Subexponential Parameterized Algorithm for Interval Completion

IVAN BLIZNETS, St. Petersburg Academic University of the Russian Academy of Sciences, Russia
FEDOR V. FOMIN, Department of Informatics, University of Bergen, Norway
MARCIN PILIPCZUK and MICHAŁ PILIPCZUK, Institute of Informatics, University of Warsaw, Poland

In the Interval Completion problem we are given an n-vertex graph G and an integer k, and the task is to transform G by making use of at most k edge additions into an interval graph. This is a fundamental graph modification problem with applications in sparse matrix multiplication and molecular biology. The question about fixed-parameter tractability of Interval Completion was asked by Kaplan et al. [FOCS 1994; SIAM J. Comput. 1999] and was answered affirmatively more than a decade later by Villanger et al. [STOC 2007; SIAM J. Comput. 2009], who presented an algorithm with running time $O(k^{2k}n^3m)$. We give the first subexponential parameterized algorithm solving Interval Completion in time $k^{O(\sqrt{k})}n^{O(1)}$. This adds Interval Completion to a very small list of parameterized graph modification problems solvable in subexponential time.

CCS Concepts: • **Theory of computation → Graph algorithms analysis**; **Fixed parameter tractability**;

Additional Key Words and Phrases: Subexponential algorithms, interval graphs, completion problems, graph modification problems

ACM Reference format:
Ivan Bliznets, Fedor V. Fomin, Marcin Pilipczuk, and Michał Pilipczuk. 2018. Subexponential Parameterized Algorithm for Interval Completion. *ACM Trans. Algorithms* 14, 3, Article 35 (June 2018), 62 pages.
https://doi.org/10.1145/3186896

1 INTRODUCTION

In the Interval Completion problem, we are asked if a given graph G can be complemented by at most k edges into an interval graph, that is, the intersection graph of intervals of the real line. This is a fundamental NP-complete problem, mentioned as problem GT35 in Garey and Johnson [17], arising naturally in different areas. In sparse matrix computations, the problem is equivalent to reordering columns and rows of a matrix, thereby reducing its profile [19]. In molecular biology, the problem models the task of building a map describing the relative positions of the clones [20, 26]. Interval Completion fits into the broader class of graph modification problems on which hundreds of papers have been written. The systematic study of the parameterized complexity

Authors' addresses: I. Bliznets, St. Petersburg Academic University of the Russian Academy of Sciences, 27 Fontanka, St.Petersburg 191023, Russia; email: ivanbliznets@tut.by; F. V. Fomin, Department of Informatics, University of Bergen, Postboks 7803, 5020 Bergen, Norway; email: fomin@ii.uib.no; M. Pilipczuk, Institute of Informatics, University of Warsaw, Banacha 2, 02-097, Warsaw, Poland; email: m.pilipczuk@mimuw.edu.pl; M. Pilipczuk, Institute of Informatics, University of Warsaw, Banacha 2, 02-097, Warsaw, Poland; email: michal.pilipczuk@mimuw.edu.pl.

https://doi.org/10.1145/3186896

of completion problems was initiated by Kaplan et al. in [24, 25], who showed that Chordal Completion, Strongly Chordal Completion, and Proper Interval Completion are fixed-parameter tractable (FPT). The parameterized complexity of Interval Completion remained open until 2007, when Villanger et al. [22, 30] settled this long-standing open problem in the affirmative by designing an algorithm with running time $O(k^{2k} \cdot n^3 m)$. A faster algorithm, with running time $O(6^k(n+m))$, was recently given by Cao [7–9].

Our main interest in Interval Completion is due to the new developments in parameterized complexity. It is well known (see, e.g., [13]) that for most of the natural parameterized problems the existence of subexponential parameterized algorithms can be refuted, unless the Exponential Time Hypothesis (ETH) [23] fails. Until recently, the only notable exceptions of parameterized subexponential problems were problems on special classes of graphs like planar graphs, or more generally, graphs excluding some fixed graph as a minor [10], and on tournaments [1]. Luckily, the structure of the "parameterized subexponential world" is much more interesting and complicated than was anticipated for a long time. It appeared very recently that several graph modification problems, mostly problems of completing to some graph class, like Chordal Completion, Threshold Completion, Proper Interval Completion , and Trivially Perfect Completion are solvable in subexponential time $k^{O(\sqrt{k})}n^{O(1)}$, where n is the input size and k is the number of edges in the completion [4, 11, 14, 15, 18]. On the other hand, even for completion problems for a vast majority of graph classes (even very simple ones, like cographs or complements of cluster graphs), it is possible to rule out the existence of subexponential parameterized algorithms [11, 27] under plausible complexity assumptions. Thus, subexponential-time solvability is a very unusual and exceptional property of a parameterized problem.

While the examples of subexponential-time solvability show that some parameterized NP-hard problems are significantly "easier" than most of the other problems from the same complexity class, we do not know why this is the case, what the underlying difference is, and how to identify such problems. The usual "prerequisites" for all parameterized graph modification problems solvable in subexponential time prior to this work were that establishing membership in FPT is easy (in most of the cases a simple branching does the job) and, moreover, the problem admits a polynomial kernel.[1] Interval Completion absolutely does not fit into this pattern: All known FPT algorithms solving this problem are quite non-trivial [7–9, 30] (it took 13 years to design the first such algorithm) and the existence of a polynomial kernel for Interval Completion stands open for a long time. This is why we find the subexponential-time solvability of Interval Completion striking.

Another interesting point about Interval Completion is the following. Completion problems have deep connections with width measures of graphs. For example, the treewidth of a graph, one of the most fundamental graph parameters, is the minimum, over all possible completions into a chordal graph, of the maximum clique size minus one. Similarly, the pathwidth of a graph can be defined as the minimum, over all possible completions into an interval graph, of the maximum clique size minus one. See the survey of Bodlaender for more information on these parameters [5]. Another important graph parameter is the treedepth, also known as the vertex ranking number, the ordered chromatic number, and the minimum elimination tree height. This parameter appears in various settings, in particular in the theory of sparse graphs developed by Nešetřil and Ossona de Mendez [29]. Mirroring the connection between treewidth and chordal graphs, pathwidth and interval graphs, the treedepth of a graph can be defined as the largest clique size in a completion to a trivially perfect graph. Similarly, we may observe a relation between the class of proper interval

[1]Recall that a *polynomial kernel* for a parameterized problem is a polynomial-time preprocessing routine that reduces an input instance (G, k) to an equivalent one of size bounded polynomially in k, without increasing the parameter.

Fig. 1. Graph classes and corresponding graph parameters. Inequalities on the right side are with ±1 slackness.

graphs and the bandwidth of a graph, as well as the class of threshold graphs and the vertex cover number of a graph. (For definitions of these graph classeswe refer to [6].) Taking into account relations between these graph classes and parameters, we arrive at the diagram presented in Figure 1. It is interesting to note that all completion problems to the graph classes depicted in Figure 1 were established to be solvable in subexponential parameterized time [4, 11, 15]. This article provides a subexponential algorithm for the last and the most difficult piece in Figure 1, namely, Interval Completion.

Our Results and Techniques. Our main result is the following theorem.

Theorem 1.1. *Interval Completion is solvable in time* $k^{O(\sqrt{k})} \cdot n^{O(1)}$.

We now describe briefly our techniques employed to prove Theorem 1.1, together with the main obstacles making our approach significantly different from the approaches used for previous subexponential algorithms.

The natural way to proceed then would be to follow the approach which worked nicely for other completion problems: focus on the structural definition of interval graphs (as opposed to the definition via forbidden induced subgraphs) and build an interval model of the output graph via dynamic programming. On a very high level, our algorithm works in this fashion: we gradually enumerate more and more rich families of different potential "dividing" structures in the final interval model—keeping a bound of $k^{O(\sqrt{k})}n^{O(1)}$ on the size of each family—and finally obtain structures rich enough to perform a dynamic programming algorithm with running time polynomial in the size of the enumerated families.

The natural "dividing" structures in all graph classes in Figure 1 are maximal cliques and clique separators, and the core part of the known subexponential algorithms for Chordal Completion [15], Proper Interval Completion [4], and Trivially Perfect Completion [11] is a combinatorial argument that bounds the number of candidates for such structures by $n^{O(\sqrt{k})}$. This, in combination with known polynomial kernels for these problems, yields a $k^{O(\sqrt{k})}$ bound on the number of candidates for maximal cliques and clique separators. A second step is to design a dynamic programming algorithm whose states are based on these structures. As the number of states is subexponential in k, the entire algorithm would run in subexponential parameterized time.

There are two major problems with this approach in the case of Interval Completion. First, although we are able to provide a combinatorial bound of $n^{O(\sqrt{k})}$ reasonable candidates for maximal cliques and clique separators in the output interval graph (see Lemma 3.4), the existence of the second ingredient—a polynomial kernel for Interval Completion—remains a notorious open problem. Observe that a $n^{O(\sqrt{k})}$ term is unacceptable in any fixed-parameter algorithm. To cope with this obstacle, we employ a much more insightful analysis of maximal cliques in the output interval graph, and arrive at a (finally useful) improved $k^{O(\sqrt{k})}n^8$ bound on the number of candidates.

The lack of a known polynomial kernel for the problem raises also one more difficulty. One of the more popular "atomic operations" in the known subexponential algorithms is to choose one vertex

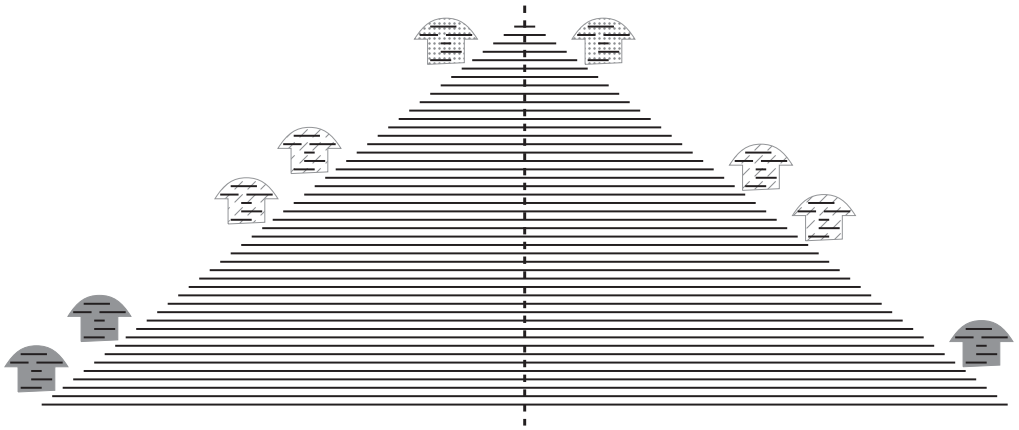

Fig. 2. An example of an interval graph with a large number of choices for left/right alignment. Within each pattern group (dotted, lined, solid), the small "mushroom" components can swap sides. A state of a dynamic programming algorithm at the middle clique marked with a dashed line would need to remember an alignment choice for each pattern group.

v and guess *all* edges from the solution incident with it, provided that there are at most $\sqrt{k}$ edges of the solution incident with v. In the presence of a polynomial kernel, such a step leads to $k^{O(\sqrt{k})}$ subcases—perfectly fine if we perform only a constant number of such steps. However, in the case of Interval Completion, such a step yields an (again) unacceptable $n^{O(\sqrt{k})}$ term in the running time. Luckily, a deep analysis of the structure of YES-instances to Interval Completion shows that there are actually only $k^{O(\sqrt{k})}n^{O(1)}$ reasonable ways to choose solution edges incident with such a "cheap" vertex, making the aforementioned "atomic operation" possible also in our case. Despite its triviality in the case of previous works, it turns out that the proof of the $k^{O(\sqrt{k})}n^{O(1)}$ bound is the most technical and involved part of our article.

The second major obstacle in our quest for a subexponential parameterized algorithm for Interval Completion appears when we try to develop a dynamic programming algorithm based on the knowledge of candidates for maximal cliques and clique separators in the output interval graph. Contrary to the case of Chordal Completion and Trivially Perfect Completion, it turns out that these structures are far from being sufficient to design a dynamic programming algorithm constructing a model of the output interval graph in a natural "left-to-right" manner. The reason is that the knowledge of a clique separator Ω in the output interval graph does not tell us much about which of the components of $G \setminus \Omega$ are to the left, and which are to the right of the separator Ω in an interval model of the output interval graph. (Recall that in an interval graph, each clique separator corresponds to a vertical line that pierces intervals belonging to the separator.) However, the knowledge of which vertices of G were already processed is crucial for constructing an interval model in a "left-to-right" manner.

An example illustrating why it is hard to deduce the alignment of the components of $G \setminus \Omega$ for a maximal clique or clique separator Ω is depicted in Figure 2. Here, a maximal clique Ω is marked with a vertical dashed line. The small "mushrooms" are components of $G \setminus \Omega$. Observe that one can swap (take a mirror image) the set of dotted mushrooms, striped mushrooms, and solid mushrooms independently of each other. Hence, a state of a dynamic programming algorithm needs to remember, apart from the maximal clique Ω, the alignment choice of each "pattern" group of mushrooms (dotted, striped, solid)—and there can be many of them.

Looking at the example in Figure 2, it is tempting to develop a different dynamic programming algorithm that processes the graph in a "top-to-bottom" manner, subsequently taking alignment decisions on each mushroom group, but not remembering the decision in the state between the groups. However, observe that if the graph locally looks like a proper interval graph (as opposed to the example in Figure 2), the "left-to-right" approach seems much more feasible. Hence, to make the dynamic programming approach work in the case of Interval Completion, we need to merge the "left-to-right" and "top-to-bottom" approaches, arriving at a quite technical definition of an actual state of dynamic programming.

We remark that the subexponential algorithm for Interval Completion cannot be obtained by modifying previous algorithms of Villanger et al. [30] and Cao [7–9]. In all these algorithms, one of the initial steps is to exhaustively branch on small induced forbidden subgraphs. That is, if the input graph contains an induced subgraph H of size at most c for some constant $c = O(1)$ such that H is not an interval graph, the algorithm branches in all possible ways to add an edge to break H. The subsequent arguments are strongly relying on the structural properties of the graphs excluding all small forbidden induced subgraphs for interval graphs (cf. the olive ring decomposition of [8, 9]). However, since such a recursive exhaustive branching cannot lead to time complexity better than single-exponential, this technique cannot be used in a subexponential algorithm.

A short comparison with the algorithm for seemingly similar Proper Interval Completion (PIC for short) is also in place. Although both algorithms follow the same general approach paved by Fomin and Villanger [15], the actual difficulties, and methods to avoid them, are completely different. First, in the PIC case a polynomial kernel is known [2], and a subexponential bound on both the number of candidates for maximal cliques Ω, and on the number of left/right choices for $G \setminus \Omega$, are not trivial, but relatively simple. The main difficulty in the PIC case lies in the fact that this information is not sufficient to perform a natural left-to-right dynamic programming, as one needs to ensure that no interval contains another in the output model; an issue non-existent in the interval case. To cope with this obstacle, in [4] the dynamic programming structure is also reengineered, and not only for a completely different reason than here, but also in a completely different manner—loosely speaking, apart from maximal cliques, the algorithm of [4] uses a type of separation similar to the classic $O^\star(10^n)$ exact algorithm for bandwidth of Feige [12].

Organization of the Article. We first introduce notation and preliminary results in Section 2, and give a more detailed, yet informal overview of the proof of Theorem 1.1 in Section 3.

Then, in Sections 4–8, we provide a full proof of Theorem 1.1. Section 4 describes a module-based reduction rule and introduces some auxiliary results on neighborhood classes in a (near) interval graph. In Section 5, we prove the subexponential bound on the number of candidates for *sections*, a technical notion close to a clique separator. In Section 6, we provide a bound of $k^{O(\sqrt{k})}n^{O(1)}$ reasonable ways to add solution edges incident to one vertex, provided that there are at most $\sqrt{k}$ such solution edges. After one additional combinatorial lemma in Section 7, we describe the final dynamic programming algorithm in Section 8.

Section 9 concludes the article and suggests directions of future research.

2 PRELIMINARIES

Graph Notation. In most cases, we follow standard graph notation. For a set of vertices $A \subseteq V$, $G[A]$ denotes the subgraph induced by A, and $G \setminus A$ is a shorthand for $G[V \setminus A]$. For a graph G, by $\mathtt{cc}(G)$ we denote the family of vertex sets of connected components of G. For a path P and two vertices $x, y \in V(P)$, by $P[x, y]$ we denote the subpath of P between x and y, inclusive. For a vertex v, we use $N_G(v)$ and $N_G[v]$ to denote the open and the closed neighborhood of v, respectively.

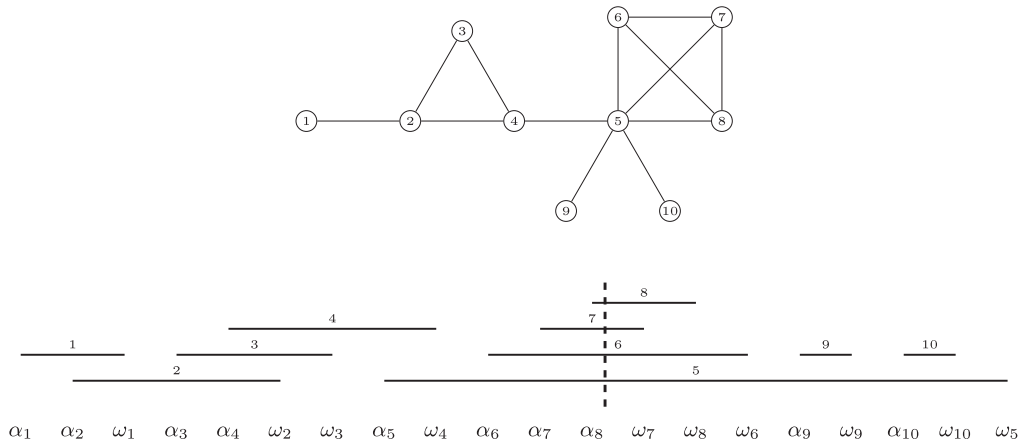

Fig. 3. An example of a graph with an interval model and its combinatorial representation. The vertical dashed line represents one of the maximal cliques of the graph, being section $\Omega_\sigma(12)$. We remark that this is not the canonical model of the represented graph (assuming the natural order on the vertex labels): for the canonical model, one should swap events ω_2 with ω_3 and ω_7 with ω_8.

For a vertex set $S \subseteq V$, we denote by $N_G(S)$ the set $\bigcup_{v \in S} N_G(v) \setminus S$. Two disjoint sets $A, B \subseteq V$ are *anti-adjacent* if there is no edge with one endpoint in A and the second endpoint in B.

For any graph G we shall speak about, we implicitly fix some arbitrary total ordering $\prec$ on $V(G)$. We shall use this ordering to break ties and canonize some objects (interval models, completion sets, solutions, etc.). Such a canonization will turn out to be helpful when handling greedy arguments in the final dynamic programming routine.

Interval Graphs. A graph G is an *interval graph* if it admits an intersection model of the following form: each vertex is assigned a closed interval on a real line, and two vertices are adjacent if and only if their intervals intersect.

We formalize the notion of a model in the following combinatorial way. For each $v \in V(G)$, we create two symbols α_v and ω_v, henceforth called *events*, and denote $\mathcal{E}(X) = \bigcup_{v \in X} \{\alpha_v, \omega_v\}$ for any $X \subseteq V(G)$. An *interval model* is a permutation (bijection) $\sigma : \mathcal{E}(V(G)) \to \{1, 2, \ldots, 2n\}$ such that

(1) for each $v \in V(G)$ we have $\sigma(\alpha_v) < \sigma(\omega_v)$ (an interval starts before it ends), and
(2) for each $u, v \in V(G)$ we have $uv \notin E(G)$ if and only if $\sigma(\omega_v) < \sigma(\alpha_u)$ or $\sigma(\omega_u) < \sigma(\alpha_v)$ (vertices are non-adjacent if and only if their intervals are disjoint).

The numbers $1, 2, \ldots, 2n$ in the codomain of a model σ are called *positions*.

Informally speaking, the aforementioned combinatorial notion of an interval model corresponds to a "real" model, where no two endpoints of intervals coincide (which we can assume without loss of generality). The permutation σ corresponds to the order of endpoints of intervals: α_v represents the starting (left) endpoint of the interval associated with v, and ω_v represents the ending (right) endpoint. See Figure 3 for an example.

Given an interval model σ of a graph G, we say that an event ε_1 is *before* or *to the left* of an event ε_2 iff $\sigma(\varepsilon_1) < \sigma(\varepsilon_2)$. In this situation, we also say that ε_2 is *later* or *to the right* of ε_1.

For an interval model σ of a graph G and a set $X \subseteq V(G)$, we denote by $\alpha_\sigma(X)$ and $\omega_\sigma(X)$, respectively, the first and last positions where events of $\mathcal{E}(X)$ appear in σ.

For an interval model σ of a graph G and an integer p, the set

$$\Omega_\sigma(p) = \{v \in V(G) : \sigma(\alpha_v) \leq p < \sigma(\omega_v)\}$$

is called a *section at position p*. By somehow abusing the notation, for an event ε we write $\Omega_\sigma(\varepsilon)$ for $\Omega_\sigma(\sigma(\varepsilon))$, and call it a *section at event* ε. We omit the subscript if it is clear from the context. Note that every section is a clique in G.

Intuitively speaking, a section is a set of vertices whose intervals become "pinned down" by a vertical line drawn *just after* event $\sigma^{-1}(p)$ (see Figure 3). Thus, all these intervals share a common point, so they are pairwise adjacent in the graph.

It is worth observing that for any integers $p \leq q$ and an interval model σ of G, the set $\Omega_\sigma(p) \cap \Omega_\sigma(q)$ consists of all vertices $v \in V(G)$ whose intervals start not later than p and end after q (i.e., with $\alpha_v \leq p$ and $\omega_v > q$), while $\Omega_\sigma(p) \cup \Omega_\sigma(q)$ consists of all vertices $v \in V(G)$ whose intervals contain a point just after p or a point just after q (i.e., either $\alpha_v \leq p < \omega_v$ or $\alpha_v \leq q < \omega_v$).

We refer to an inclusion-wise maximal clique of a graph G as to a maximal clique. It is well-known [21] that $\Omega \subseteq V(G)$ is a maximal clique in an interval graph G with model σ if and only if it is a section drawn between a starting and ending event: there exist $v_1, v_2 \in V(G)$ (possibly $v_1 = v_2$) such that $\Omega = \Omega_\sigma(\alpha_{v_2})$ and $\sigma(\alpha_{v_2}) + 1 = \sigma(\omega_{v_1})$.

We also use the following notions of maximality and minimality in interval models. Let $X \subseteq V(G)$, where G is an interval graph with a fixed model σ. We say that $v \in X$ is *interval-maximal* in X (w.r.t. σ) if for no other $w \in X$ it holds that $\sigma(\alpha_w) < \sigma(\alpha_v) < \sigma(\omega_v) < \sigma(\omega_w)$. Analogously, $v \in X$ is *interval-minimal* in X (w.r.t. σ) if for no other $w \in X$ it holds that $\sigma(\alpha_v) < \sigma(\alpha_w) < \sigma(\omega_w) < \sigma(\omega_v)$. Clearly, each non-empty set of vertices has an interval-maximal and interval-minimal vertex, but these vertices may not be defined uniquely.

We recall that in linear time we can check if a given graph G is an interval graph, and if this is the case, find an interval model of G [21]. In our work, we will need a slightly stronger statement.[2]

Lemma 2.1 (♠). *Given an interval graph G and two cliques $\Omega_1, \Omega_2 \subseteq V(G)$, one can in polynomial time check whether there exists an interval model of G that starts with all starting events of $\mathcal{E}(\Omega_1)$ and ends with all ending events of $\mathcal{E}(\Omega_2)$.*

For the final dynamic programming routine, we need to "canonize" a model of an interval graph G. Recall that we have fixed a total order $\prec$ on $V(G)$; assume $V(G) = \{v_1, v_2, \ldots, v_n\}$ where $v_1 \prec v_2 \prec \cdots \prec v_n$. For a model σ of G, we consider a tuple

$$(\sigma(\alpha_{v_1}), \sigma(\alpha_{v_2}), \ldots, \sigma(\alpha_{v_n}), \sigma(\omega_{v_n}), \sigma(\omega_{v_{n-1}}), \ldots, \sigma(\omega_{v_1}))$$

and define a *canonical model* of G to be the model with the aforementioned tuple being lexicographically minimum among all models of G.

We note two properties of a canonical model σ that are of our interest. The first one is straightforward.

Lemma 2.2. *Assume σ is the canonical model of an interval graph G. Then, for each $u, v \in V(G)$, if $\sigma(\alpha_u) + 1 = \sigma(\alpha_v)$, then $u \prec v$ and if $\sigma(\omega_u) + 1 = \sigma(\omega_v)$, then $u \succ v$. That is, the canonical model orders consecutive starting/ending points of the intervals according to $\prec$.*

The second one says that canonizing a model fixes an order in which modules with the same neighborhood appear in the model.

Lemma 2.3 (♠). *Let σ be the canonical model of an interval graph G. Let $X \subseteq V(G)$ be a clique, and let $C_1, C_2, \ldots, C_s$ be components of $G \setminus X$ (not necessarily all of them) such that $N_G(v) \setminus C_i = X$ for every $1 \leq i \leq s$ and every $v \in C_i$. Since the components C_i are pairwise non-adjacent, $\omega_\sigma(C_i) < \alpha_\sigma(C_j)$ or $\omega_\sigma(C_j) < \alpha_\sigma(C_i)$ for any $i \neq j$. Without loss of generality, assume that*

$$\alpha_\sigma(C_1) < \omega_\sigma(C_1) < \alpha_\sigma(C_2) < \omega_\sigma(C_2) < \cdots < \alpha_\sigma(C_s) < \omega_\sigma(C_s).$$

[2]Proofs marked with ♠ are straightforward and were moved to the Appendix in order not to disturb the reasoning.

For each $1 \leq i \leq s$, *let* $x_i \in C_i$ *be the first vertex of* C_i *in the order* $\prec$. *Then*

$$x_1 \prec x_2 \prec \cdots \prec x_s.$$

That is, σ *sorts the components* C_i *according to the order of their* $\prec$*-minimum vertices.*

Interval Completion. For a graph G, a *completion* of G is a set $F \subseteq \binom{V(G)}{2} \setminus E(G)$ such that $G + F := (V(G), E(G) \cup F)$ is an interval graph. A completion is *minimal* if it is inclusion-wise minimal, and *minimum* if it has minimum possible cardinality. In the Interval Completion problem the input consists of a graph G and an integer k, and we ask for a completion of G of size at most k. For an instance (G, k) of Interval Completion, a completion of cardinality at most k is called a *solution*. The notions of *minimal* and *minimum solutions* are defined naturally.

For a completion F in a graph G, we say that v is *touched* by F if there is an edge in F incident with v; otherwise, v is *untouched*. A set of vertices X is *touched* if it contains a touched vertex, and *untouched* otherwise. We also say that a vertex $v \in V(G)$ is *cheap* (with respect to the completion F) if at most $\sqrt{k}$ edges of F are incident with v; a vertex is *expensive* if it is not cheap. Note that there are at most $2k$ touched vertices and at most $2\sqrt{k}$ expensive ones. For a completion F and a vertex $v \in V(G)$, by $F(v)$ we denote the set of edges $e \in F$ that are incident with v.

We now canonize solutions F to an Interval Completion instance (G, k). Given a partial order $\prec$ on a finite set U, we define a partial order on the family of subsets of U as follows: if $A, B \subseteq U$, then we first sort the elements of A and B according to $\prec$, and then compare the obtained sequences lexicographically. By somehow abusing the notation, we denote by $\prec$ the imposed order on the subsets of U as well.

This definition automatically extends the partial order $\prec$ on $V(G)$ first onto $\binom{V(G)}{2}$, and then onto the family of completions of G. We define the canonical solution to (G, k) to be the minimum solution in the order $\prec$ among all minimum solutions to (G, k).

Given an instance (G, k) of Interval Completion, we start with augmenting it in the following way. We add a universal vertex $\mathfrak{r}$ adjacent to all vertices of $V(G)$, and two vertices $\mathfrak{r}_L$ and $\mathfrak{r}_R$, adjacent only to $\mathfrak{r}$, obtaining a graph G'. We assume $\mathfrak{r} \prec \mathfrak{r}_L \prec v \prec \mathfrak{r}_R$ for any $v \in V(G)$. Note that for any completion F of G, F is also a completion of G': given a model of $G + F$, we may construct a model of $G' + F$ by preceding the events of $\mathcal{E}(V(G))$ with $\alpha_{\mathfrak{r}}, \alpha_{\mathfrak{r}_L}, \omega_{\mathfrak{r}_L}$ and succeeding them with $\alpha_{\mathfrak{r}_R}, \omega_{\mathfrak{r}_R}, \omega_{\mathfrak{r}}$. Consequently, in every minimal completion of G', the vertices $\mathfrak{r}$, $\mathfrak{r}_L$, and $\mathfrak{r}_R$ are untouched. Thus, henceforth we assume that, whenever we consider an instance (G, k) to Interval Completion, G already contains vertices $\mathfrak{r}$, $\mathfrak{r}_L$, and $\mathfrak{r}_R$. By Lemmas 2.2 and 2.3 (applied to $X = \{\mathfrak{r}\}$), the canonical model of any completion of G starts with $\alpha_{\mathfrak{r}}, \alpha_{\mathfrak{r}_L}, \omega_{\mathfrak{r}_L}$ and ends with $\alpha_{\mathfrak{r}_R}, \omega_{\mathfrak{r}_R}, \omega_{\mathfrak{r}}$.

A short informal rationale for this augmentation is that in some places of the algorithm we would like to pick the "first/last untouched vertex whose interval ends/starts after/before position p" or "an untouched vertex whose interval contains the interval of v"; note that $\mathfrak{r}_L/\mathfrak{r}_R$ is always a good candidate for the first choice, and $\mathfrak{r}$ for the second one. Also, the addition of $\mathfrak{r}$ makes G connected and makes all (interesting to us) sections at positions $1 \leq p < 2n$ non-empty.

3 OVERVIEW OF THE ALGORITHM

In this section, we provide an informal overview on the proof of Theorem 1.1.

3.1 Module Reduction Rule

We start with a simple module-based reduction rule. Recall that $M \subseteq V(G)$ is a *module* in a graph G if $N(v_1) \setminus M = N(v_2) \setminus M$ for any $v_1, v_2 \in M$. (Equivalently, for any $v \notin M$ we have either $M \subseteq N(v)$ or $M \cap N(v) = \emptyset$.)

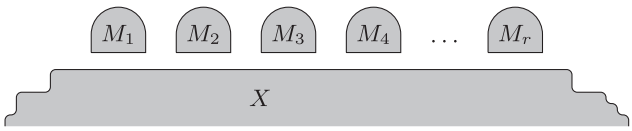

Fig. 4. The alignment of most of the components M_i in the model σ of the interval graph $G + F$.

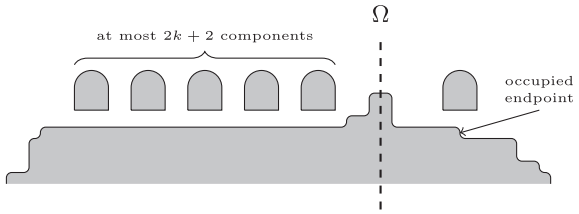

Fig. 5. Motivation for the Module Reduction Rule.

Assume that in a YES-instance (G, k) of Interval Completion we have recognized a set $X \subseteq V(G)$ such that many (significantly more than $2k$) connected components $M_1, M_2, \ldots, M_r$ of $G \setminus X$ are modules, fully adjacent to X. Then, it is easy to observe that any solution F to (G, k) needs to yield an ordering σ of $G + F$ similar to the one depicted in Figure 4: X becomes a clique, and most of the components M_i are drawn one after another on the "plateau" formed by all the intervals of the vertices of X. Moreover, note that all but at most k components M_i need to induce interval graphs, and all but at most $2k$ components M_i are left untouched by the solution F.

However, if $r \geq 2k + 2$, then there are at least two such untouched M_i's—say M_1 and M_2—and, in the interval graph $G + F$ they force X to be a clique, reserving space between M_1 and M_2 for any other M_i with $G[M_i]$ being an interval graph. Thus, we may reduce the number of such M_i's to $2k + 2$, without changing the answer to the instance (G, k).

Reduction Rule 3.1 (Module Reduction Rule). Let (G, k) be an instance of Interval Completion. Assume there exist $X \subseteq V(G)$ and connected components $M_1, M_2, \ldots, M_{2k+3}$ of $G \setminus X$ that are modules in G and, moreover, $N(M_i) = N(M_1)$ for each $1 \leq i \leq 2k + 3$. Then proceed as follows. If for more than k indices i the subgraph $G[M_i]$ is not an interval graph, return that (G, k) is a NO-instance. Otherwise, pick arbitrary j such that $G[M_j]$ is an interval graph and remove M_j from G.

We remark here that the Module Reduction Rule can be applied exhaustively in polynomial time, using the module decomposition of the graph G: It is easy to observe that, if the rule is applicable, then all components M_i are children of a single *union* node in the module decomposition tree.

Let us now explain our motivation for introducing the Module Reduction Rule (see also Figure 5). In many steps of the algorithm, we analyze some clique Ω of the interval graph $G + F$, and we would like to control the number of connected components of $G \setminus \Omega$. There are two types of such components: the ones that are modules, and the ones that are not modules. If the Module Reduction Rule has been applied exhaustively, then we have a bound on the number of components of the first type for a fixed neighborhood $X \subseteq \Omega$; observe that there are only $2(|\Omega| + 1)$ choices for such neighborhood.

Consider now a component C that is not a module, that is, there are vertices $v_1, v_2 \in C$ s.t. $N(v_1) \setminus C \neq N(v_2) \setminus C$. If C is not touched by the solution, then an endpoint of a vertex of $(N(v_1) \triangle N(v_2)) \setminus C$ needs to be "occupied" by C, that is, one of the events of w needs to lie between

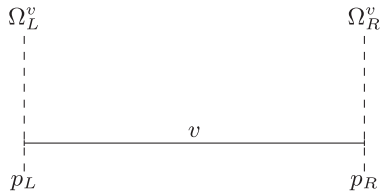

Fig. 6. Notation around a cheap vertex v.

the first and the last of the events of $\mathcal{E}(\{v_1, v_2\}) \subseteq \mathcal{E}(C)$. Furthermore, no two such untouched components C can "occupy" the same event of $\mathcal{E}((N(v_1) \triangle N(v_2)) \setminus C)$ as they are nonadjacent. As $N(v_1) \cup N(v_2) \subseteq \Omega \cup C$, there are at most $2|\Omega|$ such events, and every component C that is neither touched nor a module uses one such event. Together with at most $2k$ touched components, we have a bound of $2k + 2|\Omega|$ on the number of components of C that are not modules.

3.2 Dynamic Programming

We now sketch the dynamic programming algorithm for Interval Completion.

Let us first try a somewhat naive approach, based on the subexponential algorithm for the Chordal Completion problem. We would like to construct an interval completion F of G by building an interval model for the final graph $G + F$; the completion F is formed exactly by the edges that need to be added to make the built model correct. This interval model will be constructed from left to right, intuitively by a "sweeping" dynamic programming algorithm. In such an approach, a state of the dynamic programming algorithm would need to consist of (i) the current section Ω (clique) of the interval graph, and (ii) information for which components of $G \setminus \Omega$ we have already drawn their interval model and which are still pending. To make the approach work, we would need to have a subexponential number of candidates for both pieces of a state.

