PTAS for Densest k-Subgraph in Interval Graphs

Tim Nonner

IBM Research - Zurich
`tno@zurich.ibm.com`

Abstract. Given an interval graph and integer k, we consider the problem of finding a subgraph of size k with a maximum number of induced edges, called *densest k-subgraph problem in interval graphs*. It has been shown that this problem is NP-hard even for chordal graphs [17], and there is probably no PTAS for general graphs [12]. However, the exact complexity status for interval graphs is a long-standing open problem [17], and the best known approximation result is a 3-approximation algorithm [16]. We shed light on the approximation complexity of finding a densest k-subgraph in interval graphs by presenting a polynomial-time approximation scheme (PTAS), that is, we show that there is an $(1+\epsilon)$-approximation algorithm for any $\epsilon > 0$, which is the first such approximation scheme for the densest k-subgraph problem in an important graph class without any further restrictions.

1 Introduction

The *densest k-subgraph problem* is defined as follows: given a graph with n vertices and an integer $k \leq n$, find a subgraph with k vertices that maximizes the number of induced edges. The NP-hardness is a direct consequence of the well-known fact that finding a maximum clique is NP-hard. This contrasts to the problem of finding an arbitrary-sized subgraph of maximum density, i.e., a subgraph with maximum average degree, which is polynomially solvable [14]. The first approximation algorithm for the densest k-subgraph problem with an approximation guarantee of $\mathcal{O}(n^{0.3885})$ was given by Kortsarz and Peleg [13]. Feige et al. [8] improved this factor to $\mathcal{O}(n^{\delta})$ for some $\delta < 1/3$. Until recently, when Bhaskara et al. [5] presented an $\mathcal{O}(n^{1/4+\epsilon})$-approximation algorithm for any $\epsilon > 0$, it has been a long-standing open problem whether the factor $\mathcal{O}(n^{1/3})$ can be significantly beaten. Ashahiro et al. [3] showed that a simple greedy strategy yields an approximation ratio of $\mathcal{O}(n/k)$, and Feige and Langberg [7] showed that n/k is achievable using semidefinite programming. On the other hand, using a random sampling technique, Arora et al. [2] presented a polynomial-time approximation scheme (PTAS) for dense graphs with $k = \Omega(n)$, that is, if the number of edges is $\Omega(n^2)$ and $k = \Omega(n)$, then there is an $(1+\epsilon)$-approximation algorithm for any $\epsilon > 0$. However, there is probably no PTAS for general graphs [12].

These weak approximation guarantees for general graphs motivated the study of special graph classes, like perfect graphs, chordal graphs, and interval graphs.

Unfortunately, although a maximum clique of a perfect graph can be found in polynomial time [10], finding a densest k-subgraph remains NP-hard even for chordal graphs and bipartite graphs [17], two important subclasses of perfect graphs. However, the complexity status of finding a densest k-subgraph in interval graphs, a subclass of chordal graphs, has been a prominent open problem over the last three decades [17]. In an *interval graph*, each vertex corresponds to an interval, and two vertices are connected via an edge if their corresponding intervals overlap. Such a geometric structure occurs frequently in scheduling, VLSI-design, and biology, see for instance [9]. This unknown complexity status gave rise to the search for approximation results. Liazi et al. [16] presented a 3-approximation algorithm for interval graphs and chordal graphs, and there is moreover a PTAS if the clique graph is a star [15], i.e., if the intersection graph of the maximal cliques is a star. Recently, Chen et al. [6] showed that a large family of intersection graphs, including interval graphs, admit constant factor approximation algorithms. Finally, Backer and Keil [4] gave a 3/2-approximation algorithm for proper interval graphs [4], a subclass of interval graphs where no interval is allowed to contain another one.

Contributions. We significantly improve upon the known approximation results by presenting a PTAS for finding a densest k-subgraph in interval graphs. It is worth mentioning here that this is the first PTAS for an import graph class without any further restrictions, since so far only PTASs for dense graphs with $k = \Omega(n)$ [2] and a quite restricted subclass of chordal graphs [15] are known. We conjecture that finding a densest k-subgraph in interval graphs is NP-hard, but proving this (or giving a polynomial time algorithm instead) remains a challenging open problem [17].

Technique and outline. If k is constant, then we can easily find a densest k-subgraph in polynomial time by simply enumerating all $\binom{n}{k} = \mathcal{O}(n^k)$ subgraphs of size k. Hence, the difficulty of the problem stems from the fact that k is part of the input. Now, let V be the vertices of the input interval graph, and let $V^* \subseteq V$ with $|V^*| = k$ be a subset of vertices that induce a densest k-subgraph. Consider some clique $C \subseteq V$, and let $C^* = C \cap V^*$ be the subclique of C contained in V^*. Assume then that we already know the vertices $V^* \backslash C$, and hence we only need to pick some $k - |V^* \backslash C|$ vertices from C that maximize the number of induced edges in combination with $V^* \backslash C$. Clearly, C^* will solve this simplified problem. However, we might need to enumerate $\binom{|C|}{|C^*|}$ subcliques of C in order to find C^* (or another subclique of similar quality), which is not possible in polynomial time if $|C^*|$ is not constant. Therefore, we show in Section 4 that, by losing an $(1 - \epsilon)$-factor in the number of induced edges for an arbitrary small $\epsilon > 0$, it suffices to only consider a polynomial number of subcliques (Lemma 4). To extend this search space restriction to the whole interval graph, we first need to introduce the notion of a sequence representation in Section 3 which allows us to decompose any densest k-subgraph into a sequence of cliques. Finally, we present a backward dynamic program in Section 5 that finds a densest k-subgraph in this restricted

