Extremal Separation Problems for Temporal Instance Queries

Jean Christoph Jung[1] , **Vladislav Ryzhikov**[2] , **Frank Wolter**[3] and **Michael Zakharyaschev**[2]

[1]Department of Computer Science, TU Dortmund University, Germany
[2]School of Computing and Mathematical Sciences, Birkbeck, University of London, UK
[3]Department of Computer Science, University of Liverpool, UK

jean.jung@tu-dortmund.de, {v.ryzhikov,m.zakharyaschev}@bbk.ac.uk, wolter@liverpool.ac.uk

Abstract

The separation problem for a class $\mathcal{Q}$ of database queries is to find a query in $\mathcal{Q}$ that distinguishes between a given set of 'positive' and 'negative' data examples. Separation provides explanations of examples and underpins the query-by-example paradigm to support database users in constructing and refining queries. As the space of all separating queries can be large, it is helpful to succinctly represent this space by means of its most specific (logically strongest) and general (weakest) members. We investigate this extremal separation problem for classes of instance queries formulated in linear temporal logic *LTL* with the operators conjunction, 'next', and 'eventually'. Our results range from tight complexity bounds for verifying and counting extremal separators to algorithms computing them.

1 Introduction

The separation (aka fitting or consistency) problem for a class $\mathcal{Q}$ of queries is to find some $q \in \mathcal{Q}$ that separates a given set $E = (E^+, E^-)$ of positive and negative data examples in the sense that $\mathcal{D} \models q$ for all $\mathcal{D} \in E^+$, and $\mathcal{D} \not\models q$ for all $\mathcal{D} \in E^-$. Separation underpins the query-by-example approach, which aims to support database users in constructing queries and schema mappings with the help of data examples [Alexe *et al.*, 2011; Martins, 2019], inductive logic programming [Cropper *et al.*, 2022], and, more recently, automated feature extraction, where separating queries are proposed as features in classifier engineering [Kimelfeld and Ré, 2018; Barceló *et al.*, 2021]. Separating queries (and, more generally, formulas) also underpin recent logic-based approaches aiming to explain positive and negative data examples given by applications [Sarker *et al.*, 2017; Camacho and McIlraith, 2019; Raha *et al.*, 2022].

The space of all separating queries, denoted $\mathsf{s}(E, \mathcal{Q})$, forms a convex subset of $\mathcal{Q}$ under the containment (or logical entailment) relation $q \models q'$ between queries q, q', and is an instance of the more general version spaces [Mitchell, 1982]. If finite, $\mathsf{s}(E, \mathcal{Q})$ can be represented by its extremal elements: the most specific (logically strongest) and most general (logically weakest) separators in $\mathcal{Q}$. In fact, many known algorithms check the existence of separators by looking for a most specific one [ten Cate and Dalmau, 2015; Barceló and Romero, 2017; Gutiérrez-Basulto *et al.*, 2018; Funk *et al.*, 2019]. This is not surprising as query classes are often closed under $\wedge$, and so the conjunction of all separators gives the unique most specific one. Dually, the unique most general separator is given by the disjunction of all separators if the query class is closed under $\vee$, which is a less common assumption. For the case of first-order queries constructed using $\wedge$ and existential quantifiers (conjunctive queries or CQs), a systematic study of extremal separation has recently been conducted in the award winning [ten Cate *et al.*, 2023].

Here, we study extremal separation for temporal instance queries. Data instances take the form $(\delta_0, \ldots, \delta_n)$ describing temporal evolutions, where the δ_i are the sets of atomic propositions that are true at time i. Queries are formulated in the fragment of linear temporal logic *LTL* with the operators $\wedge$, $\bigcirc$ (next), and $\Diamond$ (eventually). These $\bigcirc\Diamond$-*queries* and its subclass of $\Diamond$-*queries* without $\bigcirc$, are obtained by restricting CQs to propositional temporal data and form the core of most temporal query languages proposed in the database and knowledge representation literature [Chomicki and Toman, 2018; Baader *et al.*, 2015; Borgwardt *et al.*, 2015; Artale *et al.*, 2021]. We are particularly interested in subclasses of the classes of $\bigcirc\Diamond$-and $\Diamond$-queries that are not closed under $\wedge$ and take the form of path queries:

$$q = \rho_0 \wedge \boldsymbol{o}_1(\rho_1 \wedge \boldsymbol{o}_2(\rho_2 \wedge \cdots \wedge \boldsymbol{o}_n \rho_n)) \qquad (1)$$

with $\boldsymbol{o}_i \in \{\bigcirc, \Diamond\}$ and conjunctions ρ_i of atoms. The temporal patterns expressed by path queries correspond to common subsequences, subwords, and combinations thereof, which have been investigated in the string pattern matching literature for more than 50 years. Their applications range from computational linguistics to bioinformatics and revision control systems [Bergroth *et al.*, 2000; Chowdhury *et al.*, 2010; Abboud *et al.*, 2015; Blum *et al.*, 2021]. In fact, our results can also be interpreted and applied in that research tradition and our techniques combine both logic and automata-based methods with pattern matching. We note that not admitting $\vee$ in our query languages is crucial for finding this type of patterns and adding it would often trivialise separation.

Example 1. Suppose the first two sequences of events shown below are 'positive', the third one is 'negative', and our task is to explain this phenomenon using path queries. The space of
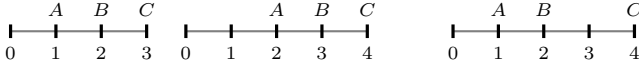
possible explanations includes $q_1 = \Diamond(A \wedge \bigcirc(B \wedge \bigcirc C))$, $q_2 = \Diamond(A \wedge \bigcirc\bigcirc C)$ and $q_3 = \Diamond(B \wedge \bigcirc C)$, all of which are true at 0 in the positive examples and false in the negative one. In fact, q_1 is the unique most specific explanation, while q_2 and q_3 are the non-equivalent most general ones. On the other hand, there exists no explanation in terms of $\Diamond$ only.

The (bounded-size) separator existence problem for various classes $\mathcal{Q}$ of *LTL*-queries has recently been studied in [Fijalkow and Lagarde, 2021; Raha *et al.*, 2022; Fortin *et al.*, 2023]. Our aim here is to investigate systematically the separator spaces $\mathsf{s}(E, \mathcal{Q})$ by determining the complexity of verifying and counting most specific and general separators and giving algorithms computing them. On our way, we determine the complexity of entailment (aka containment in database theory) between queries and computing weakening and strengthening frontiers, which are the key to understanding $\mathsf{s}(E, \mathcal{Q})$. Intuitively, a weakening/strengthening frontier of $q \in \mathcal{Q}$ is a set of queries properly weaker/stronger than q that form a boundary between q and all of its weakenings/strengthenings in $\mathcal{Q}$.

In detail, we first prove that query containment is in P for the class of $\Diamond$-queries and all of our classes of path queries. Based on this result, we show that strengthening and weakening frontiers can be computed in polytime for path queries. This is also the case for $\Diamond$-queries and weakening frontiers but not for strengthening ones. It follows that checking whether a path query is a most specific/general separator and whether a $\Diamond$-query is a most general one are both in P. In contrast, we establish CONP-completeness of checking whether a $\Diamond$-query is a most specific separator and whether a path query is the unique most specific/general one. Using frontiers, we show for path queries that the existence of unique most specific/general separators is in the complexity class US (for which unique SAT is complete) and that counting the number of most general/specific separators is in $\sharp$P. We show that these upper bounds are tight, sometimes using the rich literature on algorithms for sequences (e.g., longest common subsequences). Our lower bounds mostly require only a bounded number of atomic propositions and an unbounded number of either negative or positive examples.

These complexity results are complemented with algorithms for computing extremal separators in the majority of our query classes. The algorithms use a graph encoding of the input example set, associating (extremal) separators with certain paths in the graph. Complexity-wise, they are optimal, running in polytime if the number of examples is bounded and exponential time otherwise.

1.1 Related Work

Being inspired by the investigation of extremal separation for first-order conjunctive queries (CQs) [ten Cate *et al.*, 2023], our results turn out to be very different. For instance, while separability by CQs is NEXPTIME-complete and separating queries are exponential in the size of the examples (the extremal ones even larger), the extremal separation problems for *LTL*-queries are often complete for SAT-related complex-

ity classes, with separating queries being of polynomial size. For work on separation in the query-by-example paradigm we refer the reader to [Zhang *et al.*, 2013; Weiss and Cohen, 2017; Kalashnikov *et al.*, 2018; Deutch and Gilad, 2019; Staworko and Wieczorek, 2012; Barceló and Romero, 2017; Cohen and Weiss, 2016; Arenas *et al.*, 2016] in the database context and to [Gutiérrez-Basulto *et al.*, 2018; Ortiz, 2019; Cima *et al.*, 2021; Jung *et al.*, 2022] in the context of KR.

Our contribution is also closely related to work on synthesising *LTL*-formulas that explain the positive and negative data examples coming from an application [Lemieux *et al.*, 2015; Neider and Gavran, 2018; Camacho and McIlraith, 2019; Raha *et al.*, 2022; Fortin *et al.*, 2022; Fortin *et al.*, 2023]. While concerned with separability of temporal data instances and, in particular, separability by *LTL*-formulas of small size, the separator spaces $\mathsf{s}(E, \mathcal{Q})$ themselves have not yet been investigated in this context.

2 Data and Queries in *LTL*

Fix some countably-infinite set of unary predicate symbols, called *atoms*. A *signature*, σ, is any finite set of atoms. A *(temporal) data instance* is any finite set $\mathcal{D} \neq \emptyset$ of *facts* $A(\ell)$ with an atom A and a *timestamp* $\ell \in \mathbb{N}$, saying that A happened at ℓ. The *size* $|\mathcal{D}|$ of $\mathcal{D}$ is the number of symbols in it, with the timestamps given in *unary*. Let $\max \mathcal{D}$ be the maximal timestamp in $\mathcal{D}$. Where convenient, we also write $\mathcal{D}$ as the word $\delta_0 \dots \delta_{\max \mathcal{D}}$ with $\delta_i = \{A \mid A(i) \in \mathcal{D}\}$. The *signature* $\mathsf{sig}(\mathcal{D})$ of $\mathcal{D}$ is the set of atoms occurring in it.

We query data instances by means of *LTL*-formulas, called *queries*, that are built from atoms (treated as propositional variables) and the logical constant $\top$ (truth) using $\wedge$ and the temporal operators $\bigcirc$ (next time) and $\Diamond$ (sometime in the future). We consider the following classes of queries:

$\mathcal{Q}[\bigcirc\Diamond]$: all $\bigcirc\Diamond$-*queries*;

$\mathcal{Q}[\Diamond]$: all $\Diamond$-*queries* (not containing $\bigcirc$);

$\mathcal{Q}_p[\bigcirc\Diamond]$: *path* $\bigcirc\Diamond$-*queries* of the form (1), where the conjunctions of atoms ρ_i are often treated as sets and the empty conjunction as $\top$;

$\mathcal{Q}_p[\Diamond]$: all *path* $\Diamond$-*queries* (not containing $\bigcirc$);

$\mathcal{Q}_{in}$: *interval-queries* of the form (1) with $\rho_0 = \top$, $\rho_1 \neq \top$, $o_1 = \Diamond$, and $o_i = \bigcirc$, for $i > 1$.

Queries in $\mathcal{Q}_{in}$ single out an interval of a fixed length starting at some time-point ≥ 1; q_1–q_3 from Example 1 are in $\mathcal{Q}_{in}$. $\mathcal{Q}^\sigma$ is the restriction of a class $\mathcal{Q}$ to a signature σ. The *temporal depth* $tdp(q)$ of q is the maximum number of nested temporal operators in q. The *signature* $\mathsf{sig}(q)$ of q is the set of atoms in q; the *size* $|q|$ of q is the number of symbols in it.

The *truth-relation* $\mathcal{D}, n \models q$—saying that q is true in $\mathcal{D}$ at moment $n \in \mathbb{N}$—is defined as usual in temporal logic under the *strict semantics*: $\mathcal{D}, n \models \top$ for all $n \in \mathbb{N}$; $\mathcal{D}, n \models A_i$ iff $A_i(n) \in \mathcal{D}$; $\mathcal{D}, n \models \bigcirc q'$ iff $\mathcal{D}, n+1 \models q'$; and $\mathcal{D}, n \models \Diamond q'$ iff $\mathcal{D}, m \models q'$, for some $m > n$. A data instance $\mathcal{D}$ is called a *positive example* for a query q if $\mathcal{D}, 0 \models q$; otherwise, $\mathcal{D}$ is a *negative example* for q. Checking whether $\mathcal{D}$ is a positive (negative) example for our queries q can obviously be done in polytime in $|\mathcal{D}|$ and $|q|$.

We write $\boldsymbol{q} \models \boldsymbol{q}'$ if $\mathcal{D}, 0 \models \boldsymbol{q}$ implies $\mathcal{D}, 0 \models \boldsymbol{q}'$ for all instances $\mathcal{D}$, and $\boldsymbol{q} \equiv \boldsymbol{q}'$ if $\boldsymbol{q} \models \boldsymbol{q}'$ and $\boldsymbol{q}' \models \boldsymbol{q}$, in which case $\boldsymbol{q}$ and $\boldsymbol{q}'$ are *equivalent*. For example, for any query $\boldsymbol{q}$, we have $\Diamond\bigcirc\boldsymbol{q} \equiv \bigcirc\Diamond\boldsymbol{q} \equiv \Diamond\Diamond\boldsymbol{q}$. It follows that every path query in $\mathcal{Q}_p[\bigcirc\Diamond]$ is equivalent to a query of the form (1), in which $\rho_n \neq \top$ and whenever $\rho_i = \top$, $0 < i < n$, then $\boldsymbol{o}_i = \boldsymbol{o}_{i+1}$; in this case we say that $\boldsymbol{q}$ is in *normal form*. Unless indicated otherwise, we assume all path queries to be in normal form.

Sequences. There is a close link between evaluating path queries and algorithms for finding patterns in strings [Crochemore *et al.*, 2007]. A *sequence* is a data instance $\mathcal{D} = \delta_0 \dots \delta_n$ with $\delta_0 = \emptyset$ and $|\delta_i| = 1$, for $i > 0$; a *sequence query* is a path query of the form (1) with $|\rho_0| = 0$ and $|\rho_i| = 1$, for $i > 0$. Querying sequences using sequence queries corresponds to the following matching problems:

- for any sequence query $\boldsymbol{q} \in \mathcal{Q}_p[\Diamond]$ of the form (1), we have $\mathcal{D}, 0 \models \boldsymbol{q}$ iff $\rho_1 \dots \rho_m$ is a *subsequence* of $\mathcal{D}$;
- for any sequence query $\boldsymbol{q} \in \mathcal{Q}_{in}$ of the form (1), we have $\mathcal{D}, 0 \models \boldsymbol{q}$ iff $\rho_1 \dots \rho_m$ is a *subword* of $\mathcal{D}$.

3 Query Containment

The *query containment problem* for a class $\mathcal{Q}$ of queries is to decide whether $\boldsymbol{q} \models \boldsymbol{q}'$, for any given $\boldsymbol{q}, \boldsymbol{q}' \in \mathcal{Q}$. In contrast to conjunctive queries, where query containment is NP-complete [Chandra and Merlin, 1977], query containment is tractable for the majority of query classes defined above:

Theorem 2. *The query containment problems for $\mathcal{Q}_p[\bigcirc\Diamond]$, $\mathcal{Q}[\Diamond]$ (and their subclasses) are all in* P.

To prove Theorem 2, suppose first that we are given two queries $\boldsymbol{q}, \boldsymbol{q}' \in \mathcal{Q}_p[\bigcirc\Diamond]$, where $\boldsymbol{q}$ takes of the form (1) and $\boldsymbol{q}' = \rho_0' \wedge \boldsymbol{o}_1'(\rho_1' \wedge \dots \wedge \boldsymbol{o}_m'\rho_m')$. Denote by $[m]$ the closed interval $[0, m] \subseteq \mathbb{N}$. A function $h \colon [m] \to [n]$ is *monotone* if $h(i) < h(j)$ whenever $i < j$. Then tractability of containment for path queries follows from the criterion below, which is proved in the appendix by induction on $tdp(\boldsymbol{q}')$:

Lemma 3. *Let $\boldsymbol{q}, \boldsymbol{q}' \in \mathcal{Q}_p[\bigcirc\Diamond]$. Then $\boldsymbol{q} \models \boldsymbol{q}'$ iff there is a monotone function $h \colon [m] \to [n]$ such that $h(0) = 0$ and, for all $i \in [m]$, we have $\rho_i' \subseteq \rho_{h(i)}$ and if $\boldsymbol{o}_{i+1}' = \bigcirc$, then $\boldsymbol{o}_{h(i+1)} = \bigcirc$ and $h(i+1) = h(i) + 1$.*

We refer to any function h defined in Lemma 3 as a *containment witness* for the pair $\boldsymbol{q}, \boldsymbol{q}' \in \mathcal{Q}_p[\bigcirc\Diamond]$.

Suppose now $\boldsymbol{q} \in \mathcal{Q}[\Diamond]$. As shown in [Fortin *et al.*, 2022], we can convert $\boldsymbol{q}$ in polytime to an equivalent query in the *normal form* $\rho \wedge \boldsymbol{q}_1 \wedge \dots \wedge \boldsymbol{q}_n$, where ρ is a conjunction of atoms and each $\boldsymbol{q}_i$ is in $\mathcal{Q}_p[\Diamond]$ and starts with $\Diamond$. Tractability of containment for $\mathcal{Q}[\Diamond]$-queries follows from:

Lemma 4. *If $\boldsymbol{q} = \rho \wedge \boldsymbol{q}_1 \wedge \dots \wedge \boldsymbol{q}_n \in \mathcal{Q}[\Diamond]$ is in normal form, $\boldsymbol{q}' \in \mathcal{Q}_p[\Diamond]$ and $\boldsymbol{q} \models \boldsymbol{q}'$, then there is $\boldsymbol{q}_i$, $1 \leq i \leq n$, with $\rho \wedge \boldsymbol{q}_i \models \boldsymbol{q}'$.*

Proof. In the detailed proof given in the appendix, we show that, assuming $\rho \wedge \boldsymbol{q}_i \not\models \boldsymbol{q}'$ for all i, we can convert the $\boldsymbol{q}_i$ into a data instance $\mathcal{D}$ with $\mathcal{D}, 0 \models \boldsymbol{q}$ and $\mathcal{D}, 0 \not\models \boldsymbol{q}'$. ⊣

For queries $\boldsymbol{q} \in \mathcal{Q}[\bigcirc\Diamond]$, Lemma 4 does not hold:

Example 5. Let $\boldsymbol{q} = \boldsymbol{q}_1 \wedge \boldsymbol{q}_2 \wedge \boldsymbol{q}_3 \wedge \boldsymbol{q}_4$, where

$$\boldsymbol{q}_1 = \Diamond(a \wedge \bigcirc((a \wedge b) \wedge \bigcirc a)),$$
$$\boldsymbol{q}_2 = \Diamond(b \wedge \bigcirc((a \wedge b) \wedge \bigcirc b)),$$
$$\boldsymbol{q}_3 = \Diamond(a \wedge \bigcirc((a \wedge b) \wedge \bigcirc b)),$$
$$\boldsymbol{q}_4 = \Diamond(a \wedge \Diamond(b \wedge \Diamond(a \wedge \Diamond(b \wedge \Diamond(a \wedge \Diamond b))))),$$

and $\boldsymbol{q}' = \Diamond(b \wedge \Diamond((a \wedge b) \wedge \Diamond a))$. Then $\boldsymbol{q} \models \boldsymbol{q}'$ but $\boldsymbol{q}_i \not\models \boldsymbol{q}'$, for any i, $1 \leq i \leq 4$.