Our first combinatorial result is that there is actually a subexponential number of candidates for sections.

Theorem 3.1. *Given an Interval Completion instance (G, k) where the Module Reduction Rule is not applicable, one can in $k^{O(\sqrt{k})}n^{O(1)}$ time enumerate a family $\mathcal{S}$ of $k^{O(\sqrt{k})}n^{17}$ subsets of $V(G)$ such that for any minimal solution F to (G, k), in the canonical model σ of $G + F$ all sections of σ belong to $\mathcal{S}$.*

We sketch the proof of Theorem 3.1 in Section 3.3 and provide a full proof in Section 5.

However, as already explained in Figure 2 in the Introduction, a number of reasonable choices for the second ingredient—the number of partitions of components of $G \setminus \Omega$ between the left and right side of a section Ω—can be exponential in $|\Omega|$, and we were unable to design an argument decreasing this number of choices. Observe that the example in Figure 2 suggests a different direction: a dynamic programming algorithm should rather sweep such a "pyramid" from top to bottom, as then it does not need to remember the left/right alignment of already processed "mushrooms."

What should a state in such a dynamic programming algorithm look like? Consider a vertex $v \in V(G)$ and let σ be the canonical model of some minimal solution F. Let $p_L = \sigma(\alpha_v)$ and $p_R = \sigma(\omega_v)$ be the positions of endpoints of the interval of v and let $\Omega_L^v = \Omega_\sigma(p_L)$ and $\Omega_R^v = \Omega_\sigma(p_R)$ be the sections and these positions (see Figure 6). Note that, by Theorem 3.1, there are only $k^{O(\sqrt{k})}n^{O(1)}$ candidates for the defined objects. In the model σ, every connected component C of $G \setminus (\Omega_L^v \cup \Omega_R^v)$ lives either before p_L, between p_L and p_R, or after p_R; in the second case, C is fully adjacent to v in $G + F$, and in the other cases C is anti-adjacent to v in $G + F$. If we knew $F(v)$ (or, have a subexponential number of candidates for it), we could classify the components of $G \setminus (\Omega_L^v \cup \Omega_R^v)$

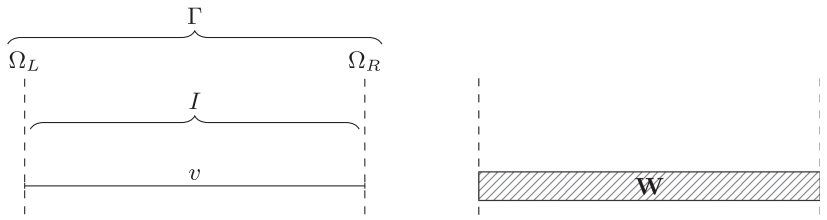

Fig. 7. A world $\mathbf{W}$ with its most important elements (to the left) and its symbolic notation used in subsequent figures (to the right).

between the ones that live on top of v in σ (i.e., between p_L and p_R) and the ones that live before or after v in σ. Note that such a partition would precisely correspond to the intended dynamic programming state for the graph in Figure 2: for a vertex v in the main pyramid, it would partition the graph into the part "above" and "below" v, allowing us to sweep the pyramid top to bottom.

It seems hard to obtain a subexponential number of candidates for $F(v)$ for expensive vertices v. On the other hand, for a cheap vertex v there is a trivial bound of $n^{\sqrt{k}}$ candidates for $F(v)$ that, in addition, would become $k^{O(\sqrt{k})}$ if a polynomial kernel for Interval Completion was known. Our second combinatorial result is that one can circumvent the need for a polynomial kernel in the above reasoning.

Theorem 3.2. *Given an Interval Completion instance (G, k) where the Module Reduction Rule is not applicable, and a designated vertex $v \in V(G)$, one can in $k^{O(\sqrt{k})}n^{O(1)}$ time enumerate a family $\mathcal{F}$ of at most $k^{O(\sqrt{k})}n^{70}$ subsets of $V(G)$ such that for any minimal solution F to (G, k) for which v is cheap w.r.t. F, the set $\{w \in V(G) : vw \in F\}$ belongs to $\mathcal{F}$.*

We sketch the proof of Theorem 3.2 in Section 3.4 and provide a full proof in Section 6.

Armed with Theorem 3.2, for a cheap vertex v w.r.t. a minimal solution F, we define a world to be tuple $\mathbf{W}$ consisting of v, p_L, p_R, Ω_L^v, Ω_R^v, and $F(v) \in \mathcal{F}$ as defined above; thus, we have only $k^{O(\sqrt{k})}n^{O(1)}$ worlds to consider. See Figure 7 for a schematic picture of a world.

In the dynamic programming algorithm, with a world $\mathbf{W}$ it is natural to associate the following computational task: what is the optimum way to arrange the events between positions p_L and p_R, that is, among vertices of $N_{G+F(v)}[v]$? That is, the area between p_L and p_R is the "important area" of a world $\mathbf{W}$. In the pyramid example (Figure 2), this would correspond to computing optimum arrangements of events for the upper parts of the pyramid; the task becomes more complex as we move down the pyramid. Observe that a world does not distinguish which vertices of $G \setminus N_{G+F(v)}[v]$ are before or after v in the model σ. In particular, in the pyramid example it does not make a decision (yet) on the left/right arrangement of "mushrooms" below the cheap vertex v, and it does not remember the exact decision on the left/right arrangement of "mushrooms" above the cheap vertex v.

However, the family of worlds is not rich enough to make transitions from one world to a second one in the dynamic programming algorithm. Consider a world $\mathbf{W} = (v, p_L, p_R, \Omega_L^v, \Omega_R^v, F(v))$ and a second world $\mathbf{W}' = (w, q_L, q_R, \Omega_L^w, \Omega_R^w, F(w))$ that is in some sense close to $\mathbf{W}$; say, the cheap vertex w of $\mathbf{W}'$ is the cheap vertex in Ω_L^v with the rightmost endpoints among the cheap vertices in Ω_L^v. The situation now differs depending on the relative position of the right endpoints p_R and q_R.

If the right endpoint q_R of w is to the left of p_R, then the worlds $\mathbf{W}$ and $\mathbf{W}'$ naturally partition the positions into segments from q_L to p_L, p_L to q_R, and q_R to p_R (see the left panel of Figure 8).

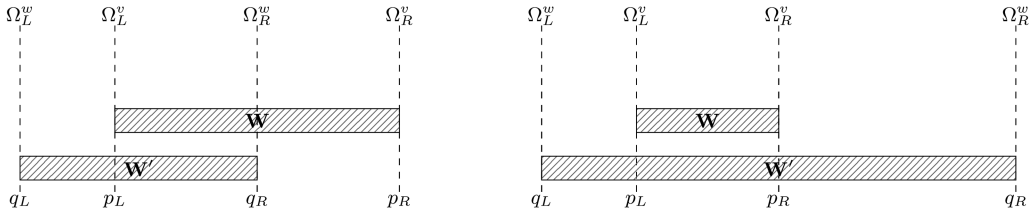

Fig. 8. Two cases of how two close worlds interact.

Furthermore, from all four sections Ω_L^v, Ω_R^v, Ω_L^w, and Ω_R^w and sets $F(v)$ and $F(w)$ one can deduce which events lie in which segments. However, now to compute the optimum arrangement of events between p_L and p_R, the dynamic programming algorithm would need to use information about optimum arrangements between p_L and q_R and q_R and p_R. In other words, our dynamic programming algorithm would need to use not a single world as a state, but a pair of worlds, and ask for optimum arrangements between two consecutive endpoints of the cheap vertices of these worlds.

The situation becomes even more complicated in the case when q_R is to the right of p_R, that is, the worlds are "nested" (see the right panel of Figure 8). In particular, we should expect $N_{G+F(v)}[v] \subseteq N_{G+F(w)}[w]$. Then, the vertices of $G \setminus (\Omega_L^v \cup \Omega_R^v \cup \Omega_L^w \cup \Omega_R^w)$ can be partitioned into three types:

(1) the vertices $u \notin N_{G+F(w)}[w]$ live in σ either before q_L or after q_R, and we do not worry about them at the moment;
(2) the vertices $u \in N_{G+F(v)}[v]$ live in σ between p_L and p_R, and the entry of the dynamic programming algorithm for the world $\mathbf{W}$ should take care of the optimum arrangement of their endpoints; and
(3) the vertices $u \in N_{G+F(w)}[w] \setminus N_{G+F(v)}[v]$ can live either between q_L and p_L or between p_R and q_R.

The last case causes problems for two reasons.

Issue A: It is unknown which such vertices u lie between q_L and p_L, and which between p_R and q_R.

Issue B: While the choice of w implies that only expensive vertices of Ω_L^v can start between q_L and p_L (and there are at most $2\sqrt{k}$ of them), the gap between p_R and q_R can be arbitrarily large, and thus we expect the dynamic programming algorithm to inspect some already computed entries to determine the optimum arrangement of the events there.

We remedy Issue B by taking into consideration not one "close" world $\mathbf{W}'$, but two of them: one $\mathbf{W}'$ as defined above, and another one defined symmetrically at the endpoint p_R. More formally, a *terrace* is defined as a tuple of three worlds $\mathbf{W}^{\text{in}}$, $\mathbf{W}_1^{\text{out}}$, and $\mathbf{W}_2^{\text{out}}$ with their respective cheap vertices v, v_1, and v_2 where

- $\sigma(\alpha_{v_i}) < \sigma(\alpha_v) < \sigma(\omega_v) < \sigma(\omega_{v_i})$ for $i = 1, 2$, and
- among the cheap vertices satisfying the previous condition, v_1 has the rightmost starting event in the model σ, whereas v_2 has the leftmost ending event.

See Figure 9 for some additional notation. Note that we allow $\mathbf{W}_1^{\text{out}} = \mathbf{W}_2^{\text{out}}$. In a terrace, we are interested in the optimum way to arrange events in one of the dotted areas in Figure 9. Observe that each vertex whose interval is fully contained in one of these areas belongs to $I := (N_{G+F}(v_1) \cap N_{G+F}(v_2)) \setminus (N_{G+F}(v) \cup \Omega_L^1 \cup \Omega_R^2)$.

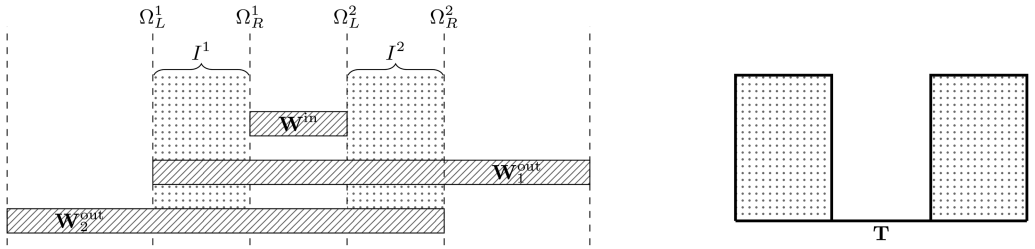

Fig. 9. A terrace with its most important notation (to the left) and its symbolic notation used in subsequent figures (to the right). The dotted areas are the "important" areas for a terrace: the left one has borders Ω_L^1, Ω_R^1, and interior I^1, and the right one has borders Ω_L^2, Ω_R^2, and interior I^2.

The crucial observation is that, by the choice of v_1 and v_2, each vertex of $\Omega_R^1 \cap \Omega_L^2$ that has an endpoint in the dotted areas (i.e., does not belong to $\Omega_L^1 \cap \Omega_R^2$) needs to be expensive and, consequently, there are at most $2\sqrt{k}$ such vertices. Denote the set of these vertices as K, that is, $K = (\Omega_R^1 \cap \Omega_L^2) \setminus (\Omega_L^1 \cap \Omega_R^2)$.

By the above observation, the notion of a terrace with two "outer" worlds $\mathbf{W}_1^{\text{out}}$ and $\mathbf{W}_2^{\text{out}}$ solves Issue B (uncontrolled complexity of the segment from p_R to q_R). That is, the space between the left endpoints of v_1 and v is constrained in the following sense: every cheap vertex with the left endpoint in this space needs to have a right endpoint before the right endpoint of v. A symmetric statement holds for right endpoints in the space between the right endpoint of v and the right endpoint of v_2.

Issue A, in the language of terraces, means that we would like to reason how the vertices of I are split between areas I^1 and I^2. In Section 7, we provide a combinatorial argument showing that the number of left/right choices between I^1 and I^2 is $k^{O(\sqrt{k}+|K|)}$; with $|K| = O(\sqrt{k})$ this gives the desired subexponential bound. In this overview, we provide a simplified argument in Section 3.5 with a worse bound of $k^{O(k^{3/4})}$.

To sum up, we have $k^{O(\sqrt{k})}n^{O(1)}$ reasonable choices for a terrace, together with the partition of the set I into dotted areas I^1 and I^2.

Recall that a terrace was a method to cope with one of the cases with relative position of the right endpoints of the worlds $\mathbf{W}$ and $\mathbf{W}'$: when the segment of the cheap vertex v of $\mathbf{W}$ is contained in the segment of the cheap vertex w of $\mathbf{W}'$. In the other case, we argued that it should be sufficient to consider a state consisting of two worlds and ask for optimum arrangement of events between two consecutive endpoints of the cheap vertices of the worlds.

This approach meets the notion of a terrace in our final definition of a state of dynamic programming. Namely, the states of the final dynamic programming algorithm are *pairs* of terraces or worlds, together with their "important areas"; for such a state, we ask for the best way to arrange events in the *intersection* of the important areas (see Figure 10). As the number of such dynamic programming states is bounded by $k^{O(\sqrt{k})}n^{O(1)}$, the number of states fits within the promised time bound.

Let us now briefly elaborate why such a definition of a state is rich enough. First, let us be a bit more precise in the definition. Recall that a terrace $\mathbf{T} = (\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}})$ has two important areas, to the left and to the right of the world $\mathbf{W}^{\text{in}}$. We fix the notion of the *first important area* as the area between $p_L^1(\mathbf{T}) := p_L(\mathbf{W}_1^{\text{out}})$ and $p_R^1(\mathbf{T}) := p_L(\mathbf{W}^{\text{in}})$, and the *second important area* as the area between $p_L^2(\mathbf{T}) := p_R(\mathbf{W}^{\text{in}})$ and $p_R^2(\mathbf{T}) := p_R(\mathbf{W}_2^{\text{out}})$. For a world $\mathbf{W}$, we define both the first

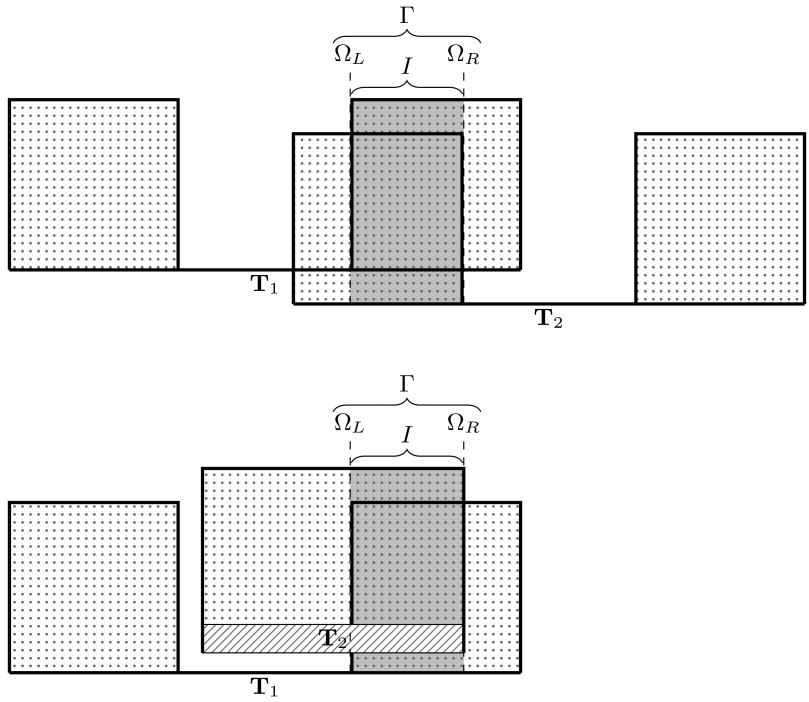

Fig. 10. A final DP state defined by two terraces (above), and by a terrace and a world (below). In this DP state we ask for the optimal way to arrange events in the gray area.

and the second important areas to be equal to the area "above" the world, that is, this is the area between $p_L^1(\mathbf{W}) := p_L^2(\mathbf{W}) := p_L(\mathbf{W})$ and $p_R^1(\mathbf{W}) := p_R^2(\mathbf{W}) = p_R(\mathbf{W})$.

To be consistent with the notation in Section 8, we define a *state* to be a pair $\mathrm{S} = (\mathbf{T}_1, \mathbf{T}_2)$, where every $\mathbf{T}_i$ is a terrace or a world. We require that

$$p_L^1(\mathbf{T}_2) \le p_L^2(\mathbf{T}_1) < p_R^1(\mathbf{T}_2) \le p_R^2(\mathbf{T}_1).$$

Note that this definition allows $\mathbf{T}_1 = \mathbf{T}_2$. For every state $\mathrm{S} = (\mathbf{T}_1, \mathbf{T}_2)$, the dynamic programming algorithm asks for an optimum way to arrange events between $p_L(\mathrm{S}) := p_L^2(\mathbf{T}_1)$ and $p_R(\mathrm{S}) := p_R^1(\mathbf{T}_2)$, that is, in the intersection of the *second* important area of $\mathbf{T}_1$ and the *first* important area of $\mathbf{T}_2$. We call this area the *important area* of the state S. Note that the knowledge of a state S gives us full knowledge about which events should be arranged there, as well as about the sections at the boundary of the important area.

The dynamic programming algorithm computes the values for states in the order of increasing size of their important areas. As a base case, if the important area of a state consists of $O(\sqrt{k})$ events, we can check all permutations of events there by brute force. On the other hand, an optimum ordering of the whole graph will be computed at the cell for the state consisting of twice the world for the vertex r. At every step, given a state S with $\Omega(\sqrt{k})$ events in its important area, the dynamic programming algorithm tries to partition the important area of S into two or three smaller important areas of other states (while keeping the condition that in adjacent states the boundary sections are equal). Note that states with smaller important areas have already precomputed value by our choice of the order of computations. To prove that the algorithm is correct, we

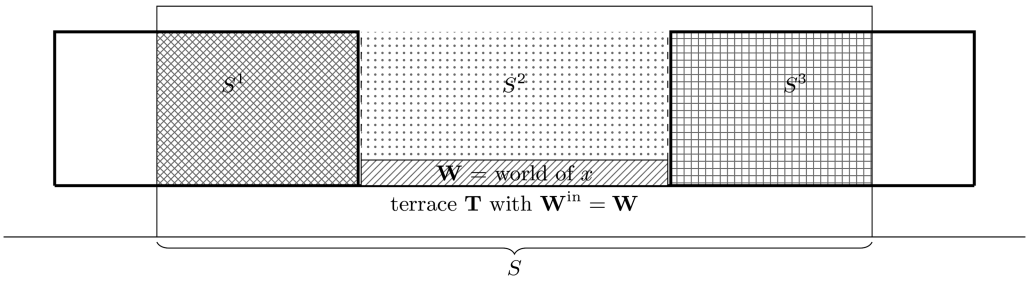

Fig. 11. A computation of the value for DP state S when both endpoints of x lie in the important area of the state and we glue values from three substates.

need to show that in the canonical model σ of the canonical solution, every state present there can be decomposed into two or three smaller states.

Consider now a state $\mathrm{S} = (\mathbf{T}_1, \mathbf{T}_2)$ that is present in σ and that asks for an arrangement of $\Omega(\sqrt{k})$ events; by taking the constant in the Ω-notation to be large, we observe that most of these events are of cheap vertices.

The crux now lies in the following choice: let x be a cheap vertex that has an endpoint in the important area of S, and maximizes the span of its interval in the model σ intersected with the important area of S. More precisely, x maximizes

$$\min(\sigma(\omega_x), p_R(\mathrm{S}) + 1) - \max(\sigma(\alpha_x), p_L(\mathrm{S})).$$

We consider two cases, depending on whether in σ both endpoints of x lie in the important area of S, or only one of them.

Both endpoints of x lie in the important area of S. Let $\mathbf{W}$ be the world of x and let $\mathbf{T}$ be the terrace with $\mathbf{W}^{\text{in}} = \mathbf{W}$; note that, given σ, the worlds $\mathbf{W}_1^{\text{out}}$ and $\mathbf{W}_2^{\text{out}}$ in $\mathbf{T}$ are defined uniquely. The maximality of x ensures that $\mathbf{W}_1^{\text{out}}$ starts not later than $p_L(\mathrm{S})$, while $\mathbf{W}_2^{\text{out}}$ ends not earlier than $p_R(\mathrm{S})$, as otherwise the underlying cheap vertex of $\mathbf{W}_1^{\text{out}}$ or $\mathbf{W}_2^{\text{out}}$ would be a better candidate for x. Here, the assumption that the important area of $\mathbf{W}^{\text{in}}$ is a subset of important areas of $\mathbf{W}_1^{\text{out}}$ and $\mathbf{W}_2^{\text{out}}$ plays a crucial role. Consequently, the important area of S decomposes into the important areas of states $\mathrm{S}^1 = (\mathbf{T}_1, \mathbf{T})$, $\mathrm{S}^2 = (\mathbf{W}, \mathbf{W})$, and $\mathrm{S}^3 = (\mathbf{T}, \mathbf{T}_2)$. See Figure 11 for an illustration.

Only one endpoint of x lies in the important area of S. By symmetry, assume that this is the right endpoint; the other case is symmetric. As in the previous case, let $\mathbf{W}$ be the world of x and let $\mathbf{T}$ be the terrace with $\mathbf{W}^{\text{in}} = \mathbf{W}$. Again, the maximality of x ensures that the world $\mathbf{W}_2^{\text{out}}$ in $\mathbf{T}$ ends not earlier than $p_R(\mathrm{S})$, as otherwise its underlying cheap vertex would be a better candidate for x. Consequently, the important area of S decomposes into the important areas of states $\mathrm{S}^1 = (\mathbf{T}_1, \mathbf{W})$ and $\mathrm{S}^2 = (\mathbf{T}, \mathbf{T}_2)$. See Figure 12 for an illustration.

Since the dynamic programming algorithm computes the value for every of the $k^{O(\sqrt{k})}n^{O(1)}$ states by looking into pairs and triples of previously computed states, it runs within the promised time bound. This concludes the overview of the dynamic programming algorithm in the proof of Theorem 1.1.

3.3 Candidates for Sections and Maximal Cliques

In this section, we sketch the proof of Theorem 3.1. As an intermediate step, we provide an enumeration algorithm for potential maximal cliques in the Interval Completion problem, showing the following.

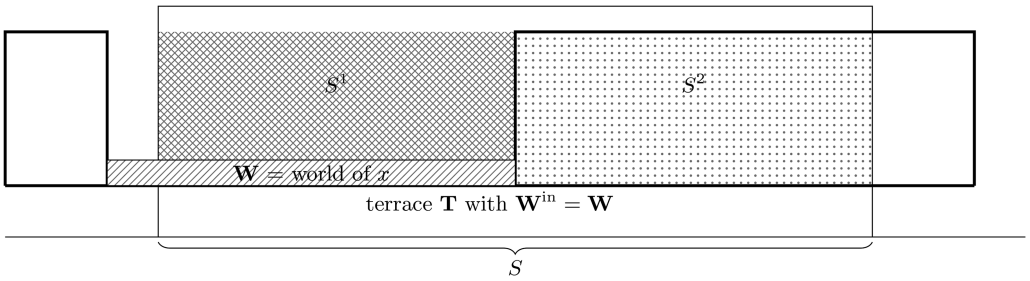

Fig. 12. A computation of the value for DP state S when only one endpoint of x lies in the important area of the state and we glue values from two substates.

Theorem 3.3. *Given an Interval Completion instance (G, k) where the Module Reduction Rule is not applicable, one can in $k^{O(\sqrt{k})}n^{O(1)}$ time enumerate a family $\mathcal{K}$ of $k^{O(\sqrt{k})}n^8$ subsets of $V(G)$ such that for any minimal solution F to (G, k), all maximal cliques of $G + F$ belong to $\mathcal{K}$.*

It is not hard to see that Theorem 3.3 implies Theorem 3.1.

Proof of Theorem 3.1. Let (G, k) be an Interval Completion instance, and F be a minimal solution to (G, k) with σ being the canonical model of $G + F$. Clearly, $\emptyset$, $\{\mathrm{r}\}$, $\{\mathrm{r}, \mathrm{r}_L\}$, and $\{\mathrm{r}, \mathrm{r}_R\}$ are sections of σ; we include them into $\mathcal{S}$ at the beginning.

Let $\Omega_\sigma(p)$ be a section of σ. Without loss of generality, assume that $\Omega_\sigma(p)$ is not one of the four aforementioned "obvious" sections. Let $p_1 \leq p$ be the largest integer such that $\Omega_\sigma(p_1)$ is a maximal clique of $G + F$; such p_1 always exists as $p_1 = 2$ with $\Omega_\sigma(2) = \{\mathrm{r}, \mathrm{r}_L\}$ is a candidate value. Symmetrically, we define p_2 to be the smallest integer with $p_2 \geq p$ such that $\Omega_\sigma(p_2)$ is a maximal clique of $G + F$.

Let $r = |\Omega_\sigma(p_1) \setminus \Omega_\sigma(p_2)|$. We infer that σ places events of $\{\omega_v : v \in \Omega_\sigma(p_1) \setminus \Omega_\sigma(p_2)\}$ on positions $p_1 + 1, p_1 + 2, \ldots, p_1 + r$, and then it places events of $\{\alpha_v : v \in \Omega_\sigma(p_2) \setminus \Omega_\sigma(p_1)\}$ on positions $p_1 + r + 1, p_1 + r + 2, \ldots, p_2$; otherwise there would be a section between sections $\Omega_\sigma(p_1)$ and $\Omega_\sigma(p_2)$ that would yield a maximal clique, contradicting the choice of p_1 or of p_2. Moreover, by Lemma 2.2 the events of $\{\omega_v : v \in \Omega_\sigma(p_1) \setminus \Omega_\sigma(p_2)\}$ are sorted according to the reversed total order $\prec$, while the events of $\{\alpha_v : v \in \Omega_\sigma(p_2) \setminus \Omega_\sigma(p_1)\}$ are sorted according to the total order $\prec$. Consequently, the set $\Omega_\sigma(p)$ can be deduced from the maximal cliques $\Omega_\sigma(p_1)$ and $\Omega_\sigma(p_2)$ (both belonging to the set $\mathcal{K}$ given by Theorem 3.3) and the value of $p - p_1$, for which we have $n + 1$ choices. Indeed,

- if $p - p_1 \leq r$, then $\Omega_\sigma(p)$ equals $\Omega_\sigma(p_1)$ without the $(p - p_1)$ vertices of $\Omega_\sigma(p_1) \setminus \Omega_\sigma(p_2)$ that are latest in $\prec$;
- if $p - p_1 > r$, then $\Omega_\sigma(p)$ equals $\Omega_\sigma(p_2)$ without the $(p_2 - p)$ vertices of $\Omega_\sigma(p_2) \setminus \Omega_\sigma(p_1)$ that are latest in $\prec$.

Theorem 3.1 follows. □

Hence, we now sketch the proof of Theorem 3.3. We first start with an $n^{O(\sqrt{k})}$ bound, and then argue how to obtain the actual FPT bound of Theorem 3.3.

Let us fix an Interval Completion instance (G, k), its minimal solution F, a model σ of $G + F$, and a maximal clique $\Omega = \Omega_\sigma(p)$. Recall that $\sigma(\alpha_{v_2}) = p$ and $\sigma(\omega_{v_1}) = p + 1$ for some vertices v_1 and v_2. Without loss of generality, assume that Ω is different from the two "obvious" maximal cliques $\{\mathrm{r}, \mathrm{r}_L\}$ and $\{\mathrm{r}, \mathrm{r}_R\}$ and, consequently, $3 < p < 2n - 3$ and $v_1, v_2 \notin \{\mathrm{r}, \mathrm{r}_L, \mathrm{r}_R\}$.

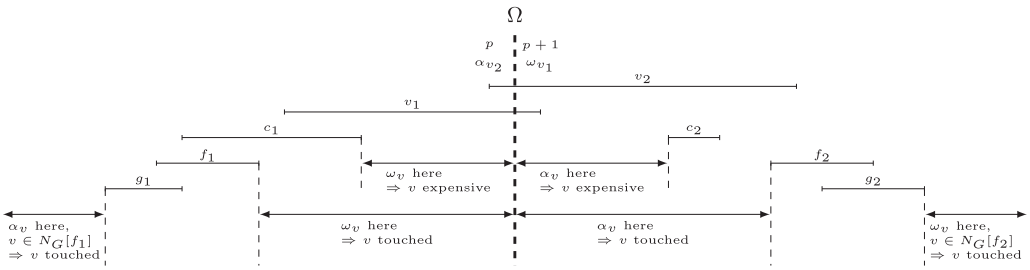

Fig. 13. The anatomy of a maximal clique Ω, with eight important vertices guessed by the algorithm.

We define the following vertices (see also Figure 13):

(1) c_1 is the cheap vertex with the rightmost position of ω_{c_1}, among all cheap vertices c satisfying $\sigma(\omega_c) \leq \sigma(\omega_{v_1}) = p + 1$;
(2) c_2 is the cheap vertex with the leftmost position of α_{c_2}, among all cheap vertices c satisfying $\sigma(\alpha_c) \geq \sigma(\alpha_{v_2}) = p$;
(3) f_1 is the untouched vertex with the rightmost position of ω_{f_1}, among all untouched vertices f satisfying $\sigma(\omega_f) \leq \sigma(\omega_{v_1}) = p + 1$;
(4) f_2 is the untouched vertex with the leftmost position of α_{f_2}, among all untouched vertices f satisfying $\sigma(\alpha_f) \geq \sigma(\alpha_{v_2}) = p$;
(5) g_1 is the untouched vertex with the leftmost position of α_{g_1}, among all untouched vertices of $N_G[f_1] \setminus \{\Omega \setminus \{v_1\}\}$;
(6) g_2 is the untouched vertex with the rightmost position of ω_{g_2}, among all untouched vertices of $N_G[f_2] \setminus \{\Omega \setminus \{v_2\}\}$.

Let us remark that some of these vertices can in fact be equal. We also remark that all quantifications in the aforementioned definitions are done on non-empty sets: $\mathfrak{r}_L$ is a good candidate for both c_1 and f_1, $\mathfrak{r}_R$ is a good candidate for both c_2 and f_2, f_1 is a good candidate for g_1, and f_2 is a good candidate for g_2. Hence, all these vertices are well-defined.

Also, observe that $v_1 \in N_G[v_2]$, as otherwise $v_1v_2 \in F$ and, by swapping the events ω_{v_1} and α_{v_2} in the model σ, we obtain a model for $G + (F \setminus \{v_1v_2\})$, contradicting the minimality of F.

We say that a vertex v *lies to the left* of the clique Ω if $\sigma(\omega_v) \leq p + 1$, and *lies to the right* if $\sigma(\alpha_v) \geq p$. Clearly, v_1, c_1, f_1, g_1 lie to the left of Ω and v_2, c_2, f_2, g_2 lie to the right of Ω. Note that, perhaps a bit counterintuitively, if $v = v_1 = v_2$, then v lies both to the left and to the right of Ω.

Let w be any vertex of the graph. Observe that if some vertex of $N_{G+F}[w]$ lies to the left of Ω, then $\sigma(\alpha_w) \leq p$. Similarly, if some vertex of $N_{G+F}[w]$ lies to the right of Ω, then $\sigma(\omega_w) \geq p + 1$. In particular, if both these events happen, then w belongs to Ω.

Define now the following sets.

$$F_i^\circ = \{v \in V(G) : vc_i \in F\} \text{ for } i = 1, 2;$$
$$X_1^\circ = \{v \in V(G) : \sigma(\omega_{c_1}) < \sigma(\omega_v) \leq p + 1\};$$
$$X_2^\circ = \{v \in V(G) : p \leq \sigma(\alpha_v) < \sigma(\alpha_{c_2})\}.$$

As c_1 and c_2 are cheap, $|F_1^\circ|, |F_2^\circ| \leq \sqrt{k}$. By the definition of c_1 and c_2, all vertices of $X_1^\circ \cup X_2^\circ$ are expensive. Note that $|X_1^\circ \cap X_2^\circ| \leq 1$ and $X_1^\circ \cap X_2^\circ$ is non-empty only if it consists of $v_1 = v_2$. Therefore, $|X_1^\circ| + |X_2^\circ| \leq 2\sqrt{k} + 1$.

We now show the main combinatorial observation: the knowledge of vertices v_1, v_2, c_1, c_2 and sets F_i° and X_i° for $i = 1, 2$ already uniquely defines the clique Ω.

Lemma 3.4.

$$\Omega = (N_G[\{v_1, c_1\} \cup X_1^\circ] \cup F_1^\circ) \cap (N_G[\{v_2, c_2\} \cup X_2^\circ] \cup F_2^\circ).$$

Proof. The inclusion "$\supseteq$" is immediate from the previous discussion: every vertex $v \in N_G[\{v_1, c_1\} \cup X_1^\circ] \cup F_1^\circ$ is either to the left of Ω in $G + F$, or at least one neighbor of v is to the left of Ω. Similarly, for every $u \in N_G[\{v_2, c_2\} \cup X_2^\circ] \cup F_2^\circ$, at least one vertex from $N_{G+F}[u]$ is to the right of Ω in $G + F$. Hence, we now focus on the other inclusion.

Without loss of generality, assume there exists a vertex $v \in \Omega$ that does not belong to F_2° or to $N_G[\{v_2, c_2\} \cup X_2^\circ]$. In particular, $v \notin \{v_1, v_2, c_2\}$, and hence $\alpha_v < p$. As $v \notin F_2^\circ$ and $vc_2 \notin E(G)$, we have $\sigma(\omega_v) < \sigma(\alpha_{c_2})$. Moreover, by the definition of X_2°, v is not adjacent in G to any vertex whose starting event lies between positions p and $\sigma(\alpha_{c_2}) - 1$. Hence, v is not adjacent in G to any vertex whose starting event lies on or after position p.

Consider an ordering σ' that is created from the model σ by moving the event ω_v to the position just before the event α_{v_2} (i.e., we move ω_v to the position p and shift all events on positions p and later by one to the right). By our previous arguments, σ' is a valid interval model of some completion F' of G. As $v \in \Omega$, the event ω_v has been moved to the left during this operation, and $F' \subseteq F$. Moreover, $vv_2 \in F \setminus F'$, which contradicts the minimality of F. □

As the sets F_i° and X_i° are of size $O(\sqrt{k})$, Lemma 3.4 already gives us an $n^{O(\sqrt{k})}$ bound on the number of candidates for maximal cliques in $G + F$. However, in the absence of a polynomial kernel for Interval Completion, we need to work further to obtain the bound promised in Theorem 3.3. In this quest, we will make use of the vertices f_i and g_i.

The choice of vertices v_i, c_i, f_i, and g_i for $i = 1, 2$ contributes with factor n^8 to the bound of Theorem 3.3; our goal is to produce $k^{O(\sqrt{k})}$ candidates for a fixed choice of these eight vertices. To this end, we develop a branching algorithm that maintains a choice of candidate sets X_1, X_2, F_1, and F_2 for X_1°, X_2°, F_1°, and F_2°, respectively, and a guess K on the clique Ω. At each step of the recursion, the algorithm outputs the current set K as a possible choice, and branches into $k^{O(1)}$ number of subcases, choosing one additional vertex to include into one of the sets X_i or F_i, updating K accordingly.[3] As the depth can be bounded by $O(\sqrt{k})$, we obtain the promised bound of $k^{O(\sqrt{k})}$ candidates for the clique Ω.