search space in polynomial time. The principle of reducing the search space in order to make it treatable by dynamic programming in polynomial time has been succesfully applied during the last decades [11,1] to obtain approximation schemes. However, finding the right reduction and dynamic program is a highly problem-specific challenge.

2   Preliminaries

Let $G = (V, E)$ be an *interval graph*, i.e., each vertex in V corresponds to an interval, and two vertices are connected via an edge in E if their corresponding intervals overlap. Hence, the we can also think of the vertices in V as intervals. For an interval subset $V' \subseteq V$, let $E(V')$ denote the number of edges of the subgraph of G induced by V', i.e., the number of overlaps between intervals in V'. Hence, our goal is to find an interval set $V^* \subseteq V$ of size k that maximizes $E(V^*)$. We also refer to such a subset as a *densest k-interval subset*. Let OPT $:= E(V^*)$ denote this maximal number of overlaps. Finally, for two disjoint interval sets $V', V'' \subseteq V$, let $E(V', V'')$ denote the number of edges of the bipartite subgraph of G induced by these two interval sets, i.e., the number of overlaps between intervals in V' and intervals in V''. Observe that a *clique* $C \subseteq V$ is a set of intervals with $\cap_{I \in C} I \neq \emptyset$.

Let $l_I \in \mathbb{R}$ and $r_I \in \mathbb{R}$ denote the left and right endpoint of an interval $I \in V$, respectively, and assume that all endpoints of intervals are distinct, which can be easily ensured without changing the overlap structure. This also ensures that all intervals are distinct. We write $I_1 < I_2$ for two intervals $I_1, I_2 \in V$ if $r_{I_1} < l_{I_2}$. In this case, we also say that the pair I_1, I_2 is *consecutive*. Analogously, we call a sequence of intervals $I_1, I_2, \ldots, I_s \in V$ with $I_1 < I_2 < \ldots < I_s$ a *consecutive interval sequence*. Finally, for an interval $I \in V$, let $V_I \subseteq V$ be the set of all intervals $I' \in V$ with $I \subseteq I'$.

Lemma 1. *For any densest k-interval subset $V^* \subseteq V$, we may assume for each interval $I \in V^*$ that $V_I \subseteq V^*$.*

Proof. Assume for contradiction that there is an interval $I \in V^*$ and another interval $I' \in V_I \backslash V^*$. In this case, since $I \subset I'$, we could replace interval I by interval I' without modifying the size of V^* and without decreasing $E(V^*)$. Therefore, iterating this scheme terminates and gives us a densest k-interval subset V^* that satisfies the property from the claim.   □

3   Sequence Representations

For a consecutive interval pair $I_1, I_2 \in V$, let $C_{I_1 I_2}$ denote the clique of all intervals $I \in V$ with $l_{I_1} < l_I < r_{I_1}$ and $r_{I_1} < r_I < r_{I_2}$. Moreover, let $C'_{I_1 I_2}$ denote the clique of all intervals $I \in V$ with $r_{I_1} < l_I < r_{I_2}$ and $r_{I_2} \leq r_I$. Note that $I_2 \in C'_{I_1 I_2}$. We illustrate the cliques $C_{I_1 I_2}$ and $C'_{I_1 I_2}$ in Figure 1.
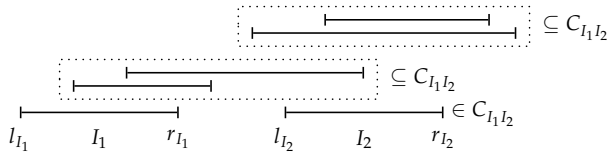
Fig. 1. Example intervals in the cliques $C_{I_1 I_2}$ and $C'_{I_1 I_2}$

To avoid case distinctions, we add a dummy interval I_∞ to V with $I < I_\infty$ for any other interval $I \in V \backslash \{I_\infty\}$. Our goal is then to find a densest $(k+1)$-interval subset $V^* \subseteq V$ with $I_\infty \in V^*$. Since I_∞ does not overlap with any other interval in V, this slightly modified problem is equivalent to our original problem of simply finding a densest k-interval subset. Specifically, removing I_∞ from a solution $V' \subseteq V$ for this modified problem yields a solution for our original problem with the same number of interval overlaps $E(V')$. Hence, OPT denotes the maximal number of interval overlaps in both cases.

