the size of the instance and k the parameter of the instance.

Polynomial-Time Approximation Scheme

A *PTAS* for a minimization problem is an algorithm A_{ϵ} such that for any fixed $\epsilon > 0$, A_{ϵ} runs in polynomial time and outputs a solution of cost $< (1 + \epsilon)OPT$

k-Sparsest Subgraph Problem (k-SS)

Input: a graph G = (V, E), $k \le |V|$. **Output:** a set $S \subseteq V$ of size exactly k. **Goal:** minimize E(S) (the number of edges induced by S)

In this talk:

- FPT algorithm in interval graphs (parameterized by the cost of the solution)
- PTAS in proper interval graphs

k-Sparsest Subgraph Problem (k-SS)

Input: a graph G = (V, E), $k \le |V|$. **Output:** a set $S \subseteq V$ of size exactly k. **Goal:** minimize E(S) (the number of edges induced by S)

In this talk:

- FPT algorithm in interval graphs (parameterized by the cost of the solution)
- PTAS in proper interval graphs

Interval graphs = intersection graphs of intervals in the real line.



k-Sparsest Subgraph Problem (k-SS)

Input: a graph G = (V, E), $k \le |V|$. **Output:** a set $S \subseteq V$ of size exactly k. **Goal:** minimize E(S) (the number of edges induced by S)

In this talk:

- FPT algorithm in interval graphs (parameterized by the cost of the solution)
- PTAS in proper interval graphs

Interval graphs = intersection graphs of intervals in the real line.



proper interval graph = no interval contains properly another one = unit interval graphs

k-Sparsest Subgraph Problem (k-SS)

Input: a graph G = (V, E), $k \le |V|$. **Output:** a set $S \subseteq V$ of size exactly k. **Goal:** minimize E(S) (the number of edges induced by S)

In this talk:

- FPT algorithm in interval graphs (parameterized by the cost of the solution)
- PTAS in proper interval graphs

Interval graphs = intersection graphs of intervals in the real line.



proper interval graph = no interval contains properly another one = unit interval graphs

Contents



2 FPT Algorithm in Interval Graphs

3 PTAS in Proper Interval Graphs



Given a set \mathcal{I} of intervals, $k \leq |\mathcal{I}|$ and a cost C^*

Given a set \mathcal{I} of intervals, $k \leq |\mathcal{I}|$ and a cost C^*

Idea of the algorithm:

Given a set \mathcal{I} of intervals, $k \leq |\mathcal{I}|$ and a cost C^*

Idea of the algorithm:

• we sort intervals according to their right endpoints

Given a set \mathcal{I} of intervals, $k \leq |\mathcal{I}|$ and a cost C^*

Idea of the algorithm:

- we sort intervals according to their right endpoints
- parameters of the dynamic programming:

 $s \leftarrow$ left endpoint of the leftmost interval, $k' \leftarrow k, \ C' \leftarrow C^*$

Given a set \mathcal{I} of intervals, $k \leq |\mathcal{I}|$ and a cost C^*

Idea of the algorithm:

- we sort intervals according to their right endpoints
- parameters of the dynamic programming:
 s ← left endpoint of the leftmost interval, k' ← k, C' ← C*
- given the parameters, we construct all subsets T s.t.

Given a set $\mathcal I$ of intervals, $k \leq |\mathcal I|$ and a cost $\mathcal C^*$

Idea of the algorithm:

- we sort intervals according to their right endpoints
- parameters of the dynamic programming:
 s ← left endpoint of the leftmost interval, k' ← k, C' ← C*
- given the parameters, we construct all subsets T s.t.

```
(i) T is connected
```

(ii) T starts after s (i.e. to the right of s)

(iii)
$$E(T) \leq C'$$

Given a set $\mathcal I$ of intervals, $k \leq |\mathcal I|$ and a cost $\mathcal C^*$

Idea of the algorithm:

- we sort intervals according to their right endpoints
- parameters of the dynamic programming:
 s ← left endpoint of the leftmost interval, k' ← k, C' ← C*
- given the parameters, we construct all subsets T s.t.
 - (i) T is connected (ii) T starts after s (i.e. to the right of s) (iii) $E(T) \le C'$

recursive call with :

- $\blacktriangleright k' \leftarrow k' |T|$
- $C' \leftarrow C E(T)$
- $s \leftarrow$ left endpoint of the rightmost interval after T

Given a set $\mathcal I$ of intervals, $k \leq |\mathcal I|$ and a cost $\mathcal C^*$

Idea of the algorithm:

- we sort intervals according to their right endpoints
- parameters of the dynamic programming:
 s ← left endpoint of the leftmost interval, k' ← k, C' ← C*
- given the parameters, we construct all subsets T s.t.
 - (i) T is connected
 - (ii) T starts after s (i.e. to the right of s)
 - (iii) $E(T) \leq C'$
- recursive call with :
 - $k' \leftarrow k' |T|$

•
$$C' \leftarrow C - E(T)$$

- $s \leftarrow$ left endpoint of the rightmost interval after T
- ⇒ at most k.C*.n different inputs what about the running time of one call ?