Obviously, the main technical difficulty lies in the argumentation that there are only $k^{O(1)}$ reasonable choices in each step of the recursion. Here the guess on the vertices f_i and g_i helps: we carefully analyze the structure of connected components of $G \setminus (X_1 \cup X_2 \cup K \cup \{v_1, v_2, c_1, c_2, f_1, f_2, g_1, g_2\})$ and argue that only a limited number of vertices may possibly live between f_1 and f_2 in the model σ of $G + F$. Moreover, in this argument we heavily rely on the fact that the Module Reduction Rule is not applicable, which in various places enables us to bound the number of components that are considered. For all the details of the reasoning, we refer to Section 5.

3.4 Guessing Fill-In Edges with Fixed Endpoint

Armed with the bound on the number of possible sections (Theorem 3.1), we move to sketch the most technical result of our work, namely, Theorem 3.2.

[3] This statement is not completely true; in some cases we are able only to guess a *neighborhood* of a vertex in X_i°, without indicating the vertex itself. However, this is sufficient for the purpose of the reasoning of Lemma 3.4.

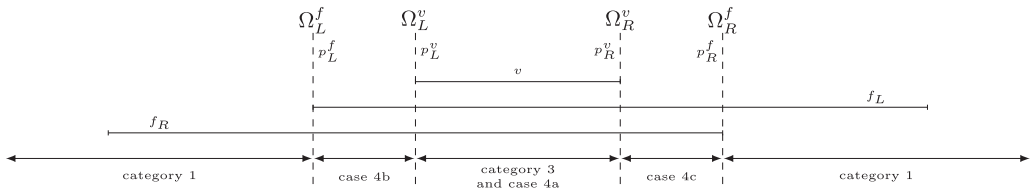

Fig. 14. Situation around the vertex v in the proof of Theorem 3.2, together with categories and cases of Lemma 3.5.

Let (G, k) and $v \in V(G)$ be as in the statement. Fix a minimal completion F of the Interval Completion instance (G, k), and fix a model σ of $G + F$. We define the following (see also Figure 14).

(1) Denote $p_L^v = \sigma(\alpha_v)$ and $p_R^v = \sigma(\omega_v)$.
(2) Let f_L be the untouched vertex with the rightmost starting endpoint among untouched vertices f satisfying $\sigma(\alpha_f) \leq p_L^v < p_R^v \leq \sigma(\omega_f)$.
(3) Let f_R be the untouched vertex with the leftmost ending endpoint among untouched vertices f satisfying $\sigma(\alpha_f) \leq p_L^v < p_R^v \leq \sigma(\omega_f)$.
(4) Denote $p_L^f = \sigma(\alpha_{f_L})$ and $p_R^f = \sigma(\omega_{f_R})$.
(5) Denote $\Omega_L^f = \Omega_\sigma(p_L^f)$, $\Omega_L^v = \Omega_\sigma(p_L^v)$, $\Omega_R^v = \Omega_\sigma(p_R^v - 1)$, and $\Omega_R^f = \Omega_\sigma(p_R^f - 1)$.

Note that r is a good candidate for both f_L and f_R, thus these vertices exist. We remark also that it may happen that $v = f_L$, $v = f_R$, or $f_L = f_R$. However, we may say the following about the order of these vertices.

$$\sigma(\alpha_{f_R}) \leq p_L^f \leq p_L^v < p_R^v \leq p_R^f \leq \sigma(\omega_{f_L}).$$

We start by enumerating all possible choices of vertices f_L, f_R and sections Ω_L^f, Ω_L^v, Ω_R^v, Ω_R^f, using the family $\mathcal{S}$ of Theorem 3.1. By the bound of Theorem 3.1, there are at most $k^{O(\sqrt{k})}n^{70}$ subcases (henceforth called *branches*) to consider. In the rest of the proof we aim to compute a single set B of size $O(k^5)$ for a single choice of the aforementioned two vertices and four sections, such that B contains $\{w : vw \in F\}$ for any minimal solution F to (G, k) for which the choice of f_L, f_R and Ω_L^f, Ω_L^v, Ω_R^v, Ω_R^f is correct. When the set B is computed, we insert all its subsets of size at most $\sqrt{k}$ into the family $\mathcal{F}$.

Thus, henceforth we fix a choice of f_L, f_R and Ω_L^f, Ω_L^v, Ω_R^v, Ω_R^f and we assume that the guess of these vertices and sets is correct for a minimal solution F with model σ of $G + F$. Observe that we should expect the following:

$$v \in \Omega_L^v \cap \Omega_R^v,$$
$$f_L, f_R \in \Omega_L^f \cap \Omega_R^f,$$
$$\Omega_L^f \cap \Omega_R^f \subseteq \Omega_L^f \cap \Omega_R^v \subseteq \Omega_L^v \cap \Omega_R^v,$$
$$\Omega_L^f \cap \Omega_R^f \subseteq \Omega_L^v \cap \Omega_R^f \subseteq \Omega_L^v \cap \Omega_R^v.$$

We maintain also a set B^{sure} of vertices w for which we deduce that $vw \in F$ is implied by the choice of f_L, f_R and Ω_L^f, Ω_L^v, Ω_R^v, Ω_R^f. We start with $B^{\mathrm{sure}} = (\Omega_L^v \cup \Omega_R^v) \setminus N_G(v)$. If at any point the size of B^{sure} exceeds k, we discard the current branch.

We start with the following observation, directly implied by the assumption that f_L and f_R are untouched and $|F| \leq k$.

Lemma 3.5. *For any connected component C of $G \setminus (\Omega_L^f \cup \Omega_L^v \cup \Omega_R^v \cup \Omega_R^f)$ the following holds:*

1. *If $C \cap N_G(f_L) \cap N_G(f_R) = \emptyset$, then $\omega_\sigma(C) < p_L^f$ or $\alpha_\sigma(C) > p_R^f$. In particular, $vw \notin E(G) \cup F$ for every $w \in C$.*
2. *If C contains a vertex of $N_G(f_L) \cap N_G(f_R)$, then $p_L^f < \alpha_\sigma(C) < \omega_\sigma(C) < p_R^f$ and $C \subseteq N_G(f_L) \cap N_G(f_R)$.*
3. *If, moreover, C contains a neighbor of v in G, then $p_L^v < \alpha_\sigma(C) < \omega_\sigma(C) < p_R^v$ and $vw \in E(G) \cup F$ for every $w \in C$.*
4. *In the last case, if $C \subseteq (N_G(f_L) \cap N_G(f_R)) \setminus N_G(v)$, then one of the following cases holds:*
   (a) *$p_L^v < \alpha_\sigma(C) < \omega_\sigma(C) < p_R^v$ and $vw \in F$ for every $w \in C$. Moreover, in this case $N_G(C) \subseteq \Omega_L^v \cup \Omega_R^v$.*
   (b) *$p_L^f < \alpha_\sigma(C) < \omega_\sigma(C) < p_L^v$ and $vw \notin F$ for every $w \in C$. Moreover, in this case $N_G(C) \subseteq \Omega_L^f \cup \Omega_L^v$.*
   (c) *$p_R^v < \alpha_\sigma(C) < \omega_\sigma(C) < p_R^f$ and $vw \notin F$ for every $w \in C$. Moreover, in this case $N_G(C) \subseteq \Omega_R^f \cup \Omega_R^v$.*

   Moreover, if $|C| > k$, then the first option does not happen.

By Lemma 3.5, we can sort the connected components of $G \setminus (\Omega_L^f \cup \Omega_L^v \cup \Omega_R^v \cup \Omega_R^f)$ into three *categories*, depending on whether they fall into point 1, 3, or 4. Naturally, we may ignore the components from the first category in the construction of B, while for every C from the second category, we need to include $C \setminus N(v)$ in B^{sure}. The last category is the most interesting, as we are not able to directly decide whether the vertices of the component should be inserted into B^{sure} or not. Instead, we resort to constructing a set B of $O(k^5)$ size that contains all components that may fall into case 4a of the last category.

The subpoints of this category (i.e., 4a, 4b, and 4c) are henceforth called *cases*. Note that for each connected component C we know its category, but we do not know its case if it falls into category 4.

We now perform some cleaning. If there exists a component $C \in \mathsf{cc}(G \setminus (\Omega_L^f \cup \Omega_L^v \cup \Omega_R^v \cup \Omega_R^f))$ that does not fall into any category (e.g., we have $C \not\subseteq N_G(f_L) \cap N_G(f_R)$, but C contains a common neighbor of f_L and f_R), we discard the current branch. Moreover, we may include into B^{sure} all non-neighbors of v that lie in a connected component C that falls into category 3 of Lemma 3.5, that is, that contains a neighbor of v.

Clearly, only at most k components fall into case 4a of Lemma 3.5, since each such component induces at least one fill edge incident to v. However, we do not know which of the components falling into category 4 are in fact those interesting ones. Hence, our main task now is to pinpoint a set of $O(k^4)$ potential components falling into category 4 for which case 4a may possibly happen. As each such component is of size at most k, this would conclude the proof of Theorem 3.2.

Let $\mathcal{C}$ be the family of all connected components C of $G \setminus (\Omega_L^f \cup \Omega_L^v \cup \Omega_R^v \cup \Omega_R^f)$ that fall into category 4 of Lemma 3.5, that is, $C \subseteq (N_G(f_L) \cap N_G(f_R)) \setminus N_G(v)$. We distinguish the following subfamilies that correspond to the subcases of category 4.

$$\mathcal{C}_v = \{C \in \mathcal{C} : N_G(C) \subseteq \Omega_L^v \cup \Omega_R^v\},$$
$$\mathcal{C}_L = \{C \in \mathcal{C} : N_G(C) \subseteq \Omega_L^f \cup \Omega_L^v\},$$
$$\mathcal{C}_R = \{C \in \mathcal{C} : N_G(C) \subseteq \Omega_R^f \cup \Omega_R^v\}.$$

If $\mathcal{C}_v \cup \mathcal{C}_L \cup \mathcal{C}_R \neq \mathcal{C}$, we discard the current branch. Moreover, for any $C \in \mathcal{C}_v \setminus (\mathcal{C}_L \cup \mathcal{C}_R)$ we include all vertices of C into B^{sure}, as such a component will surely fall into case 4a.

Our goal now is to focus on C_L and pinpoint a small set of components of $C_L \cap C_v$ that may possibly fall into case 4a of Lemma 3.5. The arguments for C_R will be symmetrical.

To this end, we will construct a family $\mathcal{T} \subseteq C_L$ of *troublesome* components. Informally speaking, a component is troublesome if it is highly unclear where or how it should live in the model σ. We will argue that there is a bounded number of troublesome components (strictly speaking, $O(k^2)$ of them) and any component that falls into case 4a of Lemma 3.5 is in some sense "close" to a troublesome component.

We first focus on components $C \in C_L \cap C_R$. Observe that for such a component we have $N_G(C) \subseteq \Omega_L^v \cap \Omega_R^v$. Denote $P = \Omega_L^f \cap \Omega_R^f$ and $K = (\Omega_L^v \cap \Omega_R^v) \setminus P$. By the choice of f_L and f_R, each vertex in K is touched by the solution F and, consequently, $|K| \leq 2k$. If there exists a vertex $v \in C$ with $P \not\subseteq N_G(v)$, then necessarily C is touched by the solution. Otherwise, $P \subseteq N_G(v) \subseteq P \cup K$ for any $v \in C$ and, since the Module Reduction Rule is not applicable, we infer that there are only $O(k^2)$ components of $C_L \cap C_R$. We treat all of them as troublesome ones, and put them into $\mathcal{T}$.

Furthermore, we put into $\mathcal{T}$ all connected components $C \in C_L$ that cannot be drawn in the model of a completion of G between sections Ω_L^f and Ω_L^v without adding a fill-in edge. More formally, we denote $F_L = \binom{\Omega_L^v}{2} \setminus E(G) \subseteq F$ and define the following.

Definition 3.6. A component $C \in C_L \cap C_v$ is *freely drawable* if there exists an interval model σ_C of $(G + F_L)[C \cup \Omega_L^v]$ that starts with all starting events of $\mathcal{E}(\Omega_L^v \cap \Omega_L^f)$ and ends with all ending events of $\mathcal{E}(\Omega_L^v)$.

Observe that one can recognize freely drawable components in polynomial time using Lemma 2.1.

It is easy to see that each component that is not freely drawable either is touched by the solution F, or falls into case 4c. However, in the latter case we have $C \in C_L \cap C_R$, and all such components have already been considered troublesome. Hence, we expect at most $2k$ not freely drawable components of $(C_L \cap C_v) \setminus C_R$, and we put all of them into $\mathcal{T}$.

We now inspect the possible order of the starting endpoints of the vertices of $\Omega_L^v \setminus \Omega_L^f$; all these endpoints appear between positions p_L^f and p_L^v. We denote

$$X = \bigcup_{C \in C_L \setminus C_v} N_G(C) \cap \Omega_L^v.$$

It turns out that any component $C \in (C_L \cap C_v) \setminus C_R$ that contains a vertex $w \in C$ with $X \not\subseteq N_G(w)$ is necessarily touched by F: the solution F needs to make w adjacent either to the entire X, or to some vertices of the connected component of $C_L \setminus C_v$ that neighbors a vertex of $X \setminus N_G(w)$. Thus, we may treat all such components C as troublesome, and assume henceforth that each remaining component $C \in (C_L \cap C_v) \setminus C_R$ is both freely drawable and fully adjacent to X. We refer to Figure 15 for an illustration.

Now observe that if a component $C \in (C_L \cap C_v) \setminus C_R$ is freely drawable, then there exist vertices $v_1, v_2 \in C$ with

$$N_G(v_1) \cap \Omega_L^v = \Phi_1(C) := \bigcap_{w \in C} N_G(w) \cap \Omega_L^v,$$
$$N_G(v_2) \cap \Omega_L^v = \Phi_2(C) := \bigcup_{w \in C} N_G(w) \cap \Omega_L^v.$$

Consider now two components $C_1, C_2 \in (C_L \cap C_v) \setminus C_R$. If neither of them is touched by F (in particular, neither of them falls into case 4a), then we should expect $\Phi_2(C_1) \subseteq \Phi_1(C_2)$ or $\Phi_2(C_2) \subseteq \Phi_1(C_1)$, depending on the relative order of C_1 and C_2 in the model σ. Hence, if this is not

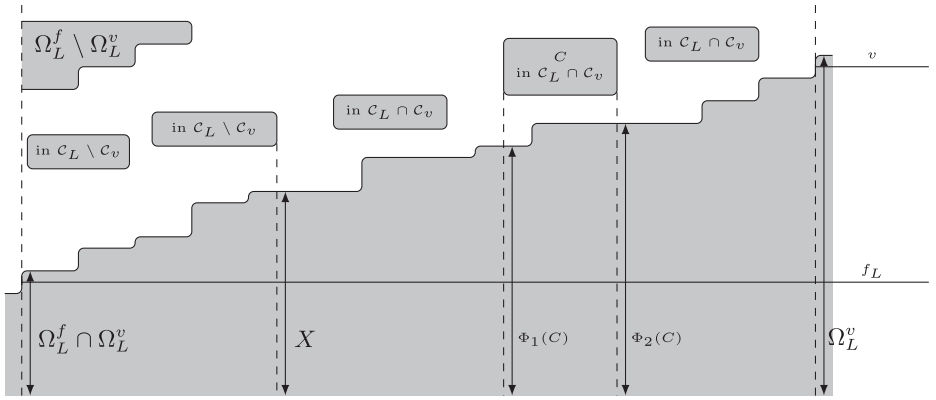

Fig. 15. The situation between Ω_L^f and Ω_L^v.

the case, we have a *conflict* between C_1 and C_2: one of these two components needs to be touched by F.

We construct an auxiliary conflict graph, where each vertex corresponds to a not-yet-troublesome component of $(C_L \cap C_v) \setminus C_R$, and each edge corresponds to a conflict; by the previous argumentation, the components touched by the solution need to form a vertex cover of this auxiliary conflict graph. Hence, we may compute a 2-approximate vertex cover of the conflict graph, and consider all components of this vertex cover as troublesome.

This step concludes the recognition of troublesome components $\mathcal{T}$.

We now observe that

$$(G + F_L)\left[\Omega_L^v \cup \bigcup ((C_L \cap C_v) \setminus (C_R \cup \mathcal{T}))\right]$$

is an interval graph and, moreover, it admits an interval model that starts with the starting events of X and ends with the ending events of Ω_L^v. The crucial observation now is the following: if for some $C \in (C_L \cap C_v) \setminus (C_R \cup \mathcal{T})$, the sets $\Phi_1(C)$ and $\Phi_2(C)$ differ significantly from sets $\Phi_1(D)$ and $\Phi_2(D)$ for all $D \in \mathcal{T}$, then no troublesome component will interfere with the representation of C between positions p_L^f and p_L^v and, consequently, C is untouched by the solution and falls into case 4b. The exhaustive application of the Module Reduction Rule ensures that only a bounded number of components C may have sets $\Phi_1(C)$ and $\Phi_2(C)$ similar to some troublesome component. As there are only $O(k^2)$ troublesome components, we are left only with a bounded number of candidates for case 4a. This concludes the sketch of the proof of Theorem 3.2.

3.5 A Simplified Argument for the Subexponential Number of Left/Right Choices for a Terrace

Recall that we are considering a terrace consisting of worlds $\mathbf{W}^{\text{in}}$, $\mathbf{W}_1^{\text{out}}$, and $\mathbf{W}_2^{\text{out}}$ with their respective cheap vertices v, v_1, and v_2. All other notation is depicted in Figure 9. We are interested in the optimum way to arrange events in one of the dotted areas in Figure 9. Observe that each vertex whose interval is fully contained in one of these areas belongs to $I := (N_{G+F}(v_1) \cap N_{G+F}(v_2)) \setminus (N_{G+F}(v) \cup \Omega_L^1 \cup \Omega_R^2)$. Our goal is to construct a subexponential number of reasonable partitions $I = I^1 \uplus I^2$ between the left and the right dotted area, with a guarantee that we may restrict our attention only to those partitions.

The critical observation made in Section 3.2 is that the set $K = (\Omega_R^1 \cap \Omega_L^2) \setminus (\Omega_L^1 \cap \Omega_R^2)$ consists only of expensive vertices due to the choice of v_1 and v_2.

Consider a connected component C of $G[I]$. We distinguish two cases for the alignment of C in the interval graph $G + F$: either there exist two vertices $v_1, v_2 \in C$ with $N_{G+F}(v_1) \cap K \neq N_{G+F}(v_2) \cap K$, or all of the vertices of C have the same neighborhood in K in the graph $G + F$. In the latter case, we argue that the component C may choose its place in the model σ in a greedy manner, and there are only $k^{O(\sqrt{k})}$ ways to arrange such components. In the first case, observe that such a component C "occupies" an endpoint event of a vertex of K and, if two components C_1 and C_2 occupy the same endpoint, they need to be connected by an edge of F. Since $|K| \leq 2\sqrt{k}$, we have at most $4\sqrt{k}$ endpoint events of vertices of K. If an endpoint event $\varepsilon \in \mathcal{E}(K)$ is occupied by a_ε components, then this means that we need to add at least $\binom{a_\varepsilon}{2}$ fill-in edges between these components. Then we have that $|\mathcal{E}(K)| \leq 4\sqrt{k}$ and $\sum_{\varepsilon \in \mathcal{E}(K)} \binom{a_\varepsilon}{2} \leq k$, and a simple application of the Cauchy-Schwarz inequality shows that $\sum_{\varepsilon \in \mathcal{E}(K)} a_\varepsilon = O(k^{3/4})$, that is, there are only $O(k^{3/4})$ components that fall into the first case. Moreover, the exhaustive application of the Module Reduction Rule ensures us that there are only $k^{O(1)}$ components of $G[I]$ in total.

Hence, we have $k^{O(k^{3/4})}$ guesses which components fall into the first case, $2^{O(k^{3/4})}$ guesses about their alignment to I^1 or I^2, and then the remaining components can be processed greedily. In Section 7, we develop a more careful reasoning that bounds the number of components falling into the first case by $O(\sqrt{k})$, instead of $O(k^{3/4})$ as presented in the argument above.

4 MODULES AND NEIGHBORHOOD CLASSES

Sections 4–8 contain a full proof of Theorem 1.1. We start with a study of possible neighborhood classes in an (almost) interval graph G, and provide the aforementioned module-based reduction rule in full detail.

4.1 Modules and Module-Based Reduction Rule

Recall that $M \subseteq V(G)$ is a *module* in a graph G if $N(v_1) \setminus M = N(v_2) \setminus M$ for any $v_1, v_2 \in M$. (Equivalently, for any $v \notin M$ we have either $M \subseteq N(v)$ or $M \cap N(v) = \emptyset$.) A module M is *connected* if $G[M]$ is connected. Cao proved the following.

Lemma 4.1 (Theorem 4.2 of [7]). *If M is a connected module in G, and F is a minimum completion of G, then M is a module in $G + F$ as well.*

Motivated by Lemma 4.1, we formulate the following reduction rule.

Reduction Rule 4.1 (Module Reduction Rule). Let (G, k) be an instance of Interval Completion. Assume there exist $X \subseteq V(G)$ and connected components $M_1, M_2, \ldots, M_{2k+3}$ of $G \setminus X$ that are modules in G and, moreover, $N(M_i) = N(M_1)$ for each $1 \leq i \leq 2k + 3$. Then proceed as follows. If for more than k indices i the subgraph $G[M_i]$ is not an interval graph, return that (G, k) is a NO-instance. Otherwise, pick arbitrary j such that $G[M_j]$ is an interval graph and remove M_j from G.

Clearly, if $G[M_i]$ is not an interval graph, any completion of G needs to contain an edge with both endpoints in M_i. Hence, the size of a minimum completion of G is lower bounded by the number of M_i's such that $G[M_i]$ is not an interval graph. Consequently, if the Module Reduction Rule concludes that (G, k) is a NO-instance, then the conclusion is correct.

Moreover, observe that any solution to Interval Completion in G naturally projects to a solution in $G \setminus M_j$ of at most the same size: if $G + F$ is an interval graph, so is $(G + F) \setminus M_j$. The

following lemma shows that the deletion of M_j in the Module Reduction Rule actually does not change our task at all.

Lemma 4.2. *Assume that the Module Reduction Rule is applicable to graph G, and its application deletes a module M_j. Then any solution to $(G \setminus M_j, k)$ is a solution to (G, k) as well.*

Proof. Without loss of generality, assume that $j = 2k + 3$. Let $G' = G \setminus M_j$, let F be a solution to (G', k), and let σ be an interval model of $G' + F$. As $|F| \leq k$, there are at least two modules M_i $(1 \leq i \leq 2k + 2)$ untouched by F; w.l.o.g. assume M_1 and M_2 are untouched by the solution. In the following, we show that M_1 and M_2 "reserve" a space in the model σ where we can insert M_j without any further cost.

As M_1 and M_2 are two connected components of $G \setminus X$ and both are untouched by F, all events of $\mathcal{E}(M_1)$ lie before all events of $\mathcal{E}(M_2)$, or all events of $\mathcal{E}(M_1)$ lie after all events of $\mathcal{E}(M_2)$ in the model σ; w.l.o.g., assume the first case. Denote $p_1 = \omega_\sigma(M_1)$ and $p_2 = \alpha_\sigma(M_2)$; note that $p_1 < p_2$. Let $Y = N(M_1) = N(M_2) \subseteq X$. As both M_1 and M_2 are untouched by F, we infer that $\Omega_\sigma(p_1) = \Omega_\sigma(p_2 - 1) = Y$, and Y is a clique in $G' + F$.

Let $\hat{\sigma}$ be an interval model of $G[M_j]$. Consider a model σ' created from σ by inserting all events of $\mathcal{E}(M_j)$ after position p_1 in σ, in the order according to model $\hat{\sigma}$. As $\Omega_\sigma(p_1) = N_G(M_j) = Y$, this is an interval model of $G + F$, and the lemma is proven. □

We now describe how to apply the Module Reduction Rule efficiently. To this end, we recall the module decomposition theorem, introduced by Gallai [16].

A module decomposition of a graph G is a rooted tree T, where each node t is labeled by a module $M^t \subseteq V(G)$, and is one of four types:

leaf: t is a leaf of T, and M^t is a singleton;
union: $G[M^t]$ is disconnected, and the children of t are labeled with different connected components of $G[M^t]$;
join: the complement of $G[M^t]$ is disconnected, and the children of t are labeled with different connected components of the complement of $G[M^t]$;
prime: neither of the above holds, and the children of t are labeled with different modules of G that are proper subsets of M^t, and are inclusion-wise maximal with this property.

Moreover, we require that the root of T is labeled with the module $V(G)$. We need the following properties of the module decomposition.

Theorem 4.3 (see [28]). *For a graph G, the following holds.*

(1) A module decomposition $(T, (M^t)_{t \in V(T)})$ of G exists, is unique, and computable in linear time.
(2) At any prime node t of T, the labels of the children form a partition of M^t. In particular, for each vertex v of G there exists exactly one leaf node with label $\{v\}$.
*(3) Each module M of G is either a label of some node of T, or there exists a **union** or **join** node t such that M is a union of labels of some children of G.*

We now show that the Module Reduction Rule can be applied efficiently using the module decomposition of a graph.

Lemma 4.4. *There is a polynomial-time algorithm that, given an instance (G, k) finds sets $X, M_1, \ldots, M_{2k+3} \subseteq V(G)$ on which the Module Reduction Rule is applicable, or correctly concludes that no such sets exist.*

Proof. We claim that, if the Module Reduction Rule is applicable to sets $X, M_1, \ldots, M_{2k+3}$, then there exists a **union** node t such that each set M_i is a label of some child of t.

From the last property of Theorem 4.3, we infer that, for any two modules M, M' of G, we have $M \subseteq M'$, $M' \subseteq M$, or $M \cap M' = \emptyset$ unless there exists a **union** or **join** node t in the module decomposition of G such that both M and M' are unions of labels of some children of t.

Notice now that a union of arbitrary number of sets M_i is a module in G as well. By applying the conclusion of the last paragraph to the modules $\bigcup_{i=1}^{2k+2} M_i$ and $\bigcup_{i=2}^{2k+3} M_i$, and using the fact that all M_i's are connected and pairwise non-adjacent, we infer that M_i's must be, in fact, children of the same **union** node t.

Therefore, to look for an application of the Module Reduction Rule it suffices to inspect all **union** nodes of the module decomposition of G, and for each such node t, classify the labels of the children of t according to their neighborhood. The Module Reduction Rule is applicable if and only if for some **union** node t at least $2k + 3$ children of t have labels with equal neighborhood. □

By Lemma 4.2, an application of the Module Reduction Rule does not change the answer to the input instance (G, k). Lemma 4.4 shows that the rule can be applied in polynomial time. Thus, we may apply the Module Reduction Rule exhaustively and henceforth we assume, sometimes implicitly, that it is no longer applicable.

4.2 Neighborhood Classes

We now provide some auxiliary structural lemmas about neighborhood classes in the input graph G.

For a graph G and a set $A \subseteq V(G)$, we say that two vertices $v_1, v_2 \notin A$ have *the same neighborhood with respect to* A if $N_G(v_1) \cap A = N_G(v_2) \cap A$. Clearly, this is an equivalence relation on $V(G) \setminus A$; each equivalence class of this relation is called *a neighborhood class w.r.t. A*.

The motivation for the results in this section is the following. In many places the algorithm makes some branching, choosing some vertex or a connected subgraph. In a straightforward analysis, each such branching will have around n options. With a branching of depth $\sqrt{k}$, and without a polynomial kernel for Interval Completion, this would lead to undesirable $n^{\sqrt{k}}$ factor in the running time. The structural results developed here limit the number of options in such branchings to polynomial in k; in some sense they are "local" kernelization results.

Lemma 4.5. *Assume G is a graph with completion set F, and let $A \subseteq V(G)$. Then in G there are at most $(2|A| + 1)^2 + |F|$ neighborhood classes w.r.t. A. In particular, if (G, k) is a YES-instance of Interval Completion, then there are at most $(2|A| + 1)^2 + k$ neighborhood classes w.r.t. A.*

Proof. Let $X \subseteq V(G) \setminus A$ be the set of vertices such that there exists some fill-in edge $xa \in F$ with $x \in X$ and $a \in A$. Clearly, $|X| \leq |F|$. To prove the lemma it suffices to show that there are at most $(2|A| + 1)^2$ neighborhood classes w.r.t. A in the graph $G \setminus X$.

Let σ be an interval model of the graph $G + F$. Pick any $v \in V(G) \setminus (A \cup X)$. As $v \notin X$, the edges between v and A in G are defined by the interval model σ, that is, $va \notin E(G)$ for $a \in A$ iff $\sigma(\omega_a) < \sigma(\alpha_v)$ or $\sigma(\alpha_a) > \sigma(\omega_v)$. Consider the model σ restricted to $\mathcal{E}(A)$, and note that there are $|\mathcal{E}(A)| + 1 = 2|A| + 1$ ways to insert the event α_v into this model, and at most this number of ways to insert ω_v. Consequently, there at most $(2|A| + 1)^2$ possible neighborhood classes w.r.t. A for vertices $v \in V(G) \setminus (A \cup X)$ and the lemma follows. □

Lemma 4.6. *Assume (G, k) is a YES-instance of Interval Completion, and the Module Reduction Rule is not applicable to (G, k). Let r be a positive integer and let $A \subseteq V(G)$. Then the number of connected components C of $G \setminus A$ for which there exists $v_C \in C$ with $|A \setminus N_G(v_C)| \leq r$ is at most $12kr + 4k + 18r + 4$.*

Proof. Let F be a solution to (G, k), and let σ be a model of $G + F$. Let $\mathcal{C}$ be the set of all connected components C of $G \setminus A$ that are untouched by F and for which there exists $v_C \in C$ with

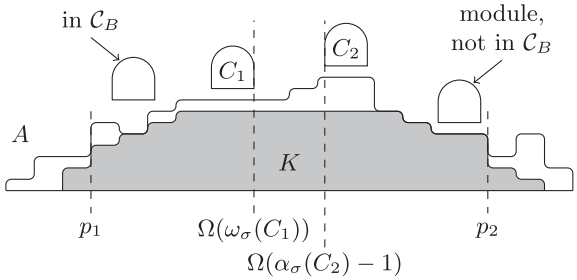

Fig. 16. Notation used in the proof of Lemma 4.6.

$|A \setminus N_G(v_C)| \leq r$. We aim to show that $|C| \leq (6r+1)(2k+2)+6r+2$, which will settle the claim since at most $2k$ components of $G \setminus A$ are touched by F.

If $|C| \leq 1$, then there is nothing to show, so assume otherwise. Let $C_1, C_2 \in C$. As both C_1 and C_2 are untouched, and there are no edges between the vertices of C_1 and the vertices of C_2, in the model σ all events of $\mathcal{E}(C_1)$ lie before or after all events of $\mathcal{E}(C_2)$; without loss of generality, assume that $\omega_\sigma(C_1) < \alpha_\sigma(C_2)$. Denote $K = A \cap N_G(v_{C_1}) \cap N_G(v_{C_2})$. Note that $|K| \geq |A| - 2r$ and $K \subseteq \Omega(\omega_\sigma(C_1))$, $K \subseteq \Omega(\alpha_\sigma(C_2)-1)$. Consequently, K is a clique in $G+F$. We refer to Figure 16 for an illustration of the notation used in this proof.

Denote $B = A \setminus K$, we have $|B| \leq 2r$. Let $\mathcal{E} \subseteq \mathcal{E}(K)$ be the set of the last $r+1$ starting events of $\mathcal{E}(K)$ and the first $r+1$ ending events of $\mathcal{E}(K)$ in the model σ (or $\mathcal{E} = \mathcal{E}(K)$ in case $|K| \leq r+1$). Recall that K is a clique in $G+F$ and $K \subseteq \Omega(\omega_\sigma(C_1))$, so all starting events of $\mathcal{E}(K)$ appear before position $\omega_\sigma(C_1)$, and all ending events of $\mathcal{E}(K)$ appear after this position.

Let $\mathcal{C}_B$ be the set of these connected components $C \in C$ for which there exists $\varepsilon \in \mathcal{E} \cup \mathcal{E}(B)$ with

$$\alpha_\sigma(C) < \sigma(\varepsilon) < \omega_\sigma(C). \tag{4.1}$$

As the components of C are untouched by F and pairwise non-adjacent in G, no two components of C can satisfy Equation (4.1) with the same event ε. Consequently,

$$|\mathcal{C}_B| \leq |\mathcal{E} \cup \mathcal{E}(B)| \leq 6r+2.$$

Denote by p_1 and p_2 the positions of the first and last events of $\mathcal{E}$, respectively. By the definition of $\mathcal{E}$, all events of $\mathcal{E}(A)$ that lie between p_1 and p_2 belong to $\mathcal{E} \cup \mathcal{E}(B)$.

Let $C \in C \setminus \mathcal{C}_B$. We have $\omega_C > p_1$, as otherwise C would be nonadjacent to the $r+1$ vertices of K that have their starting points on or after p_1, which contradicts $|A \setminus N_G(v_C)| \leq r$ (or C is non-adjacent to the whole A in case of $|K| \leq r+1$, which contradicts the connectivity of G). Similarly, we have $\alpha_C < p_2$. Since $C \notin \mathcal{C}_B$, we have actually $\alpha_C > p_1$ and $\omega_C < p_2$, that is, in the model σ all events of $\mathcal{E}(C)$ lie between the first and the last event of $\mathcal{E}$. Consequently, by the definition of $\mathcal{C}_B$, C is a module in $G+F$; as C is untouched by F, C is a module in G as well. Moreover, if for two components $C, C' \in C \setminus \mathcal{C}_B$ the events of $\mathcal{E}(C)$ and $\mathcal{E}(C')$ lie between the same two events of $\mathcal{E} \cup \mathcal{E}(B)$, then $N_G(C) = N_G(C')$. Therefore, if more than $2k+2$ such components lie between two consecutive events of $\mathcal{E} \cup \mathcal{E}(B)$, the Module Reduction Rule would be applicable. Consequently, $|C \setminus \mathcal{C}_B| \leq (6r+1)(2k+2)$, and the lemma is proven. □

5 LISTING POTENTIAL MAXIMAL CLIQUES AND SECTIONS

The main result of this section is Theorem 3.1, which we recall here for completeness.

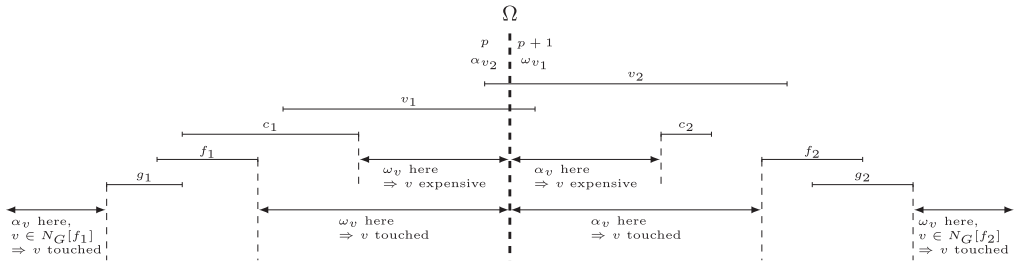

Fig. 17. The anatomy of a maximal clique Ω, with eight vertices guessed by the algorithm.