At the moment, the question whether containment of $\mathcal{Q}[\bigcirc\Diamond]$-queries is tractable remains open.

4 Example Sets and Separating Queries

An *example set* E is a pair (E^+, E^-) of finite sets $E^+ \neq \emptyset$ and E^- of data instances. We say that a query $\boldsymbol{q}$ *separates* E if all $\mathcal{D} \in E^+$ are positive examples and all $\mathcal{D} \in E^-$ are negative examples for $\boldsymbol{q}$. Denote by $\mathsf{s}(E, \mathcal{Q})$ the set of queries in a class $\mathcal{Q}$ separating E and call E $\mathcal{Q}$-*separable* if $\mathsf{s}(E, \mathcal{Q}) \neq \emptyset$. Our general aim is to understand the structure of $\mathsf{s}(E, \mathcal{Q})$ for various important query classes $\mathcal{Q}$.

We consider queries in $\mathcal{Q}$ *modulo equivalence*, not distinguishing between $\boldsymbol{q} \equiv \boldsymbol{q}'$. In this case, the relation $\models$ is a partial order on $\mathcal{Q}$. For any $\boldsymbol{q} \in \mathsf{s}(E, \mathcal{Q})$, we clearly have

- $tdp(\boldsymbol{q}) \leq \min\{\max \mathcal{D} \mid \mathcal{D} \in E^+\}$,
- $\mathsf{sig}(\boldsymbol{q}) \subseteq \bigcap\{\mathsf{sig}(\mathcal{D}) \mid \mathcal{D} \in E^+\}$,

so $\mathsf{s}(E, \mathcal{Q})$ is finite. We refer to the $\models$-minimal elements of $\mathsf{s}(E, \mathcal{Q}) \neq \emptyset$ as *most specific $\mathcal{Q}$-separators of E* and to the $\models$-maximal elements as *most general $\mathcal{Q}$-separators of E* (modulo $\equiv$); they comprise the sets $\mathsf{mss}(E, \mathcal{Q})$ and $\mathsf{mgs}(E, \mathcal{Q})$, respectively. If these sets are singletons, we call their only element the *unique most specific* and, respectively, *unique most general $\mathcal{Q}$-separator of E*. Note that the former always exists if $\mathsf{s}(E, \mathcal{Q}) \neq \emptyset$ and $\mathcal{Q}$ is closed under $\wedge$.

Example 6. (i) Suppose $\mathcal{D}^+ = \{A(0), B(0), A(1), B(1)\}$, $\mathcal{D}^- = \{A(0)\}$ and $E = (\{\mathcal{D}^+\}, \{\mathcal{D}^-\})$. The separator space $\mathsf{s}(E, \mathcal{Q}_p[\Diamond])$ is shown below as a Hasse diagram with arrows indicating the partial order $\models$ (and $\top, A \notin \mathsf{s}(E, \mathcal{Q}_p[\Diamond])$), so
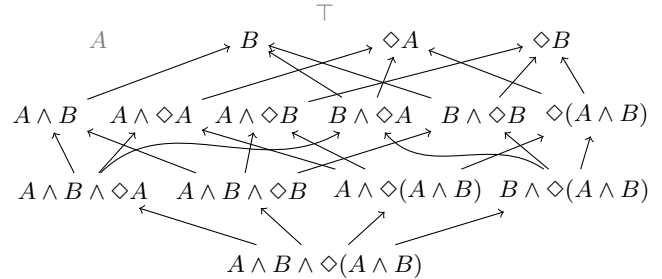
the most general $\mathcal{Q}_p[\Diamond]$-separators of E comprise the set $\mathsf{mgs}(E, \mathcal{Q}_p[\Diamond]) = \{B, \Diamond A, \Diamond B\}$ and the unique most specific one is $A \wedge B \wedge \Diamond(A \wedge B)$.

(ii) For the example set $E = (\{\mathcal{D}_1^+, \mathcal{D}_2^+\}, \{\mathcal{D}_1^-\})$ with

$$\mathcal{D}_1^+ = \{A(1), B(2)\}, \mathcal{D}_2^+ = \{B(1), A(2)\}, \mathcal{D}_1^- = \{C(0)\},$$

$\mathsf{s}(E, \mathcal{Q}_p[\Diamond]) = \{\Diamond A, \Diamond B\} = \mathsf{mss}/\mathsf{mgs}(E, \mathcal{Q}_p[\Diamond])$ but there is no unique most specific/general separator. In contrast, $\mathsf{s}(E, \mathcal{Q}[\Diamond]) = \{\Diamond A, \Diamond B, \Diamond A \wedge \Diamond B\}$ has the unique most specific separator $\Diamond A \wedge \Diamond B$ but no unique most general one.

(*iii*) One can show that $\mathsf{mss}/\mathsf{mgs}(E, \mathcal{Q})$ always contains a *longest/shortest separator* of E in $\mathcal{Q}$ (of largest/smallest temporal depth). To illustrate, let $E = (\{\mathcal{D}_1^+\}, \{\mathcal{D}_1^-, \mathcal{D}_2^-, \mathcal{D}_3^-\})$,

$$\mathcal{D}_1^+ = \{B(0), C(0), A(1)\},$$
$$\mathcal{D}_1^- = \{B(0)\}, \ \mathcal{D}_2^- = \{C(0)\}, \ \mathcal{D}_3^- = \{A(1)\}.$$

Then $\mathsf{mgs}(E, \mathcal{Q}_p[\diamond]) = \{B \wedge C, B \wedge \diamond A, C \wedge \diamond A\}$, where $B \wedge C$ of depth 0 is the shortest separator of E in $\mathcal{Q}_p[\diamond]$.

Our main concern is the following three algorithmic problems for query classes $\mathcal{Q} \subseteq \mathcal{Q}[\bigcirc\diamond]$ with input E and $\boldsymbol{q} \in \mathcal{Q}$:

(most specific/general separator verification): decide whether $\boldsymbol{q}$ is an element of $\mathsf{mss}(E, \mathcal{Q}) / \mathsf{mgs}(E, \mathcal{Q})$;

(counting most specific/general separators): count the elements of $\mathsf{mss}(E, \mathcal{Q}) / \mathsf{mgs}(E, \mathcal{Q})$;

(computing a most specific/general separator): construct some query in $\mathsf{mss}(E, \mathcal{Q}) / \mathsf{mgs}(E, \mathcal{Q})$.

We are particularly interested in deciding whether there is a unique most specific/general separator and computing it. To achieve our aims, we obviously should be able to decide whether $\boldsymbol{q} \in \mathsf{s}(E, \mathcal{Q})$ (*separator verification*) and whether $\mathsf{s}(E, \mathcal{Q}) \neq \emptyset$ (*separator existence*). As mentioned in Section 2, separator verification is in P. We are also interested in the case when the number of positive or negative examples in E is bounded. The table below summarises the complexities of separator existence for our query classes, where $\mathsf{b}^+ / \mathsf{b}^-$

separator existence	$\mathsf{b}^+, \mathsf{b}^-$	b^+	b^- or unbounded
$\mathcal{Q}_p[\diamond] / \mathcal{Q}_p[\bigcirc\diamond]$	in P	NP-c.	NP-c.
$\mathcal{Q}[\diamond] / \mathcal{Q}[\bigcirc\diamond] / \mathcal{Q}_{in}$	in P	in P	NP-c.

means that $|E^+| / |E^-|$ is bounded, and one can assume a bounded signature. Except for $\mathcal{Q}_{in}$, these results are shown in [Fortin *et al.*, 2023]. The proofs use techniques developed for the *longest common subsequence problem* (given $k > 0$ and a set E^+ of sequences, is there a sequence query in $\mathcal{Q}_p[\diamond]$ of depth $\geq k$ that is a subsequence of all $\mathcal{D} \in E^+$; see Section 2) and separator existence for sequence queries in $\mathcal{Q}_p[\diamond]$ [Maier, 1978; Fraser, 1996]. The NP-upper bound for $\mathcal{Q}_{in}$ is trivial; the lower one follows from the proof of Theorem 21, and tractability for b^+ is ensured by the observation that there are polynomially-many relevant intervals in the E^+-examples.

This NP-lower bound shows that great care is needed when transferring techniques from the literature on algorithms for sequences to our framework as separability of example sets of sequences using sequence queries in $\mathcal{Q}_{in}$ is easily seen to be in P even for unbounded example sets and signatures.

A key to the extremal separator problems above is the following notions of strengthening and weakening frontiers.

5 Strengthening and Weakening Frontiers

Let $\mathcal{Q}$ be a class of queries and $\boldsymbol{q} \in \mathcal{Q}$. A set $\mathcal{F} \subseteq \mathcal{Q}$ is called a *strengthening frontier* for $\boldsymbol{q}$ in $\mathcal{Q}$ if

– for any $\boldsymbol{q}' \in \mathcal{F}$, we have $\boldsymbol{q}' \models \boldsymbol{q}$ and $\boldsymbol{q}' \not\equiv \boldsymbol{q}$;
– for any $\boldsymbol{q}'' \in \mathcal{Q}$, if $\boldsymbol{q}'' \models \boldsymbol{q}$ and $\boldsymbol{q}'' \not\equiv \boldsymbol{q}$, then there is $\boldsymbol{q}' \in \mathcal{F}$ such that $\boldsymbol{q}'' \models \boldsymbol{q}'$.

A set $\mathcal{F} \subseteq \mathcal{Q}$ is called a *weakening frontier* for $\boldsymbol{q}$ in $\mathcal{Q}$ if

– for any $\boldsymbol{q}' \in \mathcal{F}$, we have $\boldsymbol{q} \models \boldsymbol{q}'$ and $\boldsymbol{q}' \not\equiv \boldsymbol{q}$;

– for any $\boldsymbol{q}'' \in \mathcal{Q}$, if $\boldsymbol{q} \models \boldsymbol{q}''$ and $\boldsymbol{q}'' \not\equiv \boldsymbol{q}$, then there is $\boldsymbol{q}' \in \mathcal{F}$ with $\boldsymbol{q}' \models \boldsymbol{q}''$.

Trivial strengthening/weakening frontiers for $\boldsymbol{q}$ comprise all queries that are properly stronger/weaker than $\boldsymbol{q}$ in $\mathcal{Q}$; our concern, however, is finding small frontiers. We show now that, for $\mathcal{Q}_p^\sigma[\diamond]$ and $\mathcal{Q}_p^\sigma[\bigcirc\diamond]$, one can compute a strengthening/weakening frontier for any given $\boldsymbol{q}$ in polytime.

Theorem 7. *Let* $\boldsymbol{q} \in \mathcal{Q}_p^\sigma[\bigcirc\diamond]$ *be in normal form* (1)*. A strengthening frontier for* $\boldsymbol{q}$ *in* $\mathcal{Q}_p^\sigma[\bigcirc\diamond]$ *can be computed in polytime by applying once to* $\boldsymbol{q}$ *one of the following operations, for* $i \in [n]$ *and* $\boldsymbol{q}_i = \rho_i \wedge \boldsymbol{o}_{i+1}(\rho_{i+1} \wedge \cdots \wedge \boldsymbol{o}_n \rho_n)$:

1. *extend some* ρ_i *in* $\boldsymbol{q}$ *by some* $A \in \sigma \setminus \rho_i$;
2. *replace some* $\diamond \boldsymbol{q}_i$ *in* $\boldsymbol{q}$ *by* $\diamond(\top \wedge \diamond \boldsymbol{q}_i)$;
3. *replace some* $\boldsymbol{o}_i = \diamond$ *in* $\boldsymbol{q}$ *by* $\bigcirc$ *provided that the resulting query is in normal form*;
4. *add* $\boldsymbol{o}_{n+1}\rho_{n+1}$ *at the end of* $\boldsymbol{q}$*, where* $\boldsymbol{o}_{n+1} = \diamond$ *and* $\rho_{n+1} = A$*, for some* $A \in \sigma$.

If $\boldsymbol{q} \in \mathcal{Q}_p^\sigma[\diamond]$*, a strengthening frontier for* $\boldsymbol{q}$ *in* $\mathcal{Q}_p^\sigma[\diamond]$ *can be computed in polytime using operations 1, 2, and 4.*

Proof. Let $\mathcal{F}$ be the set of queries obtained by a single application of one of these operations to $\boldsymbol{q}$. By Lemma 3, $\boldsymbol{q}' \models \boldsymbol{q}$ and $\boldsymbol{q} \not\equiv \boldsymbol{q}'$ for all $\boldsymbol{q}' \in \mathcal{F}$. Let $\boldsymbol{q}' \models \boldsymbol{q}$ and $\boldsymbol{q} \not\equiv \boldsymbol{q}'$, for some $\boldsymbol{q}' \in \mathcal{Q}_p^\sigma[\bigcirc\diamond]$ of the form $\boldsymbol{q}' = \rho_0' \wedge \boldsymbol{o}_1'(\rho_1' \wedge \cdots \wedge \boldsymbol{o}_m'(\rho_m'))$. Take a containment witness $h : [n] \to [m]$ for $\boldsymbol{q}', \boldsymbol{q}$. If h is surjective, then $n = m$ and $h(i) = i$ for all $i \in [n]$, and so $\rho_i \subseteq \rho_i'$. As $\boldsymbol{q} \not\equiv \boldsymbol{q}'$, either $\boldsymbol{o}_i' = \bigcirc$ and $\boldsymbol{o}_i = \diamond$, for some $i \in [n]$, or $\rho_i \subsetneq \rho_i'$, for some $i \in [n]$. In the former case, operation 3 gives $\boldsymbol{q}'' \in \mathcal{F}$ with $\boldsymbol{q}' \models \boldsymbol{q}''$; in the latter one, operation 1 gives such a $\boldsymbol{q}''$.

Suppose h is not surjective. If there is $i < n$ such that $h(i+1) - h(i) \geq 2$, then $\boldsymbol{q}' \models \boldsymbol{q}''$ for a $\boldsymbol{q}'' \in \mathcal{F}$ given by operation 2. Otherwise m is not in the range of h, and we get such a $\boldsymbol{q}'' \in \mathcal{F}$ by operation 4. ⊣

Example 8. For $\sigma = \{A, B\}$ and $\boldsymbol{q} = \diamond(A \wedge \bigcirc B)$, operations 1–4 give the following strengthening frontier for $\boldsymbol{q}$:

$\diamond(A \wedge B \wedge \bigcirc B), \diamond(A \wedge \bigcirc(A \wedge B)), \diamond(\top \wedge \diamond(A \wedge \bigcirc B))$, $\bigcirc(A \wedge \bigcirc B), \diamond(A \wedge \bigcirc(B \wedge \diamond A)), \diamond(A \wedge \bigcirc(B \wedge \diamond B))$.

A weakening frontier can be constructed by reversing operations 1–3 from Theorem 7 and using a similar argument:

Theorem 9. *Let* $\boldsymbol{q} \in \mathcal{Q}_p^\sigma[\bigcirc\diamond]$ *be in normal form* (1)*. A weakening frontier for* $\boldsymbol{q}$ *in* $\mathcal{Q}_p^\sigma[\bigcirc\diamond]$ *can be computed in polytime by applying once to* $\boldsymbol{q}$ *one of the following operations, for* $i \in [n]$ *and* $\boldsymbol{q}_i = \rho_i \wedge \boldsymbol{o}_{i+1}(\rho_{i+1} \wedge \cdots \wedge \boldsymbol{o}_n \rho_n)$:

1. *drop some atom from* ρ_i;
2. *replace some* $\diamond(\top \wedge \diamond \boldsymbol{q}_i)$ *in* $\boldsymbol{q}$ *by* $\diamond \boldsymbol{q}_i$;
3. *replace some* $\boldsymbol{o}_i = \bigcirc$ *by* $\diamond$.

If $\boldsymbol{q} \in \mathcal{Q}_p^\sigma[\diamond]$*, a weakening frontier for* $\boldsymbol{q}$ *in* $\mathcal{Q}_p^\sigma[\diamond]$ *can be computed in polytime using operations 1 and 2.*

Example 10. For $\sigma = \{A, B\}$ and $\boldsymbol{q} = \diamond(A \wedge \bigcirc B)$, operations 1-3 give the following weakening frontier for $\boldsymbol{q}$:

$$\diamond(\top \wedge \diamond B), \quad \diamond A, \quad \diamond(A \wedge \diamond B).$$

Note that the computed weakening frontier can be made smaller by omitting $\diamond A$, which is weaker than $\diamond(A \wedge \diamond B)$.

We next consider frontiers for queries in $\mathcal{Q}^\sigma[\diamond]$. We say that $\boldsymbol{q} = \rho \wedge \boldsymbol{q}_1 \wedge \cdots \wedge \boldsymbol{q}_n \in \mathcal{Q}^\sigma[\diamond]$ in normal form is *redundancy-free* if it does not contain $\boldsymbol{q}_i, \boldsymbol{q}_j$ with $i \neq j$ and $\boldsymbol{q}_i \models \boldsymbol{q}_j$. Clearly, for any $\boldsymbol{q} \in \mathcal{Q}^\sigma[\diamond]$, we can compute an equivalent redundancy-free $\boldsymbol{q}' \in \mathcal{Q}^\sigma[\diamond]$ in polytime.

Theorem 11. *Let* $\boldsymbol{q} = \rho \wedge \boldsymbol{q}_1 \wedge \cdots \wedge \boldsymbol{q}_n \in \mathcal{Q}^\sigma[\diamond]$ *be redundancy-free. A weakening frontier for* $\boldsymbol{q}$ *in* $\mathcal{Q}^\sigma[\diamond]$ *can be computed in polytime by a single application of one of the following operations to* $\boldsymbol{q}$:

1. *drop some atom from* ρ;
2. *replace some* $\boldsymbol{q}_i$ *by* $\bigwedge \mathcal{F}_i$, *where* $\mathcal{F}_i$ *is the weakening frontier for* $\boldsymbol{q}_i$ *in* $\mathcal{Q}_p^\sigma[\diamond]$ *provided by Theorem 9.*

Proof. Let $\mathcal{F}$ be the set of queries defined above. Clearly, $\boldsymbol{q} \models \boldsymbol{q}'$; as $\boldsymbol{q}$ is redundancy-free and in view of Lemma 4, $\boldsymbol{q} \not\equiv \boldsymbol{q}'$ for all $\boldsymbol{q}' \in \mathcal{F}$. Suppose $\boldsymbol{q} \models \boldsymbol{q}'$ and $\boldsymbol{q} \not\equiv \boldsymbol{q}'$, for some $\boldsymbol{q}' = \rho' \wedge \boldsymbol{q}'_1 \wedge \cdots \wedge \boldsymbol{q}'_m$ in $\mathcal{Q}^\sigma[\diamond]$. By Lemma 4, $\rho' \subseteq \rho$ and, for each j, $1 \leq j \leq m$, there exists $f(j)$, $1 \leq f(j) \leq n$, with $\boldsymbol{q}_{f(j)} \models \boldsymbol{q}'_j$. If $\rho' \subsetneq \rho$, then operation 1 gives a $\boldsymbol{q}'' \in \mathcal{F}$ with $\boldsymbol{q}'' \models \boldsymbol{q}'$. Otherwise, as $\boldsymbol{q} \not\equiv \boldsymbol{q}'$, there is $\boldsymbol{q}_i$ such that $\boldsymbol{q}'_j \not\models \boldsymbol{q}_i$ for all $\boldsymbol{q}'_j$, $1 \leq j \leq m$. Let $\boldsymbol{q}''$ be the query obtained from $\boldsymbol{q}$ by replacing $\boldsymbol{q}_i$ with $\bigwedge \mathcal{F}_i$ by operation 2. To establish $\boldsymbol{q}'' \models \boldsymbol{q}'$, it suffices to show that $f(j) = i$ implies $\bigwedge \mathcal{F}_i \models \boldsymbol{q}_j$. So suppose $f(j) = i$. Then $\boldsymbol{q}'_j \not\models \boldsymbol{q}_i$, and so, by the definition of $\mathcal{F}_i$, there is $\boldsymbol{q}''_i \in \mathcal{F}_i$ with $\boldsymbol{q}''_i \models \boldsymbol{q}_j$. $\dashv$

However, strengthening frontiers for queries in $\mathcal{Q}^\sigma[\diamond]$ are not necessarily of polynomial size as shown by the following:

Example 12. We represent queries in $\mathcal{Q}^\sigma[\diamond]$ of the form (1) as $\rho_0 \ldots \rho_n$. For $\sigma = \{A_1, A_2, B_1, B_2\}$, let $\boldsymbol{q}_1 = \emptyset(\boldsymbol{q}\sigma)^n \boldsymbol{q}$, $\boldsymbol{q}_2 = \emptyset\sigma^{2n+1}$, and $\boldsymbol{q} = \{A_1, A_2\}\{B_1, B_2\}$. Using [Fortin *et al.*, 2022, Example 18], one can show that any strengthening frontier for the query $\boldsymbol{q}_1 \wedge \boldsymbol{q}_2$ in $\mathcal{Q}^\sigma[\diamond]$ is of size $O(2^n)$.