We say that an interval subset $V' \subseteq V$ admits a *sequence representation* if there is a consecutive interval sequence $I_1, I_2, \ldots, I_s$ and a sequence of cliques $Q_1, Q_2, \ldots, Q_{s-1}$ with $Q_i \subseteq C_{I_i I_{i+1}}$ for each $1 \leq i < s$ such that $V' = \cup_{i=1}^{s} V_{I_i} \cup \cup_{i=1}^{s-1} Q_i$. In words, for each interval $I \in V'$, we either have that there is an index $1 \leq i \leq s$ with $I_i \subseteq I$, or there is an index $1 \leq i < s$ with $I \in Q_i$. We schematically depict the overlap structure of the cliques $Q_1, Q_2, \ldots, Q_{s-1}$ in Figure 2. If $I_\infty \in V'$, note that it must then hold for any such sequence representation that $I_s = I_\infty$.
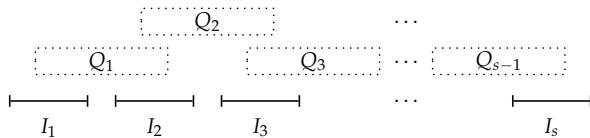

Fig. 2. Overlap structure of the cliques $Q_1, Q_2, \ldots, Q_{s-1}$

The following lemma says that we only have to consider interval subsets that admit a sequence representation.

Lemma 2. *We may assume that any densest $(k+1)$-interval subset $V^* \subseteq V$ with $I_\infty \in V^*$ admits a sequence representation.*

Proof. Given a densest $(k+1)$-interval subset $V^* \subseteq V$ with $I_\infty \in V^*$, we inductively construct the claimed sequence representation. During each iteration, we want to conserve the invariant that the currently constructed interval sequence $I_1, I_2, \ldots, I_i$ and cliques $Q_1, Q_2, \ldots, Q_{i-1}$ are a sequence representation of $\{I \in V^* \mid l_I \leq l_{I_i}\}$. Since we extend this interval sequence by one interval

in each iteration, this already gives that we end up with the claimed sequence representation $I_1, I_2, \ldots, I_s$ and $Q_1, Q_2, \ldots, Q_{s-1}$ of V^*. Note that we obtain that finally $I_s = I_\infty$.

To start this inductive construction, let $I_1 \in V^*$ be the interval with leftmost left endpoint l_{I_1} subject to the constraint that there is no interval $I \in V^*$ with $I \subset I_1$. Recall that we know from Lemma 1 that we may assume that $V_{I_1} \subseteq V^*$. Hence, to see that the invariant described above holds, we only have to show that also $\{I \in V^* \mid l_I < l_{I_1}\} \subseteq V_{I_1}$. Recall here the assumption that all left endpoints of intervals are distinct, and thus $I = I_1$ if $l_I = l_{I_1}$. Now, assume for contradiction that there is an interval $I \in V^*$ with $l_I < l_{I_1}$ and $I \notin V_{I_1}$. Hence, we obtain that $r_I < r_{I_1}$. Let then $I' \in V^*$ be an interval with the properties that $I' \subseteq I$ and there is no interval $I'' \in V^*$ with $I'' \subset I'$. If $l_{I'} > l_{I_1}$, then $I' \subset I_1$, which contradicts the selection of I_1. On the other hand, if $l_{I'} < l_{I_1}$, then I_1 is not an interval with leftmost left endpoint, which contradicts the selection of I_1 as well. This shows that the invariant holds after the first iteration. Now assume that, for some $i \geq 1$, we have a sequence representation $I_1, I_2, \ldots, I_i$ and $Q_1, Q_2, \ldots, Q_{i-1}$ of $\{I \in V^* \mid l_I \leq l_{I_i}\}$. If $I_i = I_\infty$, then we are done. Otherwise, let $I_{i+1} \in V^*$ be the interval with leftmost left endpoint $l_{I_{i+1}}$ subject to the constraints that $I_i < I_{i+1}$ and there is no interval $I \in V^*$ with $I \subset I_{i+1}$. Moreover, define $Q_i := V^* \cap C_{I_i I_{i+1}}$. Using nearly the same arguments as for the first iteration shows that the invariant is conserved during each iteration.   □

4   Simple Sequence Representations

Consider some fixed but arbitrary small $\epsilon > 0$, and assume that $1/\epsilon$ is integral and $2/\epsilon + 2 \geq 4$. Moreover, let $C \subseteq V$ be some clique, and let $I'_1, I'_2, \ldots, I'_u$ with $r_{I'_1} > r_{I'_2} > \ldots > r_{I'_u}$ be an ordering of the intervals in C according to their right endpoints. Recall here the assumption that all interval endpoints are distinct. We will show how the following *inputs* define a subclique $Q \subseteq C$:

(1) an integer v with $2 \leq v \leq 2/\epsilon + 2$,
(2) an integer sequence $j_1, j_2, \ldots, j_v$ with $j_1 = 1 < j_2 < \ldots < j_v = u + 1$,
(3) an integer sequence $h_1, h_2, \ldots, h_{v-1}$ with $0 \leq h_t \leq j_{t+1} - j_t$ for each $1 \leq t < v$.