Given a set $\mathcal I$ of intervals, $k \leq |\mathcal I|$ and a cost $\mathcal C^*$

Idea of the algorithm:

- we sort intervals according to their right endpoints
- parameters of the dynamic programming:
 s ← left endpoint of the leftmost interval, k' ← k, C' ← C*
- given the parameters, we construct all subsets T s.t.
 - (i) T is connected
 - (ii) T starts after s (i.e. to the right of s)
 - (iii) $E(T) \leq C'$
- recursive call with :
 - ► $k' \leftarrow k' |T|$

•
$$C' \leftarrow C - E(T)$$

- $s \leftarrow$ left endpoint of the rightmost interval after T
- ⇒ at most k.C*.n different inputs what about the running time of one call ?

Let $\Omega_s(C')$ be the set of all subsets satisfying (i), (ii) and (iii)

• given the parameters, we construct all subsets T s.t.

(i) *T* is connected
(ii) *T* starts after *s* (i.e. to the right of *s*)
(iii) *E*(*T*) ≤ *C*'

Let $\Omega_s(C')$ be the set of all subsets satisfying (i), (ii) and (iii)

Lemma

• given the parameters, we construct all subsets T s.t.

```
(i) T is connected
(ii) T starts after s (i.e. to the right of s)
(iii) E(T) \le C'
```

Let $\Omega_s(C')$ be the set of all subsets satisfying (i), (ii) and (iii)




















FPT Algorithm in Interval Graphs



We now bound the size of $\Gamma_s(C^*)$:

• $y_i = B \Rightarrow$ there exists a clique of size B in the solution

We now bound the size of $\Gamma_s(C^*)$:

• $y_i = B \Rightarrow$ there exists a clique of size B in the solution $\Rightarrow y_i \le \sqrt{2C^*} + 2$

We now bound the size of $\Gamma_s(C^*)$:

- $y_i = B \Rightarrow$ there exists a clique of size *B* in the solution $\Rightarrow y_i \le \sqrt{2C^*} + 2$
- for each step i ∈ {0,..., (t − 1)} and corresponding s, we can find a pair of intervals (crossing or at the right of s) overlapping such that in the next step, one of them is at the left of s (no multiple counts of the same pair)

We now bound the size of $\Gamma_s(C^*)$:

- $y_i = B \Rightarrow$ there exists a clique of size *B* in the solution $\Rightarrow y_i \le \sqrt{2C^*} + 2$
- for each step i ∈ {0,..., (t − 1)} and corresponding s, we can find a pair of intervals (crossing or at the right of s) overlapping such that in the next step, one of them is at the left of s (no multiple counts of the same pair)
 ⇒ t ≤ C* + 1

We now bound the size of $\Gamma_s(C^*)$:

- $y_i = B \Rightarrow$ there exists a clique of size *B* in the solution $\Rightarrow y_i \le \sqrt{2C^*} + 2$
- for each step i ∈ {0,..., (t − 1)} and corresponding s, we can find a pair of intervals (crossing or at the right of s) overlapping such that in the next step, one of them is at the left of s (no multiple counts of the same pair)
 ⇒ t ≤ C* + 1

Thus:

$$|\Gamma_{s}(C^{*})| \leq (\sqrt{2C^{*}}+2)^{C^{*}+1}$$

and each step of the dynamic programming runs in FPT time.

We now bound the size of $\Gamma_s(C^*)$:

- $y_i = B \Rightarrow$ there exists a clique of size *B* in the solution $\Rightarrow y_i \le \sqrt{2C^*} + 2$
- for each step i ∈ {0,..., (t − 1)} and corresponding s, we can find a pair of intervals (crossing or at the right of s) overlapping such that in the next step, one of them is at the left of s (no multiple counts of the same pair)
 ⇒ t ≤ C* + 1

Thus:

$$|\Gamma_{s}(C^{*})| \leq (\sqrt{2C^{*}}+2)^{C^{*}+1}$$

and each step of the dynamic programming runs in FPT time.

Theorem

k-Sparsest Subgraph in Interval Graphs is FPT parameterized by the cost of the solution.

Contents



- 2 FPT Algorithm in Interval Graphs
- 3 PTAS in Proper Interval Graphs
 - Open Problems and Future Work

Idea of the algorithm:

• sorting intervals according to their right endpoints

- sorting intervals according to their right endpoints
- greedy decomposition of the graph into a path of separators

- sorting intervals according to their right endpoints
- greedy decomposition of the graph into a path of separators
- re-structuration of an optimal solution into a near optimal solution such that all near optimal solutions can be enumerated in polynomial time

- sorting intervals according to their right endpoints
- greedy decomposition of the graph into a path of separators
- re-structuration of an optimal solution into a near optimal solution such that all near optimal solutions can be enumerated in polynomial time
- dynamic programming processes the graph through the decomposition, enumerating all possible solutions.