Note also that weakening frontiers for $\boldsymbol{q} \in \mathcal{Q}_{in}^\sigma$ can be computed in polytime using operation 1 in Theorem 9 (if $i = 1$ and $|\rho_i| = 1$, we drop $\top$ and take $\diamond(\rho_2 \wedge \bigcirc(\rho_3 \wedge \cdots \wedge \bigcirc\rho_n))$). On the other hand, strengthening frontiers can be infinite:

Example 13. All $\diamond(A \wedge \bigcirc^n A)$, $n > 0$, are in any strengthening frontier for $\boldsymbol{q} = \diamond A$ in $\mathcal{Q}_{in}^\sigma$, where $\sigma = \{A\}$.

As shown by Theorem 17, the lack of polytime computable strengthening frontiers in $\mathcal{Q}[\diamond]$ affects the complexity of verifying most specific separators. In contrast, the lack of polytime computable strengthening frontiers in $\mathcal{Q}_{in}$ turns out to be harmless. For $n \geq tdp(\boldsymbol{q})$, call $\mathcal{F} \subseteq \mathcal{Q}$ an n-*bounded weakening/strengthening frontier* for $\boldsymbol{q}$ in $\mathcal{Q}$ if $\mathcal{F}$ is a weakening/strengthening frontier for $\boldsymbol{q}$ in the class $\{\boldsymbol{q}' \in \mathcal{Q} \mid tdp(\boldsymbol{q}') \leq n\}$ (we assume n to be given in unary).

Theorem 14. *Weakening frontiers and* n-*bounded strengthening frontiers in* $\mathcal{Q}_{in}^\sigma$ *can be computed in polytime.*

Theorems 7, 9 give an alternative way of polytime computing unique characterisations of $\mathcal{Q}_p^\sigma[\bigcirc\diamond]$-queries by examples [Fortin *et al.*, 2022], opening another route to studying unique characterisations and exact learning of temporal queries.

6 Complexity

Now we show complexity bounds for the decision and counting problems from Section 4, starting with verification and observing that polytime computable n-bounded frontiers imply tractable verification of most specific/general separators:

Lemma 15. *If an* n-*bounded strengthening/weakening frontier for* $\boldsymbol{q} \in \mathcal{Q}$ *is polytime computable in* $|\boldsymbol{q}|$ *and* n, *then most specific/general separator verification for* $\mathcal{Q}$ *is in* P.

Proof. Let $n = \min\{\max \mathcal{D} \mid \mathcal{D} \in E^+\}$. We compute an n-bounded strengthening frontier $\mathcal{F}$ for $\boldsymbol{q}$ in $\mathcal{Q}$ and use that $\boldsymbol{q} \in \mathsf{mss}(E, \mathcal{Q})$ iff $\boldsymbol{q} \in \mathsf{s}(E, \mathcal{Q})$ and $\boldsymbol{q}' \notin \mathsf{s}(E, \mathcal{Q})$ for all $\boldsymbol{q}' \in \mathcal{F}$. The case of $\mathsf{mgs}(E, \mathcal{Q})$ is similar. $\dashv$

Corollary 16. *Most specific/general separator verification is in* P *for* $\mathcal{Q}_p[\diamond]$, $\mathcal{Q}_p[\bigcirc\diamond]$, *and* $\mathcal{Q}_{in}$. *Most general separator verification is in* P *for* $\mathcal{Q}[\diamond]$.

Proof. It follows from Lemma 15 and Theorems 7, 9, 14 that most specific/general separator verification is in P for $\mathcal{Q}_p[\diamond]$, $\mathcal{Q}_p[\bigcirc\diamond]$, and $\mathcal{Q}_{in}$. By Theorem 11, most general separation verification is also in P for $\mathcal{Q}[\diamond]$. $\dashv$

Most specific separator verification is harder for $\mathcal{Q}[\diamond]$:

Theorem 17. *For* $\mathcal{Q}[\diamond]$, *the most specific separator verification problem coincides with the unique most specific separator verification problem and is* CONP-*complete.*

Proof. As we know, E has a unique most specific separator in $\mathcal{Q}[\diamond]$ iff $\mathsf{s}(E, \mathcal{Q}[\diamond]) \neq \emptyset$. Hence most specific separator verification coincides with unique most specific separator verification. Given $\boldsymbol{q}$ and E, we can check in NP that either $\boldsymbol{q} \notin \mathsf{s}(E, \mathcal{Q}[\diamond])$ (which is in P) or that there is $\boldsymbol{q}' \in \mathsf{s}(E, \mathcal{Q}[\diamond])$ with $\boldsymbol{q}' \models \boldsymbol{q}$ and $\boldsymbol{q}' \not\equiv \boldsymbol{q}$ (which is in NP). This gives the CONP upper bound. The more involved proof of the lower one (in the appendix) is based on ideas similar to those in the proof of Theorem 21 below. $\dashv$

In the cases when most specific/general separators are not necessarily unique, we obtain the following:

Theorem 18. *Unique most specific/general separator verification in* $\mathcal{Q}_p[\bigcirc\diamond]$, $\mathcal{Q}_p[\diamond]$, $\mathcal{Q}_{in}$ *as well as unique most general separator verification in* $\mathcal{Q}[\diamond]$ *are* CONP-*complete.*

Proof. The upper bounds follow from Corollary 16. The lower ones are by reduction of the problem to decide whether, for a Boolean formula φ and a satisfying assignment $\mathfrak{a}$, there is a satisfying assignment for φ different from $\mathfrak{a}$. We use ideas similar to those in the proof of Theorem 21. $\dashv$

We now turn to the existence and counting problems. Recall that UNIQUE SAT is the problem to decide whether there is exactly one satisfying assignment for a propositional formula [Blass and Gurevich, 1982]. It is in Δ_2^p, CONP-hard, and complete for the class US of problems solvable by a non-deterministic polynomial-time Turing machine M that accepts iff M has exactly one accepting path. A counting problem is in the class $\sharp$P if there is such M whose number of accepting paths coincides with the number of solutions to the problem [Arora and Barak, 2009].

Theorem 19. *For $\mathcal{Q}_p[\bigcirc\Diamond]$, $\mathcal{Q}_p[\Diamond]$, $\mathcal{Q}_{in}$, counting most specific/general separators is $\sharp$P-complete; the existence of a unique most specific/general separator is* US-*complete. The same holds for most general separators in $\mathcal{Q}[\Diamond]$.*

Theorem 19 follows from the very general and fine-grained complexity results provided by Theorems 20 and 21 below.

Theorem 20. *Let $\mathcal{Q} \subseteq \mathcal{Q}_p[\bigcirc\Diamond]$ be a class with polytime decidable membership and polytime computable n-bounded strengthening/weakening frontiers for queries in $\mathcal{Q}$. Then*

- *counting most specific/general $\mathcal{Q}$-separators is in $\sharp$P;*
- *the existence of a unique most specific/general $\mathcal{Q}$-separator is in* US.

This also holds for most general separators in classes of conjunctions of $\bigcirc\Diamond$-path queries closed under dropping conjuncts and having polytime computable weakening frontiers.

Proof. Given E, construct a TM M that guesses a $\boldsymbol{q} \in \mathcal{Q}$ with $tdp(\boldsymbol{q}) \leq n = \min\{\max \mathcal{D} \mid \mathcal{D} \in E^+\}$ and accepts if $\boldsymbol{q}$ separates E and no $\boldsymbol{q}'$ in the n-bounded frontier for $\boldsymbol{q}$ in $\mathcal{Q}$ separates E. As we know, the required checks are in P. In the second claim, M guesses a query $\boldsymbol{q} = \rho \wedge \Diamond\boldsymbol{q}_1 \wedge \cdots \wedge \Diamond\boldsymbol{q}_l$ with $\boldsymbol{q}_i \in \mathcal{Q}_p[\bigcirc\Diamond]$ and $l \leq |E^-|$. $\dashv$

For the lower bounds, we take into account the boundedness of $\mathsf{sig}(E)$ and the cardinalities $|E^+|$ and $|E^-|$ of positive and negative examples in E. The next result provides matching lower bounds for Theorem 19 even if the signature is bounded and only one of $|E^+|$ and $|E^-|$ is unbounded, except for $\mathcal{Q}_{in}$ and $\mathcal{Q}[\Diamond]$, where $|E^+|$ has to be unbounded.

Theorem 21. *Let $\mathcal{Q} \subseteq \mathcal{Q}_p[\bigcirc\Diamond]$ be any class of queries containing all $\Diamond\rho$ with a conjunction of atoms ρ. Then*

- *counting most specific/general $\mathcal{Q}$-separators is $\sharp$P-hard;*
- *the existence of unique most specific/general separator is* US-*hard*

even for $E = (E^+, E^-)$ and $\sigma = \mathsf{sig}(E)$ with (a) $|E^-| \leq 1$ or (b) $|E^-| \leq 1$, $|\sigma| = 2$ and $\mathcal{Q} = \mathcal{Q}_{in}$ or $\mathcal{Q} \supseteq \mathcal{Q}_p[\Diamond]$, or (c) $|E^+| \leq 4$, $|\sigma| = 3$, $\mathcal{Q} \supseteq \mathcal{Q}_p[\Diamond]$. This result also holds for most general separators in $\mathcal{Q}[\Diamond]$ and $|E^-| = 1$, $|\sigma| = 2$.

Proof. We sketch the proof of the first claim by a parsimonious reduction from SAT for an unbounded number of negative examples (which can be easily merged into one). Take a CNF $\varphi = \psi_1 \wedge \cdots \wedge \psi_k$ with clauses ψ_i over variables $x_1, \ldots, x_n$. We construct $E = (E^+, E^-)$ such that there is a bijection between the satisfying assignments for φ and the separators for E in $\mathcal{Q}[\bigcirc\Diamond]$ (even queries of the form $\Diamond\rho$). The claim follows from the fact that the separating queries are mutually $\models$-incomparable. Define the positive examples $E^+ = \{\mathcal{D}_0, \mathcal{D}'_0, \mathcal{D}_1, \ldots, \mathcal{D}_n\}$ with $2n$ atoms $A_1, \bar{A}_1, \ldots, A_n, \bar{A}_n$ by taking

$$\mathcal{D}_0 = \{A_i(1), \bar{A}_i(1) \mid 1 \leq i \leq n\},$$
$$\mathcal{D}'_0 = \{A_i(2), \bar{A}_i(2) \mid 1 \leq i \leq n\},$$
$$\mathcal{D}_i = \{A_i(1), \bar{A}_i(2), A_j(1), \bar{A}_j(1), A_j(2), \bar{A}_j(2) \mid i \neq j\},$$

for $1 \leq i \leq n$. Let $E^- = \{\mathcal{D}_1^1, \ldots, \mathcal{D}_k^1, \mathcal{D}_1^2, \ldots, \mathcal{D}_n^2\}$, where $\mathcal{D}_i^1$, $1 \leq i \leq k$, comprises $\bar{A}_j(1)$ if x_j does not occur negatively in ψ_i, and $A_j(1)$ if x_j does not occur positively in ψ_i,

and $\mathcal{D}_i^2 = \{A_j(1), \bar{A}_j(1) \mid j \neq i\}$. For an assignment $\mathfrak{a}$ for $x_1, \ldots, x_n$, let $\rho_\mathfrak{a}$ contain A_i if $\mathfrak{a}(x_i) = 1$ and $\bar{A}_i$, otherwise. Now our claim follows from the following: (i) if $\mathfrak{a}$ satisfies φ, then $\Diamond\rho_\mathfrak{a}$ separates E; (ii) if $\boldsymbol{q} \in \mathcal{Q}_p[\bigcirc\Diamond]$ separates E, then $\boldsymbol{q} \equiv \Diamond\rho_\mathfrak{a}$, for some $\mathfrak{a}$ satisfying φ.

To bound σ and/or E^+, we give parsimonious reductions from SAT and employ techniques for the longest common subsequence problem and separator existence for sequence queries in $\mathcal{Q}_p[\Diamond]$ [Blin *et al.*, 2012; Fraser, 1996]. $\dashv$

Again, re-using techniques from the literature on algorithms for sequences needs some care. Similar to separability, for example sets of sequences, counting most general/specific sequence $\mathcal{Q}_{in}$-separators is easily seen to be in P even for unbounded signatures and example sets, but Theorem 21 shows that this is not so for $\mathcal{Q}_{in}$ on non-sequence data instances.

Surprisingly, the complexities of counting most general separators and deciding the existence of a unique one diverge if we bound the number of examples, cf. Theorems 23 and 27.

Theorem 22. *Let $\mathcal{Q} = \mathcal{Q}[\Diamond]$ or $\mathcal{Q} \subseteq \mathcal{Q}_p[\bigcirc\Diamond]$ be any class containing all $\Diamond\rho$ with a conjunction of atoms ρ. Then counting most general $\mathcal{Q}$-separators is $\sharp$P-hard even for example sets $E = (E^+, E^-)$ with $|E^+| = 2$ and $|E^-| = 1$.*

Proof. The proof is by reduction from counting satisfying assignments for monotone formulas: E^- is as in the proof of Theorem 21 and the positive examples are $\mathcal{D}_0, \mathcal{D}'_0$. $\dashv$

Theorem 22 does not hold for counting most specific separators in $\mathcal{Q}_{in}$, which is easily seen to be in P if $|E^+|$ is bounded. It remains open whether some version of this theorem holds for most specific separators in $\mathcal{Q}_p[\bigcirc\Diamond]$ or $\mathcal{Q}_p[\Diamond]$.

7 Algorithms

The frontiers defined in Section 5 give a polytime algorithm for computing a most specific/general separator starting from *any* given separator. Suppose $\mathcal{Q} \in \{\mathcal{Q}_p[\Diamond], \mathcal{Q}_p[\bigcirc\Diamond], \mathcal{Q}_{in}\}$, we are given $\boldsymbol{q} \in \mathsf{s}(E, \mathcal{Q})$ and need a most specific separator. By Theorems 7, 9, the length of the longest $\models$-chain in $\mathsf{s}(E, \mathcal{Q}_p[\bigcirc\Diamond])$ is polynomial in $|E|$.[1] We take, if possible, some $\boldsymbol{q}' \in \mathsf{s}(E, \mathcal{Q})$ in the strengthening frontier for $\boldsymbol{q}$, then $\boldsymbol{q}'' \in \mathsf{s}(E, \mathcal{Q})$ in the strengthening frontier for $\boldsymbol{q}'$, etc. This process terminates after polynomially-many steps, returning a most specific $\mathcal{Q}$-separator (and so the unique one, if any).

Thus, we can focus on algorithms deciding the existence of a (unique most specific/general) separator and constructing it. In the next theorem, we compute not just a random input separator, but a longest/shortest one. As strengthening/weakening frontiers contain queries that are not shorter/longer than the input, the procedure above will compute a longest most specific/shortest most general separator.

Theorem 23. *Let $E = (E^+, E^-)$, $\sigma = \mathsf{sig}(E)$, t_+/t_- be the maximum timestamp in E^+/E^-, $c_+ = |E^+|$, $c_- = |E^-|$, and $\mathcal{Q} \in \{\mathcal{Q}_p[\Diamond], \mathcal{Q}_p[\bigcirc\Diamond]\}$. The following can be done in time $O(t_+^{c_+} t_-^{c_-})$:*

(a) *deciding whether $\mathsf{s}(E, \mathcal{Q}) \neq \emptyset$;*

---

[1] In contrast, the proof of Theorem 21 shows the size of the maximal antichain in $\mathsf{s}(E, \mathcal{Q}_p[\Diamond])$ is in general exponential in $|E|$.

(b) *computing a longest/shortest separator in* $\mathsf{s}(E, \mathcal{Q})$;

(c) *deciding the existence of a unique most specific $\mathcal{Q}$-separator and a unique most general $\mathcal{Q}_p[\diamond]$-separator, and constructing such a separator.*

For bounded c_+ and c_-, problem (a) is in NL.

*Proof.* We only sketch the construction for $\mathcal{Q} = \mathcal{Q}_p[\diamond]$.

(a) First, we define a directed labelled rooted graph $\mathfrak{P}$ whose paths from the root represent $\mathcal{Q}_p[\diamond]$-queries with positive examples $E^+ = \{\mathcal{D}_1, \ldots, \mathcal{D}_{c_+}\}$. Its nodes are vectors $(n_1, \ldots, n_{c_+})$ with $0 \leq n_i \leq \max \mathcal{D}_i$, which are labelled by the sets $\{A \in \sigma \mid A(n_i) \in \mathcal{D}_i \text{ for all } i\}$, and the edges are $(n_1, \ldots, n_{c_+}) \rightarrow (n'_1, \ldots, n'_{c_+})$ with $n_i < n'_i$ for all i. The *root* of $\mathfrak{P}$ is $\bar{0} = (0, \ldots, 0)$. To illustrate, let $E = (E^+, E^-)$, where $E^+ = \{\mathcal{D}_1^+, \mathcal{D}_2^+\}$, $E^- = \{\mathcal{D}_1^-, \mathcal{D}_2^-, \mathcal{D}_3^-\}$,

$$\mathcal{D}_1^+ = \{A(0), C(1), D(1), B(2)\},$$
$$\mathcal{D}_2^+ = \{A(0), C(1), B(2), D(2)\},$$
$$\mathcal{D}_1^- = \{A(0), C(1)\}, \ \mathcal{D}_2^- = \{A(0), D(1)\}, \ \mathcal{D}_3^- = \{B(0)\}.$$

Graph $\mathfrak{P}$ is shown on the left-hand side of the picture below:
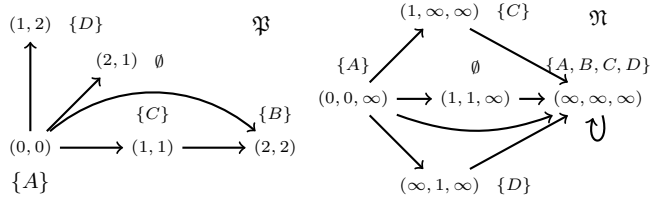
Each path starting at $\bar{0}$ gives rise to a $\mathcal{Q}_p^\sigma[\diamond]$-query with positive examples in E^+: e.g., $(0, 0) \rightarrow (1, 1) \rightarrow (2, 2)$ gives rise to the query $A \wedge \diamond(C \wedge \diamond B)$.