We define Q as follows: for each $1 \leq t < v$, Q contains exactly the h_t intervals in $\{I'_{j_t}, I'_{j_t+1}, \ldots, I'_{j_{t+1}-1}\}$ with leftmost left endpoints. Thus, any such combination of inputs defines a subclique $Q \subseteq C$. Let then $P_\epsilon(C)$ be the set of all subcliques constructed in this way for all possible such inputs. Since $|C| = u \leq n$, we immediately obtain the polynomial bound $|P_\epsilon(C)| \leq (2/\epsilon + 1) \cdot n^{2/\epsilon} \cdot n^{2/\epsilon+1}$.

Example. Consider the clique C of intervals $I'_1, I'_2, \ldots, I'_7$ depicted in Figure 3 which are ordered according to their right endpoints. Then, for the inputs $v = 4$, $j_1 = 1 < j_2 = 4 < j_3 = 7 < j_4 = 8$, $h_1 = h_2 = 2$, and $h_3 = 0$, we add the set $Q = \{I'_1, I'_3, I'_5, I'_6\}$ to $P_\epsilon(C)$, which are the intervals with solid lines in Figure 3. To see this, note that I'_1 and I'_3 are the h_1 intervals in $\{I'_1, I'_2, I'_3\}$ with leftmost left endpoints, and I'_5 and I'_6 are the h_2 intervals in $\{I'_4, I'_5, I'_6\}$ with leftmost left endpoints. Since $h_3 = 0$, Q does not contain the single interval in $\{I'_7\}$.
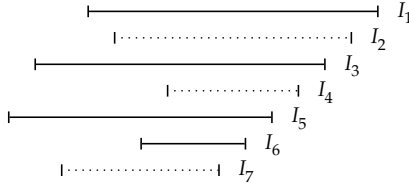
Fig. 3. Example construction of a clique $Q \in P_\epsilon(C)$

Let $V' \subseteq V$ be an interval subset with a sequence representation $I_1, I_2, \ldots, I_s$ and $Q_1, Q_2, \ldots, Q_{s-1}$. We then say that V' admits a *simple sequence representation* if additionally $Q_i \in P_\epsilon(C_{I_i I_{i+1}})$ for each $1 \leq i < s$. The following lemma is critical for the correctness of the PTAS, since it allows us to trade the size of the search space for accuracy.

Lemma 3. *There is an interval subset $V' \subseteq V$ with $|V'| = k+1$, $I_\infty \in V'$, and $E(V') \geq (1 - 8\epsilon)\mathrm{OPT}$ that admits a simple sequence representation.*

To prove Lemma 3, we need one preliminary lemma.

Lemma 4. *Consider an interval subset $V' \subseteq V$ with a sequence representation $I_1, I_2, \ldots, I_s$ and $Q_1, Q_2, \ldots, Q_{s-1}$. Then, for any $1 \leq i < s$, there exists a clique $Q \in P_\epsilon(C_{I_i I_{i+1}})$ with $|Q| = |Q_i|$ such that replacing the clique Q_i by Q decreases $E(V')$ by at most $\epsilon |Q_i||Q_i'|$, where $Q_i' := V' \cap C'_{I_i I_{i+1}}$. Moreover, if $|Q_i| = 1$, then $E(V')$ does not decrease at all.*

Proof. We abbreviate $C = C_{I_i I_{i+1}}$ and $C' = C'_{I_i I_{i+1}}$ throughout this proof, and let $I_1', I_2', \ldots, I_u'$ with $r_{I_1'} > r_{I_2'} > \ldots > r_{I_u'}$ be an ordering of the intervals in C according to their right endpoints. First, consider the case that $|Q_i| = 1$. In this case, there is one input that defines a set $Q \in P_\epsilon(C)$ with exactly $Q = Q_i$. Specifically, if $Q_i = \{I_t'\}$ for $t < u$, then we use the input $j_1 = 1 < j_2 = t < j_3 = t + 1 < j_4 = u + 1$, $h_1 = h_3 = 0$, and $h_2 = 1$. On the other hand, if $Q_i = \{I_u'\}$, then we use the input $j_1 = 1 < j_2 = u < j_3 = u + 1$, $h_1 = 0$, and $h_2 = 1$. This proves the second part of the claim. Hence, we may assume that $|Q_i| > 1$ in what follows.