Theorem 3.1. *Given an Interval Completion instance (G, k) where the Module Reduction Rule is not applicable, one can in $k^{O(\sqrt{k})}n^{O(1)}$ time enumerate a family $\mathcal{S}$ of $k^{O(\sqrt{k})}n^{17}$ subsets of $V(G)$ such that for any minimal solution F to (G, k), in the canonical model σ of $G + F$ all sections of σ belong to $\mathcal{S}$.*

As already shown in Section 3, Theorem 3.1 follows from the following enumeration algorithm for potential maximal cliques in the Interval Completion problem.

Theorem 3.3. *Given an Interval Completion instance (G, k) where the Module Reduction Rule is not applicable, one can in $k^{O(\sqrt{k})}n^{O(1)}$ time enumerate a family $\mathcal{K}$ of $k^{O(\sqrt{k})}n^8$ subsets of $V(G)$ such that for any minimal solution F to (G, k), all maximal cliques of $G + F$ belong to $\mathcal{K}$.*

The rest of this section is devoted to the proof of Theorem 3.3.

5.1 Eight Important Vertices and the Structure of the Clique

Let us fix an Interval Completion instance (G, k), its minimal solution F, a model σ of $G + F$, and a maximal clique $\Omega = \Omega_\sigma(p)$. Recall that $\sigma(\alpha_{v_2}) = p$ and $\sigma(\omega_{v_1}) = p + 1$ for some vertices v_1 and v_2 (possibly $v_1 = v_2$). Without loss of generality, assume that Ω is different from the two "obvious" maximal cliques $\{\mathrm{r}, \mathrm{r}_L\}$ and $\{\mathrm{r}, \mathrm{r}_R\}$ and, consequently, $3 < p < 2n - 3$ and $v_1, v_2 \notin \{\mathrm{r}, \mathrm{r}_L, \mathrm{r}_R\}$.

Define the following vertices (see also Figure 17):

(1) c_1 is the cheap vertex with rightmost position of ω_{c_1}, among the cheap vertices c satisfying $\sigma(\omega_c) \leq \sigma(\omega_{v_1}) = p + 1$;
(2) c_2 is the cheap vertex with leftmost position of α_{c_2}, among the cheap vertices c satisfying $\sigma(\alpha_c) \geq \sigma(\alpha_{v_2}) = p$;
(3) f_1 is the untouched vertex with rightmost position of ω_{f_1}, among the untouched vertices f satisfying $\sigma(\omega_f) \leq \sigma(\omega_{v_1}) = p + 1$;
(4) f_2 is the untouched vertex with leftmost position of α_{f_2}, among the untouched vertices f satisfying $\sigma(\alpha_f) \geq \sigma(\alpha_{v_2}) = p$;
(5) g_1 is the untouched vertex with leftmost position of α_{g_1}, among all untouched vertices of $N_G[f_1] \setminus \{\Omega \setminus \{v_1\}\}$;
(6) g_2 is the untouched vertex with rightmost position of ω_{g_2}, among all untouched vertices of $N_G[f_2] \setminus \{\Omega \setminus \{v_2\}\}$.

Let us remark that some of these vertices can in fact be equal. We also remark that all quantifications in the aforementioned definitions are done on non-empty sets: r_L is a good candidate for both c_1 and f_1, r_R is a good candidate for both c_2 and f_2, f_1 is a good candidate for g_1, and f_2 is a good candidate for g_2. Hence, all these vertices are well-defined.

We observe the following relations between the positions of endpoints of the previously defined vertices.

Lemma 5.1. *The following inequalities hold:*

$$\sigma(\omega_{g_1}) \le \sigma(\omega_{f_1}) \le \sigma(\omega_{c_1}) \le \sigma(\omega_{v_1}) = p+1,$$
$$\sigma(\alpha_{g_2}) \ge \sigma(\alpha_{f_2}) \ge \sigma(\alpha_{c_2}) \ge \sigma(\alpha_{v_2}) = p.$$

Proof. Recall that, according to our definitions, an untouched vertex is cheap as well. The first inequality in each line follows from the definition of f_1 and f_2, as otherwise g_1 or g_2 would be a better candidate for f_1 or f_2, respectively. The remaining inequalities follow directly from the definitions of the corresponding vertices. □

We also need the following observation.

Lemma 5.2. $v_1 \in N_G[v_2]$ *and* $v_2 \in N_G[v_1]$.

Proof. If $v_1 = v_2$, then the claim is obvious, so assume otherwise. For the sake of contradiction, suppose $v_1v_2 \notin E(G)$, so $v_1v_2 \in F$ since $v_1v_2 \in E(G+F)$. Note that by swapping the events ω_{v_1} and α_{v_2} in the model σ we obtain a model for $G + (F \setminus \{v_1v_2\})$, contradicting the minimality of F. □

We say that a vertex v *lies to the left* of the clique Ω if $\sigma(\omega_v) \le p+1$, and *lies to the right* if $\sigma(\alpha_v) \ge p$. Clearly, v_1, c_1, f_1, g_1 lie to the left of Ω and v_2, c_2, f_2, g_2 lie to the right of Ω. Note that, perhaps a bit counterintuitively, if $v = v_1 = v_2$, then v lies both to the left and to the right of Ω.

We note the following straightforward observation.

Lemma 5.3. *For every vertex $w \in V(G)$ the following holds. If some vertex of $N_{G+F}[w]$ lies to the left of Ω, then $\sigma(\alpha_w) \le p$. If some vertex of $N_{G+F}[w]$ lies to the right of Ω, then $\sigma(\omega_w) \ge p+1$. In particular, if both these events happen, w belongs to Ω.*

Define now the following sets.

$$F_i^\circ = \{v \in V(G) : vc_i \in F\} \text{ for } i = 1, 2;$$
$$X_1^\circ = \{v \in V(G) : \sigma(\omega_{c_1}) < \sigma(\omega_v) \le p+1\};$$
$$X_2^\circ = \{v \in V(G) : p \le \sigma(\alpha_v) < \sigma(\alpha_{c_2})\}.$$

As c_1 and c_2 are cheap, $|F_1^\circ|, |F_2^\circ| \le \sqrt{k}$. By the definition of c_1 and c_2, all vertices of $X_1^\circ \cup X_2^\circ$ are expensive. Note that $|X_1^\circ \cap X_2^\circ| \le 1$ and $X_1^\circ \cap X_2^\circ$ is non-empty only if it consists of $v_1 = v_2$. Therefore, $|X_1^\circ| + |X_2^\circ| \le 2\sqrt{k} + 1$.

The following lemma characterizes Ω in terms of previously defined vertices and sets, and is a starting point of our algorithm.

Lemma 5.4.

$$\Omega = (N_G[\{v_1, c_1, f_1\} \cup X_1^\circ] \cup F_1^\circ) \cap (N_G[\{v_2, c_2, f_2\} \cup X_2^\circ] \cup F_2^\circ).$$

Proof. The inclusion "$\supseteq$" follows directly from Lemma 5.3: vertices of $N_G[\{v_1, c_1, f_1\} \cup X_1^\circ] \cup F_1^\circ$ either are or have at least one neighbor on the left of Ω in $G+F$, while vertices $N_G[\{v_2, c_2, f_2\} \cup X_2^\circ] \cup F_2^\circ$ either are or have at least one neighbor on the right of Ω in $G+F$. Hence, we now focus on the other inclusion.

Without loss of generality, assume there exists a vertex $v \in \Omega$ that does not belong to F_2° or to $N_G[\{v_2, c_2, f_2\} \cup X_2^\circ]$. In particular, $v \notin \{v_1, v_2, c_2\}$ by Lemma 5.2, and hence $\alpha_v < p$. As $v \notin F_2^\circ$ and $vc_2 \notin E(G)$, we have $\sigma(\omega_v) < \sigma(\alpha_{c_2})$. Moreover, by the definition of X_2°, v is not adjacent in G to any vertex whose starting event lies between positions p and $\sigma(\alpha_{c_2}) - 1$. Hence, v is not adjacent in G to any vertex whose starting event lies on or after position p.

Consider an ordering σ' that is created from the model σ by moving the event ω_v to the position just before the event α_{v_2} (i.e., we move ω_v to the position p and shift all events on positions p

and later by one to the right). By our previous arguments, σ' is a valid interval model of some completion F' of G. As $v \in \Omega$, the event ω_v has been moved to the left during this operation, and $F' \subseteq F$. Moreover, $vv_2 \in F \setminus F'$, which contradicts the minimality of F. □

We note that, if a polynomial kernel for Interval Completion had been known, Lemma 5.4 would have finished the proof of Theorem 3.3, as it provides us with a way to enumerate $n^{O(\sqrt{k})}$ candidates for maximal cliques in $G + F$, by guessing the vertices v_i, c_i, f_i and sets F_i°, X_i° for $i = 1, 2$.[4] However, the question of existence of such a kernel is widely open. Hence, we need to employ a careful and involved analysis of the structure of the clique Ω and the sets defined above to show the following: we may replace brute-force guessing of sets F_i°, X_i° with a branching procedure that selects each vertex of F_i°, X_i° among $\mathrm{poly}(k)$ potential candidates, instead of n.

5.2 Structure of the Recursion

We now proceed to the description of the algorithm of Theorem 3.3. The algorithm first iterates through all possible choices of the vertices v_i, c_i, f_i, g_i for $i = 1, 2$; for each choice, we seek for maximal cliques where the chosen vertices correspond to their definitions in the previous section. This step yields the promised n^8 factor in the bound on the size of the family $\mathcal{K}$.

Hence, for fixed choice of vertices v_i, c_i, f_i, g_i, we aim to output $k^{O(\sqrt{k})}$ sets in the family $\mathcal{K}$. The algorithm now becomes a branching algorithm: at each recursive call, in polynomial time we will insert at most one set into the family $\mathcal{K}$, invoke at most $\mathrm{poly}(k)$ recursive calls, and the depth of the recursion will be bounded by $O(\sqrt{k})$. Intuitively, we aim to guess the sets F_i° and X_i°, and at each step we want to identify a set of only $\mathrm{poly}(k)$ candidate vertices, such that one of the candidates certainly belongs to one of the sets F_i°, X_i°. Thus, we describe the algorithm in the language of "guessing" the maximal clique Ω.

More formally, during the course of the recursive branching algorithm we keep five sets $X_1, X_2, F_1, F_2, K \subseteq V(G)$, and we are looking for maximal cliques Ω satisfying the following:

(1) $\{v_1, c_1, f_1\} \subseteq X_1 \subseteq X_1^\circ \cup \{v_1, c_1, f_1\}$ and $\{v_2, c_2, f_2\} \subseteq X_2 \subseteq X_2^\circ \cup \{v_2, c_2, f_2\}$.
(2) $F_1 \subseteq F_1^\circ$ and $F_2 \subseteq F_2^\circ$.
(3) $(N_G[X_1] \cup F_1) \cap (N_G[X_2] \cup F_2) \subseteq K \subseteq \Omega$.

The set X_i is our "current guess" on the set $X_i^\circ \cup \{v_i, c_i, f_i\}$ and the set F_i is our "current guess" on the set F_i°. By Lemma 5.4, already properties 1 and 2 imply $(N_G[X_1] \cup F_1) \cap (N_G[X_2] \cup F_2) \subseteq \Omega$; the set K is our "current guess" for the clique Ω.

However, in some cases we will not be able to guess a vertex of X_1 or X_2, but instead we will be guessing its *neighborhood class* with respect to Ω. The results of Section 4.2 help us to limit the number of choices in such a step. For this reason, we allow the set K to be a proper superset of $(N_G[X_1] \cup F_1) \cap (N_G[X_2] \cup F_2)$, that is, to contain more than the vertices definitely included in Ω by Lemma 5.4.

We initially define $X_1 = \{v_1, c_1, f_1\}$, $X_2 = \{v_2, c_2, f_2\}$, $F_1 = F_2 = \emptyset$, and $K = N_G[X_1] \cap N_G[X_2]$. It is straightforward to verify that these sets satisfy all aforementioned properties. We note the following.

Lemma 5.5.

$$|\Omega \setminus (N_G[v_1] \cap N_G[v_2])| \leq k.$$

Proof. Note that for any $v \in \Omega \setminus (N_G[v_1] \cap N_G[v_2])$, either vv_1 or vv_2 belongs to F. □

[4] Actually, one may observe that the vertices f_1 and f_2 are not needed for the argumentation of Lemma 5.4. We include them for convenience, as they will be needed in further arguments.

Let us now focus on one recursive call, where the sets X_1, X_2, F_1, F_2, K are given. We consider connected components of $G \setminus (X_1 \cup X_2 \cup K)$ and classify them into four classes, depending on whether they contain a vertex of $N_G(X_1) \cup F_1$ and whether they contain a vertex of $N_G(X_2) \cup F_2$. That is, we partition the set $\mathtt{cc}(G \setminus (X_1 \cup X_2 \cup K))$ into four classes C_{ab} for $a, b \in \{0, 1\}$: $C \in C_{10} \cup C_{11}$ iff $C \cap (N_G(X_1) \cup F_1) \neq \emptyset$ and $C \in C_{01} \cup C_{11}$ iff $C \cap (N_G(X_2) \cup F_2) \neq \emptyset$. Formally,

$$C \in C_{00} \Leftrightarrow C \cap (N_G(X_1) \cup F_1) = \emptyset \wedge C \cap (N_G(X_2) \cup F_2) = \emptyset,$$
$$C \in C_{01} \Leftrightarrow C \cap (N_G(X_1) \cup F_1) = \emptyset \wedge C \cap (N_G(X_2) \cup F_2) \neq \emptyset,$$
$$C \in C_{10} \Leftrightarrow C \cap (N_G(X_1) \cup F_1) \neq \emptyset \wedge C \cap (N_G(X_2) \cup F_2) = \emptyset,$$
$$C \in C_{11} \Leftrightarrow C \cap (N_G(X_1) \cup F_1) \neq \emptyset \wedge C \cap (N_G(X_2) \cup F_2) \neq \emptyset.$$

5.3 Case One: Components Knowing Both Sides of the Clique

Assume there exists $C \in C_{11}$. Note that $v_1, v_2 \notin C$, since $v_1 \in X_1$ and $v_2 \in X_2$. Hence, by Lemma 5.3, C contains a vertex whose interval starts before position p in the model σ, and a vertex whose interval ends after position $p + 1$. As $G[C]$ is connected, $C \cap (\Omega \setminus K) \neq \emptyset$.

Let P be a shortest path between $N_G(X_1) \cup F_1$ and $N_G(X_2) \cup F_2$ in the subgraph $G[C]$. Note that P contains at least two vertices, as otherwise the single vertex of P should be included in K. We note the following.

Lemma 5.6. *Either $V(P) \subseteq \Omega$ or $V(P)$ contains a vertex of $(F_1^\circ \setminus F_1) \cup (X_1^\circ \setminus X_1) \cup (F_2^\circ \setminus F_2) \cup (X_2^\circ \setminus X_2)$.*

Proof. Assume there exists $v \in V(P) \setminus \Omega$. Without loss of generality, assume that v is to the right of Ω, that is, $\sigma(\alpha_v) > p + 1$ (as $v \notin \{v_1, v_2\}$). Moreover, assume that v is the first vertex on the path P (when traversed from $N_G(X_1) \cup F_1$ to $N_G(X_2) \cup F_2$) that lies to the right of Ω.

As the first vertex of P belongs to $N_G(X_1) \cup F_1$, v is not the first vertex of P. Let w be the predecessor of v on the path P. Since w does not lie to the right of Ω (by the choice of v), and $vw \in E(G)$, we infer that $w \in \Omega$. As P is a shortest path between $N_G(X_1) \cup F_1$ and $N_G(X_2) \cup F_2$, we have $w \notin F_2$ and $wc_2 \notin E(G)$.

If $\sigma(\omega_w) \geq \sigma(\alpha_{c_2})$, then $wc_2 \in F$, but $w \notin F_2$. Hence, $w \in F_2^\circ \setminus F_2$. Otherwise, if $\sigma(\omega_w) < \sigma(\alpha_{c_2})$, then we have $p + 1 < \sigma(\alpha_v) < \sigma(\omega_w) < \sigma(\alpha_{c_2})$. By the choice of c_2, we infer that $v \in X_2^\circ$. Clearly, $v \notin X_2$, so $v \in X_2^\circ \setminus X_2$ and the lemma is proven. □

Lemma 5.6 enables us to do a good branching providing that P is short. Luckily, this is always the case.

Lemma 5.7. $|V(P)| \leq 3k$.

Proof. Denote $H = G + F$. Let R be a shortest path between the first and the last vertex of P in the graph $H[V(P)]$. We first claim that each vertex on R is touched by the solution F and, consequently, $|V(R)| \leq 2k$.

Clearly, each vertex $v \in V(R) \cap \Omega$ is touched by F, as vv_1 or vv_2 needs to belong to F. Consider then $v \in V(R) \setminus \Omega$ and, without loss of generality, assume that v lies to the left of Ω, that is, $\sigma(\omega_v) < p$. We now show that $\sigma(\omega_v) > \sigma(\omega_{f_1})$; this would prove the claim as then v is touched by the definition of f_1. Assume otherwise. Clearly, v is not the last vertex of P (and R), and the vertex w succeeding v on R needs to satisfy $\sigma(\alpha_w) \leq \sigma(\omega_{f_1})$. Consequently, there exists a vertex w' on R that lies later than v on R, and which neighbors f_1 in H. As f_1 is untouched, we have that $w'f_1 \in E(G)$, which means that $w' \in N_G(X_1)$. Since w' is not the first vertex of P, this contradicts the choice of P.

To finish the proof, we now show that $|V(P)| - |V(R)| \leq |F| \leq k$. Let $s = |V(P)|$ and $x_1, x_2, \ldots, x_s$ be the vertices of P in the order of their appearance. The essence of the proof lies in the fact

that whenever R uses some edge $x_a x_b \in F$, $a < b$, then F needs to contain a triangulation of the cycle $x_a - x_{a+1} - \cdots - x_b - x_a$, consisting of $(b - a - 2)$ edges. Thus, we need to "pay" with $(b - a - 1)$ edges of F (including $x_a x_b$) to shorten the length of P by, again, $(b - a - 1)$. The formal argumentation follows.

Define the sequence $a_1, a_2, \ldots, a_r$ as follows. Let $a_1 = 1$ and, given $1 \leq a_i < s$, define a_{i+1} to be such an index, such that $x_{a_{i+1}}$ is the vertex from the set $\{x_{a_i+1}, x_{a_i+2}, \ldots, x_s\}$ that appears earliest on the path R. Clearly, by the definition, $x_{a_{i+1}}$ appears on R later than x_{a_i} and $a_i < a_{i+1}$. This definition ends when $a_r = s$ for some index r.

Consider now an edge $e_{i+1} := x_{b_{i+1}} x_{a_{i+1}}$ on the path R, that is, $x_{b_{i+1}}$ is the predecessor of $x_{a_{i+1}}$ on R. Clearly, $b_{i+1} \leq a_i$, since otherwise b_{i+1} would be a better candidate for a_{i+1}. If $e_{i+1} \in E(G)$, then we have $b_{i+1} = a_i = a_{i+1} - 1$ since P is an induced path in G. Otherwise, $e_{i+1} \in F$. By the definition of a_{i+1}, all internal vertices x_b of $R[x_{a_i}, x_{a_{i+1}}]$ satisfy $b < a_i$,[5] as otherwise they would be better candidates for a_{i+1}. Hence, as P is an induced path in G and R is an induced path in $H = G + F$, F needs to contain a triangulation of the cycle consisting of the subpath $R[x_{a_i}, x_{a_{i+1}}]$ and the subpath $P[x_{a_i}, x_{a_{i+1}}]$. This triangulation consists of at least $(a_{i+1} - a_i - 2)$ edges. Moreover, since R is an induced path in $H = G + F$, all the edges of the triangulation need to have at least one endpoint in the set $\{x_{a_i+1}, x_{a_i+2}, \ldots, x_{a_{i+1}-1}\}$; note that the second endpoint always lies in the set $\{x_1, x_2, \ldots, x_{a_{i+1}}\}$. Together with the edge e_{i+1}, we infer that there are at least $(a_{i+1} - a_i - 1)$ edges $x_a x_b$ of F such that $a < b$ and $a_i < b \leq a_{i+1}$. Note that this statement also trivially holds in the first case, when $e_{i+1} \in E(G)$.

Observe that the specified set of edges of F are pairwise disjoint for different edges e_{i+1}. We infer that

$$|V(P)| - |V(R)| \leq s - r = \sum_{i=1}^{r-1} (a_{i+1} - a_i - 1) \leq |F| \leq k,$$

and the lemma is proven. □

Lemmas 5.6 and 5.7 enable us to perform the following branching strategy. In a loop, as long as C_{11} is not empty, we pick arbitrary $C \in C_{11}$, compute a shortest path P in $G[C]$ between $N_G(X_1) \cup F_1$ and $N_G(X_2) \cup F_2$, and proceed as follows. First, if the bound of Lemma 5.7 does not hold, that is, if $|V(P)| > 3k$, then we conclude that the current guesses are incorrect and we terminate the current branch. Second, we invoke at most $4|V(P)|$ recursive calls (branches), in each branch assigning one of the vertices $v \in V(P)$ to one of the sets F_1, F_2, X_1, X_2 that does not contain v already. Third, we put the entire $V(P)$ into K and go back to the beginning of the loop. By Lemma 5.5, we may terminate the current branch if the size of the set K increased by more than k since the root of the recursion. Consequently, by the bound of Lemma 5.7, the aforementioned loop produces $O(k^2)$ recursive calls, and leaves us with a situation where $C_{11} = \emptyset$.

5.4 Case Two: Components not Knowing Any Side of the Clique

We now focus on a component $C \in C_{00}$, that is, a connected component of $G \setminus (X_1 \cup X_2 \cup K)$ that does not contain any vertices of $N_G(X_1 \cup X_2) \cup F_1 \cup F_2$. In particular, note that for any such component it holds that $N_G(C) \subseteq K \setminus \{v_1, v_2\}$.

We now prove a few properties of such components C, assuming $C \cap \Omega \neq \emptyset$. Our goal is to prove that each such component contains a vertex of $F_1^\circ \cup X_1^\circ \cup F_2^\circ \cup X_2^\circ$, and, moreover, both the sizes and the number of candidates for such components are bounded polynomially in k.

Lemma 5.8. *If $C \in C_{00}$ and $C \cap \Omega \neq \emptyset$, then $\sigma(\omega_{f_1}) < \alpha_\sigma(C) < \omega_\sigma(C) < \sigma(\alpha_{f_2})$.*

[5] Recall that for a path P and two vertices x, y on P, by $P[x, y]$ we denote the subpath of P between x and y, inclusive.

Proof. Recall that f_1 and f_2 are untouched by the solution F, both belong to $X_1 \cup X_2$, and C does not contain any neighbor of $X_1 \cup X_2$. □

Lemma 5.9. *If $C \in \mathcal{C}_{00}$ and $C \cap \Omega \neq \emptyset$, then all vertices of C are touched by the solution, and, consequently, $|C| \leq 2k$.*

Proof. Let $v \in C$. If $v \in \Omega$, v is touched by F as $vv_1, vv_2 \in F$. If v lies to the left of Ω then, by Lemma 5.8, $\sigma(\omega_v) > \sigma(\omega_{f_1})$, and v is touched by the choice of f_1. The case of v lying to the right of Ω is symmetrical. □

Lemma 5.10. *If $C \in \mathcal{C}_{00}$ and $C \cap \Omega \neq \emptyset$, then there exists $v \in C$ such that $|K \setminus N_G(v)| \leq k$.*

Proof. Observe that any vertex of $C \cap \Omega$ needs to be adjacent to all vertices of K in $G + F$, and $|F| \leq k$. □

Lemma 5.11. *If $C \in \mathcal{C}_{00}$ and $C \cap \Omega \neq \emptyset$, then C contains a vertex of $(F_1^\circ \setminus F_1) \cup (X_1^\circ \setminus X_1) \cup (F_2^\circ \setminus F_2) \cup (X_2^\circ \setminus X_2)$.*

Proof. We first show that $C \not\subseteq \Omega$. Assume the contrary. Let $|C| = s$ and $x_1, x_2, \ldots, x_s$ be the vertices of C. Consider a model σ' created from σ by taking out all events of $\mathcal{E}(C)$ and inserting them, in the order $\alpha_{x_1}, \alpha_{x_2}, \ldots, \alpha_{x_s}, \omega_{x_s}, \omega_{x_{s-1}}, \ldots, \omega_{x_1}$ between positions $p-1$ and p (i.e., just before the event α_{v_2} at position p. As $N_G(C) \subseteq K \setminus \{v_1, v_2\}$, σ' is a valid interval model of some completion F' of G. As $C \subseteq \Omega \setminus \{v_1, v_2\}$ and, in particular, C is a clique in $G + F$, for any $x_i \in C$ we have $\sigma(\alpha_{x_i}) < p < \sigma(\omega_{x_i})$ and, consequently, $F' \subseteq F$. Moreover, $x_i v_2 \in F \setminus F'$ for any $x_i \in C$, contradicting the minimality of F.

Since C is connected in G, we may pick $v, w \in C$ such that $vw \in E(G)$, $v \in \Omega$ and $w \notin \Omega$; w.l.o.g. assume that w lies to the left of Ω. If $\sigma(\alpha_v) \leq \sigma(\omega_{c_1})$, then $vc_1 \in F$ and $v \in F_1^\circ \setminus F_1$. Otherwise, we have $\sigma(\omega_{c_1}) < \sigma(\omega_w) < p$ and $w \in X_1^\circ \setminus X_1$. This finishes the proof of the lemma. □

By Lemmas 5.9 and 5.10, all components $C \in \mathcal{C}_{00}$ that may have a non-empty intersection with Ω need to (a) be of size at most $2k$ and (b) have a vertex with at most k non-neighbors in K. By Lemma 4.6, applied to the set $A := K$ and parameter $r := k$, in a YES-instance we expect $\mathcal{O}(k^2)$ components satisfying the second requirement. (Formally, we conclude that (G, k) is a NO-instance and return $\mathcal{K} = \emptyset$ if the bound of Lemma 4.6 turns out to be violated.) Consequently, all components satisfying both requirements (a) and (b) have $\mathcal{O}(k^3)$ vertices in total. This, together with Lemma 5.11, motivates the following branching step. First, we invoke $\mathcal{O}(k^3)$ recursive calls, in each call picking a vertex from a component satisfying both (a) and (b) and inserting it into one of the sets F_1, X_1, F_2, X_2. Finally, we pass the instance to the next case, assuming that no component of $\mathcal{C}_{00}$ contains a vertex of Ω.

5.5 Case Three: Components Knowing One Side of the Clique

We are left with the components of $\mathcal{C}_{01} \cup \mathcal{C}_{10}$. By symmetry, we may focus on $\mathcal{C}_{10}$ only.

Consider $C \in \mathcal{C}_{10}$. The main obstacle we obtain in this section is that an analog of Lemma 5.9 does not hold (in particular, C may contain a lot of vertices in $N_G(f_1)$) and, consequently, C may be large. To apply arguments similar to the previous case, we need to further analyze the structure of such component C.

To this end, we define $\mathcal{D}_1 = \mathsf{cc}(G[\bigcup \mathcal{C}_{10} \setminus N_G(f_1)])$. Now, for each $D \in \mathcal{D}_1$ we have not only $D \cap (N_G(X_2) \cup F_2) = \emptyset$ but also $D \cap N_G(f_1) = \emptyset$, and we can state analogs of Lemmas 5.8 and 5.9.

Lemma 5.12. *For any $D \in \mathcal{D}_1$, either $\omega_\sigma(D) < \sigma(\alpha_{f_1})$ or $\sigma(\omega_{f_1}) < \alpha_\sigma(D) < \omega_\sigma(D) < \sigma(\alpha_{f_2})$. Moreover, if the second option happens, then all vertices of D are touched by F and $|D| \leq 2k$.*

Proof. As D is connected and does not contain any neighbor of the untouched vertices f_1 and f_2, we need only to exclude the possibility $\alpha_\sigma(D) > \sigma(\omega_{f_2})$. However, this clearly follows from the fact that there exists a connected component $C \in \mathcal{C}_{10}$ containing D: $N_{G+F}(C)$ contains a vertex of X_1 and does not contain f_2. This proves the first assertion of the lemma.

Assume now that $\sigma(\omega_{f_1}) < \alpha_\sigma(D) < \omega_\sigma(D) < \sigma(\alpha_{f_2})$. Pick any $v \in D$. If $v \in \Omega$, then v is touched by F as $vv_2 \in F$. Otherwise, $\sigma(\omega_{f_1}) < \sigma(\omega_v) < p$ or $\sigma(\alpha_{f_2}) > \sigma(\alpha_v) > p+1$. In both cases v is touched by the choice of f_1 or f_2. □

The following lemma shows formally why we are interested in components of $\mathcal{D}_1$.

Lemma 5.13. *A component $C \in \mathcal{C}_{10}$ contains an element of Ω if and only if there exists $D \in \mathcal{D}_1$, $D \subseteq C$, such that $D \cap \Omega \neq \emptyset$ or $\alpha_\sigma(D) > p+1$. In particular, such a component D satisfies the second option of Lemma 5.12.*

Proof. Assume first that such a component D exists for some $C \in \mathcal{C}_{10}$. If D contains a vertex of Ω, then clearly so does C, so assume $\alpha_\sigma(D) > p+1$. Then $N_{G+F}(D) \cap X_1 = \emptyset$ but $N_{G+F}(C) \cap X_1 \neq \emptyset$. Hence, as $G[C]$ is connected and D is a connected component of $G[C] \setminus N_G(f_1)$, we infer that there exists some $z \in N_G(D) \cap N_G(f_1)$. Such a z clearly belongs to Ω by Lemma 5.3.

In the other direction, assume that $C \cap \Omega \neq \emptyset$. Suppose first that there exists $x \in C$ with $\sigma(\alpha_x) > p+1$. Then $x \notin N_G(f_1)$ and $x \in D$ for some $D \in \mathcal{D}_1$. If $D \cap \Omega \neq \emptyset$, we are done. Otherwise, by the connectivity of D we have $\alpha_\sigma(D) > p+1$ and the claim is proven.

So we have $\sigma(\alpha_x) < p$ for any $x \in C$, as $v_1, v_2 \notin C$. Consider an interval model σ' created from σ by taking all events of $\mathcal{E}(C)$ that are placed at positions at least p, and putting them (in the same order) just before position p (i.e., between positions $p-1$ and p). As $N_G(C) \subseteq (X_1 \cup K) \setminus \{v_2\}$, this is a valid interval model of $G + F'$ for some completion F'. As $\sigma(\alpha_x) < p$ for any $x \in C$, we have $F' \subseteq F$. Moreover, $xv_2 \in F \setminus F'$ for any $x \in C \cap \Omega$. By the minimality of F, we have $C \cap \Omega = \emptyset$, which contradicts our assumption about C and concludes the proof. □

Hence, we now focus on components D and try to deduce which of them may possibly satisfy one of the conditions imposed in Lemma 5.13. We first make use of the untouched vertex g_1 to filter out some clearly "useless" components of $\mathcal{D}_1$.

Lemma 5.14. *If for $D \in \mathcal{D}_1$, we have $D \cap N_G(g_1) \neq \emptyset$, then $\omega_\sigma(D) < \sigma(\alpha_{f_1})$ (i.e., the first option of Lemma 5.12 happens).*

Proof. Follows directly from the inequality $\sigma(\omega_{g_1}) \leq \sigma(\omega_{f_1})$ (Lemma 5.1). □

We denote $\mathcal{D}_2 = \{D \in \mathcal{D}_1 : g_1 \notin N_G(D)\}$ and define $Z = \bigcup_{D \in \mathcal{D}_2} N_G(D) \setminus (K \cup X_1)$. Note that $N_G(D) \subseteq X_1 \cup K \cup N_G(f_1)$ by the definition of $\mathcal{C}_{10}$ and $\mathcal{D}_1$. Consequently, $Z \subseteq N_G(f_1) \cap \bigcup \mathcal{C}_{10}$. The following observation is the main reason to introduce the vertex g_1 and "filter out" components of $\mathcal{D}_1 \setminus \mathcal{D}_2$ in Lemma 5.14.

Lemma 5.15. *All vertices of Z are touched by F and, consequently, $|Z| \leq 2k$.*

Proof. Let $z \in Z$ and let $D \in \mathcal{D}_2$ such that $z \in N_G(D)$. If $z \in \Omega$, then $zv_2 \in F$ and z is touched, so assume otherwise. As $z \in N_G(f_1)$, we infer that $\sigma(\omega_z) < p$.

Consider two cases for component D given by Lemma 5.12. If $\omega_\sigma(D) < \sigma(\alpha_{f_1})$, then, as $D \in \mathcal{D}_2$ and $g_1 \in N_G[f_1]$, we have actually $\omega_\sigma(D) < \sigma(\alpha_{g_1})$. Hence, $\sigma(\alpha_z) < \sigma(\alpha_{g_1})$. As $z \in N_G(f_1)$ and $z \notin \Omega$, we infer that z is touched by the choice of g_1. In the second case, if $\sigma(\omega_{f_1}) < \alpha_\sigma(D)$, then $\sigma(\omega_z) > \sigma(\omega_{f_1})$. As $\sigma(\omega_z) < p$, we infer that z is touched by the choice of f_1. □

Formally, if the bound of Lemma 5.15 does not hold, we terminate the current branch. Otherwise, any $D \in \mathcal{D}_2$ satisfies $N_G(D) \subseteq K \cup X_1 \cup Z$, and $|Z| + |X_1| \leq 2k + O(\sqrt{k})$.

We now focus on the possibility of $D \cap \Omega \neq \emptyset$ for some $D \in \mathcal{D}_2$.

Lemma 5.16. *If $D \cap \Omega \neq \emptyset$ for some $D \in \mathcal{D}_2$, then $D \cap ((F_2^\circ \setminus F_2) \cup (X_2^\circ \setminus X_2)) \neq \emptyset$.*

Proof. We first show that if $D \cap \Omega \neq \emptyset$, then there exists $w \in D$ with $\sigma(\alpha_w) > p+1$. Assume the contrary, and consider a model σ' created from σ by taking all events of $\mathcal{E}(D)$ that are placed by σ on positions to the right of Ω (i.e., at positions with numbers at least p) and move them just before position p (i.e., the event α_{v_2}), in the same order as they appear in σ. As $N_G(D) \subseteq X_1 \cup N_G(f_1)$, σ' is an interval model of some completion F' of G. Since we supposed that no vertex of D starts in σ after position p, we have $F' \subseteq F$. Moreover, $vv_2 \in F \setminus F'$ for any $v \in D \cap \Omega$, a contradiction to the minimality of F.

By the connectivity of D, there exist $v, w \in D$ such that $vw \in E(G)$, $v \in \Omega$, and $\sigma(\alpha_w) > p+1$. Consider two cases. If $\sigma(\omega_v) \geq \sigma(\alpha_{c_2})$, then $vc_2 \in F$ and $v \in F_2^\circ \setminus F_2$. Otherwise, we have $\sigma(\alpha_w) < \sigma(\omega_v) < \sigma(\alpha_{c_2})$, and hence, by the choice of c_2, w is expensive. Consequently, $w \in X_2^\circ \setminus X_2$. □

We now note that if $D \cap \Omega \neq \emptyset$, then any $v \in D \cap \Omega$ needs to satisfy $|K \setminus N_G(v)| \leq k$. Let $\mathcal{D}_3 \subseteq \mathcal{D}_2$ be the family of these connected components D of $\mathcal{D}_2$ that (a) have size at most $2k$, and (b) contain a vertex v that has at most k non-neighbors in K. By Lemma 5.12, if $D \cap \Omega \neq \emptyset$, then $D \in \mathcal{D}_3$. By Lemma 4.6 applied to the set $A := K \cup X_1 \cup Z$ and $r = k + |Z| + |X_1| = \mathcal{O}(k)$, we infer that in a YES-instance we expect $|\mathcal{D}_3| = \mathcal{O}(k^2)$ (formally, we terminate the algorithm and return $\mathcal{K} = \emptyset$ if this is not the case). Consequently, $|\bigcup \mathcal{D}_3| = \mathcal{O}(k^3)$. Hence, Lemma 5.16 allows us to branch into $\mathcal{O}(k^3)$ recursive calls: in each call we put one of the vertices of $\bigcup \mathcal{D}_3$ into one of the sets F_2, X_2. We proceed further with the assumption that no vertex of $\bigcup \mathcal{D}_2$ belongs to Ω, and we focus on the possibility that $\alpha_\sigma(D) > p+1$ for some $D \in \mathcal{D}_2$.