Next, define another graph $\mathfrak{N}$ for $E^- = \{\mathcal{D}_1, \ldots, \mathcal{D}_{c_-}\}$. Its nodes are $(n_1, \ldots, n_{c_-})$, where $n_i \in [0, \max \mathcal{D}_i] \cup \{\infty\}$, including $\bar{\infty} = (\infty, \ldots, \infty)$, which is labelled by σ. The label of $(n_1, \ldots, n_{c_-})$ is $\{A \mid A(n_i) \in \mathcal{D}_i, \text{ for all } n_i \neq \infty\}$. The edges are defined in the same way as for $\mathfrak{P}$, with $n_i < \infty$, for any n_i. The root of $\mathfrak{N}$ is $(n_1, \ldots, n_l)$, where $n_i = 0$ if $\{A(0) \mid A(0) \in \mathcal{D}, \text{ for all } \mathcal{D} \in E^+\} \subseteq \mathcal{D}_i$ and $n_i = \infty$ otherwise (see the picture above). Let q be a $\mathcal{Q}_p^\sigma[\diamond]$-query of the form (1) with ρ_0 contained in the root's label. Then all $\mathcal{D}_i \in E^-$ are negative examples for q iff every path starting at the root and having labels $\rho'_0, \ldots, \rho'_n$ with $\rho'_i \supseteq \rho_i$, for all $i \leq n$, comes through node $\bar{\infty}$. In our example, every such path for $q = A \wedge \diamond B$ and $q = A \wedge \diamond(C \wedge \diamond B)$ involves $\bar{\infty}$. However, this is not the case for $q = A \wedge \diamond C$.

Consider now a graph $\mathfrak{P} \otimes \mathfrak{N}$ with nodes (n, m), where n is a node in $\mathfrak{P}$ and m a node in $\mathfrak{N}$ with $l(n) \subseteq l(m)$, for the labels $l(n)$ and $l(m)$ of n and m. We have an edge $(n, m) \rightarrow (n', m')$ in $\mathfrak{P} \otimes \mathfrak{N}$ iff $n \rightarrow n'$ in $\mathfrak{P}$, $m \rightarrow m'$ in $\mathfrak{N}$, and $\mathfrak{P} \otimes \mathfrak{N}$ has no (n', m'') with $m \rightarrow m''$, $m''_i < m'_i$ and $m''_i \neq \infty$, for some i, m_i being the ith coordinate of m. One can see that, for any (n, m) in $\mathfrak{P} \times \mathfrak{N}$ and n' in $\mathfrak{P}$, there exists at most one edge $(n, m) \rightarrow (n', m')$. The root of $\mathfrak{P} \times \mathfrak{N}$ comprises the roots of $\mathfrak{P}$ and $\mathfrak{N}$. In our example, the edges from the root $(\bar{0}, (0, 0, \infty))$ of $\mathfrak{P} \otimes \mathfrak{N}$ lead to $((1, 2), (\infty, 1, \infty))$, $((2, 1), (1, 1, \infty))$, $((1, 1), (1, \infty, \infty))$ and $((2, 2), \bar{\infty})$. Given a path

$$\pi = (\bar{0} = \boldsymbol{n}_0, \boldsymbol{m}_0), \ldots, (\boldsymbol{n}_n, \boldsymbol{m}_n) \tag{2}$$

let $\boldsymbol{q}_\pi$ be the $\mathcal{Q}_p[\diamond]$-query of the form (1) with $\rho_i = l(\boldsymbol{n}_i)$ (note that $\boldsymbol{q}_\pi$ is not necessarily in normal form). We call π a *separating path* for E if $l(\boldsymbol{n}_n) \neq \emptyset$ and $\boldsymbol{m}_n = \infty$.

Lemma 24. $\mathsf{s}(E, \mathcal{Q}) \neq \emptyset$ *iff* $\mathfrak{P} \otimes \mathfrak{N}$ *contains a separating path π, with $\boldsymbol{q}_\pi$ separating E.*

In our running example, $\mathfrak{P} \otimes \mathfrak{N}$ has two *separating paths*: $\pi_1 = (\bar{0}, (0, 0, \infty)), ((2, 2), \bar{\infty})$ and $\pi_2 = (\bar{0}, (0, 0, \infty)), ((1, 1), (1, \infty, \infty)), ((2, 2), \bar{\infty})$, which give rise to the separators $\boldsymbol{q}_{\pi_1} = A \wedge \diamond B$ and $\boldsymbol{q}_{\pi_2} = A \wedge \diamond(C \wedge \diamond B)$.

The existence of a separating path in $\mathfrak{P} \otimes \mathfrak{N}$ can be checked in time $O(t_+^{c_+} t_-^{c_-})$. If c_+ and c_- are bounded, given $(\boldsymbol{n}, \boldsymbol{m})$ and $(\boldsymbol{n}', \boldsymbol{m}')$, we can check in logspace whether $(\boldsymbol{n}, \boldsymbol{m})$ is the root of $\mathfrak{P} \otimes \mathfrak{N}$ and $(\boldsymbol{n}, \boldsymbol{m}) \rightarrow (\boldsymbol{n}', \boldsymbol{m}')$, and so the existence of a separating path can be decided in NL.

The proof of point (b) relies on the following observation:

Lemma 25. *If $\boldsymbol{q}$ of the form (1) separates E, then there is a separating path of the form (2) with $\rho_i \subseteq l(\boldsymbol{n}_i)$, for all $i \leq n$.*

It follows that the length of a longest/shortest separator for E coincides with the length of a longest/shortest separating path in $\mathfrak{P} \otimes \mathfrak{N}$, which can be found in polytime.

The proof of point (c) is based on the following criterion:

Lemma 26. *A $\mathcal{Q}_p[\diamond]$-query $\boldsymbol{q}$ is a unique most specific $\mathcal{Q}_p[\diamond]$-separator for E iff there is a separating path π such that $\boldsymbol{q} = \boldsymbol{q}_\pi$ and $\boldsymbol{q}_\pi \models \boldsymbol{q}_\nu$, for every separating path ν.*

In our example, $\boldsymbol{q}_{\pi_2}$ is a unique most specific separator. We show that the criterion of Lemma 26 can be checked in polytime in $\mathfrak{P} \times \mathfrak{N}$. For unique most general separators, the seemingly obvious inversion of $\boldsymbol{q}_\pi \models \boldsymbol{q}_\nu$ does not give a criterion, and a different type of graph is required. ⊣

Finally, we show how Theorem 23 can be used to check the existence of and construct unique most general separators in $\mathcal{Q}[\diamond]$ (the case of unique most specific ones is trivial).

Theorem 27. *An example set $E = (E^+, E^-)$ has a unique most general separator in $\mathcal{Q}[\diamond]$ iff $\boldsymbol{q} = \bigwedge_{\mathcal{D} \in S} \boldsymbol{q}_\mathcal{D}$ separates E, where S is the set of $\mathcal{D} \in E^-$ such that $(E^+, \{\mathcal{D}\})$ has a unique most general separator, $\boldsymbol{q}_\mathcal{D}$, in $\mathcal{Q}_p[\diamond]$. In this case, $\boldsymbol{q}$ is a unique most general separator of E in $\mathcal{Q}[\diamond]$.*

8 Conclusions

We have conducted a comprehensive complexity analysis of extremal separators in the spaces $\mathsf{s}(E, \mathcal{Q})$ with $\mathcal{Q}$ ranging from various classes of temporal path $\bigcirc\diamond$-queries to arbitrary $\diamond$-queries. For arbitrary $\bigcirc\diamond$-queries, we only know more or less straightforward upper bounds such as Π_2^p for (unique) most general separator verification and Σ_3^p for unique most general separator existence. Establishing tight bounds remains a challenging open problem, which requires a deeper understanding of query containment for these queries.

We also plan to analyse the shortest and longest separators. For instance, we show in the appendix that verifying such separators in $\mathsf{s}(E, \mathcal{Q}_p[\bigcirc\diamond])$ is CONP-complete—harder than verifying most specific/general ones. An empirical evaluation of our algorithms and more expressive query languages (say, with non-strict $\diamond$ and 'until') are left for future work.

References

[Abboud *et al.*, 2015] A. Abboud, A. Backurs, and V. V. Williams. Tight hardness results for LCS and other sequence similarity measures. In FOCS, pages 59–78. IEEE Computer Society, 2015.

[Alexe *et al.*, 2011] B. Alexe, B. ten Cate, Ph. Kolaitis, and W. Ch. Tan. Characterizing schema mappings via data examples. *ACM Trans. Database Syst.*, 36(4):23, 2011.

[Arenas *et al.*, 2016] M. Arenas, G. I. Diaz, and E. Kostylev. Reverse engineering SPARQL queries. In *Proc. of WWW*, pages 239–249, 2016.

[Arora and Barak, 2009] S. Arora and B. Barak. *Computational Complexity - A Modern Approach*. Cambridge University Press, 2009.

[Artale *et al.*, 2021] A. Artale, R. Kontchakov, A. Kovtunova, V. Ryzhikov, F. Wolter, and M. Zakharyaschev. First-order rewritability of ontology-mediated queries in linear temporal logic. *Artif. Intell.*, 299:103536, 2021.

[Baader *et al.*, 2015] F. Baader, S. Borgwardt, and M. Lippmann. Temporal query entailment in the description logic $\mathcal{SHQ}$. *J. Web Semant.*, 33:71–93, 2015.

[Barceló and Romero, 2017] P. Barceló and M. Romero. The complexity of reverse engineering problems for conjunctive queries. In *Proc. of ICDT*, pages 7:1–7:17, 2017.

[Barceló *et al.*, 2021] P. Barceló, A. Baumgartner, V. Dalmau, and B. Kimelfeld. Regularizing conjunctive features for classification. *J. Comput. Syst. Sci.*, 119:97–124, 2021.

[Bergroth *et al.*, 2000] L. Bergroth, H. Hakonen, and T. Raita. A survey of longest common subsequence algorithms. In *Proc. of SPIRE*, pages 39–48, 2000.

[Blass and Gurevich, 1982] A. Blass and Y. Gurevich. On the unique satisfiability problem. *Inf. Control.*, 55(1-3):80–88, 1982.

[Blin *et al.*, 2012] G. Blin, L. Bulteau, M. Jiang, P. Tejada, and S. Vialette. Hardness of longest common subsequence for sequences with bounded run-lengths. In *Proc. of CPM*, volume 7354 of *Lecture Notes in Computer Science*, pages 138–148. Springer, 2012.

[Blum *et al.*, 2021] Ch. Blum, M. Djukanovic, A. Santini, H. Jiang, Chu-Min Li, F. Manyà, and G. Raidl. Solving longest common subsequence problems via a transformation to the maximum clique problem. *Comput. Oper. Res.*, 125:105089, 2021.

[Borgwardt *et al.*, 2015] S. Borgwardt, M. Lippmann, and V. Thost. Temporalizing rewritable query languages over knowledge bases. *J. Web Semant.*, 33:50–70, 2015.

[Camacho and McIlraith, 2019] A. Camacho and Sh. McIlraith. Learning interpretable models expressed in linear temporal logic. In *Proc. of ICAPS 2018*, pages 621–630. AAAI Press, 2019.

[Chandra and Merlin, 1977] A. Chandra and P. Merlin. Optimal implementation of conjunctive queries in relational data bases. In *Conference Record of the Ninth Annual ACM Symposium on Theory of Computing*, pages 77–90. ACM, 1977.

[Chomicki and Toman, 2018] J. Chomicki and D. Toman. Temporal logic in database query languages. In Ling Liu and M. Tamer Özsu, editors, *Encyclopedia of Database Systems, Second Edition*. Springer, 2018.

[Chowdhury *et al.*, 2010] R. A. Chowdhury, Hai-Son Le, and V. Ramachandran. Cache-oblivious dynamic programming for bioinformatics. *IEEE/ACM Trans. on Computational Biology and Bioinformatics*, 7(3):495–510, 2010.

[Cima *et al.*, 2021] G. Cima, F. Croce, and M. Lenzerini. Query definability and its approximations in ontology-based data management. In *Proc. of CIKM'21*, pages 271–280. ACM, 2021.

[Cohen and Weiss, 2016] S. Cohen and Y. Y. Weiss. The complexity of learning tree patterns from example graphs. *ACM Trans. Database Syst.*, 41(2):14:1–14:44, 2016.

[Crochemore *et al.*, 2007] M. Crochemore, Ch. Hancart, and T. Lecroq. *Algorithms on strings*. Cambridge University Press, 2007.

[Cropper *et al.*, 2022] A. Cropper, S. Dumancic, R. Evans, and S. H. Muggleton. Inductive logic programming at 30. *Mach. Learn.*, 111(1):147–172, 2022.

[Deutch and Gilad, 2019] D. Deutch and A. Gilad. Reverse-engineering conjunctive queries from provenance examples. In *Proc. of EDBT*, pages 277–288, 2019.

[Fijalkow and Lagarde, 2021] N. Fijalkow and G. Lagarde. The complexity of learning linear temporal formulas from examples. In *Proc. of the 15th Int. Conf. on Grammatical Inference*, volume 153 of *Proc. of Machine Learning Research*, pages 237–250. PMLR, 2021.

[Fortin *et al.*, 2022] M. Fortin, B. Konev, V. Ryzhikov, Y. Savateev, F. Wolter, and M. Zakharyaschev. Unique characterisability and learnability of temporal instance queries. In *Proc. of KR*, 2022.

[Fortin *et al.*, 2023] M. Fortin, B. Konev, V. Ryzhikov, Y. Savateev, F. Wolter, and M. Zakharyaschev. Reverse engineering of temporal queries mediated by LTL ontologies. In *Proc. of IJCAI*, pages 3230–3238. ijcai.org, 2023.

[Fraser, 1996] C. Fraser. Consistent subsequences and supersequences. *Theor. Comput. Sci.*, 165(2):233–246, 1996.

[Funk *et al.*, 2019] M. Funk, J.-Ch. Jung, C. Lutz, H. Pulcini, and F. Wolter. Learning description logic concepts: When can positive and negative examples be separated? In *Proc. of IJCAI*, pages 1682–1688, 2019.

[Gutiérrez-Basulto *et al.*, 2018] V. Gutiérrez-Basulto, J. Ch. Jung, and L. Sabellek. Reverse engineering queries in ontology-enriched systems: The case of expressive Horn description logic ontologies. In *Proc. of IJCAI-ECAI*, 2018.

[Jung *et al.*, 2022] J. Ch. Jung, C. Lutz, H. Pulcini, and F. Wolter. Logical separability of labeled data examples under ontologies. *Artif. Intell.*, 313:103785, 2022.

[Kalashnikov *et al.*, 2018] D. Kalashnikov, L. Lakshmanan, and D. Srivastava. Fastqre: Fast query reverse engineering. In *Proc. of SIGMOD*, pages 337–350, 2018.

[Kimelfeld and Ré, 2018] B. Kimelfeld and Ch. Ré. A relational framework for classifier engineering. *ACM Trans. Database Syst.*, 43(3):11:1–11:36, 2018.

[Lemieux *et al.*, 2015] C. Lemieux, D. Park, and I. Beschastnikh. General LTL specification mining (T). In *Proc. of ASE*, pages 81–92. IEEE, 2015.

[Maier, 1978] D. Maier. The complexity of some problems on subsequences and supersequences. *J. ACM*, 25(2):322–336, 1978.

[Martins, 2019] D. Martins. Reverse engineering database queries from examples: State-of-the-art, challenges, and research opportunities. *Inf. Syst.*, 83:89–100, 2019.

[Mitchell, 1982] T. Mitchell. Generalization as search. *Artif. Intell.*, 18(2):203–226, 1982.

[Neider and Gavran, 2018] D. Neider and I. Gavran. Learning linear temporal properties. In *Proc. of FMCAD 2018*, pages 1–10. IEEE, 2018.

[Ortiz, 2019] M. Ortiz. Ontology-mediated queries from examples: a glimpse at the DL-Lite case. In *Proc. of GCAI*, pages 1–14, 2019.

[Raha *et al.*, 2022] R. Raha, R. Roy, N. Fijalkow, and D. Neider. Scalable anytime algorithms for learning fragments of linear temporal logic. In *Proc. of TACAS*, volume 13243 of *Lecture Notes in Computer Science*, pages 263–280. Springer, 2022.

[Sarker *et al.*, 2017] Md. K. Sarker, Ning Xie, D. Doran, M. Raymer, and P. Hitzler. Explaining trained neural networks with semantic web technologies: First steps. In *Proc. of NeSy*, 2017.

[Staworko and Wieczorek, 2012] S. Staworko and P. Wieczorek. Learning twig and path queries. In *Proc. of ICDT*, pages 140–154, 2012.

[ten Cate and Dalmau, 2015] B. ten Cate and V. Dalmau. The product homomorphism problem and applications. In *Proc. of ICDT*, pages 161–176, 2015.

[ten Cate *et al.*, 2023] B. ten Cate, V. Dalmau, M. Funk, and C. Lutz. Extremal fitting problems for conjunctive queries. In *Proc. of PODS*, pages 89–98. ACM, 2023.

[Weiss and Cohen, 2017] Y. Y. Weiss and S. Cohen. Reverse engineering SPJ-queries from examples. In *Proc. of PODS*, pages 151–166, 2017.

[Zhang *et al.*, 2013] M. Zhang, H. Elmeleegy, C. M. Procopiuc, and D. Srivastava. Reverse engineering complex join queries. In *Proc. of SIGMOD*, pages 809–820, 2013.

Appendix: Proofs

A Proofs for Section 3

We assume that queries q are in normal form. Recall that $[n] = \{0, \dots, n\}$. We start by introducing a helpful tool for checking whether $\mathcal{D}, 0 \models q$. Let q be a path $\bigcirc\Diamond$-query of the form (1), that is,

$$q = \rho_0 \wedge \boldsymbol{o}_1(\rho_1 \wedge \boldsymbol{o}_2(\rho_2 \wedge \cdots \wedge \boldsymbol{o}_n \rho_n)),$$

where $\boldsymbol{o}_i \in \{\bigcirc, \Diamond\}$ and ρ_i is a conjunction of atoms. A *satisfying function* f for q into a data instance $\mathcal{D}$ maps $[n]$ to $[\max \mathcal{D}]$ such that

- $f(0) = 0$;
- $\rho_i \subseteq \mathcal{D}_{f(i)} := \{A \mid A(f(i)) \in \mathcal{D}\}$;
- $f(i) < f(i+1)$;
- $f(i+1) = f(i) + 1$ if $\boldsymbol{o}_{i+1} = \bigcirc$.