To define the set $Q \in P_\epsilon(C)$ for the first part of the claim, we have to define the respective inputs of Q used during the construction of $P_\epsilon(C)$. To this end, for each $1 \leq j \leq u$, let $b_j := E(\{I_j'\}, Q_i')$ denote the number of intervals in Q_i' which overlap with I_j'. Consequently, since the intervals $I_1', I_2', \ldots I_u' \in C$ are ordered according to their right endpoints and all intervals in $Q_i' \subseteq C'$ are right of these intervals as schematically depicted in Figure 1, we obtain that $|Q_i'| \geq b_1 \geq b_2 \geq \ldots \geq b_u$. Using this, we now inductively construct a sequence $j_1, j_2, \ldots, j_v$ with $j_1 = 1 < j_2 < \ldots < j_v = u + 1$ as follows:

Since $j_1 = 1$, the start of this inductive construction is well-defined. Now assume that we have already defined some integers $j_1, j_2, \ldots, j_t$. Then, if $b_{j_t} - b_{j_t+1} > \epsilon|Q_i'|$, set $j_{t+1} := j_t + 1$. Otherwise, let $j_{t+1} > j_t$ be the maximal index such that still $b_{j_t} - b_{j_{t+1}} \leq \epsilon|Q_i'|$. If this index is the last index u, then set $j_{t+1} := u + 1$ instead, and terminate this inductive construction. This gives the integer sequence $j_1, j_2, \ldots, j_v$.

Observe that it holds for any $1 \leq t \leq v - 2$ that $b_{j_t} - b_{j_{t+2}} > \epsilon|Q_i'|$. Consequently, since $b_1 \leq |Q_i'|$, we obtain that $v \leq 2/\epsilon + 2$. Moreover, for each $1 \leq t < v$, we either have that $j_{t+1} = j_t + 1$ or $b_{j_t} - b_{j_{t+1}} \leq \epsilon|Q_i'|$.

Example. Assume that C is the clique from Figure 3, and moreover assume that $\epsilon|Q_i'| = 5$, $b_1 = 10, b_2 = 8, b_3 = 6, b_4 = 6, b_5 = 0, b_6 = 0$, and $b_7 = 0$. In any case, we have $j_1 = 1$. Consequently, it holds that $j_2 = 4$ is the maximal index such that still $b_{j_1} - b_{j_2} = 4 \leq \epsilon|Q_i'|$. Next, since $b_{j_2} - b_{j_2+1} = 6 > \epsilon|Q_i'|$, we set $j_3 := j_2 + 1$. Finally, $j_4 = 7$ is the maximal index such that still $b_{j_3} - b_{j_4} = 0 < \epsilon|Q_i'|$. However, since this is the last index, we set $j_4 := 8$ instead.

Finally, for each $1 \leq t < v$, define $C_t := \{I_{j_t}', I_{j_t+1}', \ldots, I_{j_{t+1}-1}'\}$, and let $h_t := |Q_i \cap C_t|$. Hence, $\sum_{t=1}^{v-1} h_t = |Q_i|$. The sequences $j_1, j_2, \ldots, j_v$ and $h_1, h_2, \ldots, h_{v-1}$ are the inputs required to define the claimed set $Q \in P_\epsilon(C)$. Because of the definition of the sequence $h_1, h_2, \ldots, h_{v-1}$, we immediately obtain $|Q| = |Q_i|$. For each $1 \leq t < v$, observe that the construction of Q implies that $Q \cap C_t$ contains the h_t intervals in C_t with leftmost left endpoints.

We still have to show that the interval set Q constructed above has the claimed property. First, since $|Q| = |Q_i|$, we have $E(Q) = E(Q_i) = E(V' \cap C)$. Therefore, to bound the decrease of $E(V')$ due to replacing Q_i by Q, we only need to show that $E(Q, V' \backslash C) \geq E(Q_i, V' \backslash C) - \epsilon|Q_i||Q_i'|$. To this end, we partition $V' \backslash C$ into four pairwise disjoint parts:

$$V^< := \{I \in V' \backslash C \mid r_I < r_{I_i}\},$$
$$V^= := \{I \in V' \backslash C \mid l_I < r_{I_i} \leq r_I\},$$
$$V^> := \{I \in V' \backslash C \mid r_{I_i} < l_I < r_{I_{i+1}}\},$$
$$V^\gg := \{I \in V' \backslash C \mid l_I > r_{I_{i+1}}\}.$$

Since these sets are pairwise disjoint, we may consider them separately:

Case $V^<$: Consider some index $1 \leq t < v$, and recall that $Q \cap C_t$ contains the h_t intervals in C_t with leftmost left endpoints. On the other hand, $Q_i \cap C_t$ also contains h_t from C_t, but these are not necessarily the ones with leftmost left endpoints. Consequently, since $V^<$ are the intervals left of r_{I_i} and it holds for any interval $I \in C_t$ that $r_{I_i} \in I$, this shows that $E(Q \cap C_t, V^<) \geq E(Q_i \cap C_t, V^<)$. Combining this for all indices $1 \leq t < v$ finally gives that $E(Q, V^<) \geq E(Q_i, V^<)$.

Case $V^=$: Since $|Q| = |Q_i|$ and $r_{I_i} \in I$ for each interval $I \in C$, we have that $E(Q, V^=) = E(Q_i, V^=)$.

Case $V^{\gg}$: Since no interval $V^{\gg}$ overlaps with an interval in C, we trivially obtain $E(Q, V^{\gg}) = 0 = E(Q_i, V^{\gg})$. Therefore, combining this case with the last two cases, we see that the next case is the only case where $E(V')$ might decrease.