Lemma 5.17. *If $\alpha_\sigma(D) > p+1$ for some $D \in \mathcal{D}_2$, then either $Z \cap (F_2^\circ \setminus F_2) \neq \emptyset$ or there exists $w \in D \cap (X_2^\circ \setminus X_2)$ such that $N_G(w) \cap Z = N_G(w) \cap (\Omega \setminus K) \neq \emptyset$.*

Proof. First note that, as $\alpha_\sigma(D) > p+1$, then $N_G(D) \subseteq K \cup Z$, and D does not contain any vertex of F_1°. Moreover, as $D \subseteq C$ for some $C \in \mathcal{C}_{10}$, we have that $N_G(D) \cap Z \neq \emptyset$.

Pick any $z \in N_G(D) \cap Z$. As $zf_1 \in E(G)$ and $\alpha_\sigma(D) > p+1$, we have $z \in \Omega \setminus K$. If $\sigma(\omega_z) \geq \sigma(\alpha_{c_2})$, then we have $z \in F_2^\circ \setminus F_2$ and we are done. Otherwise, any neighbor $w \in N_G(z) \cap D$ satisfies $\sigma(\alpha_w) < \sigma(\omega_z) < \sigma(\alpha_{c_2})$ and, by the choice of c_2, we infer that $w \in X_2^\circ \setminus X_2$. As $N_G(w) \subseteq D \cup K \cup Z$, such w satisfies the requirements of the lemma; the fact that $N_G(w) \cap Z = N_G(w) \cap (\Omega \setminus K)$ follows easily from the assumptions about D and the definition of Z. □

Lemma 5.17, together with the bound $|Z| \leq 2k$ of Lemma 5.15, allows us to perform the following branching. In the first $|Z|$ recursive calls we pick a vertex of Z and insert it into F_2. Then, we invoke Lemma 4.5 on the set $A := Z$, expecting $\mathcal{O}(k^2)$ neighborhood classes w.r.t. Z in the graph G (formally, if this is not the case, we conclude that (G, k) is a NO-instance and return an empty set $\mathcal{K}$). We branch into $\mathcal{O}(k^2)$ subcases, in each recursive call picking a neighborhood class R w.r.t. Z with non-empty neighborhood $N_G(R) \cap Z$ and inserting this neighborhood into K.

Finally, we are left with the case where the conclusion is that no component $D \in \mathcal{D}_2$ satisfies $\alpha_\sigma(D) > p+1$; recall that we have already concluded before that no component $D \in \mathcal{D}_2$ has a non-empty intersection with Ω. By Lemma 5.13, we infer that in fact there are no vertices of Ω at all in the components of $\mathcal{C}_{10}$.

Therefore, we pass the instance to the symmetric case of $\mathcal{C}_{01}$ and we perform all the symmetric branchings. In the remaining subcase, we can finally conclude that $K = \Omega$: We have $\mathcal{C}_{11} = \emptyset$, and we have already concluded that there are no vertices of Ω in the components of $\mathcal{C}_{00}$, of $\mathcal{C}_{10}$, nor of $\mathcal{C}_{01}$. Hence, we insert the set K into the constructed family $\mathcal{K}$.

It remains to argue that we output $k^{\mathcal{O}(\sqrt{k})}$ sets for each choice of the vertices v_i, c_i, f_i, g_i, $i = 1, 2$. Clearly, each step of the recursion invokes $\mathrm{poly}(k)$ recursive calls. To see that the depth of the

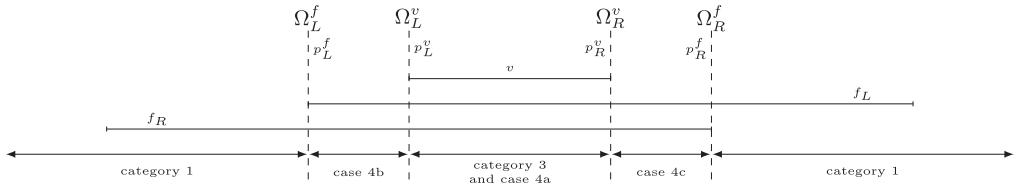

Fig. 18. Situation around the vertex v in the proof of Theorem 3.2, together with categories and cases of Lemma 6.3.

recursion can be bounded by $O(\sqrt{k})$, note that whenever we make a recursive call, we either insert a new vertex into one of the sets F_1, X_1, F_2, X_2, or we put into K all vertices of a non-empty set $N_G(w) \cap (\Omega \setminus K)$ for some $w \in (X_1^\circ \setminus X_1) \cup (X_2^\circ \setminus X_2)$—hence, this step can be done at most once for every $w \in X_1^\circ \cup X_2^\circ$ during the whole branching process. As $|F_1^\circ|, |F_2^\circ| \leq \sqrt{k}$ and $|X_1^\circ| + |X_2^\circ| \leq 2\sqrt{k} + 1$, we can prune the recursion tree at depth $6\sqrt{k} + 2$, obtaining the claimed bound on the size of $\mathcal{K}$. This concludes the proof of Theorem 3.3.

6 GUESSING FILL-IN EDGES WITH FIXED ENDPOINT

In this section, we prove Theorem 3.2, which we restate here for completeness.

Theorem 3.2. *Given an Interval Completion instance (G, k) where the Module Reduction Rule is not applicable, and a designated vertex $v \in V(G)$, one can in $k^{O(\sqrt{k})}n^{O(1)}$ time enumerate a family $\mathcal{F}$ of at most $k^{O(\sqrt{k})}n^{70}$ subsets of $V(G)$ such that for any minimal solution F to (G, k) for which v is cheap w.r.t. F, the set $\{w \in V(G) : vw \in F\}$ belongs to $\mathcal{F}$.*

We will mostly use Theorem 3.2 to guess the incident fill-in edges of a cheap vertex.

Corollary 6.1. *Given an Interval Completion instance (G, k), where the Module Reduction Rule is not applicable, and a designated vertex $v \in V(G)$, one can in $k^{O(\sqrt{k})}n^{O(1)}$ time enumerate a family $\mathcal{F}'$ of at most $k^{O(\sqrt{k})}n^{70}$ subsets of $V(G)$, such that for any minimal solution F to (G, k) for which v is cheap w.r.t. F, the set $\{w \in V(G) : vw \in F\}$ belongs to $\mathcal{F}'$.*

Proof. We first enumerate the family $\mathcal{F}$ of Theorem 3.2 and then define

$$\mathcal{F}' = \{A \subseteq V(G) : |A| \leq \sqrt{k} \wedge \exists_{B \in \mathcal{F}} A \subseteq B\}.$$

The correctness and the size bound follows directly from Theorem 3.2. □

We remark that, similarly to the arguments of the previous section, a polynomial kernel for Interval Completion would save us a lot of effort. In fact, Theorem 3.2 becomes obvious as we could then return $\mathcal{F} = \{V(G)\}$ (possibly worsening the polynomial bound on the size of a single element of $\mathcal{F}$). However, the question of existence of a polynomial kernel for Interval Completion remains widely open, and we need to employ a careful analysis to obtain the promised results.

6.1 Important Vertices and Sections

We fix a minimal completion F of the Interval Completion instance (G, k), and a model σ of $G + F$. We define the following (see also Figure 18).

(1) Denote $p_L^v = \sigma(\alpha_v)$ and $p_R^v = \sigma(\omega_v)$.
(2) Let f_L be the untouched vertex with the rightmost starting endpoint among untouched vertices f satisfying $\sigma(\alpha_f) \leq p_L^v < p_R^v \leq \sigma(\omega_f)$.

(3) Let f_R be the untouched vertex with the leftmost ending endpoint among untouched vertices f satisfying $\sigma(\alpha_f) \leq p_L^v < p_R^v \leq \sigma(\omega_f)$.
(4) Denote $p_L^f = \sigma(\alpha_{f_L})$ and $p_R^f = \sigma(\omega_{f_R})$.
(5) Denote $\Omega_L^f = \Omega_\sigma(p_L^f)$, $\Omega_L^v = \Omega_\sigma(p_L^v)$, $\Omega_R^v = \Omega_\sigma(p_R^v - 1)$ and $\Omega_R^f = \Omega_\sigma(p_R^f - 1)$.

Note that r is a good candidate for both f_L and f_R, thus these vertices exist. We remark also that it may happen that $v = f_L$, $v = f_R$, or $f_L = f_R$. However, we may say the following about the order of these vertices.

Lemma 6.2. $\sigma(\alpha_{f_R}) \leq p_L^f \leq p_L^v < p_R^v \leq p_R^f \leq \sigma(\omega_{f_L})$.

Proof. The first and the last inequalities follow from the fact that f_R is a good candidate for f_L and vice versa. The remaining inequalities are straightforward from the definition. □

We start by enumerating all possible choices of vertices f_L, f_R and sections Ω_L^f, Ω_L^v, Ω_R^v, Ω_R^f, using the family $\mathcal{S}$ of Theorem 3.1. By the bound of Theorem 3.1, there are at most $k^{O(\sqrt{k})}n^{70}$ subcases (henceforth called *branches*) to consider. In the rest of the proof, we aim to output a single set B of size $O(k^5)$ for a single choice of the aforementioned two vertices and four sections. That is, given f_L, f_R and Ω_L^f, Ω_L^v, Ω_R^v, Ω_R^f we show how to deduce a set $B \subseteq V(G)$ of size $O(k^5)$, such that B contains $\{w : vw \in F\}$ for any minimal solution F to (G, k) for which the choice of f_L, f_R and Ω_L^f, Ω_L^v, Ω_R^v, Ω_R^f is correct.

Thus, henceforth we fix a choice of f_L, f_R and Ω_L^f, Ω_L^v, Ω_R^v, Ω_R^f and we assume that the guess of these vertices and sets is correct for a minimal solution F with model σ of $G + F$. We note that, by Lemma 6.2, we should expect that

$$v \in \Omega_L^v \cap \Omega_R^v,$$
$$f_L, f_R \in \Omega_L^f \cap \Omega_R^f,$$
$$\Omega_L^f \cap \Omega_R^f \subseteq \Omega_L^f \cap \Omega_R^v \subseteq \Omega_L^v \cap \Omega_R^v,$$
$$\Omega_L^f \cap \Omega_R^f \subseteq \Omega_L^v \cap \Omega_R^f \subseteq \Omega_L^v \cap \Omega_R^v.$$

If this is not the case, we discard the branch in question.

Moreover, we maintain a set B^{sure} of vertices w for which we deduce that $vw \in F$ is implied by the choice of f_L, f_R and Ω_L^f, Ω_L^v, Ω_R^v, Ω_R^f. We start with $B^{\text{sure}} = (\Omega_L^v \cup \Omega_R^v) \setminus N_G(v)$. If at any point the size of B^{sure} exceeds k, we discard the current branch.

6.2 Preliminary Observations and Categories of Connected Components

We start with the following observation, directly implied by the assumption that f_L and f_R are untouched and $|F| \leq k$.

Lemma 6.3. *For any connected component C of $G \setminus (\Omega_L^f \cup \Omega_L^v \cup \Omega_R^v \cup \Omega_R^f)$ the following holds:*

(1) If $C \cap N_G(f_L) \cap N_G(f_R) = \emptyset$, then $\omega_\sigma(C) < p_L^f$ or $\alpha_\sigma(C) > p_R^f$. In particular, $vw \notin E(G) \cup F$ for every $w \in C$.
(2) If C contains a vertex of $N_G(f_L) \cap N_G(f_R)$, then $p_L^f < \alpha_\sigma(C) < \omega_\sigma(C) < p_R^f$ and $C \subseteq N_G(f_L) \cap N_G(f_R)$.
(3) If, moreover, C contains a neighbor of v in G, then $p_L^v < \alpha_\sigma(C) < \omega_\sigma(C) < p_R^v$ and $vw \in E(G) \cup F$ for every $w \in C$.
(4) In the last case, if $C \subseteq (N_G(f_L) \cap N_G(f_R)) \setminus N_G(v)$, then one of the following cases holds:

(a) $p_L^v < \alpha_\sigma(C) < \omega_\sigma(C) < p_R^v$ *and* $vw \in F$ *for every* $w \in C$. *Moreover, in this case* $N_G(C) \subseteq \Omega_L^v \cup \Omega_R^v$.
(b) $p_L^f < \alpha_\sigma(C) < \omega_\sigma(C) < p_L^v$ *and* $vw \notin F$ *for every* $w \in C$. *Moreover, in this case* $N_G(C) \subseteq \Omega_L^f \cup \Omega_L^v$.
(c) $p_R^v < \alpha_\sigma(C) < \omega_\sigma(C) < p_R^f$ *and* $vw \notin F$ *for every* $w \in C$. *Moreover, in this case* $N_G(C) \subseteq \Omega_R^f \cup \Omega_R^v$.

Moreover, if $|C| > k$, *then the first option does not happen.*

By Lemma 6.3, we can sort the connected components of $G \setminus (\Omega_L^f \cup \Omega_L^v \cup \Omega_R^v \cup \Omega_R^f)$ into three *categories*, depending on whether they fall into point (1), (3), or (4). Obviously, the last category is the most interesting, as we are not able to directly decide whether the vertices of the component should be inserted into B or not. The subpoints of this category (i.e., (4a), (4b), and (4c)) are henceforth called *cases*. Note that for each connected component C we know its category, but we do not know its case if it falls into category (4).

We now perform some cleaning. If there exists a component $C \in \mathsf{cc}(G \setminus (\Omega_L^f \cup \Omega_L^v \cup \Omega_R^v \cup \Omega_R^f))$ that does not fall into any category (e.g., we have $C \not\subseteq N_G(f_L) \cap N_G(f_R)$, but C contains a common neighbor of f_L and f_R), we discard the current branch. Moreover, we may include into B^{sure} all non-neighbors of v that lie in a connected component C that falls into category (3) of Lemma 6.3, that is, that contains a neighbor of v.

Clearly, only at most k components fall into case (4a) of Lemma 6.3, since each such component induces at least one fill edge incident to v. However, we do not know which of the components falling into category (4) are in fact those interesting ones. Hence, our main task now is to pinpoint a set of roughly $O(k^4)$ potential components falling into category (4) for which case (4a) may possibly happen. As each such component is of size at most k, this would conclude the proof of Theorem 3.2.

Let $\mathcal{C}$ be the family of all connected components C of $G \setminus (\Omega_L^f \cup \Omega_L^v \cup \Omega_R^v \cup \Omega_R^f)$ that fall into category (4) of Lemma 6.3, that is, $C \subseteq (N_G(f_L) \cap N_G(f_R)) \setminus N_G(v)$. We distinguish the following subfamilies that correspond to the subcases of category (4).

$$\mathcal{C}_v = \{C \in \mathcal{C} : N_G(C) \subseteq \Omega_L^v \cup \Omega_R^v\},$$
$$\mathcal{C}_L = \{C \in \mathcal{C} : N_G(C) \subseteq \Omega_L^f \cup \Omega_L^v\},$$
$$\mathcal{C}_R = \{C \in \mathcal{C} : N_G(C) \subseteq \Omega_R^f \cup \Omega_R^v\}.$$

If $\mathcal{C}_v \cup \mathcal{C}_L \cup \mathcal{C}_R \neq \mathcal{C}$, we discard the current branch. Moreover, for any $C \in \mathcal{C}_v \setminus (\mathcal{C}_L \cup \mathcal{C}_R)$ we include all vertices of C into B^{sure}, as such a component will surely fall into case (4a).

In the sequel, we will consider components that belong to different combinations of sets $\mathcal{C}_v, \mathcal{C}_L, \mathcal{C}_R$. The following fact, used often implicitly, follows directly from the definitions of $\mathcal{C}_v, \mathcal{C}_L, \mathcal{C}_R$ and inclusion relations between $\Omega_L^f, \Omega_L^v, \Omega_R^v, \Omega_R^f$.

Lemma 6.4. *The following holds:*

- *If* $C \in \mathcal{C}_L \cap \mathcal{C}_v$, *then* $N_G(C) \subseteq \Omega_L^v$. *If, moreover,* $C \notin \mathcal{C}_R$, *then* $N_G(C) \cap (\Omega_L^v \setminus \Omega_R^v) \neq \emptyset$.
- *If* $C \in \mathcal{C}_R \cap \mathcal{C}_v$, *then* $N_G(C) \subseteq \Omega_R^v$. *If, moreover,* $C \notin \mathcal{C}_L$, *then* $N_G(C) \cap (\Omega_R^v \setminus \Omega_L^v) \neq \emptyset$.
- *If* $C \in \mathcal{C}_L \cap \mathcal{C}_R$, *then* $N_G(C) \subseteq \Omega_L^v \cap \Omega_R^v$ *and in particular* $C \in \mathcal{C}_v$.

6.3 Troublesome Components

Our goal now is to focus on $\mathcal{C}_L$ and pinpoint a small set of components of $\mathcal{C}_L \cap \mathcal{C}_v$ that may possibly fall into case (4a) of Lemma 6.3. The arguments for $\mathcal{C}_R$ will be symmetrical.

To this end, we will construct a family $\mathcal{T} \subseteq C_L$ of *troublesome* components. Informally speaking, a component is troublesome if it is highly unclear where or how it should live in the model σ. We will argue that there is a bounded number of troublesome components (strictly speaking, $O(k^2)$ of them) and any component that falls into case (4a) of Lemma 6.3 is in some sense "close" to a troublesome component.

We start by putting into $\mathcal{T}$ all connected components $C \in C_L$ that cannot be drawn in the model of a completion of G between sections Ω_L^f and Ω_L^v without an incident edge of the solution. More formally, we denote $F_L = \binom{\Omega_L^v}{2} \setminus E(G) \subseteq F$ and define the following.

Definition 6.5. A component $C \in C_L \cap C_v$ is *freely drawable* if there exists an interval model σ_C of $(G + F_L)[C \cup \Omega_L^v]$ that starts with all starting events of $\mathcal{E}(\Omega_L^v \cap \Omega_L^f)$ and ends with all ending events of $\mathcal{E}(\Omega_L^v)$.

We now state the formerly informal motivation for this definition.

Lemma 6.6. *If $C \in (C_L \cap C_v) \setminus C_R$ is not freely drawable, then it is touched by F.*

Proof. As $C \notin C_R$, it cannot fall into case (4c) of Lemma 6.3. If C falls into case (4a), then it is touched due to the fill-in edges incident to v. Otherwise, unless C is touched, the model σ restricted to $C \cup \Omega_L^v$ witnesses that C is freely drawable. □

Finally, we remark that we may recognize freely drawable components in polynomial time.

Lemma 6.7. *Given $C \in C_L \cap C_v$, we can recognize if C is freely drawable in polynomial time.*

Proof. We simply use Lemma 2.1 for the graph $(G + F_L)[C \cup \Omega_L^v]$ and cliques $\Omega_L^v \cap \Omega_L^f$ and Ω_L^v. □

Using Lemma 6.7, we recognize all components of $(C_L \cap C_v) \setminus C_R$ that are not freely drawable. If there are more than $2k$ of them, by Lemma 6.6 we may discard the current branch. Otherwise, we put all not freely drawable components of $(C_L \cap C_v) \setminus C_R$ into $\mathcal{T}$.

We remark that if C is freely drawable, then $\Omega_L^v \cap \Omega_L^f \subseteq N_G(w)$ for any $w \in C$.

As we needed to exclude the components of C_R for Lemma 6.6, we now proceed to the components of $C_L \cap C_R$. Denote $P = \Omega_L^f \cap \Omega_R^f$ and $K = (\Omega_L^v \cap \Omega_R^v) \setminus P$. It turns out that the choice of f_L and f_R implies that K is small.

Lemma 6.8. *All vertices of K are touched by F and, consequently, $|K| \leq 2k$.*

Proof. Consider any $x \in K$. As $x \in \Omega_L^v \cap \Omega_R^v$, we have $\sigma(\alpha_x) \leq p_L^v < p_R^v \leq \sigma(\omega_x)$. As $x \notin \Omega_L^f \cap \Omega_R^f$, we have $\sigma(\alpha_x) > p_L^f$ or $\sigma(\omega_x) < p_R^f$. If x is untouched by F, x would be a better candidate for f_L in the first case, and a better candidate for f_R in the second case. □

Note that by Lemma 6.4, we have $N_G(C) \subseteq P \cup K$ for any $C \in C_L \cap C_R$. Lemma 6.8 allows us to use the bound of Lemma 4.6.

Lemma 6.9. $|C_L \cap C_R| = O(k^2)$.

Proof. There are at most $2k$ connected components of $C_L \cap C_R$ that are touched by F. Consider now untouched $C \in C_L \cap C_R$. As $p_L^f < \alpha_\sigma(C) < \omega_\sigma(C) < p_R^f$, we have $aw \in E(G)$ for any $w \in C$, $a \in P$. The lemma follows from an application of Lemma 4.6 to $A = P \cup K$ and $r = |K| \leq 2k$. □

Thus, if $|C_L \cap C_R|$ is too large, we discard the current branch. Moreover, we can also discard the current branch if there exists $C \in C_L \cap C_R$ with $|(C \times P) \setminus E(G)| > k$: such a component C would

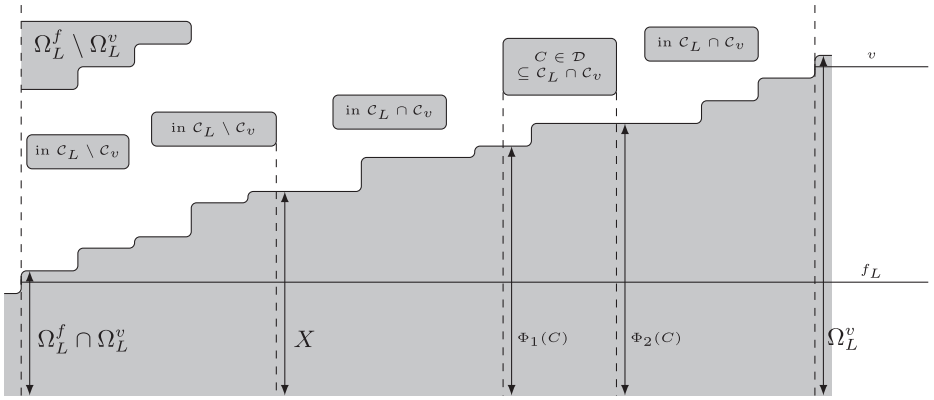

Fig. 19. A closer insight into the area between Ω_L^f and Ω_L^v.

need too much fill-in edges between itself and P. If neither of the above situations happen, we insert $C_L \cap C_R$ into $\mathcal{T}$, that is, we treat all components of $C_L \cap C_R$ as troublesome.

We now inspect the possible order of the starting endpoints of the vertices of $\Omega_L^v \setminus \Omega_L^f$; all these endpoints appear between positions p_L^f and p_L^v. We denote

$$X = \bigcup_{C \in \mathcal{C}_L \setminus \mathcal{C}_v} N_G(C) \cap \Omega_L^v$$

and observe the following.

Lemma 6.10. *For any $C \in (\mathcal{C}_L \cap \mathcal{C}_v) \setminus C_R$, if there exists $w \in C$ with $X \not\subseteq N_G(w)$, then C is touched by F.*

Proof. Consider such component C and vertex $w \in C$. As $C \notin C_R$, either case (4a) or case (4b) of Lemma 6.3 applies to C. If case (4a) applies, then $wv \in F$ and we are done, so assume otherwise.

Let $D \in \mathcal{C}_L \setminus \mathcal{C}_v$ such that there exists $x \in (N_G(D) \cap \Omega_L^v) \setminus N_G(w)$. Note that, in particular, $C \neq D$ and hence w does not have any neighbor in D in the graph G. As $D \in \mathcal{C}_L \setminus \mathcal{C}_v$, there exists some $y \in (\Omega_L^f \setminus \Omega_L^v) \cap N_G(D)$. Since $C \in \mathcal{C}_v$, then we have $y \notin N_G(C)$, so in particular, $wy \notin E(G)$.

Let P be a path in G with endpoints in x and y and all internal vertices in D; such a path exists since D is connected. Note that P contains no neighbor of w in G, but connects $y \in \Omega_L^f = \Omega_\sigma(p_L^f)$ with $x \in \Omega_L^v = \Omega_\sigma(p_L^v)$. As $p_L^f < \sigma(\alpha_w) < \sigma(\omega_w) < p_L^v$, w neighbors some vertex of P in $G + F$, and hence w is touched by F. □

By Lemma 6.10, we expect at most $2k$ components of $(\mathcal{C}_L \cap \mathcal{C}_v) \setminus C_R$ for which $X \not\subseteq N_G(w)$ for some $w \in C$. If there are more such components, we discard the current branch. Otherwise, we include all such components into $\mathcal{T}$.

We refer to Figure 19 for an illustration of some of the introduced notation.

We now define the following relation $\trianglelefteq$ on the components of $(\mathcal{C}_L \cap \mathcal{C}_v) \setminus C_R$: for two components $C_1, C_2 \in (\mathcal{C}_L \cap \mathcal{C}_v) \setminus C_R$ we have $C_1 \trianglelefteq C_2$ iff for any $v_1 \in C_1$ and for any $v_2 \in C_2$ it holds that $N_G(v_1) \cap \Omega_L^v \subseteq N_G(v_2) \cap \Omega_L^v$. Clearly, $\trianglelefteq$ is a transitive and reflexive relation on $(\mathcal{C}_L \cap \mathcal{C}_v) \setminus C_R$. Intuitively, $\trianglelefteq$ should be close to a total quasi-order, and should resemble the order in which the components of $(\mathcal{C}_L \cap \mathcal{C}_v) \setminus C_R$ that fall into case (4b) of Lemma 6.3 appear in the model σ, and components that are equivalent with respect to $\trianglelefteq$ should be interchangeable modules. This intuition is partially formalized in the following lemma.

Lemma 6.11. *If two components $C_1, C_2 \in (C_L \cap C_v) \setminus C_R$ are incomparable with respect to $\trianglelefteq$, then at least one of them is touched by F.*

Proof. If a component of C_L falls into case (4a) of Lemma 6.3, then all its vertices are touched. Hence, assume that both C_1 and C_2 fall into case (4b).

If $v_1v_2 \in F$ for some $v_1 \in C_1$, $v_2 \in C_2$, then both components are touched by F. Otherwise, $\omega_\sigma(C_1) < \alpha_\sigma(C_2)$ or $\omega_\sigma(C_2) < \alpha_\sigma(C_1)$; w.l.o.g., assume the first option. However, then for any $v_1 \in C_1$ and $v_2 \in C_2$ it holds that $N_{G+F}(v_1) \cap \Omega_L^v \subseteq N_{G+F}(v_2) \cap \Omega_L^v$. Hence, $C_1 \trianglelefteq C_2$ unless C_2 is touched. □

Consider now an auxiliary graph G_C with vertex set $(C_L \cap C_v) \setminus C_R$ and two components C_1 and C_2 being adjacent iff they are incomparable w.r.t. $\trianglelefteq$. By Lemma 6.11, the family of touched components is a vertex cover of G_C of size at most $2k$. We run a 2-approximation algorithm to find a vertex cover $\mathcal{V}$ of G_C. If $|\mathcal{V}| > 4k$, we discard the current branch. Otherwise, we insert $\mathcal{V}$ into $\mathcal{T}$.

This concludes the construction of the family $\mathcal{T}$ of troublesome components. Note that $|\mathcal{T}| = O(k^2)$ and $|\mathcal{T} \setminus (C_L \cap C_R)| = O(k)$. Let $\mathcal{D} = (C_L \cap C_v) \setminus \mathcal{T}$ be the set of not troublesome components. We summarize the properties of the components of $\mathcal{D}$.

(1) Every $C \in \mathcal{D}$ is freely drawable.
(2) $N_G(C) \subseteq \Omega_L^v$ for any $C \in \mathcal{D}$.
(3) Each component $C \in \mathcal{D}$ does not belong to C_R. That is, $N_G(C)$ contains a vertex of $\Omega_L^v \setminus \Omega_R^v$.
(4) The relation $\trianglelefteq$, restricted to $\mathcal{D}$, is a total quasi-order.
(5) For every component $C \in \mathcal{D}$ and each $w \in C$, we have $X \subseteq N_G(w)$.

6.4 Being Close and Far from a Troublesome Component

In this section, we show that any component that is *far* from all components of $\mathcal{T}$, in a specific meaning defined later, is left untouched by F. This, together with a bound on the number of components *close* to $\mathcal{T}$ will conclude the proof of Theorem 3.2.

For any component $C \in C_L$, we define the following two measures.

$$\phi_1(C) = \min_{w \in C} |N_G(w) \cap \Omega_L^v|,$$
$$\phi_2(C) = \max_{w \in C} |N_G(w) \cap \Omega_L^v|.$$

Note that $\phi_2(C_1) \leq \phi_1(C_2)$ whenever $C_1 \trianglelefteq C_2$. Observe moreover that $\phi_1(C) \geq |X|$ for each $C \in \mathcal{D}$.

Consider now some $C \in \mathcal{D}$. We first observe that $N_G(w) \cap \Omega_L^v = N_G(w) \setminus C$ for any $w \in C$. Second, note that, as C is freely drawable, for any $w_1, w_2 \in C$ we have $N_G(w_1) \cap \Omega_L^v \subseteq N_G(w_2) \cap \Omega_L^v$ or vice versa. In particular, for $C \in \mathcal{D}$, if we define sets

$$\Phi_1(C) = \bigcap_{w \in C} N_G(w) \cap \Omega_L^v,$$
$$\Phi_2(C) = \bigcup_{w \in C} N_G(w) \cap \Omega_L^v,$$

then there exist $w_1, w_2 \in C$ with $N_G(w_1) \cap \Omega_L^v = \Phi_1(C)$ and $N_G(w_2) \cap \Omega_L^v = \Phi_2(C)$. In particular, $|\Phi_1(C)| = \phi_1(C)$ and $|\Phi_2(C)| = \phi_2(C)$.

Enumerate now $\mathcal{D} = \{C^1, C^2, \ldots, C^{|\mathcal{D}|}\}$ such that

$$C^1 \trianglelefteq C^2 \trianglelefteq \ldots \trianglelefteq C^{|\mathcal{D}|}.$$

Note that the aforementioned numeration is not unique, as $\trianglelefteq$ is a quasi-order: they may exist $C_1, C_2 \in \mathcal{D}$ with $C_1 \trianglelefteq C_2$ and $C_2 \trianglelefteq C_1$. However, we note that such a situation is somehow limited by inapplicability of the Module Reduction Rule.

Lemma 6.12. *If $C_1 \trianglelefteq C_2$ and $C_2 \trianglelefteq C_1$ for some $C_1, C_2 \in \mathcal{D}$, then C_1, C_2 and $C_1 \cup C_2$ are modules in G. Moreover, if $\mathcal{D}' \subseteq \mathcal{D}$ such that $C_1 \trianglelefteq C_2$ and $C_2 \trianglelefteq C_1$ for any $C_1, C_2 \in \mathcal{D}'$, then $|\mathcal{D}'| \leq 2k+2$.*

Proof. By the definition of the relation $\trianglelefteq$, we infer that

$$N_G(v_1) \setminus C_1 = N_G(v_1) \cap \Omega_L^v = N_G(v_2) \cap \Omega_L^v = N_G(v_2) \setminus C_2$$

for any $v_1 \in C_1$, $v_2 \in C_2$. The first claim follows. For the second claim, note that if $|\mathcal{D}'| \geq 2k+3$, then the Module Reduction Rule would be applicable to any $2k+3$ components of $\mathcal{D}'$, and the set Ω_L^v. □

Corollary 6.13. *For any $1 \leq a \leq b \leq |\mathcal{D}|$, we have*

$$\phi_1(C^b) - \phi_2(C^a) \geq \left\lceil \frac{b-a}{2k+3} \right\rceil - 1.$$

Proof. Let $a < c_1 < c_2 < \cdots < c_s < b$ be the sequence of all indices $a < c < b$ such that $\phi_2(C^{c-1}) < \phi_2(C^c)$. By Lemma 6.12, $c_{i+1} - c_i \leq 2k+3$ for any $1 \leq i < s$ and $c_1 - a \leq 2k+3$, $b - c_s \leq 2k+3$. Consequently, $(2k+3)(s+1) \geq b-a$. The lemma follows from the observation that $s \leq \phi_2(C^{c_s}) - \phi_2(C^a) \leq \phi_1(C^b) - \phi_2(C^a)$. □

Given the ordering $C^1, C^2, \ldots, C^{|\mathcal{D}|}$, we can also observe the following corollary of the fact that all components of $\mathcal{D}$ are freely drawable.

Lemma 6.14. *For any $1 \leq a \leq b \leq |\mathcal{D}|$, if we define $F' = \binom{\Phi_2(C^b)}{2} \setminus E(G)$, then the graph*

$$(G + F')\left[\Phi_2(C^b) \cup \bigcup_{c=a}^{b} C^c\right]$$

is interval and admits a model that starts with the starting events of $\mathcal{E}(\Phi_1(C^a))$ and ends with the ending events of $\mathcal{E}(\Phi_2(C^b))$.

Proof. We prove the lemma by induction on $b-a$. For the base case $a = b$, observe that the claim is equivalent to the definition of C^a being freely drawable. In the induction step, pick any $a < c \leq b$ and use the induction hypothesis for components $C^a, C^{a+1}, \ldots, C^{c-1}$ and $C^c, C^{c+1}, \ldots, C^b$, obtaining models σ_1 and σ_2. Create the desired model σ_0 by concatenating

(1) the model σ_1, with removed suffix consisting of the ending events of $\mathcal{E}(\Phi_2(C^{c-1}))$,
(2) the starting events of $\mathcal{E}(\Phi_1(C^c) \setminus \Phi_2(C^{c-1}))$, and
(3) the model σ_2, with removed prefix consisting of the starting events of $\mathcal{E}(\Phi_1(C^c))$.

It is straightforward to verify that σ_0 satisfies all the promised properties. □

We now turn our attention to the troublesome components and inspect how they interact with the family $\mathcal{D}$. For each $T \in \mathcal{T}$, define the following.

$$a_1(T) = \min\{x : \phi_2(C^x) \geq \phi_1(T)\},$$
$$a_2(T) = \max\{x : \phi_1(C^x) \leq \phi_2(T)\},$$
$$b_1(T) = \min\{x : \phi_1(C^x) > \phi_1(T) + k\},$$
$$b_2(T) = \max\{x : \phi_2(C^x) < \phi_2(T)\}.$$

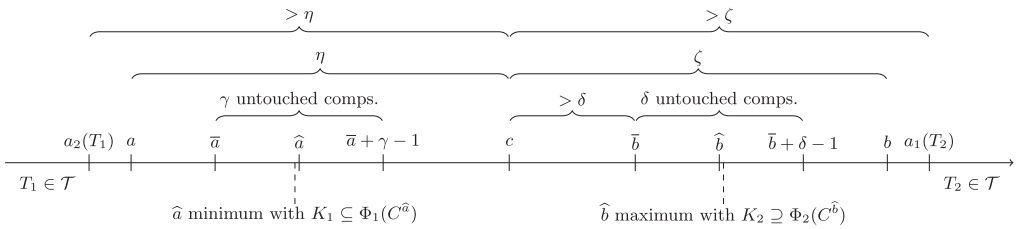

Fig. 20. The indices defined in the proof of Lemma 6.16.