It is easy to show that $\mathcal{D}, 0 \models q$ iff there exists a satisfying function for q into $\mathcal{D}$.

$$q' = \rho'_0 \wedge \boldsymbol{o}'_1(\rho'_1 \wedge \cdots \wedge \boldsymbol{o}'_m(\rho'_m)). \tag{3}$$

A function $g : [m] \to [n]$ is a *containment witness* for q, q' if the following conditions holds:

- $\rho'_i \subseteq \rho_{g(i)}$;
- $g(0) = 0$;
- $g(i) < g(i+1)$;
- if $\boldsymbol{o}'_{i+1} = \bigcirc$, then $\boldsymbol{o}_{g(i+1)} = \bigcirc$ and $g(i+1) = g(i) + 1$.

We can now state Lemma 3 as follows.

Lemma 28. $q \models q'$ *iff there is a containment witness for* q, q'.

Proof. It is easy to see that if there is a containment witness for q, q' then $q \models q'$. We prove the converse direction.

Let $\sigma = \text{sig}(q)$. The proof is by induction on the temporal depth of q'. If $tdp(q') = 0$, the claim is trivial.

Assume $q \models q'$ are given and the statement has been shown for all q'' with $tdp(q'') < tdp(q')$.

Assume first that q' starts with a nonempty sequence of $\bigcirc$. Then let $k > 0$ be maximal such that $\boldsymbol{o}'_1 = \cdots = \boldsymbol{o}'_k = \bigcirc$. Then $\rho'_k \neq \top$ (by normal form) and if $m > k$, then $\boldsymbol{o}'_{k+1} = \Diamond$.

Let r be minimal such that

$$\rho_0 \wedge \boldsymbol{o}_1(\rho_1 \wedge \cdots \wedge \boldsymbol{o}_r \rho_r)) \models \rho'_0 \wedge \boldsymbol{o}'_1(\rho'_1 \wedge \cdots \wedge \boldsymbol{o}'_k \rho'_k))$$

Claim 1. For all $i \leq k$, $\boldsymbol{o}_i = \bigcirc$ and $\rho_i \supseteq \rho'_i$. Hence $r = k$.

To prove Claim 1 observe that if there is $i \leq k$ with $\boldsymbol{o}_i = \Diamond$, then we have for

$$\mathcal{D} = \rho_0 \dots \rho_{i-1} \emptyset^m \rho_i \dots \rho_n$$

$\mathcal{D}, 0 \models q$ and $\mathcal{D}, 0 \not\models q'$, a contradiction. If there is $i \leq k$ with $\rho_i \not\supseteq \rho'_i$, then we have for

$$\mathcal{D} = \rho_0 \dots \rho_n$$

$\mathcal{D}, 0 \models q$ and $\mathcal{D}, 0 \not\models q'$, again a contradiction. This finishes the proof of Claim 1.

Next observe that if follows from $q \models q'$ that

$$\delta \models \delta'$$

for $\delta = \boldsymbol{o}_{k+1}(\rho_{k+1} \wedge \cdots \wedge \boldsymbol{o}_n \rho_n))$ and $\delta' = \boldsymbol{o}'_{k+1}(\rho'_{k+1} \wedge \cdots \boldsymbol{o}'_m \rho'_m))$. As

$$tdp(\delta') < tdp(q')$$

we can apply the IH to δ, δ' and have a containment witness g' for δ, δ'. Define g by setting $g(i) = i$ for all $i \leq k$ and $g(k+j) = g'(j) + k$ for all j with $1 \leq j \leq m - k$. Then g is a containment witness for q, q'.

Assume now that q' does not start with $\bigcirc$. Assume first there is a minimal initial subquery

$$\delta' = \rho'_0 \wedge \boldsymbol{o}'_1(\rho'_1 \wedge \cdots \wedge \boldsymbol{o}'_k \rho'_k))$$

of q' ending with $\rho'_k \neq \top$ and, moreover, $\boldsymbol{o}'_{k+1} = \Diamond$ and $k > 0$. Let

$$\delta = \rho_0 \wedge \boldsymbol{o}_1(\rho_1 \wedge \cdots \wedge \rho_r))$$

be the minimal initial subquery of q such that $\delta \models \delta'$. As $tdp(\delta') < tdp(q')$, by IH we have a containment witness g_0 for δ, δ'. As $\rho'_k \neq \top$ and r is minimal we have $g_0(k) = r$. By IH the claim also holds for $\gamma' = o'_{k+1}(\rho'_{k+1} \wedge \cdots \wedge o'_m \rho'_m))$ and $\gamma = o_{r+1}(\rho_{r+1} \wedge \cdots \wedge o_n \rho_n))$. Hence we obtain a containment witness g_1 for γ, γ'. Then we obtain a containment witness for q, q' by concatenating g_0 and g_1.

Finally assume there is no $\rho'_k \neq \top$ with $op'_{k+1} = \diamond$ and $k > 0$. Then, by normal form, q' takes the form

$$\rho'_0 \wedge \diamond^k(\rho'_k \wedge o'_{k+1}\rho'_{k+1} \cdots \wedge o'_m \rho'_m))$$

with $k > 0$, $\rho'_k \neq \top$, and $o'_{k+1} = \cdots = o'_m = \bigcirc$.

We show there exists $j \geq k$ such that $o_{j+1} = \cdots = o_{j+m-k} = \bigcirc$ and $\rho_{j+\ell} \supseteq \rho'_{k+\ell}$ for all ℓ with $0 \leq \ell \leq m - k$. Clearly then there is a containment witness g for q, q'.

To show our claim, suppose no such j exists. Then take the data instance

$$\mathcal{D} = \rho_0 w_1 \rho_1 \cdots w_n \rho_n$$

where w_i is the empty word if $o_i = \bigcirc$ and $w_i = \emptyset^m$ if $o_i = \diamond$. Then we have $\mathcal{D}, 0 \models q$ but $\mathcal{D}, 0 \not\models q'$, a contradiction. $\quad\dashv$

We consider $\mathcal{Q}[\diamond]$. We assume that $q \in \mathcal{Q}[\diamond]$ takes the from

$$\rho \wedge q_1 \wedge \cdots \wedge q_n$$

where ρ is a conjunction of atoms and each q_i is in $\mathcal{Q}_p[\diamond]$ and starts with $\diamond$.

Lemma 4. *If $q = \rho \wedge q_1 \wedge \cdots \wedge q_n \in \mathcal{Q}[\diamond]$ is in normal form, $q' \in \mathcal{Q}_p[\diamond]$ and $q \models q'$, then there is q_i, $1 \leq i \leq n$, with $\rho \wedge q_i \models q'$.*

Proof. Assume

$$q_i = \diamond(\rho_i^1 \wedge \cdots \rho_i^{n_i}))$$

for $1 \leq i \leq n$,

$$q' = \rho^0 \wedge \diamond(\rho^1 \wedge \cdots \wedge \diamond \rho^m)),$$

and $q \models q'$. Note that $\rho \supseteq \rho^0$ as otherwise $q \not\models q'$. We define functions

$$f_i : \{1, \ldots, m\} \to \{1, \ldots, n_i\} \cup \{\infty\}$$

for $1 \leq i \leq n$. The definition is by induction starting from 1.

Set $f_i(1) = j$ if j is minimal such that $\rho_i^j \supseteq \rho^1$ and $f_i(1) = \infty$ if no j with $\rho_i^j \supseteq \rho^1$ exists.

Inductively, if $f_i(\ell) = \infty$, then $f_i(\ell + 1) = \infty$. If $f_i(\ell) < \infty$, then set $f_i(\ell + 1) = j$ if j is minimal such that

- $j > f_i(\ell)$;
- $\rho_i^j \supseteq \rho^{\ell+1}$

and $f_i(\ell + 1) = \infty$ if no $j > f_i(\ell)$ such that $\rho_i^j \supseteq \rho^{\ell+1}$ exists.

Observe that if there exists $i \leq n$ such that $f_i(m) < \infty$, then $\rho \wedge q_i \models q'$, as required. Hence it remains to show that there exists such an $i \leq n$. For a proof by contradiction, assume there is no such $i \leq n$. We derive a contraction by proving that $q \not\models q'$. Let $m' \leq m$ be minimal such that $f_i(m') = \infty$ for all $i \leq n$. Define a data instance $\mathcal{D}_1$ as

$$\mathcal{D}_1 = \rho \rho_1^1 \cdots \rho_1^{k_1} \cdots \rho_n^1 \cdots \rho_n^{k_n},$$

where $k_i = \min\{n_i, f_i(1) - 1\}$ for $1 \leq i \leq n$ (we set $\infty - 1 = \infty$). If $m' = 1$, then $\mathcal{D}_1, 0 \models q$ and $\mathcal{D}_1, 0 \not\models q'$ since $\mathcal{D}_1, 0 \not\models \diamond \rho^1$, and we are done. Otherwise let

$$\delta_1 = \bigcup_{1 \leq i \leq n, f_i(1) < \infty} \rho_i^{f_i(1)}$$

We continue in this way for $\ell = 2, 3, \ldots, m'$ as follows. First define for $2 \leq \ell \leq m'$:

$$\mathcal{D}_\ell = \rho_1^{f_1(\ell)+1} \cdots \rho_1^{k_{\ell,1}} \cdots \rho_n^{f_n(\ell)+1} \cdots \rho_n^{k_{\ell,n}},$$

where $k_{\ell,i} = \min\{n_i, f_i(\ell + 1) - 1\}$ for $1 \leq i \leq n$ (note that $\rho_i^{f_i(\ell)+1} \cdots \rho_i^{k_{\ell,i}}$ is empty if $f_i(\ell) + 1 > n_i$). Then let for $2 \leq \ell < m'$:

$$\delta_\ell = \bigcup_{1 \leq i \leq n, f_i(\ell) < \infty} \rho_i^{f_i(\ell)}$$

and set

$$\mathcal{D} = \mathcal{D}_1 \delta_1 \mathcal{D}_2 \delta_2 \cdots \delta_{m'-1} \mathcal{D}_{m'}$$

It is straightforward to show that $\mathcal{D}, 0 \models q$ and $\mathcal{D}, 0 \not\models q'$. Hence $q \not\models q'$ and we have derived a contradiction. $\quad\dashv$

B Proofs for Section 4

We prove the complexity results for separability that are stated in this section for $\mathcal{Q}_{in}$.

Theorem 29. *(1) For $\mathcal{Q}_{in}$, separator existence is NP-complete. This also holds if $|E^-|$ and $\mathrm{sig}(E)$ are bounded.*
    (2) If $|E^+|$ is bounded then $\mathcal{Q}_{in}$-separator existence is in P.
    (3) For example data instances $E = (E^+, E^-)$ that are sequences, the existence of a separator that is a sequence query in $\mathcal{Q}_{in}$ is in P.

Proof. (1) The NP-upper bound follows from the fact that given E one can guess a separator of polynomial size in $\mathcal{Q}_{in}$ and check in polytime that is separates. The lower bound follows directly from the proof of Theorem 21 (b) given later in this appendix.
    (2) Assume $E^+ = \{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$. Let $m_i = \max \mathcal{D}_i$ and $m = \max\{m_i \mid 1 \le i \le n\}$. Consider for every $0 \le k < m$ and vector $\vec{k} = k_1, \ldots, k_n$ with $0 < k_i \le m_i$ the query

$$\boldsymbol{q}_{k,\vec{k}} = \Diamond(\rho^0_{k,\vec{k}} \wedge \bigcirc(\cdots \wedge \bigcirc \rho^k_{k,\vec{k}}))$$

where

$$\rho^j_{k,\vec{k}} = \bigcap_{1 \le i \le n} \{A \mid A(k_i + j) \in \mathcal{D}_i\}$$

Let Ξ denote the set of queries of the form $\boldsymbol{q}_{k,\vec{k}}$ that are in $\mathcal{Q}_{in}$. Then it is easy to see that there exists a query in $\mathcal{Q}_{in}$ that separates E iff some query in Ξ separates E. The latter can be checked in polynomial time if n is fixed.
    (3) Let $E^+ = \{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$ be a set of sequences. Assume

$$\mathcal{D}_1 = \{A_0\}\{A_1\} \cdots \{A_m\}$$

Then any possible separating sequence query in $\mathcal{Q}_{in}$ takes the form

$$\Diamond(A_{k_1} \wedge \bigcirc(\cdots \bigcirc A_{k_2}))$$

with $0 < k_1 \le k_2 \le m$. Clearly it can be checked in polynomial time whether any such query separates E.   ⊣

C Proofs for Section 5

Theorem 9. *Let $\boldsymbol{q} \in \mathcal{Q}^\sigma_p[\bigcirc\Diamond]$ be in normal form (1). A weakening frontier for $\boldsymbol{q}$ in $\mathcal{Q}^\sigma_p[\bigcirc\Diamond]$ can be computed in polytime by applying once to $\boldsymbol{q}$ one of the following operations, for $i \in [n]$ and $\boldsymbol{q}_i = \rho_i \wedge \boldsymbol{o}_{i+1}(\rho_{i+1} \wedge \cdots \wedge \boldsymbol{o}_n \rho_n)$:*

1. drop some atom from ρ_i;

2. replace some $\Diamond(\top \wedge \Diamond \boldsymbol{q}_i)$ in $\boldsymbol{q}$ by $\Diamond \boldsymbol{q}_i$;

3. replace some $\boldsymbol{o}_i = \bigcirc$ by $\Diamond$.

If $\boldsymbol{q} \in \mathcal{Q}^\sigma_p[\Diamond]$, a weakening frontier for $\boldsymbol{q}$ in $\mathcal{Q}^\sigma_p[\Diamond]$ can be constructed in polynomial time using operations 1 and 2.

Proof. Let $\mathcal{F}$ denote the set of queries defined in Theorem 9. It is easy to see using Lemma 3 that $\boldsymbol{q} \models \boldsymbol{q}'$ and $\boldsymbol{q} \not\equiv \boldsymbol{q}'$ for all $\boldsymbol{q}' \in \mathcal{F}$.
    Assume now $\boldsymbol{q} \models \boldsymbol{q}'$ and $\boldsymbol{q} \not\equiv \boldsymbol{q}'$ for some $\boldsymbol{q}' \in \mathcal{Q}^\sigma_p[\bigcirc\Diamond]$. We show that there exists a query $\boldsymbol{q}'' \in \mathcal{F}$ such that $\boldsymbol{q}'' \models \boldsymbol{q}'$. Assume

$$\boldsymbol{q}' = \rho'_0 \wedge \boldsymbol{o}'_1(\rho'_1 \wedge \cdots \wedge \boldsymbol{o}'_m(\rho'_m)).$$

and assume that $g : [m] \to [n]$ is a containment witness for $\boldsymbol{q}, \boldsymbol{q}'$.
    We distinguish the following cases.
    Case 1. g is a surjective mapping onto $[n]$. Then $n = m$ and $g(i) = i$ for all $i \in [m]$. Hence $\rho'_i \subseteq \rho_i$ for all $i \in [m]$. Then, as $\boldsymbol{q} \not\equiv \boldsymbol{q}'$, either $\boldsymbol{o}_i = \bigcirc$ and $\boldsymbol{o}'_i = \Diamond$ for some $i \in [m]$ or $\rho'_i \subsetneq \rho_i$ for some $i \in [m]$. In the first case $\boldsymbol{q}'' \models \boldsymbol{q}'$ for a $\boldsymbol{q}'' \in \mathcal{F}$ obtained in Point 3 and in the second case $\boldsymbol{q}'' \models \boldsymbol{q}'$ for a $\boldsymbol{q}'' \in \mathcal{F}$ obtained in Point 1.
    Case 2. g is not onto $[n]$. Take any $i \le n$ that is not in the range of g. If there is $i < m$ with $f(i+1) - f(i) \ge 2$, then $\boldsymbol{q}'' \models \boldsymbol{q}'$ for a $\boldsymbol{q}'' \in \mathcal{F}$ obtained in Point 2. Otherwise n is not in the range of g and $\boldsymbol{q}'' \models \boldsymbol{q}'$ for a $\boldsymbol{q}'' \in \mathcal{F}$ obtained in Point 1.   ⊣

    Example 12 We represent queries in $\mathcal{Q}^\sigma[\Diamond]$ of the form (1) as $\rho_0 \ldots \rho_n$. For $\sigma = \{A_1, A_2, B_1, B_2\}$, let $\boldsymbol{q}_1 = \emptyset(\boldsymbol{q}\sigma)^n \boldsymbol{q}$, $\boldsymbol{q}_2 = \emptyset\sigma^{2n+1}$, and $\boldsymbol{q} = \{A_1, A_2\}\{B_1, B_2\}$. We show using [Fortin *et al.*, 2022, Example 18] that any strengthening frontier for the query $\boldsymbol{q}_1 \wedge \boldsymbol{q}_2$ in $\mathcal{Q}^\sigma[\Diamond]$ is of size $O(2^n)$. Indeed, let P be the set of queries of the form $\emptyset \boldsymbol{q}_1 \ldots \boldsymbol{q}_{n+1}$, where $\boldsymbol{q}_i$ is either $\{A_1\}\{A_2\}$ or $\{B_1\}\{B_2\}$. Clearly, $|P| = 2^{n+1}$. As shown in [Fortin *et al.*, 2022], $\boldsymbol{q}_1 \wedge \boldsymbol{q}_2 \not\models \boldsymbol{q}$ for all $\boldsymbol{q} \in P$ and, for any data instance $\mathcal{D}$ with $\mathcal{D} \models \boldsymbol{q}_1 \wedge \boldsymbol{q}_2$, there is at most one $\boldsymbol{q} \in P$ with $\mathcal{D} \not\models \boldsymbol{q}$. It follows that $|\mathcal{F}| \ge 2^{n+1}$, for otherwise the pigeonhole principle would give distinct $\boldsymbol{q}^1, \boldsymbol{q}^2 \in P$ and some $\boldsymbol{q} \in \mathcal{F}$ such that $\boldsymbol{q}_1 \wedge \boldsymbol{q}_2 \wedge \boldsymbol{q}^i \models \boldsymbol{q}$, $i = 1, 2$. As $\boldsymbol{q}_1 \wedge \boldsymbol{q}_2 \not\models \boldsymbol{q}$, there is a data instance $\mathcal{D} \models \boldsymbol{q}_1 \wedge \boldsymbol{q}_2$ with $\mathcal{D} \not\models \boldsymbol{q}$, and so $\mathcal{D} \not\models \boldsymbol{q}^i$, for $i = 1, 2$, contrary to $\boldsymbol{q}^1 \ne \boldsymbol{q}^2$.

    Theorem 14 is a consequence of the following two results.

Theorem 30. *Let $q = \Diamond(\rho_1 \wedge \bigcirc(\rho_2 \wedge \cdots \wedge \bigcirc\rho_n)) \in \mathcal{Q}_{in}^\sigma$. A weakening frontier for q in $\mathcal{Q}_{in}^\sigma$ can be computed in polytime by dropping some atom from ρ_i for some $i \in [n]$; here, if $i = 1$ and ρ_i is a singleton, then also drop $\top$ and take $\Diamond(\rho_2 \wedge \bigcirc(\rho_3 \wedge \cdots \wedge \bigcirc\rho_n))$.*

Proof. Let $\mathcal{F}$ denote the set of queries defined in Theorem 30. It is easy to see using Lemma 3 that $q \models q'$ and $q \not\equiv q'$ for all $q' \in \mathcal{F}$.

Assume now $q \models q'$ and $q \not\equiv q'$ for some $q' \in \mathcal{Q}_{in}^\sigma$. We show that there exists a query $q'' \in \mathcal{F}$ such that $q'' \models q'$. Assume

$$q' = \Diamond(\rho_1' \wedge \bigcirc(\rho_2' \wedge \cdots \wedge \bigcirc\rho_m')).$$

and assume that $g : [m] \to [n]$ is a containment witness for q, q'.