Case $V^{>}$: Since the consecutive interval pair I_i, I_{i+1} is part of a sequence representation of V', there is no interval $I \in V'$ with $r_{I_i} < l_I < r_{I_{i+1}}$ and $r_I < r_{I_{i+1}}$. Therefore, we obtain that $V^{>} = Q'_i := V' \cap C'$. Hence, we only need to upper bound $E(Q_i, Q'_i) - E(Q, Q'_i) = \sum_{t=1}^{v-1} a_t$, where, for each $1 \le t < v$, we define $a_t := E(Q_i \cap C_t, Q'_i) - E(Q \cap C_t, Q'_i)$. Now note that, for each $1 \le t < v$, each interval in C_t overlaps with at least $b_{j_{t+1}}$ and at most b_{j_t} intervals in Q'_i. On the other hand, the sets $Q_i \cap C_t$ and $Q \cap C_t$ both contain exactly h_t intervals. Combining this, we find that $a_t \le h_t(b_{j_t} - b_{j_{t+1}})$, where we define $b_{j_v} := 0$. Therefore, we obtain that

$$\sum_{t=1}^{v-1} a_t = \sum_{t:j_{t+1}>j_t+1} a_t \le \sum_{t:j_{t+1}>j_t+1} h_t(b_{j_t} - b_{j_{t+1}})$$
$$\le \epsilon|Q'_i| \sum_{t:j_{t+1}>j_t+1} h_t \le \epsilon|Q'_i||Q_i|.$$

The equality in the first line is due to the fact that if $j_{t+1} = j_t + 1$, then trivially $Q_i \cap C_t = Q \cap C_t$, since $|C_t| = 1$, and consequently $a_t = 0$. Moreover, the first inequality in the second line is due to the earlier observation that either $j_{t+1} = j_t + 1$ or $b_{j_t} - b_{j_{t+1}} \le \epsilon|Q'_i|$. Hence, by combining all these arguments, we conclude that $E(Q, V^{>}) \ge E(Q_i, V^{>}) - \epsilon|Q_i||Q'_i|$.

Combining all four cases proves the claim. □

Proof (Lemma 3). Consider a densest $(k+1)$-interval subset $V^* \subseteq V$ with $I_\infty \in V^*$ and $E(V^*) = \text{OPT}$. We know from Lemma 2 that we may assume that V^* admits a sequence representation $I_1, I_2, \ldots, I_s$ and $Q_1, Q_2, \ldots, Q_{s-1}$. Moreover, Lemma 4 implies that, for each $1 \le i < s$, we can replace Q_i by a clique $Q \in P_\epsilon(C_{I_i I_{i+1}})$ with $|Q| = |Q_i|$ such that $E(V^*)$ decreases by at most $\epsilon|Q_i||Q'_i|$, and if $|Q_i| = 1$, then $E(V^*)$ does not decrease at all. Iterating these replacements yields the claimed interval subset V'. We will argue in the following paragraph that these replacements can indeed be done iteratively without amplifying these decreases.

Assume that, for some $1 \le i < s$, we have already replaced Q_t for each $t > i$ as described above. However, we obtain that the only such replacement that might affect the next replacement of Q_i by some clique $Q \in P_\epsilon(C_{I_i I_{i+1}})$ is the one with $t = i + 1$. Therefore, consider the clique $Q' \in P_\epsilon(C_{I_{i+1} I_{i+2}})$ that replaced Q_{i+1}. Because we always select intervals with leftmost left endpoints during the construction of Q' and $|Q_{i+1}| = |Q'|$, we can identify each interval $I \in Q_{i+1}$ with an interval $I' \in Q'$ with $l_{I'} \le l_I$. Hence, since $Q \subseteq C_{I_i I_{i+1}}$ is left of $C_{I_{i+1} I_{i+2}}$, we even have that $E(Q, Q') \ge E(Q, Q_{i+1})$. Consequently, replacing Q_i by Q will still decrease $E(V^*)$ by at most $\epsilon|Q_i||Q'_i|$, even if we replace Q_{i+1} by Q' first.

To bound the total decrease of $E(V^*)$, observe that

$$E(V^*) - E(V') \leq \sum_{i:|Q_i|>1} \epsilon|Q_i||Q_i'|$$

$$\leq \epsilon\left(\sum_{i=1}^{s-1} 2|Q_i|(|Q_i|-1) + \sum_{i=1}^{s-1} 2|Q_i'|(|Q_i'|-1)\right)$$

$$= \epsilon\left(\sum_{i=1}^{s-1} 4E(Q_i) + \sum_{i=1}^{s-1} 4E(Q_i')\right) \leq \epsilon 8E(V^*)$$

which completes the proof of the lemma. The first line is due to Lemma 4 and the arguments above. The second line is due to the simple observation that $ab \leq 2(a(a-1)+b(b-1))$ for any pair of integers $a > 1$ and $b \geq 1$. Moreover, the third line uses the fact that $E(C) = \binom{|C|}{2} = |C|(|C|-1)/2$ for any clique $C \subseteq V$. Finally, the fourth line is due to the fact that the cliques $Q_1, Q_2, \ldots, Q_{s-1} \subseteq V^*$ are pairwise disjoint, and hence $\sum_{i=1}^{s-1} E(Q_i) \leq E(V^*)$. The same holds for the cliques $Q_1', Q_2', \ldots, Q_{s-1}' \subseteq V^*$. $\qquad\square$