All of these values can attain $+\infty$ or $-\infty$ if the corresponding set for minimization or maximization is empty.

Clearly, $a_1(T) \leq b_1(T)$, $a_2(T) \geq b_2(T)$, and $a_1(T) \leq a_2(T) + 1$. We note that, by Corollary 6.13, we have $b_1(T) - a_1(T) = O(k^2)$ and $a_2(T) - b_2(T) = O(k)$. We claim the following.

Lemma 6.15. *If $b_2(T) - b_1(T) > 2k$, then T does not fall into case (4b) of Lemma 6.3.*

Proof. Let $x, y \in T$ such that $|N_G(x) \cap \Omega_L^v| = \phi_1(T)$ and $|N_G(y) \cap \Omega_L^v| = \phi_2(T)$. If $b_2(T) - b_1(T) > 2k$, then there exists a component C^c that is untouched by F for some $b_1(T) \leq c \leq b_2(T)$. Hence, for any $w \in C^c$ we have

$$\begin{aligned}|N_{G+F}(x) \cap \Omega_L^v| &\leq |N_G(x) \cap \Omega_L^v| + k < |N_G(w) \cap \Omega_L^v| = |N_{G+F}(w) \cap \Omega_L^v| < |N_G(y) \cap \Omega_L^v| \\ &\leq |N_{G+F}(y) \cap \Omega_L^v|.\end{aligned}$$

Summarizing, $|N_{G+F}(x) \cap \Omega_L^v| < |N_{G+F}(w) \cap \Omega_L^v| < |N_{G+F}(y) \cap \Omega_L^v|$. As T is connected in G and no edge of $G + F$ connects C^c with T, it cannot happen that both C^c and T fall into case (4b) of Lemma 6.3. However, since C^c is untouched and does not belong to C_R, C^c falls into case (4b) of Lemma 6.3. This finishes the proof of the lemma. □

Let $\mathcal{T}' = \{T \in \mathcal{T} : b_2(T) - b_1(T) \leq 2k\}$ be the set of these troublesome components for which Lemma 6.15 is not applicable. Note also that for any $T \in \mathcal{T}'$, we have $-1 \leq a_2(T) - a_1(T) = O(k^2)$.

We say that a component $C^c \in \mathcal{D}$ is *far* from a troublesome component T if either $a_2(T) < c - \eta$ or $a_1(T) > c + \zeta$, where

$$\gamma = (2k+3)(k+2)+1, \qquad \delta = 2(2k+3)+1,$$
$$\eta = \gamma \cdot (2k+2), \qquad \zeta = \delta \cdot (2k+3).$$

A component C is *close* to T if it is not far from T. Define $\mathcal{D}_0$ to be the set of these components $C^c \in \mathcal{D}$ such that C^c is far from all components of $\mathcal{T}'$ and, moreover, $\eta < c < |\mathcal{D}| - \zeta$.

With this definition, we are now ready for the crucial argumentation of this section.

Lemma 6.16. *Any component $C \in \mathcal{D}_0$ is untouched by F. Consequently, such C falls into case (4b) of Lemma 6.3.*

Proof. Let $C^c \in \mathcal{D}$ be far from all components of $\mathcal{T}'$. Denote $a = c - \eta$ and $b = c + \zeta$. By the assumptions of the lemma, $1 \leq a < b \leq |\mathcal{D}|$ and, for each $T \in \mathcal{T}'$ we have either $\phi_2(T) < \phi_1(C^a)$ or $\phi_1(T) > \phi_2(C^b)$. We refer to Figure 20 for indices defined in the course of this proof.

By the Pigeonhole Principle, there exists some $\overline{a}$, $a \leq \overline{a} \leq c - \gamma$, such that all components $C^{\overline{a}}, C^{\overline{a}+1}, \ldots, C^{\overline{a}+\gamma-1}$ are untouched by F. Symmetrically, there exists some $\overline{b}$, $c + \delta < \overline{b} \leq b - \delta + 1$, such that all components $C^{\overline{b}}, C^{\overline{b}+1}, \ldots, C^{\overline{b}+\delta-1}$ are untouched by F. By Corollary 6.13, we have

$$k < \phi_1(C^{\overline{a}+\gamma-1}) - \phi_2(C^{\overline{a}}), \tag{6.1}$$

$$0 < \phi_1(C^{\overline{b}+\delta-1}) - \phi_2(C^{\overline{b}}), \tag{6.2}$$

$$0 < \phi_1(C^{\overline{b}}) - \phi_2(C^c) \leq \phi_1(C^{\overline{b}}) - \phi_2(C^{\overline{a}+\gamma-1}). \tag{6.3}$$

Recall that an untouched component of $\mathcal{D}$ needs to fall into case (4b) of Lemma 6.3. Moreover, such components need to lie one after another in the model σ, that is, if $C_1, C_2 \in \mathcal{D}$ are untouched, then $\omega_\sigma(C_1) < \alpha_\sigma(C_2)$ or $\omega_\sigma(C_2) < \alpha_\sigma(C_1)$. Note that the first case is possible only if $C_1 \trianglelefteq C_2$, and the second one only if $C_2 \trianglelefteq C_1$.

Let $p_1 = \omega_\sigma(C^{\overline{a}+\gamma-1})$ and $p_2 = \alpha_\sigma(C^{\overline{b}})$. From Equation (6.3), we infer that $p_1 < p_2$. Denote $K_1 = \Phi_2(C^{\overline{a}+\gamma-1})$ and $K_2 = \Phi_1(C^{\overline{b}})$ and observe that $\Omega_\sigma(p_1) = K_1 \subseteq \Omega_\sigma(p_2 - 1) \subseteq K_2$.

For any $C \in \mathcal{C}_L$, we have either $\omega_\sigma(C) \leq p_1$, $\alpha_\sigma(C) \geq p_2$ or $p_1 < \alpha_\sigma(C) < \omega_\sigma(C) < p_2$. We claim the following.

Claim 6.17. *Let $C \in \mathcal{C}_L$. If $p_1 < \alpha_\sigma(C) < \omega_\sigma(C) < p_2$, then $C \in \mathcal{D}$ and $C = C^d$ for some d with $\phi_2(C^{\overline{a}+\gamma-1}) \leq \phi_1(C^d) \leq \phi_2(C^d) \leq \phi_1(C^{\overline{b}})$ (in particular, $\overline{a} < d < \overline{b} + \delta - 1$, by Corollary 6.13).*

Proof. Observe that if C satisfies $p_1 < \alpha_\sigma(C) < \omega_\sigma(C) < p_2$, then for every $w \in C$ it must hold that $K_1 \subseteq N_{G+F}(w) \cap \Omega_L^v \subseteq K_2$. Since $|F| \leq k$, we infer that $|N_G(w) \cap K_1| \geq |K_1| - k$ and $N_G(w) \subseteq K_2$, for each $w \in C$. We now consider a few cases depending on the category C belongs to.

If $C \notin \mathcal{C}_v$, then $\phi_2(C) \leq |X| \leq \phi_1(C^1)$ as $N_G(C) \cap \Omega_L^v \subseteq X$ by the definition of X. Hence, by (6.1), $\phi_2(C) + k < |K_1|$, and the edges of F cannot make C adjacent to the entire K_1.

If $C \in \mathcal{T} \setminus \mathcal{T}'$, then Lemma 6.15 implies that C cannot lie between positions p_1 and p_2. If $C \in \mathcal{T}'$ then, by the choice of C^c, $\overline{a}$, and $\overline{b}$, we have either $\phi_2(C) < \phi_1(C^{\overline{a}})$ or $\phi_1(C) > \phi_2(C^{\overline{b}+\delta-1})$. In the first case, by Equation (6.1) we infer that $\phi_2(C) + k < |K_1|$. In the second case, by Equation (6.2) we infer that $\phi_1(C) > |K_2|$. In both cases, the argumentation of the first paragraph shows that C cannot lie between positions p_1 and p_2.

We are left with the case where $C \in \mathcal{D}$ and $C = C^d$ for some $1 \leq d \leq |\mathcal{D}|$. By contradiction, assume first that $\phi_1(C^d) < \phi_2(C^{\overline{a}+\gamma-1})$. If $d \geq \overline{a}$, then C^d is untouched and the vertex $w \in C^d$ that has only $\phi_1(C^d) < |K_1|$ neighbors in Ω_L^v cannot be placed after position p_1. Otherwise, by Equation (6.1) we have $\phi_2(C^d) + k < |K_1|$, and the edges of F are not sufficient to make C^d fully adjacent to K_1. In the second case, when $\phi_2(C^d) > \phi_1(C^{\overline{b}}) = |K_2|$, clearly C^d cannot be placed before position p_2 as there exists a vertex of C^d that has more than $|K_2|$ neighbors in Ω_L^v. This finishes the proof of the claim. ⌟

Define now indices $\widehat{a}$ and $\widehat{b}$ as follows: $\widehat{a}$ is minimum such that $\phi_1(C^{\widehat{a}}) \geq |K_1|$ (equivalently, $K_1 \subseteq \Phi_1(C^{\widehat{a}})$) and $\widehat{b}$ is maximum such that $\phi_2(C^{\widehat{b}}) \leq |K_2|$ (equivalently, $K_2 \supseteq \Phi_2(C^{\widehat{b}})$). By the definition of K_1 and K_2, we have $\overline{a} < \widehat{a} \leq \overline{a} + \gamma$ and $\overline{b} - 1 \leq \widehat{b} \leq \overline{b} + \delta - 1$. Denote $F_K = \binom{K_2}{2} \setminus E(G)$; note that $F_K \subseteq F$. By Lemma 6.14, it is easy to see that there exists an interval model σ_0 of

$$(G + F_K)\left[K_2 \cup \bigcup_{d=\widehat{a}}^{\widehat{b}} C^d\right]$$

that starts with the starting events of $\mathcal{E}(K_1)$ and ends with the ending events of $\mathcal{E}(K_2)$.

Let us create a model σ' from σ by

(1) removing all events of $\bigcup_{d=\widehat{a}}^{\widehat{b}} \mathcal{E}(C^d)$ as well as all starting events of $\mathcal{E}(K_2 \setminus K_1)$; observe that, by Claim 6.4, we have, in particular, removed all events that lie in σ between positions p_1 and p_2, exclusive;

(2) inserting all events of σ_0, except for the prefix consisting of the starting events of $\mathcal{E}(K_1)$ and the ending events of $\mathcal{E}(K_2)$, in the place between former positions p_1 and p_2 in σ, in the original order.

Since $K_1 = \Omega_\sigma(p_1)$ and $K_2 = \Phi_1(C^{\overline{b}})$, we infer that σ' is an interval model of $G + F'$ for some completion F'. As $F_K \subseteq F$, we have $F' \subseteq F$. Moreover, as $\widehat{a} \leq c \leq \widehat{b}$, C^c is untouched by F'. By the inclusion-wise minimality of F, $F' = F$ and the lemma is proven.

We now show that almost all elements of $\mathcal{D}$ in fact belong to $\mathcal{D}_0$.

Lemma 6.18. $|\mathcal{D} \setminus \mathcal{D}_0| = O(k^4)$.

Proof. Clearly, a component $T \in \mathcal{T}'$ is close to $O(k^3)$ components of $\mathcal{D}$. Moreover, note that for any $T \in C_L \cap C_R$, we have that $N_G(T) \subseteq P \cup K$, but, as $|(T \times P) \setminus E(G)| \leq k$ and $|K| \leq 2k$ (Lemma 6.8), it implies $|P| - k \leq \phi_1(T) \leq \phi_2(T) \leq |P| + 2k$. Consequently, by Corollary 6.13 there are $O(k^2)$ components of $\mathcal{D}$ that are close to some $T \in C_L \cap C_R$. As $|\mathcal{T} \setminus (C_L \cap C_R)| = O(k)$, the lemma follows. □

Let C_{small} be the family of those components $C \in \mathcal{C}$ for which $|C| \leq k$. Note that a component $C \in \mathcal{C}$ can fall into case (4a) only if $C \in \mathcal{C}_v \cap C_{\text{small}}$, since each vertex of a component falling into case (4a) must have a fill-in edge to v, and the number of such edges is at most k.

Finally, denote

$$B_L = \bigcup ((\mathcal{D} \setminus \mathcal{D}_0) \cap C_{\text{small}}) \cup \bigcup (\mathcal{T} \cap \mathcal{C}_v \cap C_{\text{small}}).$$

By Lemma 6.18 and the definition of C_{small}, we have that $|\bigcup ((\mathcal{D} \setminus \mathcal{D}_0) \cap C_{\text{small}})| = O(k^5)$. Since $|\mathcal{T}| = O(k^2)$, we have $|\bigcup (\mathcal{T} \cap \mathcal{C}_v \cap C_{\text{small}})| = O(k^3)$. As a result, we obtain $|B_L| = O(k^5)$. Symmetrically, by inspecting C_R instead of C_L, we obtain a set B_R of size $O(k^5)$.

Define now $B = B^{\text{sure}} \cup B_L \cup B_R$. As $\mathcal{C}_v \setminus \mathcal{T} \subseteq \mathcal{D}$, Lemma 6.16 ensures that $\{w \in V(G) : vw \in F\} \subseteq B$. Hence, we insert B into the constructed family $\mathcal{F}$ and conclude the proof of Theorem 3.2.

7 SMALL-SEPARATION LEMMA

In this short section, we prove the following structural result.

Theorem 7.1. *Let (G, k) be a YES-instance to Interval Completion, let F be a minimum solution to (G, k), and let σ be the canonical model of $G + F$. Let $p_L < p_R$ be two integers and denote $\Omega_L = \Omega_\sigma(p_L)$, $\Omega_R = \Omega_\sigma(p_R - 1)$. Assume $K \subseteq V(G)$ is such that $K \subseteq \Omega_L \setminus \Omega_R$ or $K \subseteq \Omega_R \setminus \Omega_L$. Then there are at most $3\sqrt{k} + |K|$ connected components C of $G \setminus (\Omega_L \cup \Omega_R)$ that satisfy the following three conditions:*

(1) $N_G(C) \subseteq K \cup (\Omega_L \cap \Omega_R)$,
(2) $p_L < \alpha_\sigma(C) < \omega_\sigma(C) < p_R$, and
(3) there exists $\varepsilon \in \mathcal{E}(K)$ such that $\alpha_\sigma(C) < \sigma(\varepsilon) < \omega_\sigma(C)$.

7.1 A few words on motivation

Before we proceed to the proof of Theorem 7.1, let us now shortly elaborate on the motivation of this result.

Consider the following setting. Assume we have two vertices x and y, and we know (have guessed) that they are cheap with respect to the minimum solution F we are looking for. Moreover, in the canonical model σ of $G + F$, we have $\sigma(\alpha_x) < \sigma(\alpha_y) < \sigma(\omega_y) < \sigma(\omega_x)$. By Corollary 6.1, there are only $k^{O(\sqrt{k})}n^{70}$ choices for each of the set $F(x)$, $F(y)$, so assume we know them as well. Similarly, there is only a subexponential number of choices for the sections at the endpoints of x

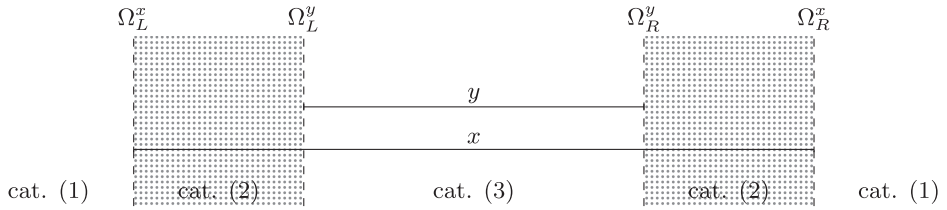

Fig. 21. Motivation for Theorem 7.1: we would like to reason about the alignment of the vertices of category (2) in the dotted areas.

and y. Hence, assume we have guessed them and denote them by Ω_L^x, Ω_L^y, Ω_R^y, and Ω_R^x. Note that we may assume that standard inclusions between these sections: $\Omega_L^x \cap \Omega_R^y \subseteq \Omega_L^y$, $\Omega_L^y \cap \Omega_R^x \subseteq \Omega_R^y$, and $\Omega_L^x \cap \Omega_R^x \subseteq \Omega_L^y \cap \Omega_R^y$.

Consider any vertex $v \in V(G) \setminus (\Omega_L^x \cup \Omega_L^y \cup \Omega_R^y \cup \Omega_R^x)$. Note that, by inspecting whether $vx \in E(G) \cup F(x)$ and whether $vy \in E(G) \cup F(y)$, we may classify v into one of three categories (see also Figure 21):

(1) $vx \notin E(G) \cup F(x)$ and $vy \notin E(G) \cup F(y)$, hence $\sigma(\omega_v) < \sigma(\alpha_x)$ or $\sigma(\alpha_v) > \sigma(\omega_x)$;
(2) $vx \in E(G) \cup F(x)$ but $vy \notin E(G) \cup F(y)$, hence $\sigma(\alpha_x) < \sigma(\alpha_v) < \sigma(\omega_v) < \sigma(\alpha_y)$ or $\sigma(\omega_y) < \sigma(\alpha_v) < \sigma(\omega_v) < \sigma(\omega_x)$;
(3) $vx \in E(G) \cup F(x)$ and $vy \in E(G) \cup F(y)$, hence $\sigma(\alpha_y) < \sigma(\alpha_v) < \sigma(\omega_v) < \sigma(\omega_y)$.

Moreover, the choice of the category needs to be homogeneous among each connected component of $G \setminus (\Omega_L^x \cup \Omega_L^y \cup \Omega_R^y \cup \Omega_R^x)$.

We will be interested mostly in the second category, and we would like to guess which components C of this category lie, in the model σ, to the left of the vertex y, and which lie to the right of it. If such a component C has a neighbor in $\Omega_L^x \setminus \Omega_L^y$ or in $\Omega_L^y \setminus \Omega_R^y$, then C needs to lie in the left; similarly, a component C with a neighbor in $\Omega_R^x \setminus \Omega_R^y$ or in $\Omega_R^y \setminus \Omega_L^y$ needs to lie in the space to the right. That is, we may deduce this choice from the neighborhood of a component C unless $N_G(C) \subseteq \Omega_L^y \cap \Omega_R^y$. In what follows, let us call a component C of the second category with $N_G(C) \subseteq \Omega_L^y \cap \Omega_R^y$ a *difficult component*.

We now argue that Theorem 7.1 is helpful if $K := (\Omega_L^y \cap \Omega_R^y) \setminus (\Omega_L^x \cap \Omega_R^x)$ is small. This is, in particular, the case if K contains only expensive vertices and thus its cardinality is bounded by $2\sqrt{k}$.

Observe that a difficult component has neighbors in K and in $\Omega_L^x \cap \Omega_L^y$. Consequently, Lemma 4.6, applied to $r = |K|$ and $A = \Omega_L^y \cap \Omega_R^y$ ensures that there are only $\mathrm{poly}(k + |K|)$ difficult components.

In the canonical model σ, a difficult component C may "occupy" an event of K (i.e., may contain an event of $\mathcal{E}(K)$ between $\alpha_\sigma(C)$ and $\omega_\sigma(C)$). For such components, Theorem 7.1 plays crucial role: it ensures that there are $O(\sqrt{k} + |K|)$ such components. In the scenario when $|K| = O(\sqrt{k})$, we can afford to guess which difficult components are of this type and to guess the left/right alignment of the chosen components. For the remaining difficult components, we provide a greedy argument: in the canonical model they do not occupy any event of any of the vertices from the cliques Ω_L^x, Ω_R^x, Ω_L^y, Ω_R^y, hence they are modules in $G + F$ and can be arranged in a greedy manner.

7.2 Proof

By symmetry, let us assume that $K \subseteq \Omega_R \setminus \Omega_L$. In particular, all starting events and no ending event of $\mathcal{E}(K)$ lie between p_L and p_R. We say that a component C *occupies* the event $\varepsilon \in \mathcal{E}(K)$

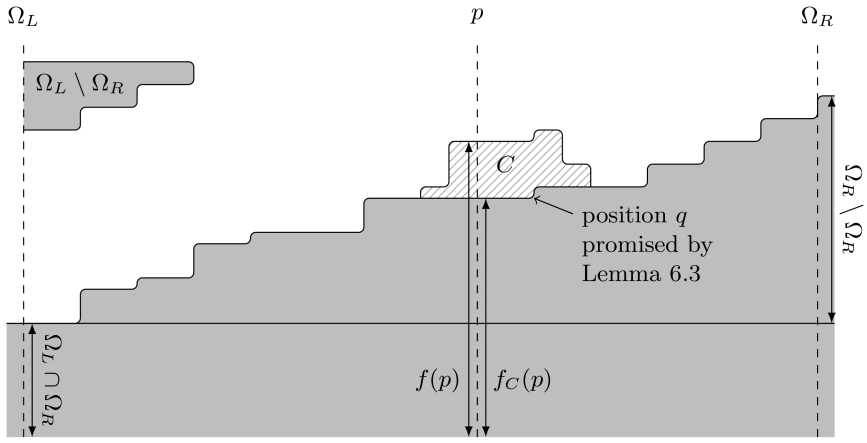

Fig. 22. Notation in proof of Theorem 7.1.

if $\alpha_\sigma(C) < \sigma(\varepsilon) < \omega_\sigma(C)$. Let $\mathcal{C}$ be the family of components of $G \setminus (\Omega_L \cup \Omega_R)$ that satisfy all conditions of Theorem 7.1, that is, we are to bound $|\mathcal{C}|$.

First, note that a much weaker bound $2k + |K|$ for Theorem 7.1 is straightforward: there are at most $2k$ components C touched by F, and no two untouched components may occupy the same event of $\mathcal{E}(K)$. However, such a bound is useless from the point of view of the aforementioned motivation.

Second, we remark that it is quite easy to obtain a bound of order $O(\sqrt{k|K|} + |K|)$. For each $C \in \mathcal{C}$, pick one endpoint $\varepsilon_C \in \mathcal{E}(K)$ occupied by C. For a starting event ε, denote $n_\varepsilon = |\{C \in \mathcal{C} : \varepsilon = \varepsilon_C\}|$. We are to bound $|\mathcal{C}| = \sum_\varepsilon n_\varepsilon$, where the number of non-zero values n_ε is bounded by $|K|$. Observe that $\sum_\varepsilon \binom{n_\varepsilon}{2} \leq |F| \leq k$, as there exists at least one edge of F between each pair of components that occupy the same endpoint. The promised bound follows from the Cauchy-Schwarz inequality.

An $O(\sqrt{k|K|} + |K|)$ bound is sufficient to establish a subexponential algorithm for Interval Completion, but the final dependency on k in the exponent would be $O(k^{2/3} \log k)$. Hence, we employ a more careful analysis of the components of $\mathcal{C}$ to obtain the bound promised in Theorem 7.1, and, consequently, reduce the dependency on k to exponential in $O(\sqrt{k} \log k)$.

For any position $p_L \leq p < p_R$ and any component $C \in \mathcal{C}$ we define

$$f(p) = |\Omega_\sigma(p)|, \qquad f_C(p) = |\Omega_\sigma(p) \setminus C|.$$

Recall that for each $C \in \mathcal{C}$, we have $p_L < \alpha_\sigma(C) < \omega_\sigma(C) < p_R$ and $N_G(C) \subseteq K \cup (\Omega_L \cap \Omega_R) \subseteq \Omega_R$. We refer to Figure 22 for an overview of the notation used in this proof.

Informally speaking, the aforementioned inclusion allows us to compare the model σ with its modification σ', where some prefix of events of $\mathcal{E}(C)$ are shifted a bit to the right, that is, $N_G(C) \subseteq \Omega_R$ ensures that σ' still represents $G + F'$ for some completion F'. If f_C for some $C \in \mathcal{C}$ has a small value at some local minimum at $p \geq \alpha_\sigma(C)$, we may shift all events of $\mathcal{E}(C)$ that lie before p to this local minimum, obtaining a smaller completion F'. We infer that f is in some sense increasing, and we need to "pay" at least one in the value of f for each component $C \in \mathcal{C}$. Theorem 7.1 will follow from an observation that the value of f cannot change by much more than $|K|$.

We proceed to a formal argumentation. In the next three lemmas, we establish the fact that f is in some sense increasing.

Lemma 7.2. *For each $C \in \mathcal{C}$ and each $\alpha_\sigma(C) \leq p < \omega_\sigma(C)$, we have $f_C(p) \geq f(\alpha_\sigma(C) - 1)$.*

Proof. Assume the contrary, and let p be the smallest position such that $\alpha_\sigma(C) \leq p < \omega_\sigma(C)$ and $f_C(p) < f(\alpha_\sigma(C) - 1)$. Note that $f(\alpha_\sigma(C) - 1) = f_C(\alpha_\sigma(C) - 1)$.

Consider a model σ' constructed from σ as follows: all events of $\mathcal{E}(C)$ that lie before or on the position p in the model σ are moved (without changing their internal order) to the place just after position p. As $N_G(C) \subseteq K \cup (\Omega_L \cap \Omega_R) \subseteq \Omega_R$, this is an interval model of $G + F'$ for some completion F' of G. We claim that $|F'| < |F|$.

Note that any $e \in F \triangle F'$ connects C with $V(G) \setminus C$ ($\triangle$ denotes the symmetric difference). Thus, it suffices to show that for each $v \in C$ we have $|\{w : vw \in F'\} \setminus C| \leq |\{w : vw \in F\} \setminus C|$, or equivalently $|F'(v)| \leq |F(v)|$, and that for at least one vertex of C the inequality is sharp.

Consider any $v \in C$. If $\sigma(\alpha_v) > p$, we have $F'(v) = F(v)$, so there is nothing to show. If $\sigma(\alpha_v) \leq p < \sigma(\omega_v)$, then, while constructing σ', we did not move ω_v while we moved α_v to the right, thus $F'(v) \subseteq F(v)$. Moreover, as p is the leftmost position with $f_C(p) < f(\alpha_\sigma(C) - 1)$, there exists $x \in V(G) \setminus C$ such that $\sigma(\omega_x) = p$. We have $vx \in F \setminus F'$ and, consequently, $F'(v) \subsetneq F(v)$. Note that there is at least one vertex that falls into the currently considered case by the connectivity of C.

We are left with the case $\sigma(\omega_v) \leq p$. However, now

$$(N_G(v) \setminus C) \uplus (\{w : vw \in F'\} \setminus C) = \Omega_\sigma(p) \setminus C,$$

whereas

$$(N_G(v) \setminus C) \uplus (\{w : vw \in F\} \setminus C) \supseteq \Omega_\sigma(\alpha_v) \setminus C;$$

here, $\uplus$ denotes a disjoint union of sets. The lemma follows from the definition of the position p:

$$|\Omega_\sigma(p) \setminus C| = f_C(p) < f_C(\sigma(\alpha_v)) = |\Omega_\sigma(\alpha_v) \setminus C|. \qquad \square$$

Lemma 7.3. *For every $C \in \mathcal{C}$ there exists an index q, $\alpha_\sigma(C) \leq q < \omega_\sigma(C)$, such that $f_C(q) > f(\alpha_\sigma(C) - 1)$.*

Proof. By Lemma 7.2, it suffices to prove that f_C is not constantly to equal $f(\alpha_\sigma(C) - 1) = f_C(\alpha_\sigma(C) - 1)$ for arguments between $\alpha_\sigma(C)$ (inclusive) and $\omega_\sigma(C)$ (exclusive). However, by the definition of $\mathcal{C}$, there exists a starting endpoint $\varepsilon \in \mathcal{E}(K)$ occupied by C. For such ε we have $f_C(\sigma(\varepsilon)) \neq f_C(\sigma(\varepsilon) - 1)$ and the lemma follows. $\square$

Lemma 7.4. *For every $C \in \mathcal{C}$ and every position p such that $\omega_\sigma(C) \leq p < p_R$, we have $f(p) > f(\alpha_\sigma(C) - 1)$.*

Proof. By contradiction, assume there exists such position p with $\omega_\sigma(C) \leq p < p_R$ and $f(p) \leq f(\alpha_\sigma(C) - 1)$. Consider a model σ' constructed from σ by taking all events of $\mathcal{E}(C)$ and putting them (without changing their internal order) between former positions p and $p + 1$. As $N_G(C) \subseteq K \cup (\Omega_L \cap \Omega_R) \subseteq \Omega_R$, this is an interval model of $G + F'$ for some completion F' of G. Again, we claim that $|F'| < |F|$.

Note that any $e \in F \triangle F'$ connects C with $V(G) \setminus C$. Thus, it suffices to show that for any $v \in C$ we have $|\{w : vw \in F'\} \setminus C| \leq |\{w : vw \in F\} \setminus C|$ and for at least one vertex of C the inequality is sharp.

Consider any $v \in C$. We have

$$(N_G(v) \setminus C) \uplus (\{w : vw \in F'\} \setminus C) = \Omega_\sigma(p),$$

whereas for any position q such that $\sigma(\alpha_v) \leq q < \sigma(\omega_v)$ we have

$$(N_G(v) \setminus C) \uplus (\{w : vw \in F\} \setminus C) \supseteq \Omega_\sigma(q) \setminus C.$$

By the definition of the position p and Lemma 7.2, we have

$$|\Omega_\sigma(p)| = f(p) \leq f(\alpha_\sigma(C) - 1) \leq f_C(q) = |\Omega_\sigma(q) \setminus C|.$$

Hence, $|F'(v)| \leq |F(v)|$.

Consider now a position q given by Lemma 7.3. By the connectivity of C, there exists $v \in C$ such that $\sigma(\alpha_v) \leq q < \sigma(\omega_v)$. For this position we have $f(\alpha_\sigma(C) - 1) < f_C(q)$ and thus $|F'(v)| < |F(v)|$. □

Concluding, we obtain the following corollary.

Corollary 7.5. *For any $C \in \mathcal{C}$ and any position $\alpha_\sigma(C) \leq p < p_R$, we have $f(p) > f(\alpha_\sigma(C) - 1)$.*

Proof. For $p < \omega_\sigma(C)$, the claim follows from Lemma 7.2 as $f_C(p) < f(p)$ for every p with $\alpha_\sigma(C) \leq p < \omega_\sigma(C)$. In the remaining case of $p \geq \omega_\sigma(C)$, the claim follows directly from Lemma 7.4. □

We now conclude the proof of Theorem 7.1 by showing that the value of f cannot change too much. A component $C \in \mathcal{C}$ is *ending expensively* if the vertex $v \in C$ with $\sigma(\omega_v) = \omega_\sigma(C)$ (i.e., ω_v is the last event of $\mathcal{E}(C)$ in the model σ) is an expensive vertex w.r.t. F, and *ending cheaply* otherwise. Note that there are at most $2\sqrt{k}$ components that end expensively. Consider a component $C \in \mathcal{C}$ with maximum $\omega_\sigma(C)$ among components that end cheaply (if there are none, the bound of Theorem 7.1 holds trivially). Let $v \in C$ satisfy $\sigma(\omega_v) = \omega_\sigma(C)$. Note that

$$f(\omega_\sigma(C)) \leq |N_G(v) \cup F(v)| \leq |\Omega_L \cap \Omega_R| + |K| + \sqrt{k},$$

as v is cheap. On the other hand, for any $p_L \leq p < p_R$ we have $\Omega_L \cap \Omega_R \subseteq \Omega_\sigma(p)$, thus

$$f(p) \geq |\Omega_L \cap \Omega_R|.$$

By Corollary 7.5, there are at most

$$f(\omega_\sigma(C)) - \min_{p_L \leq p < p_R} f(p) \leq |K| + \sqrt{k}$$

components of C that end cheaply. Together with at most $2\sqrt{k}$ components ending expensively, we obtain the bound of Theorem 7.1.

We remark here that one can obtain a slightly better $2\sqrt{2k} + |K|$ bound by redefining a cheap vertex to be one with at most $\sqrt{2k}$ incident edges from the solution. However, we prefer to stick with the thresholds defined in the preliminaries for the sake of clarity of the presentation.

8 DYNAMIC PROGRAMMING

In this final section, we describe a dynamic programming algorithm to solve Interval Completion in $O^\star(k^{O(\sqrt{k})})$ time. To this end, fix an Interval Completion instance (G, k) and, without loss of generality, assume that the Module Reduction Rule is not applicable to (G, k).

A straightforward approach, based on the subexponential algorithm for the Chordal Completion problem, would be to enumerate all possible sections via Theorem 3.1 and, for each section Ω, try to deduce (or guess) which components of $G \setminus \Omega$ lie to the left and which lie to the right to the section Ω. However, if Ω is large, there may be many such components with many different neighborhoods in Ω and, consequently, such a guessing step seems expensive. Thus, we need to employ a more involved definition of a "separation" to define a subproblem for the dynamic programming.

8.1 Worlds

We first make use of Corollary 6.1 to observe that, for a fixed vertex v that is cheap in a given minimal solution F, we can afford classifying vertices $w \in V(G) \setminus \{v\}$ depending on whether they are included in one of the sections at endpoints of v, or are incident to v.

Definition 8.1. A *world* is a tuple $\mathbf{W} = (v, \Omega_L, \Omega_R, p_L, p_R, F_v)$ where

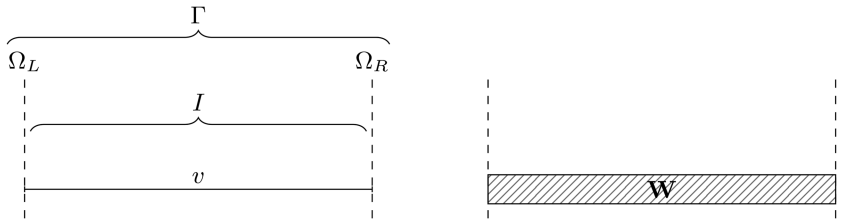

Fig. 23. A world with its most important elements (to the left) and its symbolic notation used in subsequent figures (to the right).

(1) $v \in V(G), \Omega_L, \Omega_R \subseteq V(G), F_v \subseteq (\{v\} \times (V \setminus \{v\})) \setminus E(G)$ and $1 \leq p_L \leq p_R \leq 2n-1$;
(2) $v \in \Omega_L \cap \Omega_R$;
(3) $p_R - p_L = |\Omega_L \triangle \Omega_R| + 2|N_{G+F_v}(v) \setminus (\Omega_L \cup \Omega_R)|$;
(4) for any $w \in \Omega_L \cup \Omega_R$ either $w = v$ or $vw \in E(G) \cup F_v$;
(5) for any connected component C of $G \setminus (\Omega_L \cup \Omega_R)$ either $C \subseteq N_{G+F_v}(v)$ or $C \cap N_{G+F_v}(v) = \emptyset$; and
(6) $|F_v| \leq \sqrt{k}$.