We distinguish the following cases.

Case 1. g is a surjective mapping onto $[n]$. Then $n = m$ and $g(i) = i$ for all $i \in [m]$. Hence $\rho_i' \subseteq \rho_i$ for all $i \in [m]$. Then, as $q \not\equiv q'$, $\rho_i' \subsetneq \rho_i$ for some $i \in [m]$. Then $q'' \models q'$ for a $q'' \in \mathcal{F}$ obtained by dropping an atom from ρ_i (observe that $\rho_i' \neq \top$ by the definition of interval queries, so ρ_i is not a singleton).

Case 2. g is not onto $[n]$. Then either $i = 1$ or $i = n$ are not in the range of g. If $i = 1$, then we obtain $q'' \in \mathcal{F}$ with $q'' \models q'$ by dropping an atom from ρ_1 (if ρ_1 is a singleton we also drop $\top$). If $i = n > 1$, then we obtain $q'' \in \mathcal{F}$ with $q'' \models q'$ by dropping an atom from ρ_n. $\dashv$

Theorem 31. *Let $q = \Diamond(\rho_1 \wedge \bigcirc(\rho_2 \wedge \cdots \wedge \bigcirc\rho_n)) \in \mathcal{Q}_{in}^\sigma$. A strengthening frontier for q in $\mathcal{Q}_{in}^\sigma$ can be computed by applying once to q one of the following operations:*

1. *extend some ρ_i in q by some $A \in \sigma \setminus \rho_i$;*

2. *insert $A \wedge \bigcirc^n$ before ρ_1, that is, form $\Diamond(A \wedge \bigcirc^n(\rho_1 \wedge \bigcirc(\rho_2 \wedge \cdots \wedge \bigcirc\rho_n))$ for some $A \in \sigma$ and $n \geq 1$;*

3. *add $\bigcirc^n A$ at the end of q, for some $A \in \sigma$ and $n \geq 1$.*

Proof. Let $\mathcal{F}$ be the set of queries obtained by a single application of one of these operations to q. By Lemma 3, $q' \models q$ and $q \not\equiv q'$ for all $q' \in \mathcal{F}$. Let $q' \models q$ and $q \not\equiv q'$, for some $q' \in \mathcal{Q}_{in}^\sigma$ of the form

$$q' = \Diamond(\rho_1' \wedge \bigcirc(\rho_2' \wedge \cdots \wedge \bigcirc\rho_m')).$$

Take a containment witness $h \colon [n] \to [m]$ for q', q. If h is surjective, then $n = m$ and $h(i) = i$ for all $i \in [n]$, and so $\rho_i \subseteq \rho_i'$. As $q \not\equiv q'$, $\rho_i \subsetneq \rho_i'$, for some $i \in [n]$. Then operation 1 gives a $q'' \in \mathcal{F}$ with $q' \models q''$.

Suppose h is not surjective. Then either $i = 1$ is not in the range of h or $i = m$ is not in the range of h. In the first case $q' \models q''$ for a $q'' \in \mathcal{F}$ given by operation 2. In the second case, $q' \models q''$ for a $q'' \in \mathcal{F}$ given by operation 3. $\dashv$

D Proofs for Section 6

Theorem 17. *For $\mathcal{Q}[\Diamond]$, the most specific separator verification problem coincides with the unique most specific separator verification problem and is* CONP-*complete.*

Proof. It remains to show the lower bound. We reduce SAT to the complement of most specific separator verification. Let $\varphi = \psi_1 \wedge \ldots \wedge \psi_k$ be a propositional formula with k clauses over n variables $x_1, \ldots, x_n$. We assume without loss of generality that no variable occurs both positively and negatively in a clause. We use $2n$ atoms $A_1, \bar{A}_1, \ldots, A_n, \bar{A}_n$. Define the positive examples $E^+ = \{\mathcal{D}_0, \mathcal{D}_0', \mathcal{D}_1, \ldots, \mathcal{D}_n\}$ by setting

- $\mathcal{D}_0 = \{A_i(1), \bar{A}_i(1) \mid 1 \leq i \leq n\}$,

- $\mathcal{D}_0' = \{A_i(2), \bar{A}_i(2) \mid 1 \leq i \leq n\}$,

and including in $\mathcal{D}_i$, for $1 \leq i \leq n$, the assertions:

- $A_i(1)$ and $\bar{A}_i(2)$, and

- $A_j(1), \bar{A}_j(1), A_j(2), \bar{A}_j(2)$ for all $j \neq i$.

We also add to each $\mathcal{D}_i$ the following, for $1 \leq \ell \leq k$:

- $\bar{A}_j(2 + \ell)$, if x_j does not occur negatively in ψ_ℓ, and

- $A_j(2 + \ell)$, if x_j does not occur positively in ψ_ℓ.

Now consider the query

$$q = \bigwedge_{1 \leq \ell \leq k} \Diamond\rho_\ell$$

with ρ_ℓ containing:

- $\bar{A}_j$, if x_j does not occur negatively in ψ_ℓ, and

- A_j, if x_j does not occur positively in ψ_ℓ.

The negative examples are not relevant, we simply take $E^- = \{\mathcal{D}^-\}$ with $\mathcal{D}^- = \{A_1(0)\}$.

Then clearly $\mathcal{D}, 0 \models q$ for all $\mathcal{D} \in E^+$ and $\mathcal{D}^-, 0 \not\models q$. We show that q is equivalent to the unique most specific separator if and only if φ is not satisfiable. For a variable assignment $\mathfrak{a}$ for $x_1, \ldots, x_n$, we denote with $\rho_\mathfrak{a}$ the set that contains, for all i, A_i, if $\mathfrak{a}(x_i) = 1$ and $\bar{A}_i$, otherwise.

Assume first that φ is satisfiable. Take a variable assignment $\mathfrak{a}$ witnessing this. Then $\Diamond \rho_\mathfrak{a}$ separates E but $q \not\models \Diamond \rho_\mathfrak{a}$ and so q is not a most specific separating query for E. To show that $q \not\models \Diamond \rho_\mathfrak{a}$ take the data instance $\mathcal{D}$ containing for $1 \leq \ell \leq k$:

- $\bar{A}_j(2 + \ell)$, if x_j does not occur negatively in ψ_ℓ, and

- $A_j(2 + \ell)$, if x_j does not occur positively in ψ_ℓ.

Then, by definition, $\mathcal{D}, 0 \models q$. But $\mathcal{D}, 0 \not\models \Diamond \rho_\mathfrak{a}$ since $\mathfrak{a}$ is a satisfying variable assignment for φ.

For the converse direction, assume that q' is a query in $\mathcal{Q}[\Diamond]$ that separates E but that $q \not\models q'$. We show that φ is satisfiable. By the positive examples $\mathcal{D}_0$ and $\mathcal{D}'_0$ we may assume that q' is a conjunction

$$q' = \Diamond \rho'_1 \wedge \ldots \wedge \Diamond \rho'_\ell$$

with each ρ'_j a set of atoms. It should be clear that each $\Diamond \rho'_j$ separates E and there is at least one j with $q \not\models \Diamond \rho'_j$. Thus, we can assume that q' actually takes the form $\Diamond \rho$ for some set of atoms ρ.

By construction of the positive examples $\mathcal{D}_i$, $1 \leq i \leq n$, either $\rho \subseteq \rho_\ell$ for some $1 \leq \ell \leq k$, or for every i at most one of $A_i, \bar{A}_i$ is contained in ρ. As $q \not\models q'$, the first condition does not hold and the latter holds. Thus, $\rho \subseteq \rho_\mathfrak{a}$ for some variable assignment $\mathfrak{a}$. It remains to show that $\mathfrak{a}$ is a model for φ. But this follows directly from $\rho_\mathfrak{a} \not\subseteq \rho_\ell$ for any ρ_ℓ, $1 \leq \ell \leq k$. $\dashv$

We do not know whether Theorem 17 still holds if the number of positive examples is bounded.

Theorem 18. *Unique most specific/general separator verification in $\mathcal{Q}_p[\bigcirc\Diamond]$, $\mathcal{Q}_p[\Diamond]$, $\mathcal{Q}_{in}$ as well as unique most general separator verification in $\mathcal{Q}[\Diamond]$ are* CONP-*complete.*

Proof. For the upper bound we show an NP upper bound for the complement. For given q and E, check that either $\mathcal{Q}$ is not a most general/specific separator (this is in P by Corollary 16) or there exists q' with $q' \not\equiv q$ which separates E (which can be done in NP by guessing q' and then doing the checks in P).

For the lower bounds we show an NP lower bound for the complement. We use that given a propositional formula φ and a satisfying assignment $\mathfrak{a}$ it is NP-hard to decide whether there is satisfying assignment for φ different from $\mathfrak{a}$.

Let $\varphi = \psi_1 \wedge \ldots \wedge \psi_k$ a propositional formula with k clauses over n variables $x_1, \ldots, x_n$. Let $\mathfrak{a}_0$ be a satisfying assignment. We assume without loss of generality that no variable occurs both positively and negatively in a clause. We use $2n$ atoms $A_1, \bar{A}_1, \ldots, A_n, \bar{A}_n$. Define the positive examples $E^+ = \{\mathcal{D}_0, \mathcal{D}'_0, \mathcal{D}_1, \ldots, \mathcal{D}_n\}$ by setting

- $\mathcal{D}_0 = \{A_i(1), \bar{A}_i(1) \mid 1 \leq i \leq n\}$,

- $\mathcal{D}'_0 = \{A_i(2), \bar{A}_i(2) \mid 1 \leq i \leq n\}$,

and including in $\mathcal{D}_i$, for $1 \leq i \leq n$, the assertions:

- $A_i(1)$ and $\bar{A}_i(2)$, and

- $A_j(1), \bar{A}_j(1), A_j(2), \bar{A}_j(2)$ for all $j \neq i$.

We next define $E^- = \{\mathcal{D}^1_1, \ldots, \mathcal{D}^1_k, \mathcal{D}^2_1, \ldots, \mathcal{D}^2_n\}$. In $\mathcal{D}^1_i$, $1 \leq i \leq k$, we include

- $\bar{A}_j(1)$, if x_j does not occur negatively in ψ_i, and

- $A_j(1)$, if x_j does not occur positively in ψ_i.

In $\mathcal{D}^2_i$, $1 \leq i \leq n$, we include

- $A_j(1), \bar{A}_j(1)$, for all $j \neq i$.

Set $E = (E^+, E^-)$. For a variable assignment $\mathfrak{a}$ for $x_1, \ldots, x_n$, we denote with $\rho_\mathfrak{a}$ the set that contains, for all i, A_i, if $\mathfrak{a}(x_i) = 1$ and $\bar{A}_i$, otherwise.

Lemma 32. *(i) For every model $\mathfrak{a}$ of φ, $\Diamond \rho_\mathfrak{a}$ separates E.*

(ii) If $q \in \mathcal{Q}_p[\bigcirc\Diamond]$ separates E, then $q \equiv \Diamond \rho_\mathfrak{a}$ for some model $\mathfrak{a}$ of φ.

Proof. Point (i) is straightforward.

For Point (ii), take any separator $\boldsymbol{q}$ for E. Since the positive example $\mathcal{D}_0$ satisfies $\max \mathcal{D}_0 = 1$, we can assume that $\boldsymbol{q}$ has temporal depth at most 1. It cannot be of temporal depth 0 as then it would not separate. Thus, $\boldsymbol{q}$ has temporal depth 1 and is of shape $\boldsymbol{q} = \boldsymbol{o}\rho$ for some $\rho \subseteq \{A_1, \bar{A}_1, \ldots, A_n, \bar{A}_n\}$ and $\boldsymbol{o} \in \{\Diamond, \bigcirc\}$. Since $\mathcal{D}_0'$ is a positive example, $\boldsymbol{o}$ cannot be $\bigcirc$ and $\boldsymbol{q} = \Diamond\rho$. By construction of the positive examples $\mathcal{D}_i$, $1 \leq i \leq n$, for every i at most one of $A_i, \bar{A}_i$ is contained in ρ. Since all $\mathcal{D}_i^2$ are negative examples, also at least one of $A_i, \bar{A}_i$ is contained in ρ. Thus, $\rho = \rho_{\mathfrak{a}}$ for some variable assignment $\mathfrak{a}$. It remains to show that $\mathfrak{a}$ is a model for φ. Let ψ_i be a clause of φ and assume that $\mathfrak{a} \not\models \psi_i$. Then $\mathfrak{a}(x_j) = 1$ (and hence $A_j \in \rho$ if x_j occurs negatively in ψ_i and $\mathfrak{a}(x_j) = 0$ (and hence $\bar{A}_j \in \rho$) if x_j occurs positively. It is now readily checked that $\mathcal{D}_i^1, 0 \models \Diamond\rho$, a contradiction. ⊣

Observe that the $\Diamond\rho_{\mathfrak{a}}$ are pairwise incomparable w.r.t. containment. Hence, it follows from Lemma 32 that $\Diamond\rho_{\mathfrak{a}_0}$ is not the unique most specific/general separator in $\mathcal{Q}_p[\bigcirc\Diamond]$ (equivalently, $\mathcal{Q}_p[\Diamond]$ or $\mathcal{Q}_{in}$) iff there is an assignment $\mathfrak{a}$ distinct from $\mathfrak{a}_0$ satisfying φ.

For $\mathcal{Q}[\Diamond]$ the same argument applies, but only for unique most general separators. ⊣

Theorem 21. *Let $\mathcal{Q} \subseteq \mathcal{Q}_p[\bigcirc\Diamond]$ be any class containing all $\Diamond\rho$ with ρ a conjunction of atoms. Then*

- *counting most specific/general $\mathcal{Q}$-separators is $\sharp P$-hard.*

- *the existence of unique most specific/general separator is US-hard.*

even for $E = (E^+, E^-)$ and $\sigma = \mathsf{sig}(E)$ with

1. *$|E^-| \leq 1$; or*

2. *$|E^-| \leq 1$, $|\sigma| = 2$ and $\mathcal{Q} = \mathcal{Q}_{in}$ or $\mathcal{Q} \supseteq \mathcal{Q}_p[\Diamond]$; or*

3. *$|E^+| \leq 4$, $|\sigma| = 3$, and $\mathcal{Q} \supseteq \mathcal{Q}_p[\Diamond]$.*

This result also holds for most general separator case for $\mathcal{Q}[\Diamond]$ and $|E^-| = 1$ and $|\sigma| = 2$.

Proof. We start with the proof of Point 1. For the sake of readability, we provide the proof first for an unbounded number of negative examples and argue afterwards that a single negative examples suffices.

We reduce from SAT in essentially the same way as in the lower bound proof for Theorem 18 above. Let $\varphi = \psi_1 \wedge \ldots \wedge \psi_k$ a propositional formula with k clauses over n variables $x_1, \ldots, x_n$. We assume without loss of generality that no variable occurs both positively and negatively in a clause. We use $2n$ atoms $A_1, \bar{A}_1, \ldots, A_n, \bar{A}_n$. Define the examples $E = (E^+, E^-)$ in exactly the same way as in the proof of Theorem 18. Then Lemma 32 holds. Point 1 for an unbounded number of negative examples follows directly.

To finish the proof for $|E^-| = 1$, it remains to join the negative examples into a single one. Recall that, due to the structure of the positive examples, any separator for E has temporal depth at most 1. Based on this observation, it can be verified that Lemma 32 continues to hold when E^- is replaced by $\{\mathcal{D}^-\}$ for the data instance

$$\mathcal{D}^- = \emptyset \mathcal{D}_1^1 \emptyset^2 \cdots \emptyset^2 \mathcal{D}_k^1 \emptyset^2 \mathcal{D}_1^2 \emptyset^2 \cdots \emptyset^2 \mathcal{D}_n^2.$$

We next provide the proof of Point 2. We start by considering $\mathcal{Q}_{in}$. Again, we provide the proof first for an unbounded number of negative examples and argue afterwards that a single negative examples suffices.

$\mathcal{Q}_{in}$ We reduce from SAT. Let $\varphi = \psi_1 \wedge \ldots \wedge \psi_k$ a propositional formula with k clauses over n variables $x_1, \ldots, x_n$. We use three atoms $M, A, \bar{A}$. Define the positive examples $E^+ = \{\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_n\}$ by including in $\mathcal{D}_i$, for $1 \leq i \leq n$ the assertions:

- $M(1)$ and $M(n+1)$, and

- $A(i)$ and $\bar{A}(n+i)$, and

- $A(j), \bar{A}(n+j)$ for all $j \neq i$.

In $\mathcal{D}_0$, we include

- $M(1)$, and

- $A(i), \bar{A}(i)$, for all $1 \leq i \leq n$.

We next define negative examples $E^- = \{\mathcal{D}^0, \mathcal{D}_1^1, \ldots, \mathcal{D}_k^1, \mathcal{D}^2, \mathcal{D}_1^3, \ldots, \mathcal{D}_n^3\}$. In $\mathcal{D}^0$, we include

- $M(i), A(i), \bar{A}(i)$, for all $1 \leq i \leq n-1$.

In $\mathcal{D}_i^1$, $1 \leq i \leq k$, we include

- $M(1)$, and

- $\bar{A}(j)$, if x_j does not occur negatively in ψ_i, and

- $A(j)$, if x_j does not occur positively in ψ_i.

In $\mathcal{D}^2$, we include

- $A(j), \bar{A}(j)$, for all $1 \leq j \leq n$.

In $\mathcal{D}_i^3$, $1 \leq i \leq n$, we include

- $M(1)$,

- $A(j), \bar{A}(j)$, for all $i \neq j$.

Set $E = (E^+, E^-)$. For a variable assignment $\mathfrak{a}$ for $x_1, \ldots, x_n$, we denote with $\boldsymbol{q}_{\mathfrak{a}}$ the query

$$\boldsymbol{q}_{\mathfrak{a}} = \diamondsuit(\rho_{\mathfrak{a}}^1 \wedge (\bigcirc \rho_{\mathfrak{a}}^2 \wedge (\ldots \wedge \bigcirc \rho_{\mathfrak{a}}^n) \ldots))$$

where $\rho_{\mathfrak{a}}^1 = \{M, A\}$ if $\mathfrak{a}(x_1) = 1$ and $\rho_{\mathfrak{a}}^1 = \{M, \bar{A}\}$ otherwise, and for all $i > 1$, $\rho_{\mathfrak{a}}^i = \{A\}$ if $\mathfrak{a}(x_i) = 1$ and $\rho_{\mathfrak{a}}^i = \{\bar{A}\}$, otherwise.

Lemma 33. *(i) For every model $\mathfrak{a}$ of φ, $\boldsymbol{q}_{\mathfrak{a}}$ separates E.*

(ii) If $\boldsymbol{q} \in \mathcal{Q}_{in}$ separates E, then $\boldsymbol{q} \equiv \boldsymbol{q}_{\mathfrak{a}}$ for some model $\mathfrak{a}$ of φ.

(iii) Every $\mathcal{Q}_{in}$-separator for E is both a most general and most specific separator.

Proof. Point (iii) is immediate from Points (i) and (ii) and the fact that all $\boldsymbol{q}_{\mathfrak{a}}$ are pairwise incomparable w.r.t. containment. We hence concentrate on Points (i) and (ii).

Point (i) is straightforward.