5   Dynamic Programming

We have a dynamic programming array Π with integral entries of the form $\Pi(h, s, I_{s-1}, I_s, Q_{s-1})$, where h is an integer with $0 \leq h \leq k+1$, s is an integer with $2 \leq s \leq n$, $I_{s-1}, I_s \in V$ are two consecutive intervals, and $Q_{s-1} \in P_\epsilon(C_{I_{s-1}I_s})$. The indexing of I_{s-1}, I_s, and Q_{s-1} with s is just for convenience. Our goal is to fill this array such that $\Pi(h, s, I_{s-1}, I_s, Q_{s-1}) = E(V')$ for an interval subset $V' \subseteq V$ with $|V'| = h$ that maximizes $E(V')$ subject to the constraint that V' admits a simple sequence representation $I_1, I_2, \ldots, I_s$ and $Q_1, Q_2, \ldots, Q_{s-1}$. Hence, only the last parts in this sequence representation are defined by the entry. We can bound the size of Π by $n^4 \cdot \max_{I_{s-1} < I_s} |P_\epsilon(C_{I_{s-1}I_s})|$. The second part of this product is simply the maximal size of $P_\epsilon(C_{I_{s-1}I_s})$ for any consecutive interval pair I_{s-1}, I_s. Consequently, since we already know that $P_\epsilon(C_{I_{s-1}I_s})$ has polynomial size for any such interval pair $I_{s-1}, I_s \in V$, we immediately obtain that the array Π has polynomial size as well.

We initialize Π by filling all entries of the form $\Pi(h, 2, I_1, I_2, Q_1)$ with $|Q_1 \cup V_{I_1} \cup V_{I_2}| = h$. Specifically, for any integer h with $0 \leq h \leq k+1$, any consecutive interval pair $I_1, I_2 \in V$, and any interval set $Q_1 \in P_\epsilon(I_1, I_2)$ with $|Q_1 \cup V_{I_1} \cup V_{I_2}| = h$, we set $\Pi(h, 2, I_1, I_2, Q_1) := E(Q_1 \cup V_{I_1} \cup V_{I_2})$. All other entries of Π are initialized as $-\infty$.

To define a recurrence relation, assume that we have already filled all entries of the form $\Pi(h, s-1, I_{s-2}, I_{s-1}, Q_{s-2})$, and now we want to use them to fill all entries of the form $\Pi(h, s, I_{s-1}, I_s, Q_{s-1})$, i.e., we want to increase the sequence length s by one. To this end, we can use the following recurrence relation:

$$\Pi(h, s, I_{s-1}, I_s, Q_{s-1}) =$$

$$\max_{I_{s-2} < I_{s-1}, Q_{s-2} \in P_\epsilon(C_{I_{s-2}I_{s-1}})} \left\{ \Pi(h - |Q_{s-1}| - |V_{I_s} \backslash V_{I_{s-1}}|, s - 1, I_{s-2}, I_{s-1}, Q_{s-2}) \right.$$

$$\left. + E(Q_{s-2} \cup V_{I_{s-1}}, Q_{s-1} \cup V_{I_s} \backslash V_{I_{s-1}}) + E(Q_{s-1} \cup V_{I_s} \backslash V_{I_{s-1}}) \right\}$$

In words, we take the maximum over all intervals $I_{s-2} \in V$ with $I_{s-2} < I_{s-1}$ and all cliques $Q_{s-2} \in P_\epsilon(C_{I_{s-2}I_{s-1}})$. Since $P_\epsilon(C_{I_{s-1}I_s})$ has polynomial size as already used above, this recurrence relation can be clearly implemented in polynomial time. Hence, in combination with the size of Π listed above, we find that it takes polynomial time to fill Π.