For a world $\mathbf{W} = (v, \Omega_L, \Omega_R, p_L, p_R, F_v)$ we denote (see also Figure 23):

$$\begin{aligned} v(\mathbf{W}) &= v, & F_v(\mathbf{W}) &= F_v, \\ \Omega_L(\mathbf{W}) &= \Omega_L, & \Omega_R(\mathbf{W}) &= \Omega_R, \\ p_L(\mathbf{W}) &= p_L, & p_R(\mathbf{W}) &= p_R, \\ \Gamma(\mathbf{W}) &= N_{G+F_v}[v], & I(\mathbf{W}) &= \Gamma(\mathbf{W}) \setminus (\Omega_L \cup \Omega_R). \end{aligned}$$

Definition 8.2. Let F be a completion of G and σ be a model of $G + F$. We say that the world $\mathbf{W}$ *appears* in the model σ if

(1) $F_v(\mathbf{W}) = F(v(\mathbf{W}))$,
(2) $p_L(\mathbf{W}) = \sigma(\alpha_{v(\mathbf{W})})$ and $p_R(\mathbf{W}) = \sigma(\omega_{v(\mathbf{W})}) - 1$, and
(3) $\Omega_L(\mathbf{W}) = \Omega_\sigma(p_L(\mathbf{W}))$ and $\Omega_R(\mathbf{W}) = \Omega_\sigma(p_R(\mathbf{W}))$.

The following observation is straightforward from the definition of a world.

Lemma 8.3. *For any solution F to (G, k) with model σ of $G + F$, and any vertex $v \in V(G)$ that is cheap w.r.t. F, the following tuple is in fact a world appearing in σ:*

$$(v, \Omega_\sigma(\alpha_v), \Omega_\sigma(\sigma(\omega_v) - 1), \sigma(\alpha_v), \sigma(\omega_v) - 1, F(v)).$$

We denote the world defined in Lemma 8.3 by $\mathbf{W}(\sigma, v)$.

We also remark that for a world $\mathbf{W}$ appearing in a model σ, we have for every $w \notin \Omega_L(\mathbf{W}) \cup \Omega_R(\mathbf{W})$ that

$$p_L(\mathbf{W}) < \sigma(\alpha_w) < \sigma(\omega_w) \leq p_R(\mathbf{W}) \Leftrightarrow wv(\mathbf{W}) \in E(G) \cup F_v(\mathbf{W}) \Leftrightarrow w \in I(\mathbf{W}).$$

On the other hand, Theorem 3.1 and Corollary 6.1, together with an observation that the properties of a world can be verified in polynomial time, allow us to claim the following.

Lemma 8.4. *One can in $O^\star(k^{O(\sqrt{k})})$ time enumerate a family $\mathbb{W}$ of $k^{O(\sqrt{k})}n^{106}$ worlds in G such that for any minimal solution F to (G, k), all worlds that appear in the canonical model of $G + F$ belong to $\mathbb{W}$.*

We remark that the exponent $106 = 70 + 2 \cdot 17 + 2$ (obtained by enumerating all possible choices $v, p_L, \Omega_L, \Omega_R$, and F_v) is a very rough estimation. For example, one can observe that the sections Ω_L and Ω_R were already guessed in the course of guessing F_v in the proof of Theorem 3.2. However, as the exponent in the dependency on n became unholy already a few sections ago, we refrain from optimizing it.

Worlds are first basic building blocks for our states of dynamic programming: there are only relatively few interesting worlds (Lemma 8.4) while a world $\mathbf{W}$ allows us to distinguish vertices that lie between the endpoints of $v(\mathbf{W})$ in the model we are looking for.

8.2 Terraces

Unfortunately, worlds are not sufficient to capture all relevant DP states. We need a second building block, which we call a *terrace*. Intuitively, a terrace describes the behavior either in one world (called a *flat terrace*) or in the neighborhood of a world (called a *nested terrace*).

8.2.1 *Flat Terraces.*

Definition 8.5. A *flat terrace* $\mathbf{T}$ consists of a single world $\mathbf{W}$.

For a flat terrace $\mathbf{T} = \mathbf{W}$ we denote

$$\begin{aligned} I^1(\mathbf{T}) &= I^2(\mathbf{T}) = I(\mathbf{W}),\\ \Omega_L^1(\mathbf{T}) &= \Omega_L^2(\mathbf{T}) = \Omega_L(\mathbf{W}),\\ \Omega_R^1(\mathbf{T}) &= \Omega_R^2(\mathbf{T}) = \Omega_R(\mathbf{W}),\\ p_L^1(\mathbf{T}) &= p_L^2(\mathbf{T}) = p_L(\mathbf{W}),\\ p_R^1(\mathbf{T}) &= p_R^2(\mathbf{T}) = p_R(\mathbf{W}). \end{aligned}$$

8.2.2 *Nested Terrace.*

The definition of a nested terrace is more involved. We start with the following definition.

Definition 8.6. A *nested half-terrace* $\mathbf{T}$ is a triple of worlds $(\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}})$ such that $v(\mathbf{W}_1^{\text{out}}) \neq v(\mathbf{W}^{\text{in}}) \neq v(\mathbf{W}_2^{\text{out}})$,

$$p_L(\mathbf{W}_2^{\text{out}}) \leq p_L(\mathbf{W}_1^{\text{out}}) < p_L(\mathbf{W}^{\text{in}}) \leq p_R(\mathbf{W}^{\text{in}}) < p_R(\mathbf{W}_2^{\text{out}}) \leq p_R(\mathbf{W}_1^{\text{out}}),$$

and

$$|(\Omega_L(\mathbf{W}^{\text{in}}) \cap \Omega_R(\mathbf{W}^{\text{in}})) \setminus (\Omega_L(\mathbf{W}_1^{\text{out}}) \cap \Omega_R(\mathbf{W}_2^{\text{out}}))| \leq 2\sqrt{k}.$$

Note that we allow $\mathbf{W}_1^{\text{out}} = \mathbf{W}_2^{\text{out}}$. For a nested half-terrace $\mathbf{T} = (\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}})$ we denote (see also Figure 24)

$$\begin{aligned} \Omega_L^1(\mathbf{T}) &= \Omega_L(\mathbf{W}_1^{\text{out}}), & \Omega_L^2(\mathbf{T}) &= \Omega_R(\mathbf{W}^{\text{in}}),\\ \Omega_R^1(\mathbf{T}) &= \Omega_L(\mathbf{W}^{\text{in}}), & \Omega_R^2(\mathbf{T}) &= \Omega_R(\mathbf{W}_2^{\text{out}}),\\ p_L^1(\mathbf{T}) &= p_L(\mathbf{W}_1^{\text{out}}), & p_L^2(\mathbf{T}) &= p_R(\mathbf{W}^{\text{in}}),\\ p_R^1(\mathbf{T}) &= p_L(\mathbf{W}^{\text{in}}), & p_R^2(\mathbf{T}) &= p_R(\mathbf{W}_2^{\text{out}}). \end{aligned}$$

However, to properly define $I^1(\mathbf{T})$ and $I^2(\mathbf{T})$ we need to enhance a nested half-terrace $\mathbf{T}$ with an information, for each vertex $v \in (I(\mathbf{W}_1^{\text{out}}) \cap I(\mathbf{W}_2^{\text{out}})) \setminus \Gamma(\mathbf{W}^{\text{in}})$ whether it should lie before or after $v(\mathbf{W}^{\text{in}})$ in the model σ we are looking for.

Definition 8.7. A *nested terrace* $\mathbf{T}$ is a quadruple $(\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}}, g)$ where $(\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}})$ is a nested half-terrace and $g : (I(\mathbf{W}_1^{\text{out}}) \cap I(\mathbf{W}_2^{\text{out}})) \setminus \Gamma(\mathbf{W}^{\text{in}}) \to \{1, 2\}$ is a function such that whenever

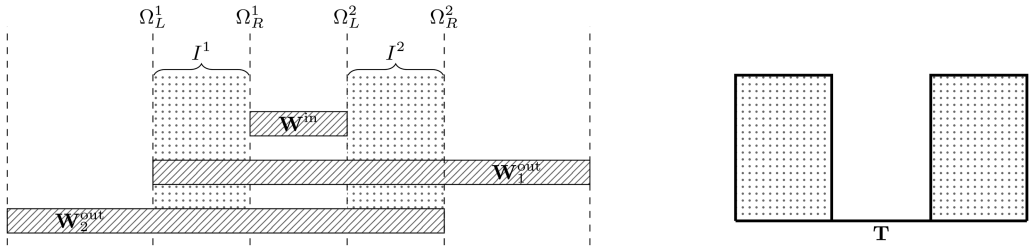

Fig. 24. A nested terrace with its most important notation (to the left) and its symbolic notation used in subsequent figures (to the right). The dotted areas are the "important" areas for a terrace: the left one has borders Ω_L^1, Ω_R^1 and interior I^1, and the right one has borders Ω_L^2, Ω_R^2 and interior I^2.

two vertices x and y in the domain of g are adjacent, then $g(x) = g(y)$ (i.e., g is constant on each connected component in the graph induced by its domain).

We may now denote for a nested terrace $\mathbf{T} = (\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}}, g)$

$$I^1(\mathbf{T}) = g^{-1}(1), \qquad I^2(\mathbf{T}) = g^{-1}(2).$$

Definition 8.8. Let F be a completion of G and σ be a model of $G + F$. We say that a nested terrace $\mathbf{T} = (\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}}, g)$ *appears* in the model σ if all $\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}}$ appear in σ and, moreover, for any $w \in (I(\mathbf{W}_1^{\text{out}}) \cap I(\mathbf{W}_2^{\text{out}})) \setminus \Gamma(\mathbf{W}^{\text{in}})$ we have $\sigma(\omega_w) < \sigma(\alpha_{v(\mathbf{W}^{\text{in}})})$ if and only if $g(w) = 1$.

A direct check from the definition shows the following.

Lemma 8.9. *Let F be a completion of G and σ be a model of $G + F$. Let $x \in V(G)$ be an arbitrary cheap vertex different than $\mathfrak{r}$. Let y_1 be the cheap vertex with rightmost $\sigma(\alpha_{y_1})$ and y_2 be the cheap vertex with leftmost $\sigma(\omega_{y_2})$ among the cheap vertices y satisfying $\sigma(\alpha_y) < \sigma(\alpha_x) < \sigma(\omega_x) < \sigma(\omega_y)$. Then $(\mathbf{W}(\sigma, x), \mathbf{W}(\sigma, y_1), \mathbf{W}(\sigma, y_2))$ is a nested half-terrace that appears in σ.*

Moreover, if we denote

$$X^1 = \{w \in V(G) : \sigma(\alpha_{y_1}) < \sigma(\alpha_w) < \sigma(\omega_w) < \sigma(\alpha_x)\},$$
$$X^2 = \{w \in V(G) : \sigma(\omega_x) < \sigma(\alpha_w) < \sigma(\omega_w) < \sigma(\omega_{y_2})\},$$
$$g = (X^1 \times \{1\}) \cup (X^2 \times \{2\}),$$

then $X^1 \cup X^2 = (I(\mathbf{W}(\sigma, y_1)) \cap I(\mathbf{W}(\sigma, y_2))) \setminus \Gamma(\mathbf{W}(\sigma, x))$ and $(\mathbf{W}(\sigma, x), \mathbf{W}(\sigma, y_1), \mathbf{W}(\sigma, y_2), g)$ is a nested terrace that appears in σ.

Proof. Note that the vertices y_1 and y_2 exist, as $\mathfrak{r}$ is a candidate for both of them. The only claim that is not straightforward is that there are at most $2\sqrt{k}$ vertices with $\sigma(\alpha_w) < \sigma(\alpha_x) < \sigma(\omega_x) < \sigma(\omega_w)$ and $\sigma(\alpha_w) > \sigma(\alpha_{y_1})$ or $\sigma(\omega_w) < \sigma(\omega_{y_2})$. However, this follows from the definition of y_1 and y_2: all such w are expensive w.r.t. F. □

We denote the nested terrace defined in Lemma 8.9 by $\mathbf{T}(\sigma, x)$. Note that the vertices y_1 and y_2 can be deduced from the model σ and vertex x; for fixed σ and x, we denote them by $y_1(\sigma, x)$ and $y_2(\sigma, x)$.

At the end of this section, we would like to include a few words about the intuition. Every terrace $\mathbf{T}$ has two "active" areas, $I^1(\mathbf{T})$ and $I^2(\mathbf{T})$, whose best possible completions we would like to compute. In a nested terrace these areas are in fact disjoint, and we have $p_L^1(\mathbf{T}) \le p_R^1(\mathbf{T}) \le p_L^2(\mathbf{T}) \le p_R^2(\mathbf{T})$. A flat terrace, however, is a degenerated case where these two areas are in fact the same. Thus, only the first and the last inequality holds, that is, we trivially have $p_L^1(\mathbf{T}) \le p_R^1(\mathbf{T})$ and

$p_L^2(\mathbf{T}) \le p_R^2(\mathbf{T})$, but not necessarily $p_R^1(\mathbf{T}) \le p_L^2(\mathbf{T})$ (and in fact this inequality will be most often false). Hence, when talking about an arbitrary terrace we will use *only* inequalities $p_L^1(\mathbf{T}) \le p_R^1(\mathbf{T})$ and $p_L^2(\mathbf{T}) \le p_R^2(\mathbf{T})$, which are true in both cases. Intuitively, in the sequel we combine pairs of terraces, and in this combination we look at only one active area of each participating terrace. Thus, we in fact have no chance of attempting using any inequality that relates the placements of two active areas of the same terrace.

8.2.3 Enumerating Terraces. We now show that we can enumerate a relatively small family of potential terraces.

Theorem 8.10. *One can in $O^\star(k^{O(\sqrt{k})})$ time enumerate a family $\mathbb{T}$ of $k^{O(\sqrt{k})}n^{318}$ terraces such that if (G,k) is a YES-instance of Interval Completion, then, for the canonical solution F and the canonical model σ of $G+F$, all terraces that appear in σ belong to $\mathbb{T}$.*

Proof. Enumeration of potential flat terraces follows directly from Lemma 8.4. Similarly, we can enumerate a family of $k^{O(\sqrt{k})}n^{318}$ nested half-terraces such that all nested half-terraces appearing in σ belong to this family. To finish the proof we need to show that, for a fixed nested half-terrace $(\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}})$, we may enumerate a family of $k^{O(\sqrt{k})}$ potential functions g. Henceforth, we assume that we have a fixed nested half-terrace $(\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}})$ that appears in σ. We describe the algorithm as a branching algorithm that generates $k^{O(\sqrt{k})}$ subcases and outputs a single function g in each subcase. We argue that in the case $(\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}})$ indeed appears in σ, the correct function g completing $(\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}})$ to a nested terrace appearing in σ will be among the enumerated candidates.

Let $\mathcal{C}_0$ be the family of these components $C \in \mathsf{cc}(G \setminus (\Omega_L(\mathbf{W}_1^{\text{out}}) \cup \Omega_L(\mathbf{W}^{\text{in}}) \cup \Omega_R(\mathbf{W}^{\text{in}}) \cup \Omega_R(\mathbf{W}_2^{\text{out}})))$ for which $C \subseteq (I(\mathbf{W}_1^{\text{out}}) \cap I(\mathbf{W}_2^{\text{out}})) \setminus \Gamma(\mathbf{W}^{\text{in}})$. That is, $\bigcup \mathcal{C}_0$ is the domain of the function g in any nested terrace $(\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}}, g)$.

Denote $P = \Omega_L(\mathbf{W}_1^{\text{out}}) \cap \Omega_R(\mathbf{W}_2^{\text{out}})$ and $K = (\Omega_L(\mathbf{W}^{\text{in}}) \cap \Omega_R(\mathbf{W}^{\text{in}})) \setminus P$. Note that we may assume $P \subseteq \Omega_L(\mathbf{W}^{\text{in}}) \cap \Omega_R(\mathbf{W}^{\text{in}})$, as otherwise clearly $(\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}})$ does not appear in σ and we may discard such a choice of a nested half-terrace. Moreover, by the definition of a nested half-terrace, $|K| \le 2\sqrt{k}$.

Pick any $C \in \mathcal{C}_0$. Note that, unless $N_G(C) \subseteq P \cup K$, we may deduce whether the vertices of C lie to the left or to the right of $v(\mathbf{W}^{\text{in}})$ in the model σ, and, consequently, fix $g(w)$ for every $w \in C$. Hence, in the rest of the proof we focus on the family $\mathcal{C} \subseteq \mathcal{C}_0$ of these components C where $N_G(C) \subseteq P \cup K$.

Claim 8.11. *Providing $(\mathbf{W}^{\text{in}}, \mathbf{W}_1^{\text{out}}, \mathbf{W}_2^{\text{out}})$ appears in σ, it holds that $|C| = O(k^2)$.*

Proof. If F is a solution to (G,k), for any $C \in \mathcal{C}$ we have $|(C \times P) \setminus E(G)| \le k$. We obtain the claim by applying Lemma 4.6 to the set $A := P \cup K$ and threshold $r := k + |K|$. □

Thus, if $|C|$ exceeds the bound of Lemma 4.6, we discard the choice of the nested half-terrace. We proceed further with the assumption $|C| = O(k^2)$.

Now we filter out components of $\mathcal{C}$ that are handled by Theorem 7.1. To this end, define $\mathcal{C}' \subseteq \mathcal{C}$ to be the family of components $C \in \mathcal{C}$ such that $\alpha_\sigma(C) < \sigma(\varepsilon) < \omega_\sigma(C)$ for some $\varepsilon \in \mathcal{E}(K)$.

Claim 8.12. $|\mathcal{C}'| \le 10\sqrt{k}$.

Proof. The claim follows from two applications of Theorem 7.1: one to the pair of sections $\Omega_L(\mathbf{W}_1^{\text{out}}), \Omega_L(\mathbf{W}^{\text{in}})$ and the set $(\Omega_L(\mathbf{W}^{\text{in}}) \cap \Omega_R(\mathbf{W}^{\text{in}})) \setminus \Omega_L(\mathbf{W}_1^{\text{out}})$ and one to the pair of sections $\Omega_R(\mathbf{W}^{\text{in}}), \Omega_R(\mathbf{W}_2^{\text{out}})$ and the set $(\Omega_L(\mathbf{W}^{\text{in}}) \cap \Omega_R(\mathbf{W}^{\text{in}})) \setminus \Omega_R(\mathbf{W}_2^{\text{out}})$. □

We guess the subfamily $\mathcal{C}'$ and for each such $C \in \mathcal{C}'$ we guess whether all vertices of C lie to the left or to the right of $v(\mathbf{W}^{\text{in}})$ in the model σ. As $|C| = O(k^2)$ and $|\mathcal{C}'| \leq 10\sqrt{k}$, such a guess leads to $k^{O(\sqrt{k})}$ subcases. We denote $\mathcal{D} = \mathcal{C} \setminus \mathcal{C}'$ the family of the remaining components.

Let $\{x_L^1, x_L^2, \ldots, x_L^{s_L-1}\}$ be the set of these $x \in K$ such that $\sigma(\alpha_x) > p_L(\mathbf{W}_1^{\text{out}})$, enumerated such that $\sigma(\alpha_{x_L^1}) < \sigma(\alpha_{x_L^2}) < \cdots < \sigma(\alpha_{x_L^{s_L-1}})$. Symmetrically, let $\{x_R^1, x_R^2, \ldots, x_R^{s_R-1}\}$ be the set of these $x \in K$ such that $\sigma(\omega_x) \leq p_R(\mathbf{W}_2^{\text{out}})$, enumerated such that $\sigma(\omega_{x_R^1}) < \sigma(\omega_{x_R^2}) < \ldots < \sigma(\omega_{x_R^{s_R-1}})$. Denote $x_L^0 = v(\mathbf{W}_1^{\text{out}})$, $x_R^{s_R} = v(\mathbf{W}_2^{\text{out}})$ and $x_R^{s_L} = x_R^0 = v(\mathbf{W}^{\text{in}})$. Recall that $|K| \leq 2\sqrt{k}$; at the cost of branching into $k^{O(\sqrt{k})}$ subcases, we guess the sequences x_L^i and x_R^i.

Let us now investigate how the components of $\mathcal{D}$ lie in the model σ.

Claim 8.13. *For any $C \in \mathcal{D}$, all events of $\mathcal{E}(C)$ are consecutive events in the model σ. That is, for any $\varepsilon \notin \mathcal{E}(C)$ either $\sigma(\varepsilon) < \alpha_\sigma(C)$ or $\sigma(\varepsilon) > \omega_\sigma(C)$.*

Proof. For the sake of contradiction, assume that there exists an event $\varepsilon \notin \mathcal{E}(C)$ such that $\alpha_\sigma(C) < \sigma(\varepsilon) < \omega_\sigma(C)$. Let $\varepsilon \in \{\alpha_w, \omega_w\}$ for some $w \notin C$. By the definition of $\mathcal{D}$, $w \notin K$. Clearly, $w \notin P = \Omega_L(\mathbf{W}_1^{\text{out}}) \cap \Omega_R(\mathbf{W}_2^{\text{out}})$. Hence, $w \notin N_G(C)$, as $C \in \mathcal{D} \subseteq \mathcal{C}$.

Take now any position p such that $\alpha_\sigma(C) - 1 \leq p \leq \omega_\sigma(C)$ and consider a model σ' created from σ by taking out all events of $\mathcal{E}(C)$ and inserting them between former positions p and $p+1$ in the original order. As every event not in $\mathcal{E}(C)$ that lies between $\alpha_\sigma(C)$ and $\omega_\sigma(C)$ is an endpoint of a non-neighbor of C, σ' is an interval model of $G + F'$ for some completion F' of G. Moreover, $F \triangle F'$ consists only of edges between C and $V(G) \setminus C$.

Pick any $v \in C$. Clearly,

$$(N_G(v) \setminus C) \uplus (\{w : vw \in F'\} \setminus C) = \Omega_\sigma(p) \setminus C.$$

On the other hand, for any position q with $\sigma(\alpha_v) \leq q < \sigma(\omega_v)$ we have

$$(N_G(v) \setminus C) \uplus (\{w : vw \in F\} \setminus C) \supseteq \Omega_\sigma(q) \setminus C.$$

Thus, if we choose p so that $|\Omega_\sigma(p) \setminus C|$ is minimum possible, we obtain $|\{w : vw \in F'\} \setminus C| \leq |\{w : vw \in F\} \setminus C|$ for every $v \in C$ and, consequently, $|F'| \leq |F|$. Consider now any $v \in C$ with $\sigma(\alpha_v) < \sigma(\varepsilon) < \sigma(\omega_v)$; let $(q, q') = (\sigma(\varepsilon) - 1, \sigma(\varepsilon))$ if ε is a closing event, and let $(q, q') = (\sigma(\varepsilon), \sigma(\varepsilon) - 1)$ if ε is an opening event. We infer that $|\Omega_\sigma(q) \setminus C| = |\Omega_\sigma(q') \setminus C| + 1$, and hence, in particular, $|\Omega_\sigma(q) \setminus C| > |\Omega_\sigma(p) \setminus C|$ by the choice of p. We thus obtain $|\{w : vw \in F'\} \setminus C| < |\{w : vw \in F\} \setminus C|$, which implies $|F'| < |F|$, a contradiction with the choice of F. ⌟

By Claim 8.13, we infer that the components of $\mathcal{D}$ are put into the model σ in a somewhat independent and greedy manner. More precisely, define for a position p a set $B(p) := \Omega_\sigma(p) \setminus (\bigcup \mathcal{D})$. On the sets $B(p)$ we define an order as follows: $B(p) \trianglelefteq B(q)$ if $|B(p)| < |B(q)|$ or $|B(p)| = |B(q)|$ and $B(p) \preceq B(q)$, where $\prec$ is the order $\prec$ on $V(G)$ extended to subsets of $V(G)$ compared lexicographically. Note that $\trianglelefteq$ is a total order.

For any $0 \leq i < s_L$, we define p_L^i to be any index $\sigma(\alpha_{x_L^i}) \leq p_L^i < \sigma(\alpha_{x_L^{i+1}})$ with minimum $B(p_L^i)$ according to the order $\trianglelefteq$. Moreover, by Claim 8.13 we can observe that for every $C \in \mathcal{D}$, the set $B(p)$ is constant for all p with $\alpha_\sigma(C) - 1 \leq p \leq \omega_\sigma(C)$. Hence, we can always choose p_L^i in such a way that $p_L^i < \alpha_\sigma(C)$ or $p_L^i \geq \omega_\sigma(C)$ for each $C \in \mathcal{D}$. Consequently, $\Omega_\sigma(p_L^i) \cap (\bigcup \mathcal{D}) = \emptyset$ and $B(p_L^i) = \Omega_\sigma(p_L^i)$. Symmetrically, we define p_R^i for $0 \leq i < s_R$; again, we can do it in such a manner that $\Omega_\sigma(p_R^i) \cap (\bigcup \mathcal{D}) = \emptyset$ and $B(p_R^i) = \Omega_\sigma(p_R^i)$ for each $0 \leq i < s_R$.

We now denote

$$P_L := P \cup (K \cap \Omega_\sigma(p_L(\mathbf{W}_1^{\text{out}}))) = \Omega_\sigma(p_L(\mathbf{W}_1^{\text{out}})) \cap \Omega_\sigma(p_R(\mathbf{W}^{\text{in}})),$$
$$P_R := P \cup (K \cap \Omega_\sigma(p_R(\mathbf{W}_2^{\text{out}}))) = \Omega_\sigma(p_L(\mathbf{W}^{\text{in}})) \cap \Omega_\sigma(p_R(\mathbf{W}_2^{\text{out}})).$$

Formally, if any of the equalities above does not hold, we may discard the choice of the half-terrace. We now claim the following.

Claim 8.14. *For every $C \in \mathcal{D}$ and for every position p with $\alpha_\sigma(C) - 1 \le p \le \omega_\sigma(C)$, the set $B(p)$ is the minimum (in the order $\trianglelefteq$) set among sets $B(q)$ for $q \in P^C$, where P^C is defined as*

$$P^C = \{p_L^i : N_G(C) \subseteq P_L \cup \{x_L^j : j \le i\}\} \cup \{p_R^i : N_G(C) \subseteq P_R \cup \{x_R^j : j > i\}\}.$$

Proof. As we already argued, the set $B(p)$ is constant for all p with $\alpha_\sigma(C) - 1 \le p \le \omega_\sigma(C)$, and equals $\Omega_\sigma(p_0) \setminus C$ for any such p_0, which we henceforth fix.

Assume that C lies to the left of $v(\mathbf{W}^{\text{in}})$ in the model σ. Let $0 \le \iota < s_L$ be such that $\sigma(\alpha_{x_L^\iota}) < \alpha_\sigma(C) < \omega_\sigma(C) < \sigma(\alpha_{x_L^{\iota+1}})$. Then, by the definition of p_L^ι we have $B(p_L^\iota) \trianglelefteq B(p_0)$. Moreover, $N_G(C) \subseteq (P \cup K) \cap B(p_0) = P_L \cup \{x_L^j : j \le \iota\}$ and hence $p_L^\iota \in P^C$. The argument for C lying on the right of $v(\mathbf{W}^{\text{in}})$ is symmetric. Hence, we infer that $\min_{q \in P^C} B(q) \trianglelefteq B(p_0)$.

In the other direction, take $q_0 \in P^C$ that yields the minimum set $B(q)$ with respect to $\trianglelefteq$; note that $B(q_0) \trianglelefteq B(p_0)$, so in particular, $|B(q_0)| \le |B(p_0)|$. Observe that we can construct a model σ' from σ by taking out all events of $\mathcal{E}(C)$ and placing them between position q_0 and $q_0 + 1$. By the definition of P^C, such a model σ' is an interval model of $G + F'$ for some completion F' of G. Observe now in $G + F$ the edges between C and $V(G) \setminus C$ constitute the whole set $B(p_0) \times C$, which in particular contains all the edges between C and $V(G) \setminus C$ that were present in the original graph G. Moreover, since $B(q_0) = \Omega_\sigma(q_0)$ because of $q_0 \in P^C$, in $G + F'$ the edges between C and $V(G) \setminus C$ constitute the whole set $B(q_0) \times C$, which again contains all the edges between C and $V(G) \setminus C$ that were present in G. Consequently, $|F'| - |F| = |B(q_0) \times C| - |B(p_0) \times C|$. By the fact that F is a minimum solution, we infer that $|B(q_0)| \ge |B(p_0)|$, which together with the previously proven reverse inequality shows that $|B(p_0)| = |B(q_0)|$. If now it happens that $B(q_0) \prec B(p_0)$, then it is easy to observe that F' is lexicographically smaller than F, a contradiction to the assumption that F is the canonical solution. This concludes the proof of the claim. ⌟

As the cost of $k^{O(\sqrt{k})}$ additional subcases, we may guess the order $\trianglelefteq$ restricted to the sections $B(p_L^i)$ and $B(p_R^i)$; note that we do not want to guess either positions p_L^i, p_R^i or sets $B(p_L^i), B(p_R^i)$ themselves, only the relative order of the sets $B(p_L^i)$ and $B(p_R^i)$ with respect to the order $\trianglelefteq$. Observe also that some of the sets $B(p_L^i)$, $B(p_R^i)$ might be actually equal (which we also guess), but this can happen only for pairs from the opposite sides: sets $B(p_L^i)$ are pairwise different because of having different intersections with $\{x_L^i : 0 \le i \le s_L\}$, and likewise sets $B(p_R^i)$ are pairwise different. Once we know the order of these sets w.r.t. $\trianglelefteq$ and the sequences x_L^i and x_R^i, Claim 8.14 allows for each component $C \in \mathcal{D}$ to choose its place in the model σ in a greedy manner.

More precisely, consider $C \in \mathcal{D}$ and the set P^C defined in Claim 8.14. Knowing the order $\trianglelefteq$, by Claim 8.14 we know that C is placed in the model σ between $\alpha_{x_L^i}$ and $\alpha_{x_L^{i+1}}$ for any $0 \le i < s_L$ such that $B(p_L^i)$ is $\trianglelefteq$-minimum in $\{B(q) : q \in P^C\}$ or between $\omega_{x_R^i}$ and $\omega_{x_R^{i+1}}$ for any $0 \le i < s_R$ such that $B(p_R^i)$ is $\trianglelefteq$-minimum in $\{B(q) : q \in P^C\}$. Hence, we know whether C lies to the left or to the right of $v(\mathbf{W}^{\text{in}})$ in the model σ unless the minimum $\{B(q) : q \in P^C\}$ is attained by some p_L^i and q_L^j at the same time.

We now inspect more closely how such a situation could happen. As $B(p_L^i) = B(p_R^j)$, we have $B(p_L^i), B(p_R^j) \subseteq \Omega_L(\mathbf{W}^{\text{in}}) \cap \Omega_R(\mathbf{W}^{\text{in}}) = P \cup K$. Hence,

$$B(p_L^i) = P_L \cup \{x_L^\ell : \ell \leq i\} = P_R \cup \{x_R^\ell : \ell > j\} = B(p_R^j).$$

In particular, for any $q \in P^C \setminus \{p_L^i, p_R^j\}$ we have $B(p_L^i) \lhd B(q)$. Recall also that for any $0 \leq i < s_L$, we have at most one $j = j(i)$ such that $B(p_L^i) = B(p_R^j)$.

Let $0 \leq i < s_L$ be such that $j(i)$ exists. Let $\mathcal{D}_i \subseteq \mathcal{D}$ be the family of such components $C \in \mathcal{D}$ such that the minimum of $\{B(q) : q \in P^C\}$ is attained at $X := B(p_L^i) = B(p_R^{j(i)})$. Note that $N_{G+F}(v) \setminus C = X$ for each $v \in C$. Hence, Lemma 2.3 applies and, as σ is the canonical model of $G + F$, the components of $\mathcal{D}_i$ are arranged according to their minimum elements in the order $\prec$. That is, for any $C_1, C_2 \in \mathcal{D}_i$ such that C_1 lies before $v(\mathbf{W}^{\text{in}})$ and C_2 lies after $v(\mathbf{W}^{\text{in}})$ in the model σ, we have that the $\prec$-minimum vertex of C_1 precedes the $\prec$-minimum vertex of C_2 in the order $\prec$. Thus, to know which components of $\mathcal{D}_i$ lie in the model σ before $v(\mathbf{W}^{\text{in}})$ it suffices to know *how many* of them lie there. As $|C| = O(k^2)$ and $s_L = O(\sqrt{k})$, guessing, for each $0 \leq i < s_L$ with defined $j(i)$, how many components of $\mathcal{D}_i$ lie before $v(\mathbf{W}^{\text{in}})$ in the model σ leads to $k^{O(\sqrt{k})}$ subcases. This concludes the proof of Theorem 8.10.

8.3 Dynamic Programming: States and Computation

8.3.1 DP States. Armed with the notion of terraces, we are ready to define the state of our dynamic programming algorithm.

Definition 8.15. A *state* $\mathbf{S}$ is a pair of terraces $(\mathbf{T}_1, \mathbf{T}_2)$ such that $p_L^1(\mathbf{T}_2) \leq p_L^2(\mathbf{T}_1) < p_R^1(\mathbf{T}_2) \leq p_R^2(\mathbf{T}_1)$ and

$$p_R^1(\mathbf{T}_2) - p_L^2(\mathbf{T}_1) = 2|I^2(\mathbf{T}_1) \cap I^1(\mathbf{T}_2)| + |\Omega_L^2(\mathbf{T}_1) \triangle \Omega_R^1(\mathbf{T}_2)|.$$

We remark that each of the terraces participating in a state might be either flat or nested. Moreover, it can happen that $\mathbf{T}_1 = \mathbf{T}_2$. For a state $\mathbf{S} = (\mathbf{T}_1, \mathbf{T}_2)$ we define (see also Figure 25)

$$\begin{aligned} \Omega_L(\mathbf{S}) &= \Omega_L^2(\mathbf{T}_1), & \Omega_R(\mathbf{S}) &= \Omega_R^1(\mathbf{T}_2), \\ p_L(\mathbf{S}) &= p_L^2(\mathbf{T}_1), & p_R(\mathbf{S}) &= p_R^1(\mathbf{T}_2), \\ I(\mathbf{S}) &= I^2(\mathbf{T}_1) \cap I^1(\mathbf{T}_2), & \Gamma(\mathbf{S}) &= I(\mathbf{S}) \cup \Omega_L(\mathbf{S}) \cup \Omega_R(\mathbf{S}). \end{aligned}$$

Definition 8.16. Let F be a completion of G and σ be a model of $G + F$. We say that a state $\mathbf{S} = (\mathbf{T}_1, \mathbf{T}_2)$ *appears* in the model σ if both $\mathbf{T}_1$ and $\mathbf{T}_2$ appear in σ.

A direct check shows the following.

Lemma 8.17. *If $\mathbf{S}$ appears in a model σ of a completion $G + F$, then the events that appear on positions p satisfying $p_L(\mathbf{S}) < p \leq p_R(\mathbf{S})$ are exactly*

$$\mathcal{E}(\mathbf{S}) := \mathcal{E}(I(\mathbf{S})) \cup \{\omega_v : v \in \Omega_L(\mathbf{S}) \setminus \Omega_R(\mathbf{S})\} \cup \{\alpha_v : v \in \Omega_R(\mathbf{S}) \setminus \Omega_L(\mathbf{S})\}.$$

Note that we have $|\mathcal{E}(\mathbf{S})| = 2|I^2(\mathbf{T}_1) \cap I^1(\mathbf{T}_2)| + |\Omega_L^2(\mathbf{T}_1) \triangle \Omega_R^1(\mathbf{T}_2)| = p_L(\mathbf{S}) - p_R(\mathbf{S})$ by the definition of a state. Observe that an immediate corollary of Theorem 8.10 is an enumeration algorithm for states.