For Point (ii), take any separator

$$\boldsymbol{q} = \diamondsuit(\rho^1 \wedge (\bigcirc \rho^2 \wedge (\ldots \wedge \bigcirc \rho^m) \ldots))$$

for E, for some $m > 0$ and sets ρ^i. Since $\mathcal{D}^0$ is a negative example for $\boldsymbol{q}$, we have $m \geq n$. Since $\mathcal{D}_0$ is a positive example for $\boldsymbol{q}$, we have that $m \leq n$, hence $m = n$. Then observe that some ρ^i has to contain M, since otherwise $\mathcal{D}^2, 0 \models \boldsymbol{q}$. By $\mathcal{D}_0$ again, it can only be ρ^1.

By construction of the positive examples, for every i at most one of $A, \bar{A}$ is contained in ρ^i. Since all $\mathcal{D}_i^3$ are negative examples, each ρ_i contains also at least one of $A, \bar{A}$. Thus, $\boldsymbol{q} = \boldsymbol{q}_{\mathfrak{a}}$ for some variable assignment $\mathfrak{a}$.

It remains to show that $\mathfrak{a}$ is a model for φ. Let ψ_i be a clause of φ and assume that $v \not\models \psi_i$ for some i. Then $\mathfrak{a}(x_j) = 1$ (and hence $A \in \rho_j$) if x_j occurs negatively in ψ_i and $\mathfrak{a}(x_j) = 0$ (and hence $\bar{A} \in \rho_j$) if x_j occurs positively. It is now readily checked that $\mathcal{D}_i^1, 0 \models \boldsymbol{q}_{\mathfrak{a}}$, a contradiction. $\dashv$

It remains to join the negative examples into a single one. Recall that, because of $\mathcal{D}_0 \in E^+$, any separator for E has temporal depth at most n. Based on this observation, it can be verified that Lemma 33 continues to hold when E^- is replaced by $\{\mathcal{D}^-\}$ for the data instance

$$\mathcal{D}^- = \emptyset \mathcal{D}^0 \emptyset^n \mathcal{D}_1^1 \emptyset^n \cdots \emptyset^n \mathcal{D}_k^1 \emptyset^n \mathcal{D}^2 \emptyset^n \mathcal{D}_1^3 \emptyset^n \cdots \emptyset^n \mathcal{D}_n^3.$$

We next assume that $\mathcal{Q}_p[\diamondsuit] \subseteq \mathcal{Q} \subseteq \mathcal{Q}_p[\bigcirc\diamondsuit]$.

We employ a reduction of 3SAT to a longest common subsequence problem, which is defined as follows: Given a set S of sequences and a number m, decide whether the sequences in S have a common subsequence of length m. We use the following result [Blin *et al.*, 2012, Proposition 6]. Given a formula $\varphi = \psi_1 \wedge \cdots \wedge \psi_k$ with k clauses over n variables $x_1, \ldots, x_n$, one can construct in polynomial time a set S of sequences over the alphabet $\{H, T\}$ such that

- S contains $k + 2$ sequences;

- there is a poly-time computable bijection f from the set L of common subsequences of S of length exactly $11n$ onto the set V of variable assignments $\mathfrak{a}$ satisfying φ;

- no common subsequence of the sequences in S of length $> 11n$ exists.

Thus, given φ as above, we use S to construct $E = (E^+, E^-)$. For any sequence $s_1 \cdots s_\ell \in S$ we include in E^+ the data instance $\emptyset\{s_1\} \cdots \{s_\ell\}$. We also include in E^+ the data instances $\mathcal{D}_1^+ = \emptyset\{H, T\}^{11n}$ and $\mathcal{D}_2^+ = \emptyset(\emptyset^{11n}\{H, T\})^{11n}$ and include in E^- the data instances $\mathcal{D}_1^- = \emptyset\{H, T\}^{11n-1}$ and $\mathcal{D}_2^- = \emptyset(\emptyset^{11n}\{H, T\})^{11n-1}$.

Claim 1. Any query $\boldsymbol{q} \in \mathcal{Q}$ separating E is a $\diamondsuit$-path sequence query and the sequence defined by $\boldsymbol{q}$ is a common subsequence of the sequences in S.

Proof of Claim 1. To prove Claim 1 suppose $\boldsymbol{q} \in \mathcal{Q}$ of the form

$$\boldsymbol{q} = \rho_0 \wedge \boldsymbol{o}_1(\rho_1 \wedge \boldsymbol{o}_2(\rho_2 \wedge \cdots \wedge \boldsymbol{o}_\ell \rho_\ell)), \tag{4}$$

separates E. By definition of separation it is clear that $\rho_0 = \top$ and $|\rho_i| \leq 1$. From $\mathcal{D}_1^+ \in E^+$ it follows that $\ell \leq 11n$ and from $\mathcal{D}_1^- \in E^-$ it follows that $\ell \geq 11n$. Hence $\ell = 11n$.

It follows from $\mathcal{D}_2^+ \in E^+$ and $\ell = 11n$ that $\boldsymbol{o}_i = \Diamond$ for all $i \leq \ell$. We obtain from $\mathcal{D}_2^- \in E^-$ that $|\rho_i| \geq 1$ for all $i \geq 1$. Thus, $\boldsymbol{q}$ is a $\Diamond$-path sequence query of length $11n$. Since we included one positive example for every sequence in S, the sequence corresponding to $\boldsymbol{q}$ is a common subsequence of the sequences in S. This finishes the proof of Claim 1.

The following claim follows directly from the definition of E.

Claim 2. For every $s_1 \ldots s_\ell \in S$, the corresponding $\Diamond$-path query $\Diamond(s_1 \wedge \Diamond(s_2 \cdots \Diamond s_\ell))$ separates E.

It follows from Claims 1 and 2 and the fact that all sequence $\Diamond$-path queries of the same temporal depth are pairwise incomparable w.r.t. containment, that there is a poly-time computable bijection f from the set L' of queries in $\mathcal{Q}$ separating E onto the set V of satisfying assignments for φ.

We now come to the proof of Point 3. We use an existing reduction showing NP-hardness of the *consistent subsequence problem*, defined as follows. Given two sets P, N of sequences, decide whether there is a sequence s *consistent with* P, N, that is, s is a subsequence of each sequence in P, but of no sequence in N.

We use the following result [Fraser, 1996, Theorem 2.1]. Given a 3CNF formula $\varphi = \psi_1 \wedge \ldots \wedge \psi_k$ with k clauses over n variables, one can construct in polynomial time two sets P, N of sequences over an alphabet $\sigma = \{\#, 0, 1\}$ such that

- P contains two sequences;

- there is a poly-time computable bijection f from the set L of sequences that are consistent with P, N of length $6k$ onto the set of models of φ;

- any sequence that is consistent with P, N has length exactly $6k$.

Thus, given φ as above, we use P, N to construct $E = (E^+, E^-)$. For any sequence $s_1 \cdots s_\ell \in P$ we include in E^+ the data instance $\emptyset\{s_1\} \cdots \{s_\ell\}$. Similarly, for any sequence $s_1 \cdots s_\ell \in N$, we include in E^- the data instance $\emptyset\{s_1\} \cdots \{s_\ell\}$. We also include in E^+ the data instances $\mathcal{D}_1^+ = \emptyset\{0, 1, \#\}^{6k}$ and $\mathcal{D}_2^+ = \emptyset(\emptyset^{6k}\{0, 1, \#\})^{6k}$, and include in E^- the data instances $\mathcal{D}_1^- = \emptyset(\{0, 1, \#\})^{6k-1}$ and $\mathcal{D}_2^- = \emptyset(\emptyset^{6k}\{0, 1, \#\})^{6k-1}$.

Assume now that a set $\mathcal{Q}$ with $\mathcal{Q}_p[\Diamond] \subseteq \mathcal{Q} \subseteq \mathcal{Q}_p[\bigcirc\Diamond]$ is given.

Claim 1. Any query $\boldsymbol{q} \in \mathcal{Q}$ separating E is a $\Diamond$-path sequence query and the sequence defined by $\boldsymbol{q}$ is consistent with P, N.

Proof of Claim 1. Suppose $\boldsymbol{q} \in \mathcal{Q}$ of the form

$$\boldsymbol{q} = \rho_0 \wedge \boldsymbol{o}_1(\rho_1 \wedge \boldsymbol{o}_2(\rho_2 \wedge \cdots \wedge \boldsymbol{o}_\ell \rho_\ell))$$

separates E. By definition of separation it is clear that $\rho_0 = \top$ and $|\rho_i| \leq 1$. From $\mathcal{D}_1^+ \in E^+$ it follows that $\ell \leq 6k$ and from $\mathcal{D}_1^- \in E^-$ it follows that $\ell \geq 6k$. Hence $\ell = 6k$. From $\mathcal{D}_2^- \in E^-$ it also follows that $\rho_i \neq \emptyset$ for all $i > 1$, hence $\boldsymbol{q}$ is a $\bigcirc\Diamond$-path sequence query.

It follows from $\mathcal{D}_2^+ \in E^+$ and $\ell = 6k$ that $\boldsymbol{o}_i = \Diamond$ for all $i \leq \ell$. We have thus shown that $\boldsymbol{q}$ is a $\Diamond$-path sequence query of length $6k$. Finally, it is easy to see that the corresponding sequence is consistent with P, N. This finishes the proof of Claim 1.

The following claim follows directly from the definition of E.

Claim 2. For every sequence consistent with P, N, the corresponding $\Diamond$-path query separates E.

It follows from Claims 1 and 2 to that we have a poly-time computable bijection f from the set L' of queries in $\mathcal{Q}$ separating E onto the set of models of φ. Moreover, the queries in L' are all $\Diamond$-path sequence queries of the same length and therefore incomparable w.r.t. $\models$.

We finally come to $\mathcal{Q}[\Diamond]$. Observe that our claim follow directly from Point 2 for $\mathcal{Q}_p[\Diamond]$: as we have only one negative example, any separating query in $\mathcal{Q}[\Diamond]$ contains a conjunct in $\mathcal{Q}_p[\Diamond]$ that separates. Hence the most general separating queries in $\mathcal{Q}[\Diamond]$ and $\mathcal{Q}_p[\Diamond]$ coincide (up to logical equivalence). So the $\sharp$P-lower bound established for $\mathcal{Q}_p[\Diamond]$ for counting most general separating queries and the US-lower bound for the existence of a unique most general separator also hold for $\mathcal{Q}[\Diamond]$. $\dashv$

Theorem 22. *Let $\mathcal{Q} = \mathcal{Q}[\Diamond]$ or $\mathcal{Q} \subseteq \mathcal{Q}_p[\bigcirc\Diamond]$ be any class containing all $\Diamond\rho$ with a conjunction of atoms ρ. Then counting most general $\mathcal{Q}$-separators is $\sharp$P-hard even for example sets $E = (E^+, E^-)$ with $|E^+| = 2$ and $|E^-| = 1$.*

Proof. By reduction from counting the number of satisfying assignments for a monotone propositional formula (that is, a formula in conjunctive normal form without negation). We re-use data instances introduced before.

Take a monotone Boolean formula $\varphi = \psi_1 \wedge \cdots \wedge \psi_k$ with clauses ψ_i over variables $x_1, \ldots, x_n$. We construct $E = (E^+, E^-)$ such that there is a bijection between the satisfying assignments for φ and the most general separators for E in $\mathcal{Q}_p[\bigcirc\Diamond]$ (even queries of the form $\Diamond\rho$). Define the positive examples $E^+ = \{\mathcal{D}_0, \mathcal{D}_0'\}$ with $2n$ atoms $A_1, \bar{A}_1, \ldots, A_n, \bar{A}_n$ by taking

$$\mathcal{D}_0 = \{A_i(1), \bar{A}_i(1) \mid 1 \le i \le n\},$$
$$\mathcal{D}_0' = \{A_i(2), \bar{A}_i(2) \mid 1 \le i \le n\}$$

Let $E^- = \{\mathcal{D}^-\}$ with

$$\mathcal{D}^- = \emptyset\mathcal{D}_1^1\emptyset^2 \cdots \emptyset^2\mathcal{D}_k^1\emptyset^2\mathcal{D}_1^2\emptyset^2 \cdots \emptyset^2\mathcal{D}_n^2.$$

where $\mathcal{D}_i^1$, $1 \le i \le k$, comprises all $\bar{A}_j(1)$ and $A_j(1)$ if x_j does not occur positively in ψ_i, and $\mathcal{D}_i^2 = \{A_j(1), \bar{A}_j(1) \mid j \ne i\}$.

For a variable assignment $\mathfrak{a}$ for $x_1, \ldots, x_n$, we denote with $\rho_\mathfrak{a}$ the set that contains, for all i, A_i, if $\mathfrak{a}(x_i) = 1$ and $\bar{A}_i$, otherwise. Now one can easily show:

(i) For every model $\mathfrak{a}$ of φ, $\Diamond\rho_\mathfrak{a}$ separates E.

(ii) If $q \in \mathcal{Q}_p[\bigcirc\Diamond]$ separates E, then $q \equiv \Diamond(\rho_\mathfrak{a} \cup \rho)$ for some model $\mathfrak{a}$ of φ and some set ρ of atoms of the form $\bar{A}_i$.

The claim for $\mathcal{Q} \subseteq \mathcal{Q}_p[\bigcirc\Diamond]$ follows. The claim for $\mathcal{Q}[\Diamond]$ also follows since we have only one negative example and so any most general separator is in $\mathcal{Q}[\Diamond]$. $\dashv$

We show the properties of $\mathcal{Q}_{in}$ that are stated in Section 6.

Theorem 34. *(1) For $\mathcal{Q}_{in}$, counting most specific separators is in P if $|E^+|$ is bounded.*
(2) For example data instances $E = (E^+, E^-)$ that contain sequences, counting most specific sequence queries in $\mathcal{Q}_{in}$ that separate E is in P.

Proof. We extend the proof of Theorem 29.
(1) Assume $E^+ = \{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$. Let Ξ be the set of queries defined in the proof of Theorem 29. Then the most specific queries in Ξ that separate E coincide with the most specific queries in $\mathcal{Q}_{in}$ that separate E, and so can be counted in polytime.
(2) $E^+ = \{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$ be a set of sequences. Assume

$$\mathcal{D}_1 = \{A_0\}\{A_1\}\cdots\{A_m\}$$

Then any possible separating sequence query in $\mathcal{Q}_{in}$ takes the form

$$\Diamond(A_{k_1} \wedge \bigcirc(\cdots\bigcirc A_{k_2}))$$

with $0 < k_1 \le k_2 \le m$. Clearly the most specific queries of this form that separate E can be counted in P. $\dashv$

E Proofs for Section 7

Let $t_\mathcal{D}(n) = \{A \mid A(n) \in \mathcal{D}\}$. It will be convenient to assume that $t_{\mathcal{E}_j}(\infty) = \sigma$, for $E_j \in E^-$, where σ is the set of symbols occurring in E^+.

Lemma 24. $s(E, \mathcal{Q}) \ne \emptyset$ iff $\mathfrak{P} \otimes \mathfrak{N}$ *contains a separating path π, with q_π separating E.*

Proof. ($\Rightarrow$) Suppose q of the form $\rho_0 \wedge \Diamond(\rho_1 \wedge (\ldots \Diamond\rho_\ell)\ldots)$ separates E. If $\ell = 0$, then $\rho_0 \subseteq l(\bar{0})$ and $\rho_0 \not\subseteq t_{\mathcal{E}_j}(0)$ for all $1 \le j \le c_-$. It follows that $l(\bar{0}) \ne \emptyset$ and $(\bar{0}, \infty)$ is in $\mathfrak{P} \times \mathfrak{N}$, which completes the proof. Suppose $\ell > 0$. Further, we assume that $\rho_0 \wedge \Diamond(\rho_1 \wedge (\ldots \Diamond\rho_{\ell-1})\ldots)$ does not separate E, which implies $\rho_\ell \ne \emptyset$. We define $n^0 = \bar{0}$. Suppose $n^i = (n_1, \ldots, n_{c_+})$ has been defined, we define $n^{i+1} = (n_1', \ldots, n_{c_+}')$ where $n_j' = \min\{n \mid n > n_j, \rho_{i+1} \subseteq t_{\mathcal{D}_j}(n)\}$, for $0 \le i < \ell$. It follows that $l(n^\ell) \ne \emptyset$. Then we take m^0 to be the root of $\mathfrak{N}$ and set m^{i+1} to be m' such that $(n^i, m^i) \to (n^{i+1}, m')$. We claim that the path $(n^0, m^0), \ldots, (n^\ell, m^\ell)$ is such that $m^\ell = \infty$. To prove the claim, we introduce some notation. For a data instance $\mathcal{E}$ and $t \ge 0$, denote by $\mathcal{E}^{\le t}$ the data instance obtained from $\mathcal{E}$ by removing all assertions $A(n)$ with $n > t$. Set $m_j^i = \min\{t \mid t \ge i \text{ and } \mathcal{E}_j^{\le t} \models \rho_0 \wedge \Diamond(\rho_1 \wedge (\ldots \Diamond\rho_i)\ldots)\}$ or $m_j^i = \infty$ if t as above does not exist, for $0 \le i \le c_-$ and $1 \le j \le l$. Let $m_-^i = (m_1^i, \ldots, m_{c_-}^i)$. It follows that $m_-^\ell = \infty$. It is easily shown by induction on i that $m_-^i = m^i$, which completes the proof.
($\Leftarrow$) Let $(n^0, m^0), \ldots, (n^\ell, m^\ell)$ be a path with $n^0 = \bar{0}$, $l(n^\ell) \ne \emptyset$ and $m^\ell = \infty$. We will show that $q = \rho_0 \wedge \Diamond(\rho_1 \wedge (\ldots \Diamond\rho_\ell)\ldots)$ with $\rho_i = l(n^i)$ separates E. Indeed, clearly $\mathcal{D} \models q$ for each $\mathcal{D} \in E^+$. By induction on i, we show that $m_-^i = m^i$. It follows that $\mathcal{E} \not\models q$ for all $\mathcal{E} \in E^-$ which completes the proof. $\dashv$

Lemma 38. *If q of the form* (1) *separates E, then there is a separating path of the form* (2) *with $\rho_i \subseteq l(n_i)$, for all $i \le n$.*

Proof. Straightforwardly follows from the construction in $(\Rightarrow)$ above. $\dashv$

Lemma 26. *A $\mathcal{Q}_p[\Diamond]$-query $\boldsymbol{q}$ is a unique most specific $\mathcal{Q}_p[\Diamond]$-separator for E iff there is a separating path π such that $\boldsymbol{q} = \boldsymbol{q}_\pi$ and $\boldsymbol{q}_\pi \models \boldsymbol{q}_\nu$, for every separating path ν.*

Proof. $(\Rightarrow)$ Let $\boldsymbol{q} = \rho_0 \wedge \Diamond(\rho_1 \wedge (\ldots \Diamond \rho_\ell)\ldots)$ be a unique most specific separator. Consider the path $\pi = (\boldsymbol{n}^0, \boldsymbol{m}^0), \ldots, (\boldsymbol{n}^\ell, \boldsymbol{m}^\ell)$ constructed in the proof of Lemma 24 $(\Rightarrow)$. Clearly, π is a separating path. Because $\boldsymbol{q}$ is a most specific separator and $\rho_i \subseteq \boldsymbol{l}(\boldsymbol{n}^i)$, it follows that $\boldsymbol{l}(\boldsymbol{n}^i) = \rho_i$, and so $\boldsymbol{q} = \boldsymbol{q}_\pi$. Let ν be an arbitrary separating path in $\mathfrak{P} \times \mathfrak{N}$. From the proof of Lemma 24 $(\Leftarrow)$ we obtain that $\boldsymbol{q}_\nu$ is a separator. It follows that $\boldsymbol{q}_\pi \models \boldsymbol{q}_\nu$.