 I_{m_1}









The decomposition

Remark

The only edges between blocks B_i and B_{i+1} are between R_i and L_{i+1} . Given $S \subseteq \mathcal{I}$ we have:

$$E(S) = \sum_{i=1}^{a} E(B_i \cap S) + \sum_{i=1}^{a-1} E(R_i \cap S, L_{i+1} \cap S)$$

Re-structuration of optimal solutions

Compaction

Let $S \subseteq I$ be a solution, and $S^c = comp(S) \subseteq I$ such that for each block $i \in \{1, ..., a\}$:

- for all $l \in L_i$, $comp(l) \in L_i$ and is at the right of l (we may have comp(l) = l)
- for all $I \in R_i$, $comp(I) \in R_i$ and is at the left of I (we may have comp(I) = I)

Re-structuration of optimal solutions

Compaction

- Let $S \subseteq I$ be a solution, and $S^c = comp(S) \subseteq I$ such that for each block $i \in \{1, ..., a\}$:
 - for all $l \in L_i$, $comp(l) \in L_i$ and is at the right of l (we may have comp(l) = l)
 - for all $I \in R_i$, $comp(I) \in R_i$ and is at the left of I (we may have comp(I) = I)



Re-structuration of optimal solutions

Compaction

- Let $S \subseteq I$ be a solution, and $S^c = comp(S) \subseteq I$ such that for each block $i \in \{1, ..., a\}$:
 - for all $l \in L_i$, $comp(l) \in L_i$ and is at the right of l (we may have comp(l) = l)
 - for all $I \in R_i$, $comp(I) \in R_i$ and is at the left of I (we may have comp(I) = I)



Re-structuration of optimal solutions

Compaction

- Let $S \subseteq I$ be a solution, and $S^c = comp(S) \subseteq I$ such that for each block $i \in \{1, ..., a\}$:
 - for all $l \in L_i$, $comp(l) \in L_i$ and is at the right of l (we may have comp(l) = l)
 - for all $I \in R_i$, $comp(I) \in R_i$ and is at the left of I (we may have comp(I) = I)



Re-structuration of optimal solutions

Compaction

Let $S \subseteq I$ be a solution, and $S^c = comp(S) \subseteq I$ such that for each block $i \in \{1, ..., a\}$:

- for all $l \in L_i$, $comp(l) \in L_i$ and is at the right of l (we may have comp(l) = l)
- for all $I \in R_i$, $comp(I) \in R_i$ and is at the left of I (we may have comp(I) = I)

Lemma

If *comp* is a compaction of a solution S such that for all block $i \in \{1, ..., a\}$, we have

$$E(comp(S \cap B_i)) \leq \rho E(S \cap B_i)$$

Then comp(S) is a ρ -approximation of S.

Re-structuration of optimal solutions

Let us built a compaction that yields a $(1 + \frac{4}{P})$ -approximation for any fixed P.

Re-structuration of optimal solutions

Let us built a compaction that yields a $(1 + \frac{4}{P})$ -approximation for any fixed P. Let $X \subseteq B_i$ be a solution. We note $X = X_L \cup X_R$. Set sizes are in lowercase.

Re-structuration of optimal solutions

Let us built a compaction that yields a $(1 + \frac{4}{P})$ -approximation for any fixed P. Let $X \subseteq B_i$ be a solution. We note $X = X_L \cup X_R$. Set sizes are in lowercase. • we divide X_L into P consecutive subsets of same size $q_L \rightarrow X_1^L, ..., X_P^L$ • we divide X_R into P consecutive subsets of same size $q_R \to X_1^R, ..., X_P^R$

Then define the compaction: for any $t \in \{1, ..., P\}$

Re-structuration of optimal solutions

Let us built a compaction that yields a $(1 + \frac{4}{P})$ -approximation for any fixed P. Let $X \subseteq B_i$ be a solution. We note $X = X_L \cup X_R$. Set sizes are in lowercase.

• we divide X_L into P consecutive subsets of same size $q_L \rightarrow X_1^L, ..., X_P^L$

• we divide X_R into P consecutive subsets of same size $q_R \rightarrow \tilde{X}_1^R, ..., \tilde{X}_P^R$ Then define the compaction: for any $t \in \{1, ..., P\}$

- Y^L_t are the q_L rightmost intervals at the left of the rightmost interval of X^L_t
 Y^L_t are the q_R leftmost intervals at the right of the leftmost interval of X^R_t



Re-structuration of optimal solutions

What do we need to construct such a solution ?