To see the correctness of this recurrence relation, consider an interval subset $V' \subseteq V$ that realizes an entry of the form $\Pi(h, s, I_{s-1}, I_s, Q_{s-1})$, and let $I_1, I_2, \dots, I_s$ and $Q_1, Q_2, \dots, Q_{s-1}$ be a simple sequence representation of V'. Let then $V'' \subseteq V'$ be the subset with the shorter simple sequence representation $I_1, I_2, \dots, I_{s-1}$ and $Q_1, Q_2, \dots, Q_{s-2}$. Hence, we have that $V'' = V' \backslash (Q_{s-1} \cup V_{I_s} \backslash V_{I_{s-1}})$. Now observe that, as schematically illustrated in Figure 4, the only intervals in V'' which might overlap with an interval in $V' \backslash V'' = Q_{s-1} \cup V_{I_s} \backslash V_{I_{s-1}}$ are the ones in $Q_{s-2} \cup V_{I_{s-1}}$. Consequently, we obtain the decomposition $E(V') = E(V'') + E(Q_{s-2} \cup V_{I_{s-1}}, Q_{s-1} \cup V_{I_s} \backslash V_{I_{s-1}}) + E(Q_{s-1} \cup V_{I_s} \backslash V_{I_{s-1}})$ as used in the recurrence relation. Thus, since V' minimizes $E(V')$, we have that also V'' minimizes $E(V'')$, since we could otherwise improve V' by replacing the intervals V'' by another interval subset. This shows that $E(V'')$ realizes the entry $\Pi(h - |Q_{s-1}| - |V_{I_s} \backslash V_{I_{s-1}}|, s - 1, I_{s-2}, I_{s-1}, Q_{s-2})$. Combining these facts shows the correctness of the recurrence relation.
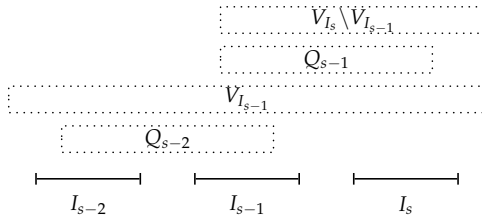

Fig. 4. Overlap structure during a recurrence relation

Theorem 1. *There is a PTAS for finding a densest k-subgraph in interval graphs.*

Proof. Use the dynamic programming approach explained above to compute all entries $\Pi(h, s, I_{s-1}, I_s, Q_{s-1})$ with $h = k + 1$ and $I_s = I_\infty$, and let V' be the interval set that realizes an optimal such entry, i.e., one that maximizes $E(V') = \Pi(h, s, I_{s-1}, I_s, Q_{s-1})$. By the definition of Π, we obtain that V' is a

densest $(k + 1)$-interval subset with $I_\infty \in V'$ subject to the constraint that V' admits a simple sequence representation. Consequently, by Lemma 3, we find that $E(V' \setminus \{I_\infty\}) = E(V') \geq (1 - 8\epsilon)\text{OPT}$. This proves the claim. □

References

1. Arora, S.: Polynomial time approximation schemes for euclidean traveling salesman and other geometric problems. Journal of the ACM 45(5), 753–782 (1998)
2. Arora, S., Karger, D.R., Karpinski, M.: Polynomial time approximation schemes for dense instances of np-hard problems. J. Comput. Syst. Sci. 58(1), 193–210 (1999)
3. Asahiro, Y., Iwama, K., Tamaki, H., Tokuyama, T.: Greedily finding a dense subgraph. J. Algorithms 34(2), 203–221 (2000)
4. Backer, J., Keil, J.M.: Constant factor approximation algorithms for the densest k-subgraph problem on proper interval graphs and bipartite permutation graphs. Inf. Process. Lett. 110(16), 635–638 (2010)
5. Bhaskara, A., Charikar, M., Chlamtac, E., Feige, U., Vijayaraghavan, A.: Detecting high log-densities: an $O(n^{1/4})$-approximation for densest k-subgraph. In: Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC 2010), pp. 201–210 (2010)
6. Chen, D.Z., Fleischer, R., Li, J.: Densest k-subgraph approximation on intersection graphs. In: Jansen, K., Solis-Oba, R. (eds.) WAOA 2010. LNCS, vol. 6534, pp. 83–93. Springer, Heidelberg (2011)
7. Feige, U., Langberg, M.: Approximation algorithms for maximization problems arising in graph partitioning. J. Algorithms 41(2), 174–211 (2001)
8. Feige, U., Peleg, D., Kortsarz, G.: The dense k-subgraph problem. Algorithmica 29(3), 410–421 (2001)
9. Golumbic, M.C., Trenk, A.N.: Tolerance Graphs. Cambridge University Press, Cambridge (2004)
10. Gröschel, M., Lovász, L., Schrijver, A.: Geometric Algorithms and Combinatorial Optimization. Springer, Heidelberg (1988)
11. Ibarra, O.H., Kim, C.E.: Fast approximation algorithms for the knapsack and sum of subset problems. Journal of the ACM 22 (1975)
12. Khot, S.: Ruling out ptas for graph min-bisection, dense k-subgraph, and bipartite clique. SIAM J. Comput. 36(4), 1025–1071 (2006)
13. Kortsarz, G., Peleg, D.: On choosing a dense subgraph. In: Proceedings of the 34th Annual Symposium on Foundations of Computer Science (FOCS 1993), pp. 692–701 (1993)
14. Lawler, E.L.: Combinatorial optimization - networks and matroids. Holt, Rinehart and Winston, New York (1976)
15. Liazi, M., Milis, I., Pascual, F., Zissimopoulos, V.: The densest k-subgraph problem on clique graphs. J. Comb. Optim. 14(4), 465–474 (2007)
16. Liazi, M., Milis, I., Zissimopoulos, V.: A constant approximation algorithm for the densest k-subgraph problem on chordal graphs. Inf. Process. Lett. 108(1), 29–32 (2008)
17. Perl, Y., Corneil, D.G.: Clustering and domination in perfect graphs. Discrete Applied Mathematics 9(1), 27–39 (1984)