Corollary 8.18. *One can in $O^\star(k^{O(\sqrt{k})})$ time enumerate a family $\mathbb{S}$ of $k^{O(\sqrt{k})} n^{636}$ states such that if (G, k) is a YES-instance of Interval Completion, then, for the canonical solution F and the canonical model σ of $G + F$, all states that appear in σ belong to $\mathbb{S}$.*

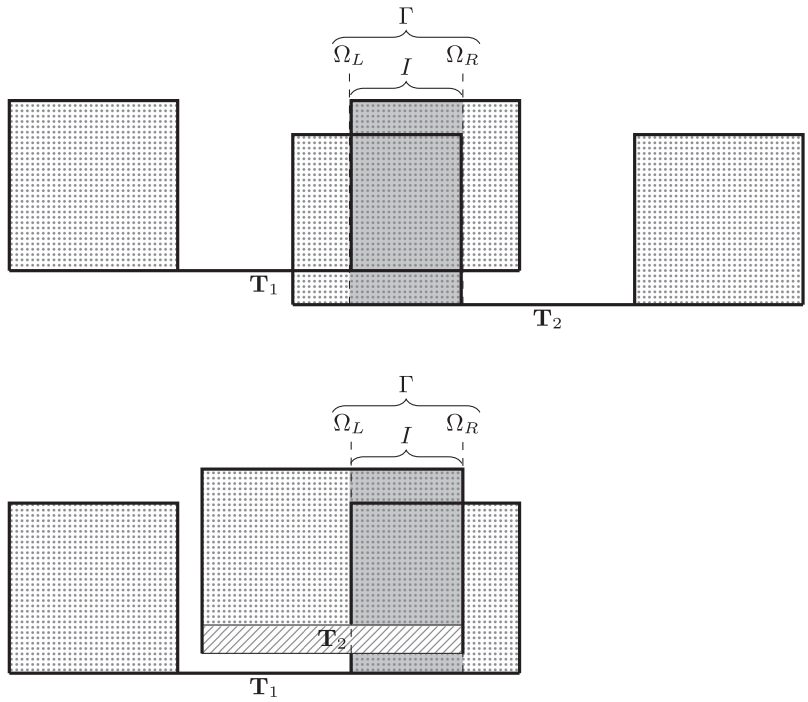

Fig. 25. A DP state defined by two nested terraces (above) and a nested terrace and a flat terrace (below). The DP state asks for the optimal way to arrange events in the gray area. Observe that the gray area is defined as an intersection of the *second* important area of the first terrace and the *first* important area of the second terrace. Furthermore, its borders are the *left* border of the *second* important area of the first terrace and the *right* border of the *first* important area of the second terrace.

8.3.2 DP Table. Thus, a state (similarly as a world and a terrace) describes which events of $\mathcal{E}(V(G))$ lie between positions $p_L(\mathbf{S})$ and $p_R(\mathbf{S})$. Moreover, there is only a subexponential number of reasonable states. However, contrary to worlds and terraces, the family of states is rich enough to allow us to perform dynamic programming on a table indexed by the family $\mathbb{S}$ of Corollary 8.18.

Formally, we say that a bijection $\pi : \mathcal{E}(\mathbf{S}) \to \{p_L(\mathbf{S}) + 1, p_L(\mathbf{S}) + 2, \ldots, p_R(\mathbf{S})\}$ is a *completion* of a state $\mathbf{S}$ if π, treated as a permutation of $\mathcal{E}(\mathbf{S})$, preceded with the starting events of $\Omega_L(\mathbf{S})$ and succeeded with the ending events of $\Omega_R(\mathbf{S})$ (in any order) is an interval model of $G[\Gamma(\mathbf{S})] + F_\pi$ for some completion F_π of $G[\Gamma(\mathbf{S})]$. With a completion π we associate a sequence $\pi(\varepsilon_1), \pi(\varepsilon_2), \ldots, \pi(\varepsilon_{|\mathcal{E}(\mathbf{S})|})$ where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{|\mathcal{E}(\mathbf{S})|}$ is the ordering of $\mathcal{E}(\mathbf{S})$ defined as follows: we first take all starting events of $\mathcal{E}(\mathbf{S})$, sorted according to $\prec$, and then all ending events of $\mathcal{E}(\mathbf{S})$, sorted according to reversed order $\prec$. For two completions π and π' of $\mathbf{S}$, we say that $\pi \lhd \pi'$ if

(1) $|F_\pi| < |F_{\pi'}|$, or
(2) $|F_\pi| = |F_{\pi'}|$ and $F_\pi \prec F_{\pi'}$, or
(3) $F_\pi = F_{\pi'}$ and the sequence associated with π is lexicographically smaller than the sequence associated with π'.

Note that $\trianglelefteq$ is a total order on completions of $\mathbf{S}$. For a state $\mathbf{S}$, we define $\pi^{\mathbf{S}}$ to be the $\trianglelefteq$-minimum completion of $\mathbf{S}$.

We also observe the following.

Lemma 8.19. *For any* S *that appears in the canonical model* σ*, we have* $\pi^{\mathrm{S}} = \sigma|_{\mathcal{E}(\mathrm{S})}$.

Proof. Clearly, $\pi := \sigma|_{\mathcal{E}(\mathrm{S})}$ is a completion of S and $F_\pi = F \cap \binom{\Gamma(\mathrm{S})}{2}$. Moreover, if we consider a model σ' defined as

$$\sigma' = \sigma|_{\mathcal{E}(V(G))\setminus\mathcal{E}(\mathrm{S})} \cup \pi^{\mathrm{S}},$$

then we obtain an interval model for $F' := (F \setminus F_\pi) \cup F_{\pi^{\mathrm{S}}}$. Observe the following:

(1) $|F_{\pi^{\mathrm{S}}}| \leq |F_\pi|$ by the minimality of π^{S}, whereas if $|F_\pi| > |F_{\pi^{\mathrm{S}}}|$, then $|F'| < |F|$, contradicting the minimality of F; hence, $|F_\pi| = |F_{\pi^{\mathrm{S}}}|$.
(2) $F_{\pi^{\mathrm{S}}} \preceq F_\pi$ by the minimality of π^{S}, whereas if $F_\pi \succ F_{\pi^{\mathrm{S}}}$, then $F' \prec F$, contradicting the fact that F is canonical; hence, $F_\pi = F_{\pi^{\mathrm{S}}}$ and $F' = F$.
(3) The sequence associated with π^{S} is lexicographically not larger than the sequence associated with π, whereas, if it would be lexicographically strictly smaller, then σ' would be a lexicographically smaller model than σ, contradicting the fact that σ is the canonical model of $G + F$. Hence, $\pi^{\mathrm{S}} = \pi$. □

In our dynamic programming algorithm, for each $\mathrm{S} \in \mathbb{S}$, we compute a value $M[\mathrm{S}]$ that is a completion of S. While we will not necessarily obtain $M[\mathrm{S}] = \pi^{\mathrm{S}}$ for every state $\mathrm{S} \in \mathbb{S}$, we will ensure that this equality is true for each S that appears in the canonical model σ.

8.3.3 DP Computation. We now proceed to the description of computation of $M[\mathrm{S}]$ for $\mathrm{S} \in \mathbb{S}$. In the base case, if $|\mathcal{E}(\mathrm{S})| \leq 4\sqrt{k} + 4$, we find $M[\mathrm{S}] = \pi^{\mathrm{S}}$ by brute-force in $O^\star(k^{O(\sqrt{k})})$ time by trying all possible bijections.

Consider now a state S where $|\mathcal{E}(\mathrm{S})| > 4\sqrt{k}$. We claim that the family of sets is rich enough so that we can compute $M[\mathrm{S}]$ by "gluing" the solution of at most three substates.

More formally, to compute $M[\mathrm{S}]$ we iterate through all possible choices of sequences $(\mathrm{S}^i)_{i=1}^s$ for $s = 2, 3$ where

(1) $p_L(\mathrm{S}^1) = p_L(\mathrm{S})$ and $\Omega_L(\mathrm{S}^1) = \Omega_L(\mathrm{S})$,
(2) $p_R(\mathrm{S}^s) = p_R(\mathrm{S})$ and $\Omega_R(\mathrm{S}^s) = \Omega_R(\mathrm{S})$,
(3) $p_R(\mathrm{S}^i) = p_L(\mathrm{S}^{i+1})$ and $\Omega_R(\mathrm{S}^i) = \Omega_L(\mathrm{S}^{i+1})$ for each $1 \leq i < s$,
(4) $\mathcal{E}(\mathrm{S}) = \biguplus_{i=1}^s \mathcal{E}(\mathrm{S}^i)$, and
(5) $p_R(\mathrm{S}^i) - p_L(\mathrm{S}^i) < p_R(\mathrm{S}) - p_L(\mathrm{S})$ for each $1 \leq i \leq s$.

For each such sequence, we consider a candidate permutation π defined as a union (concatenation) of permutations $(M[\mathrm{S}^i])_{i=1}^s$. As $M[\mathrm{S}]$, we chose the permutation π which is $\trianglelefteq$-minimum among all considered permutations that are completions of S. Note that the last condition for the states S^i ensures that, if we compute $M[\mathrm{S}]$ in the order of increasing value $p_R(\mathrm{S}) - p_L(\mathrm{S})$, then in the computation we use already known values of $M[\mathrm{S}^i]$ for $1 \leq i \leq s$.

If no candidate completion of S is found, we pick any permutation of $M[\mathrm{S}]$; as we shall see in the next lemma, such a state S cannot appear in the canonical model σ.

Lemma 8.20. *For any* S *that appears in the canonical model* σ*, we have*

$$M[\mathrm{S}] = \sigma|_{\mathcal{E}(\mathrm{S})} = \pi^{\mathrm{S}}.$$

Proof. The second equality is due to Lemma 8.19. We prove that $M[\mathrm{S}] = \sigma|_{\mathcal{E}(\mathrm{S})}$ for any state S that appears in σ, by induction on $|\mathcal{E}(\mathrm{S})| = p_R(\mathrm{S}) - p_L(\mathrm{S})$. Note that $M[\mathrm{S}]$ is defined via the same minimization condition as π^{S} but on a smaller family of permutations, so it suffices to prove that $\sigma|_{\mathcal{E}(\mathrm{S})}$ is among the candidate permutations considered when computing $M[\mathrm{S}]$. For states where $|\mathcal{E}(\mathrm{S})| \leq 4\sqrt{k} + 4$ this is clearly true, as the brute-force algorithm in fact considers all the possible candidate permutations.

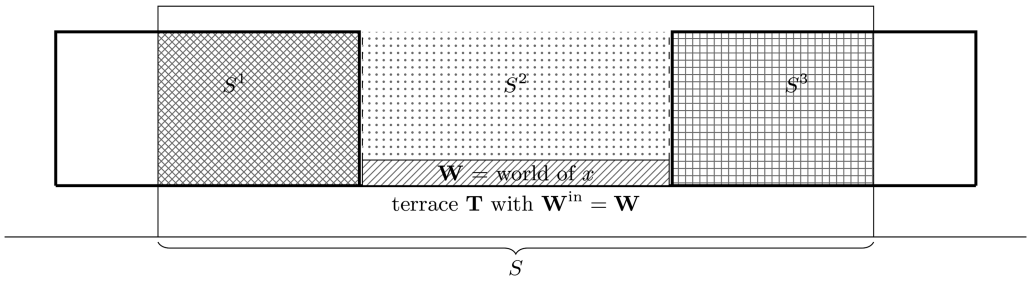

Fig. 26. A computation of the value for DP state S when $x \in I(\mathrm{S})$ and we glue values from three substates.

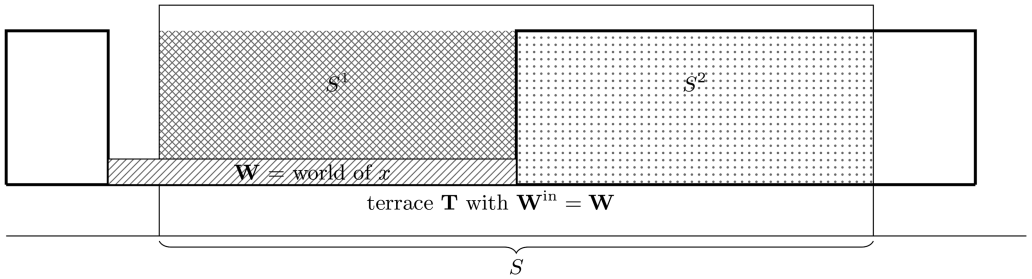

Fig. 27. A computation of the value for DP state S when $x \notin I(\mathrm{S})$ and we glue values from two substates.

Consider then $\mathrm{S} = (\mathbf{T}_1, \mathbf{T}_2)$ with $|\mathcal{E}(\mathrm{S})| > 4\sqrt{k} + 4$. Observe that in this case we have at least three vertices $x \in (\Omega_L(\mathrm{S}) \triangle \Omega_R(\mathrm{S})) \cup I(\mathrm{S})$ that are cheap w.r.t. F. Pick one such vertex with maximum possible value of

$$\min(\sigma(\omega_x), p_R(\mathrm{S}) + 1) - \max(\sigma(\alpha_x), p_L(\mathrm{S})). \tag{8.1}$$

In case of a tie, we prefer x belonging to $I(\mathrm{S})$.

We consider two cases: whether $x \in I(\mathrm{S})$ or not. If $x \in I(\mathrm{S})$, consider the flat terrace $\mathbf{T}_f = \mathbf{W}(\sigma, x)$ and the nested terrace $\mathbf{T}_n = \mathbf{T}(\sigma, x)$, with vertices $y_1 = y_1(\sigma, x)$ and $y_2 = y_2(\sigma, x)$ (see Figure 26). Observe that, by the choice of x, we have

$$\sigma(\alpha_{y_2}) \leq \sigma(\alpha_{y_1}) \leq p_L(\mathrm{S}) = p_L^2(\mathbf{T}_1) < \sigma(\alpha_x) < \sigma(\omega_x) \leq p_R(\mathrm{S}) = p_R^1(\mathbf{T}_2) < \sigma(\omega_{y_2}) \leq \sigma(\omega_{y_1}). \tag{8.2}$$

That is, the claim that $\sigma(\alpha_{y_i}) \leq p_L(\mathrm{S}) < p_R(\mathrm{S}) < \sigma(\omega_{y_i})$ for $i = 1, 2$ follows from Equation (8.1) in the choice of x, since otherwise y_i would be a better candidate for x. Consider now states $\mathrm{S}^1 = (\mathbf{T}_1, \mathbf{T}_n)$, $\mathrm{S}^2 = (\mathbf{T}_f, \mathbf{T}_f)$ and $\mathrm{S}^3 = (\mathbf{T}_n, \mathbf{T}_2)$. From Equation (8.2), we infer that

$$p_L(\mathrm{S}) = p_L(\mathrm{S}^1) < p_R(\mathrm{S}^1) = p_L(\mathrm{S}^2) \leq p_R(\mathrm{S}^2) = p_L(\mathrm{S}^3) < p_R(\mathrm{S}^3) = p_R(\mathrm{S}),$$

and, consequently, the last condition for considering states $(\mathrm{S}^i)_{i=1}^3$ holds. A direct check shows that these three states appear in σ, and the algorithm indeed considers concatenating $M[\mathrm{S}^1]$, $M[\mathrm{S}^2]$, and $M[\mathrm{S}^3]$ to obtain $M[\mathrm{S}]$. By induction hypothesis, $M[\mathrm{S}^i] = \sigma|_{\mathcal{E}(\mathrm{S}^i)}$ for $i = 1, 2, 3$ and the inductive claim follows in this case.

In the second case, without loss of generality assume that $x \in \Omega_L(\mathrm{S}) \setminus \Omega_R(\mathrm{S})$ (see Figure 27). Note that, by the criterion (8.1), x is such a cheap vertex with maximum $\sigma(\omega_x)$. Consider the flat terrace $\mathbf{T}_f = \mathbf{W}(\sigma, x)$ and the nested terrace $\mathbf{T}_n = \mathbf{T}(\sigma, x)$, with vertices $y_1 = y_1(\sigma, x)$ and $y_2 = y_2(\sigma, x)$. Observe that, by the choice of x, we have

$$\sigma(\alpha_{y_2}) \leq \sigma(\alpha_{y_1}) < \sigma(\alpha_x) \leq p_L(\mathrm{S}) = p_L^2(\mathbf{T}_1) < \sigma(\omega_x) \leq p_R(\mathrm{S}) = p_R^1(\mathbf{T}_2) < \sigma(\omega_{y_2}) \leq \sigma(\omega_{y_1}). \tag{8.3}$$

That is, the inequality $p_R(\mathbf{S}) < \sigma(\omega_{y_2})$ follows from the choice of rightmost possible $\sigma(\omega_x)$. Consider now states $\mathbf{S}^1 = (\mathbf{T}_1, \mathbf{T}_f)$, $\mathbf{S}^2 = (\mathbf{T}_n, \mathbf{T}_2)$. Using Equation (8.3), we observe that, unless $\sigma(\omega_x) = p_L(\mathbf{S}) + 1$, we have that

$$p_L(\mathbf{S}) = p_L(\mathbf{S}^1) < p_R(\mathbf{S}^1) = p_L(\mathbf{S}^2) < p_R(\mathbf{S}^2) = p_R(\mathbf{S}).$$

However, if $\sigma(\omega_x) = p_L(\mathbf{S}) + 1$, then the value of Equation (8.1) for the vertex x equals 1, and is the minimum possible. There can be at most one such $x \in \Omega_L(\mathbf{S}) \setminus \Omega_R(\mathbf{S})$ and at most one such $x \in \Omega_R(\mathbf{S}) \setminus \Omega_L(\mathbf{S})$. Since there are at least three cheap vertices in $(\Omega_L(\mathbf{S}) \triangle \Omega_R(\mathbf{S})) \cup I(\mathbf{S})$, we infer that there exists one such $x' \in I(\mathbf{S})$. As the value of Equation (8.1) for x' is at least 1, this contradicts the tie-breaking rule in the choice of x.

A direct check shows that both $\mathbf{S}^1$ and $\mathbf{S}^2$ appear in σ, and the algorithm considers concatenating $M[\mathbf{S}^1]$ with $M[\mathbf{S}^2]$ to obtain $M[\mathbf{S}]$. By induction hypothesis, $M[\mathbf{S}^i] = \sigma|_{\mathcal{E}(\mathbf{S}^i)}$ for $i = 1, 2$ and the inductive claim follows in this case as well. This concludes the proof of Lemma 8.20. □

We now observe that the world $\mathbf{W}_\mathfrak{r} := \mathbf{W}(\sigma, \mathfrak{r})$ is easy to guess:

$$\begin{aligned} v(\mathbf{W}_\mathfrak{r}) &= \mathfrak{r}, & F_v(\mathbf{W}_\mathfrak{r}) &= \emptyset, \\ p_L(\mathbf{W}_\mathfrak{r}) &= 1, & p_R(\mathbf{W}_\mathfrak{r}) &= 2n - 1, \\ \Omega_L(\mathbf{W}_\mathfrak{r}) &= \{\mathfrak{r}\}, & \Omega_R(\mathbf{W}_\mathfrak{r}) &= \{\mathfrak{r}\}. \end{aligned}$$

Hence, we may proceed as follows: we compute the table M, read the cell $M[\mathbf{S}(\mathbf{W}_\mathfrak{r}, \mathbf{W}_\mathfrak{r})]$, and add the events $\alpha_\mathfrak{r}$ and $\omega_\mathfrak{r}$ before and after the permutation found in this cell. By Lemma 8.20, if (G, k) is a YES-instance, the obtained permutation is the canonical model for $G + F$ where F is the canonical solution to (G, k). This concludes the proof of Theorem 1.1.

9 CONCLUSIONS

We would like to conclude our article with two suggestions for future research. First, in the light of our techniques, the question for a polynomial kernel for Interval Completion is appealing. We think that the techniques developed in our work to cope with the lack of a kernel, in some sense being local kernelization arguments, can help with obtaining an affirmative answer to this question. The question if Interval Completion admits a polynomial kernel is important from practical considerations too. Although the running time of our algorithm is subexponential in k, so far our result is mainly of theoretical importance due to the high degree polynomial of n. This is why the most promising approach to significantly reduce the polynomial dependency on n is to actually develop a polynomial kernel for Interval Completion. A polynomial kernel for IC would also significantly reduce the exponent in the running time by making the arguments of Section 6 obsolete. Needless to say, the argumentation of Sections 5 and 6 could be tremendously simplified if such a polynomial kernel was at our disposal. We remark here that it is also possible that the very recent techniques of Cao [8, 9], that lead to a linear dependency on the size of the graph in the "forbidden subgraph" branching algorithm, may help decrease the dependency on the size of the graph in our algorithm.

For the second suggestion, we observe that except for the case of proper interval graphs, the obtained subexponential parameterized algorithms for completion problems to graph classes present in Figure 1 run in time $k^{O(\sqrt{k})}n^{O(1)}$. As an algorithm with running time bound $2^{o(\sqrt{k})}n^{O(1)}$ would actually be a $2^{o(n)}$-time algorithm, we suspect that $2^{O(\sqrt{k})}$ or even $k^{O(\sqrt{k})}$ may be the best possible dependency on k in the running time for these problems. In an unpublished manuscript, Bliznets et al. [3] corroborate this suspicion by linking the question with the approximability of the Min Bisection problem. However, proving appropriate lower bounds under more widely believed conjectures, like the Exponential Time Hypothesis (ETH), is still open. Another, perhaps more modest

question can be asked about the necessity of the $\log k$ factor in the exponent. Maybe it is possible to solve the completion problem to at least one of the graph classes in Figure 1 within running time $2^{O(\sqrt{k})}n^{O(1)}$, thus shaving off this factor.

APPENDIX

Proof of Lemma 2.1. Without loss of generality, assume that Ω_1 and Ω_2 are non-empty, as otherwise we may with polynomial overhead guess the first or the last event of the model.

First observe that if G is disconnected, but Ω_1 and Ω_2 are in the same connected component of G, then clearly no such interval model of G exists, as any interval model of G needs to arrange connected components of G one-by-one. Hence, assume in the rest of the proof that either G is connected or Ω_1 and Ω_2 are contained in two different connected components of G. Let C_1 be the connected component containing Ω_1 and C_2 the one containing Ω_2.

Consider a graph H created from G by adding two 3-vertex paths x_1, x_2, x_3 and y_1, y_2, y_3 and making x_1 fully adjacent to Ω_1 and y_1 fully adjacent to Ω_2. We claim that there exists an interval model of G as requested in the statement of the lemma if and only if H is an interval graph. Observe that such a claim would finish the proof of the lemma, as H can be constructed in linear time.

In one direction, consider the model σ of G as in the statement of the lemma. Precede the ordering σ with events $\alpha_{x_3}, \alpha_{x_2}, \omega_{x_3}, \alpha_{x_1}, \omega_{x_2}$ and insert the event ω_{x_1} immediately after all starting events of $\mathcal{E}(\Omega_1)$. Symmetrically, succeed the ordering σ_C with events $\alpha_{y_2}, \omega_{y_1}, \alpha_{y_3}, \omega_{y_2}, \omega_{y_3}$ and insert the event α_{y_1} immediately before all ending events of $\mathcal{E}(\Omega_2)$. It is straightforward to verify that this is an interval model of the graph H.

In the other direction, let σ be an interval model of H and consider events α_{x_2} and ω_{x_2}. Observe that if $\sigma(\alpha_{x_1}) < \sigma(\alpha_{x_2})$ and simultaneously $\sigma(\omega_{x_2}) < \sigma(\omega_{x_1})$ (i.e., the interval of x_1 contains the interval of x_2), then there is no place to put the endpoints of x_3 into the model, as $x_1x_3 \notin E(H)$ but $x_2x_3 \in E(H)$. Consequently, either $\sigma(\alpha_{x_2}) < \sigma(\alpha_{x_1}) < \sigma(\omega_{x_2}) < \sigma(\omega_{x_1})$ (case (1.i)) or $\sigma(\alpha_{x_1}) < \sigma(\alpha_{x_2}) < \sigma(\omega_{x_1}) < \sigma(\omega_{x_2})$ (case (1.ii)). Assume first that the case (1.i) happens. As x_1 is adjacent to x_2 and to every vertex of Ω_1, but no vertex of $V(G)$ is adjacent to x_2, we infer that the events between ω_{x_2} and ω_{x_1} in the model σ are first all starting events of $\mathcal{E}(\Omega_1)$ and then possibly some ending events of $\mathcal{E}(\Omega_1)$, and, moreover, all other events of $\mathcal{E}(C_1)$ appear in σ to the right of ω_{x_1}. Consequently, the model σ, restricted to $\mathcal{E}(C_1)$, starts with the starting events of $\mathcal{E}(\Omega_1)$. Observe that in the case (1.ii), that is, $\sigma(\alpha_{x_1}) < \sigma(\alpha_{x_2}) < \sigma(\omega_{x_1}) < \sigma(\omega_{x_2})$, we obtain the symmetric conclusion: the model σ, restricted to $\mathcal{E}(C_1)$, ends with the ending events of $\mathcal{E}(\Omega_1)$.

An analogous reasoning can be made for the path y_1, y_2, y_3; let us denote the respective cases (2.i) and (2.ii). Consider first the case when $C_1 = C_2 = V(G)$ and G is connected, and examine the model σ restricted to $\mathcal{E}(C_1) = \mathcal{E}(C_2) = \mathcal{E}(V(G))$. From our study, we infer that this model starts with all the starting events of $\mathcal{E}(\Omega_1)$ providing that (1.i) happens, or with all the starting events of $\mathcal{E}(\Omega_2)$ providing that (2.i) happens. Moreover, this model ends with all the ending events of $\mathcal{E}(\Omega_1)$ providing that (1.ii) happens, or with all the ending events of $\mathcal{E}(\Omega_2)$ providing that (2.ii) happens. Observe, however, that if (1.i) and (2.i) happened simultaneously, then the first event of σ restricted to $\mathcal{E}(V(G))$ would be α_v for some $v \in \Omega_1 \cap \Omega_2$. In this case, we would have $\alpha_{x_1} < \alpha_v < \omega_{x_1}$ and $\alpha_{y_1} < \alpha_v < \omega_{y_1}$, which means that the intervals of x_1 and y_1 would overlap, contradicting the fact that x_1 and y_1 are not adjacent in H. Similarly, (1.ii) and (2.ii) cannot happen simultaneously. Since either (1.i) or (1.ii) happens, and either (2.i) or (2.ii) happens, we infer that either ((1.i) and (2.ii)) happens, or ((1.ii) and (2.i)) happens. In case ((1.i) and (2.ii)) we are already done, since σ restricted to $\mathcal{E}(V(G))$ has exactly the desired property. In case ((1.ii) and (2.i)) it suffices to revert the model σ restricted to $\mathcal{E}(V(G))$.

Examine now the case when $C_1 \neq C_2$. Consider model σ' of $V(G)$ constructed from σ by the following reshuffling of connected components of G: We first place the model of C_1, possibly reversing

it if (1.ii) happened instead of (1.i). Then we arrange the models of all the connected components of G other than C_1, C_2 in any order. Finally, we place the model of C_2, possibly reversing it if (2.i) happened instead of (2.ii). It is straightforward to see that this model of G has the desired property. □

Proof of Lemma 2.3. Assume otherwise, and let i be the smallest index such that $x_i \prec x_{i-1}$. Denote $p = \alpha_\sigma(C_i)$. As $i > 1$ and $N_G(v) \setminus C_j = X$ for every $1 \leq j \leq s$ and $v \in C_j$, we have that $\Omega_\sigma(p-1) = X$.

Consider a model σ' of G that is constructed as follows:

(1) First, we take all events of $\sigma^{-1}(\{1, 2, \ldots, p-1\}) \setminus \mathcal{E}(C_{i-1})$, in the order as they appear in σ.
(2) Second, we take all events of $\mathcal{E}(C_i)$, in the order as they appear in σ.
(3) Third, we take all events of $\mathcal{E}(C_{i-1})$, in the order as they appear in σ.
(4) Finally, we take all events of $\sigma^{-1}(\{p, p+1, \ldots, 2n\}) \setminus \mathcal{E}(C_i)$, in the order as they appear in σ.

A direct check shows that σ' is an interval model of G. We now claim the following: for every vertex $u \notin C_{i-1}$, we have $\sigma'(\alpha_u) \leq \sigma(\alpha_u)$. This claim is trivial for the vertices $u \in C_i$, and for the vertices $u \notin C_{i-1}$ with $\sigma(\alpha_u) < p$. Consider then any vertex $u \notin C_{i-1}$ such that $\sigma(\alpha_v) \geq p$. Since $i > 1$ and $N_G(v) \setminus C_1 = X$ for every $v \in C_1$, we infer that all the vertices of X have starting events before position p in σ, and hence $u \notin X$. Therefore, $u \notin N_G(C_i)$, so in fact $\sigma(\alpha_u) > \omega_\sigma(C_i)$. By the definition of σ', we infer that $\sigma(\alpha_u) = \sigma'(\alpha_u)$, and the claim is proven.

Now observe that

- $\sigma'(\alpha_v) \leq \sigma(\alpha_v)$ for any $v \preceq x_i$, as only for vertices $v \in C_{i-1}$ is it possible that $\sigma'(\alpha_v) > \sigma(\alpha_v)$ and all vertices of C_{i-1} are at least as late as $x_{i-1} \succ x_i$ in the order $\prec$;
- $\sigma'(\alpha_{x_i}) < \sigma(\alpha_{x_i})$, since C_{i-1} is non-empty.

Hence, σ is not the canonical model and the lemma is proven. □

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 267959. An extended abstract of this work has been presented at SODA 2016 [31].

REFERENCES

[1] Noga Alon, Daniel Lokshtanov, and Saket Saurabh. 2009. Fast FAST. In *ICALP 2009*, Lecture Notes in Computer Science, Vol. 5555. Springer, 49–58.
[2] Stephane Bessy and Anthony Perez. 2013. Polynomial kernels for proper interval completion and related problems. *Information and Computation* 231, 0 (2013), 89–108. DOI: http://dx.doi.org/10.1016/j.ic.2013.08.006
[3] Ivan Bliznets, Marek Cygan, Paweł Komosa, Lukás Mach, and Michał Pilipczuk. 2016. Lower bounds for the parameterized complexity of minimum fill-in and other completion problems. In *SODA 2016*. SIAM, 1132–1151.
[4] Ivan Bliznets, Fedor V. Fomin, Marcin Pilipczuk, and Michał Pilipczuk. 2015. A subexponential parameterized algorithm for proper interval completion. *SIAM Journal on Discrete Mathematics* 29, 4 (2015), 1961–1987.
[5] Hans L. Bodlaender. 1998. A partial k-arboretum of graphs with bounded treewidth. *Theoretial Computer Science* 209, 1–2 (1998), 1–45.
[6] Andreas Brandstädt, Van Bang Le, and Jeremy P. Spinrad. 1999. *Graph Classes. A Survey*. SIAM.
[7] Yixin Cao. 2013. An efficient branching algorithm for interval completion. *CoRR* abs/1306.3181 (2013).
[8] Yixin Cao. 2014. Linear recognition of almost interval graphs. *CoRR* abs/1403.1515 (2014).
[9] Yixin Cao. 2016. Linear recognition of almost interval graphs. In *SODA 2016*. SIAM, 1096–1115.
[10] Erik D. Demaine, Fedor V. Fomin, Mohammadtaghi Hajiaghayi, and Dimitrios M. Thilikos. 2005. Subexponential parameterized algorithms on graphs of bounded genus and H-minor-free graphs. *Journal of the ACM* 52, 6 (2005), 866–893.

[11] Pål Grønås Drange, Fedor V. Fomin, Michał Pilipczuk, and Yngve Villanger. 2015. Exploring the subexponential complexity of completion problems. *ACM Transactions on Computation Theory (TOCT)* 7, 4 (2015), 14:1–14:38.
[12] Uriel Feige. 2000. Coping with the NP-hardness of the graph bandwidth problem. In *SWAT 2000* Lecture Notes in Computer Science, Vol. 1851. Springer, 10–19.
[13] Jörg Flum and Martin Grohe. 2006. *Parameterized Complexity Theory*. Springer-Verlag, Berlin. 493 pages.
[14] Fedor V. Fomin, Stefan Kratsch, Marcin Pilipczuk, Michał Pilipczuk, and Yngve Villanger. 2014. Tight bounds for parameterized complexity of cluster editing with a small number of clusters. *Journal of Computer and System Sciences* 80, 7 (2014), 1430–1447.
[15] Fedor V. Fomin and Yngve Villanger. 2013. Subexponential parameterized algorithm for minimum fill-in. *SIAM Journal on Computing* 42, 6 (2013), 2197–2216.
[16] T. Gallai. 1967. Transitiv orientierbare graphen. *Acta Mathematica Academiae Scientiarum Hungarica* 18, 1–2 (1967), 25–66. DOI: http://dx.doi.org/10.1007/BF02020961
[17] Michael R. Garey and David S. Johnson. 1979. *Computers and Intractability, A Guide to the Theory of NP-Completeness*. W.H. Freeman and Company, New York.
[18] Esha Ghosh, Sudeshna Kolay, Mrinal Kumar, Pranabendu Misra, Fahad Panolan, Ashutosh Rai, and M. S. Ramanujan. 2015. Faster parameterized algorithms for deletion to split graphs. *Algorithmica* 71, 4 (2015), 989–1006. DOI: http://dx.doi.org/10.1007/s00453-013-9837-5
[19] Norman E. Gibbs, William G. Poole, Jr., and Paul K. Stockmeyer. 1976. A comparison of several bandwidth and profile reduction algorithms. *ACM Transactions on Mathematical Softwaaare* 2, 4 (Dec. 1976), 322–330. DOI: http://dx.doi.org/10.1145/355705.355707
[20] P. W. Goldberg, M. C. Golumbic, H. Kaplan, and R. Shamir. 1995. Four strikes against physical mapping of DNA. *Journal of Computational Biology* 2, 1 (1995), 139–152.
[21] M. C. Golumbic. 1980. *Algorithmic Graph Theory and Perfect Graphs*. Academic Press, New York.
[22] Pinar Heggernes, Christophe Paul, Jan Arne Telle, and Yngve Villanger. 2007. Interval completion with few edges. In *STOC 2007*. 374–381.
[23] Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. 2001. Which problems have strongly exponential complexity? *Journal of Computer and System Sciences* 63, 4 (2001), 512–530.
[24] Haim Kaplan, Ron Shamir, and Robert Endre Tarjan. 1994. Tractability of parameterized completion problems on chordal and interval graphs: Minimum fill-in and physical mapping. In *FOCS 1994*. IEEE Computer Society, 780–791.
[25] Haim Kaplan, Ron Shamir, and Robert E. Tarjan. 1999. Tractability of parameterized completion problems on chordal, strongly chordal, and proper interval graphs. *SIAM Journal on Computing* 28, 5 (May 1999), 1906–1922. DOI: http://dx.doi.org/10.1137/S0097539796303044
[26] Richard M. Karp. Mapping the genome: Some combinatorial problems arising in molecular biology. In *STOC 1993*. ACM, 278–285.
[27] Christian Komusiewicz and Johannes Uhlmann. 2012. Cluster editing with locally bounded modifications. *Discrete Applied Mathematics* 160, 15 (2012), 2259–2270.
[28] Ross M. McConnell and Jeremy Spinrad. 1999. Modular decomposition and transitive orientation. *Discrete Mathematics* 201, 1–3 (1999), 189–241.
[29] Jaroslav Nešetřil and Patrice Ossona de Mendez. 2012. *Sparsity - Graphs, Structures, and Algorithms*. Algorithms and Combinatorics, Vol. 28. Springer. I–XXIII, 1–457 pages.
[30] Yngve Villanger, Pinar Heggernes, Christophe Paul, and Jan Arne Telle. 2009. Interval completion is fixed parameter tractable. *SIAM Journal on Computing* 38, 5 (2009), 2007–2020.
[31] Ivan Bliznets, Fedor V. Fomin, Marcin Pilipczuk, and Michal Pilipczuk. 2016. Subexponential parameterized algorithm for Interval Completion. *SODA 2016*. SIAM, 1116–1131.

Received February 2017; revised October 2017; accepted February 2018