Re-structuration of optimal solutions

What do we need to construct such a solution ?

- the leftmost interval of X_t^L for $t \in \{1, ..., P\}$
- the rightmost interval of X_t^R for $t \in \{1, ..., P\}$
- x_R, x_L (plus remainders of divisions by P...)

 $\Rightarrow 2P + O(1)$ variables ranging in $\{0, ..., n\}$





Sketch of proof of the $(1 + \frac{4}{P})$ approximation ratio:



Sketch of proof of the $(1 + \frac{4}{P})$ approximation ratio: • $OPT = \binom{x_L}{2} + \binom{x_R}{2} + \sum_{t=1}^{a} \sum_{u=1}^{a} E(X_t^L, X_u^R)$ • $SOL = \binom{x_L}{2} + \binom{x_R}{2} + \sum_{t=1}^{a} \sum_{u=1}^{a} E(Y_t^L, Y_u^R)$



Sketch of proof of the $(1 + \frac{4}{P})$ approximation ratio: • $OPT = \binom{x_L}{2} + \binom{x_R}{2} + \sum_{t=1}^{a} \sum_{u=1}^{a} E(\frac{X_L}{t}, \frac{X_u}{u})$

•
$$SOL = \binom{x_L}{2} + \binom{x_R}{2} + \sum_{t=1}^{a} \sum_{u=1}^{a} E(Y_t^L, Y_u^R)$$

But:



Sketch of proof of the $(1 + \frac{4}{P})$ approximation ratio:

• $OPT = \begin{pmatrix} x_L \\ 2 \end{pmatrix} + \begin{pmatrix} x_R \\ 2 \end{pmatrix} + \sum_{t=1}^{a} \sum_{u=1}^{a} E(X_t^L, X_u^R)$ • $SOL = \begin{pmatrix} x_L \\ 2 \end{pmatrix} + \begin{pmatrix} x_R \\ 2 \end{pmatrix} + \sum_{t=1}^{a} \sum_{u=1}^{a} E(Y_t^L, Y_u^R)$

But:

• if some intervals of Y_t^L overlap some intervals of Y_u^R

Then:

• all intervals of X_{t+1}^{L} overlap all intervals of $\bigcup_{i=1}^{u-1} X_i^{R}$



Sketch of proof of the $(1 + \frac{4}{P})$ approximation ratio:

•
$$\begin{array}{l} OPT = \binom{x_L}{2} + \binom{x_R}{2} + \sum_{t=1}^{a} \sum_{u=1}^{a} E(X_t^L, X_u^R) \\ \bullet SOL = \binom{x_L}{2} + \binom{x_R}{2} + \sum_{t=1}^{a} \sum_{u=1}^{a} E(Y_t^L, Y_u^R) \end{array}$$

But:

• if some intervals of Y_t^L overlap some intervals of Y_u^R

Then:

• all intervals of X_{t+1}^{L} overlap all intervals of $\bigcup_{i=1}^{u-1} X_i^{R}$

Finally, we can prove that $\frac{SOL}{OPT} \leq 1 + \frac{4}{P}$
PTAS in Proper Interval Graphs

Conclusion:

Theorem

For any *P*, the previous algorithm outputs a $(1 + \frac{4}{P})$ -approximation for the *k*-Sparsest Subgraph in Proper Interval graphs in $O(n^{O(P)})$

Contents



- 2 FPT Algorithm in Interval Graphs
- 3 PTAS in Proper Interval Graphs





	Perfect				
	NP-hard				
			Bipa	rtite]
Г	Chordal		Tree	1	
			<u> </u>		ן ו
		In	terval		
١ſ	Split	Γ	Proper		
			inc.		
Ľ		_			
	_				

Perfect			
NP-hard			
	Bipartite	1	
Chordal	Tree		
	Poly.		
	Interval		
Split	Proper		
Poly.	Int.		

Perfect			
NP-hard			
	Bipar	tite	
Chordal	Troo	NP-h?	
Chordan	Poly.		
	Interval		
Split	Proper		
Poly.	Int.		

Perfect			
NP-hard			
	Bipa	rtite	
Chordal	Tree	NP-h?	
Chordan	Poly.		
] [Interval		
Split	Proper		
Poly.	Int. PTAS		

Perfect			
NP-hard			
	Bipar	tite	
Chordal	Tree	NP-h?	
	Poly.		
_			
	Interval		
Split	Proper		
Poly.	Int. PTAS		

Complexity of *k*-Sparsest Subgraph:

Perfect				
NP-hard				
			Bipa	rtite
I	Chordal		Tree	NP-h?
			Poly.	
		In Fl	nterval PT	
	Split		Proper	
	Poly. PT/		PTAS	

2 main objectives:

- extend FPT and/or approximation results to Chordal graphs
- NP-hardness for Chordal graphs

Thank you for your attention!