$(\Leftarrow)$ Let π be a separating path such that $\boldsymbol{q}_\pi \models \boldsymbol{q}_\nu$ for every separating path ν in $\mathfrak{P} \otimes \mathfrak{N}$. Let $\boldsymbol{q} = \rho_0' \wedge \Diamond(\rho_1' \wedge (\ldots \Diamond \rho_{\ell'}')\ldots)$ be a separator (for E), we only need to show that $\boldsymbol{q}_\pi \models \boldsymbol{q}'$. Consider the separating path $\nu' = (\boldsymbol{n}^0, \boldsymbol{m}^0), \ldots, (\boldsymbol{n}^{\ell'}, \boldsymbol{m}^{\ell'})$ constructed in the proof of Lemma 24 $(\Rightarrow)$ for $\boldsymbol{q}'$. Clearly, $\rho_i' \subseteq \boldsymbol{l}(\boldsymbol{n}^i)$. Since we have $\boldsymbol{q}_\pi \models \boldsymbol{q}_{\nu'}$, we also obtain $\boldsymbol{q}_\pi \models \boldsymbol{q}'$. $\dashv$

We now explain how checking the existence of a separating path π such that $\boldsymbol{q}_\pi \models \boldsymbol{q}_\nu$ for every separating path ν in $\mathfrak{P} \otimes \mathfrak{N}$ can be checked in P. We compute the required separator $\boldsymbol{q}_\pi$ by analysing the queries associated with the paths leading to each node $(\boldsymbol{n}, \boldsymbol{m})$ of $\mathfrak{P} \otimes \mathfrak{N}$ from its root $(\bar{0}, \boldsymbol{m}_0)$. We mark inductively each $(\boldsymbol{n}, \boldsymbol{m})$ either by a query $\boldsymbol{q}_\pi$, for some path π, or by ε. The root is marked by $\boldsymbol{q}_\pi$ for $\pi = (\bar{0}, \boldsymbol{m}_0)$. Now, assume that all immediate predecessors of $(\boldsymbol{n}, \boldsymbol{m})$ have already been marked. Let $\boldsymbol{q}_{\pi_i}$, $i = 1, \ldots, k$, be all of their marks different from ε. Consider the extended paths $\pi_i' = \pi_i, (\boldsymbol{n}, \boldsymbol{m})$ and the set Π of all paths to $(\boldsymbol{n}, \boldsymbol{m})$. We check whether there is a containment witness for $\boldsymbol{q}_{\pi_1'}$ and every $\boldsymbol{q}_\pi$, $\pi \in \Pi$, in which case we mark $(\boldsymbol{n}, \boldsymbol{m})$ by $\boldsymbol{q}_{\pi_1'}$. Otherwise, we do the same for $\boldsymbol{q}_{\pi_2'}$, and so on. If $(\boldsymbol{n}, \boldsymbol{m})$ has no mark after k iterations, we mark it by ε. Although Π can be of exponential size, we show that marking each node can be done in P. Indeed, consider the following problem: given $\boldsymbol{q}_{\pi_1'}$ and $(\boldsymbol{n}, \boldsymbol{m})$, decide if there exists $\pi \in \Pi$ such that there is no containment witness for $\boldsymbol{q}_{\pi_1'}$ and $\boldsymbol{q}_\pi$. This problem is in NL because such π, if exists, can be computed by a non-deterministic algorithm guessing π of the form (2) node-by-node. At each step $i \in \{0, \ldots, n\}$ we only need to remember the minimal $h(i)$ (in the set $\{0, \ldots, n'-1\}$ for the length n' of π_1') where h is a containment witness for $\boldsymbol{q}_{\pi_1'}$ and $\boldsymbol{q}_{(\boldsymbol{n}_0, \boldsymbol{m}_0), \ldots, (\boldsymbol{n}_i, \boldsymbol{m}_i)}$. If at some step it is impossible to extend $h(i)$ to $h(i+1)$ and $(\boldsymbol{n}, \boldsymbol{m})$ is reachable from $(\boldsymbol{n}_{i+1}, \boldsymbol{m}_{i+1})$, we have obtained $\pi \in \Pi$ such that there is no containment witness for $\boldsymbol{q}_{\pi_1'}$ and $\boldsymbol{q}_\pi$. Because NL = CONL and the latter problem is in NL, we conclude that checking whether there is a containment witness for $\boldsymbol{q}_{\pi_1'}$ and $\boldsymbol{q}_\pi$ for every $\pi \in \Pi$ is in NL.

Then we consider all nodes $(\boldsymbol{n}_i, \bar{\infty})$, $i = 1, \ldots, l$, with $\boldsymbol{l}(\boldsymbol{n}_i) \neq \emptyset$ that are marked by some $\boldsymbol{q}_i \neq \varepsilon$. If there are no such, there is no unique most specific separator. Otherwise, take the set Π of all paths leading to the $(\boldsymbol{n}_i, \bar{\infty})$. We check whether there is a containment witness for $\boldsymbol{q}_1$ and every $\boldsymbol{q}_\pi$, $\pi \in \Pi$, in which case $\boldsymbol{q}_1$ is returned as a unique most specific separator. Otherwise, we do the same for $\boldsymbol{q}_2$, and so on.

Theorem 23. *Let $E = (E^+, E^-)$, $\sigma = \mathsf{sig}(E)$, t_+/t_- be the maximum timestamp in E^+/E^-, $c_+ = |E^+|$, $c_- = |E^-|$, and $\mathcal{Q} \in \{\mathcal{Q}_p[\Diamond], \mathcal{Q}_p[\bigcirc\Diamond]\}$. The following can be done in time $O(t_+^{c_+} t_-^{c_-})$:*

(a) *deciding whether $\mathsf{s}(E, \mathcal{Q}) \neq \emptyset$;*

(b) *computing a longest/shortest separator in $\mathsf{s}(E, \mathcal{Q})$;*

(c) *deciding the existence of a unique most specific $\mathcal{Q}$-separator and a unique most general $\mathcal{Q}_p[\Diamond]$-separator, and constructing such a separator.*
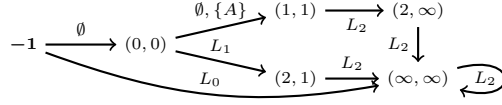
For bounded c_+ and c_-, problem (a) is in NL.

Proof. We first prove (c) for unique most general $\mathcal{Q}_p[\Diamond]$-separators. Computing unique most general separators requires a different type of graph. One might think that inverting $\boldsymbol{q}_\pi \models \boldsymbol{q}_\nu$ in Lemma 26 would give a criterion for $\boldsymbol{q}_\pi$ being such a separator. To show that this is not so, take $E^+ = \{\{A(1), B(1)\}\}$ and $E^- = \{\{C(0)\}\}$. In this case, $\mathfrak{P} \otimes \mathfrak{N}$ looks as follows:

$$\emptyset \quad ((0), (0)) \longrightarrow ((1), \bar{\infty}) \quad \{A, B\}$$

It contains only one separating path. However, there is no unique most general separator. The issue is that the labels of the paths π leading to $(\boldsymbol{n}, \bar{\infty})$ represent some of the separators, including most specific ones, but not all separators. In our example, both $\Diamond A$ and $\Diamond B$ are separators.

To handle unique most general separators, we need another type of a labelled graph. To this end, we redefine $\mathfrak{P}$ ($\mathfrak{N}$) with the tuples $\boldsymbol{n}$ (respectively, $\boldsymbol{m}$) as above as nodes and add the new node $-\bar{1}$. The edges of $\mathfrak{P}$ ($\mathfrak{N}$) are labelled with $\rho \subseteq \sigma$. Denote by $\boldsymbol{l}(\boldsymbol{n})$ ($\boldsymbol{l}(\boldsymbol{m})$) the label of $\boldsymbol{n}$ in $\mathfrak{P}$ (respectively, $\boldsymbol{m}$ in $\mathfrak{N}$) defined above. First, in $\mathfrak{P}$, we set $-\bar{1} \to_\rho \bar{0}$ for $\rho = \boldsymbol{l}(\bar{0})$. Then, we set $\boldsymbol{n} \to_\rho \boldsymbol{n}'$ for all $\boldsymbol{n}, \boldsymbol{n}'$ and ρ satisfying $\rho \subseteq \boldsymbol{l}(\boldsymbol{n})$ and $\rho \not\subseteq \boldsymbol{l}(\boldsymbol{n}'')$ for all $\boldsymbol{n}'' \in \nabla_{\boldsymbol{n}, \boldsymbol{n}'}$, where $\nabla_{\boldsymbol{n}, \boldsymbol{n}'}$ is the set of all $\boldsymbol{n}''$ satisfying $\boldsymbol{n} \to \boldsymbol{n}''$ and $n_i'' < n_i'$, $n_i'' \neq n_i'$, for some i (see the definition of $\boldsymbol{n} \to \boldsymbol{n}''$ above). In $\mathfrak{N}$, we set $-\bar{1} \to_\rho \boldsymbol{m}$, for all ρ and $\boldsymbol{m}$ satisfying $m_i = 0$ if $\rho \subseteq \{A \mid A(0) \in \mathcal{D}_i\}$, where $E^- = \{\mathcal{D}_i \mid i \in I\}$ for some I, and $m_i = \infty$ otherwise. Then, we set $\boldsymbol{m} \to_\rho \boldsymbol{m}'$ using the same definition as for $\mathfrak{P}$ with $\boldsymbol{m}, \boldsymbol{m}'$ in place of $\boldsymbol{n}, \boldsymbol{n}'$. For example, for $E^- = \{\mathcal{D}_1, \mathcal{D}_2\}$ with $\mathcal{D}_1 = \{A(1), A(2), B(2), C(2)\}$, $\mathcal{D}_2 = \{A(1), B(1), C(1)\}$, and $\sigma = \{A, B, C\}$, we have the following $\mathfrak{N}$:

$$-\mathbf{1} \xrightarrow{\ \emptyset\ } (0,0) \xrightarrow{\ L_1\ } \begin{array}{c} \emptyset, \{A\} \\ \end{array} \quad (1,1) \xrightarrow{\ L_2\ } (2,\infty)$$

The diagram shows: $-\mathbf{1} \xrightarrow{\emptyset} (0,0)$; $(0,0) \xrightarrow{\emptyset,\{A\}} (1,1) \xrightarrow{L_2} (2,\infty)$; $(1,1) \xrightarrow{L_1}$; $(0,0) \xrightarrow{L_0} (2,1)$; $(2,\infty) \xrightarrow{L_2} (\infty,\infty)$; $(2,1) \xrightarrow{L_2} (\infty,\infty)$; $(\infty,\infty) \circlearrowleft L_2$.

with $L_0 \in 2^\sigma \setminus \{\emptyset\}$, $L_1 \in \{\{B\}, \{C\}, \{B,C\}, \{A,B\}, \{A,C\}, \Sigma\}$, and $L_2 = 2^\sigma$ (we omit nodes unreachable from $-\bar{1}$ such as $(\infty,0)$). We now define $\mathfrak{P} \times \mathfrak{N}$ as a standard product of $\mathfrak{P}$ with $\mathfrak{N}$ (with a edges $(\boldsymbol{n},\boldsymbol{m}) \to_\rho (\boldsymbol{n}',\boldsymbol{m}')$ for all nodes such that $\boldsymbol{n} \to_\rho \boldsymbol{n}'$ in $\mathfrak{P}$ and $\boldsymbol{m} \to_\rho \boldsymbol{m}'$) in $\mathfrak{N}$).

Given a path

$$\pi = -\mathbf{1} = (-\bar{1}, -\bar{1}) \to_{\rho_0} (\boldsymbol{n}_0, \boldsymbol{m}_0) \cdots \to_{\rho_n} (\boldsymbol{n}_n, \boldsymbol{m}_n) \tag{5}$$
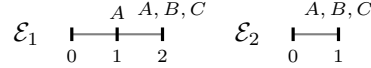
let $\boldsymbol{q}_\pi$ be the $\mathcal{Q}_p[\Diamond]$-query of the form (1) (with matching ρ_i). We call π a *separating path* for E if $\boldsymbol{m}_n = \bar{\infty}\infty$ and $\boldsymbol{l}(\boldsymbol{n}_n) \neq \emptyset$. Now, it directly follows from the construction that

- a query $\boldsymbol{q}$ is a separator iff there exists a separating path π in $\mathfrak{P} \times \mathfrak{N}$ such that $\boldsymbol{q} = \boldsymbol{q}_\pi$

and the following holds (cf. Lemma 26):

Lemma 35. *A $\mathcal{Q}_p[\Diamond]$-query $\boldsymbol{q}$ is a unique most general $\mathcal{Q}_p[\Diamond]$-separator for E iff there is a separating path π such that $\boldsymbol{q} = \boldsymbol{q}_\pi$ and $\boldsymbol{q}_\pi \models \boldsymbol{q}_\nu$, for every separating path ν.*

We could then use a marking algorithm analogous to that for finding a strongest separator. The problem is that, unless σ is fixed, $\mathfrak{P} \otimes \mathfrak{N}$ has exponentially many edges w.r.t. $|\sigma|$. To provide an algorithm with the complexity $O(|\sigma| t_+^{c+} t_-^{c-})$, we define $\mathfrak{P} \otimes \mathfrak{N}$ with the same vertices as $\mathfrak{P} \times \mathfrak{N}$ but with two sorts of labelled edges $\hookrightarrow_\rho$ and $\rightsquigarrow_\rho$ from $-\mathbf{1}$ to $(\bar{0}, \boldsymbol{m})$ and from $(\boldsymbol{n}, \boldsymbol{m})$ to $(\boldsymbol{n}', \boldsymbol{m}')$. Set $\rho_{-\mathbf{1},\boldsymbol{m}} = \bigcap_{-\mathbf{1} \to_\rho (\bar{0},\boldsymbol{m})} \rho$ and $\rho_{\boldsymbol{n},\boldsymbol{m},\boldsymbol{n}',\boldsymbol{m}'} = \bigcap_{(\boldsymbol{n},\boldsymbol{m}) \to_\rho (\boldsymbol{n}',\boldsymbol{m}')} \rho$. To define $\hookrightarrow$, we let $(\boldsymbol{n},\boldsymbol{m}) \hookrightarrow_{\rho_{\boldsymbol{n},\boldsymbol{m},\boldsymbol{n}',\boldsymbol{m}'}} (\boldsymbol{n}',\boldsymbol{m}')$ if $(\boldsymbol{n},\boldsymbol{m}) \to_{\rho_{\boldsymbol{n},\boldsymbol{m},\boldsymbol{n}',\boldsymbol{m}'}} (\boldsymbol{n}',\boldsymbol{m}')$. To define $\rightsquigarrow$, we let $(\boldsymbol{n},\boldsymbol{m}) \rightsquigarrow_{\rho_{\boldsymbol{n},\boldsymbol{m},\boldsymbol{n}',\boldsymbol{m}'}} (\boldsymbol{n}',\boldsymbol{m}')$ if $(\boldsymbol{n},\boldsymbol{m}) \to_\rho (\boldsymbol{n}',\boldsymbol{m}')$ for some ρ and $(\boldsymbol{n},\boldsymbol{m}) \not\to_{\rho_{\boldsymbol{n},\boldsymbol{m},\boldsymbol{n}',\boldsymbol{m}'}} (\boldsymbol{n}',\boldsymbol{m}')$. The definition of $-\mathbf{1} \hookrightarrow_\rho (\bar{0},\boldsymbol{m})$ and $-\mathbf{1} \rightsquigarrow_\rho (\bar{0},\boldsymbol{m})$ is analogous and left to the reader. To illustrate, let $E^- = \{\mathcal{E}_1, \mathcal{E}_2\}$ be as below

$$\mathcal{E}_1 \quad \underset{0 \quad 1 \quad 2}{\overset{A \quad A,B,C}{\longmapsto\!\!\longmapsto}} \qquad \mathcal{E}_2 \quad \underset{0 \quad 1}{\overset{A,B,C}{\longmapsto}}$$

and let $\bar{1}$ in $\mathfrak{P}$ be such that $\boldsymbol{l}(\bar{1}) = \{A,B,C,D\}$. Then, we have $(\bar{0}, (0_1, 0_2)) \hookrightarrow_\emptyset (\bar{1}, (1_1, 1_2))$ and $(\bar{0}, (0_1, 0_2)) \rightsquigarrow_\emptyset (\bar{1}, (2_1, 1_2))$. We also observe that if C was removed from $\mathcal{E}_1$ and $\mathcal{E}_2$, then it would be the case $(\bar{0}, (0_1, 0_2)) \hookrightarrow_{\{B\}} (\bar{1}, (2_1, 1_2))$. We now show:

Lemma 36. *$\boldsymbol{q}$ is a unique most general $\mathcal{Q}_p[\Diamond]$-separator for E iff $\boldsymbol{q} = \boldsymbol{q}_\pi$ for a $\hookrightarrow$-separating path π in $\mathfrak{P} \otimes \mathfrak{N}$ such that $\boldsymbol{q}_\nu \models \boldsymbol{q}_\pi$ for each separating path ν (i.e., using $\hookrightarrow$- or $\rightsquigarrow$-edges) in $\mathfrak{P} \otimes \mathfrak{N}$.*

Proof. ($\Rightarrow$) By Lemma 35, there exists a separating path π such that $\boldsymbol{q} = \boldsymbol{q}_\pi$ and $\boldsymbol{q}_\nu \models \boldsymbol{q}_\pi$ for each separating path $\boldsymbol{q}_\nu$. Let the sequence of vertices of π be $-\bar{1}, (\boldsymbol{n}_0, \boldsymbol{m}_0), \ldots, (\boldsymbol{n}_\ell, \boldsymbol{m}_\ell)$ with $(\boldsymbol{n}_0, \boldsymbol{m}_0) = (\bar{0}, \boldsymbol{m})$ and $(\boldsymbol{n}_\ell, \boldsymbol{m}_\ell) = (\boldsymbol{n}, \bar{\infty})$. Consider any path $\nu = -\mathbf{1} \to_{\rho'_0} (\boldsymbol{n}_0, \boldsymbol{m}_0) \cdots \to_{\rho'_\ell} (\boldsymbol{n}_\ell, \boldsymbol{m}_\ell)$ in $\mathfrak{P} \times \mathfrak{N}$ and let Π_π be the set of all such paths. Because by Lemma 35 $\boldsymbol{q}_\nu \models \boldsymbol{q}_\pi$ for each $\nu \in \Pi_\pi$ and $\pi \in \Pi_\pi$, it follows that $\rho_i = \bigcap_{\nu \in \Pi_\pi} \rho'_i$. Thus, π is a $\hookrightarrow$-path in $\mathfrak{P} \otimes \mathfrak{N}$. We now need to show that $\boldsymbol{q}_\nu \models \boldsymbol{q}_\pi$ for any separating path $\nu = -\mathbf{1} \Rightarrow_{\rho'_0} (\boldsymbol{n}'_0, \boldsymbol{m}'_0) \cdots \Rightarrow_{\rho'_{\ell'}} (\boldsymbol{n}'_{\ell'}, \boldsymbol{m}'_{\ell'})$ in $\mathfrak{P} \otimes \mathfrak{N}$, where $\Rightarrow_{\rho'_i} \in \{\hookrightarrow_{\rho'_i}, \rightsquigarrow_{\rho'_i}\}$. Let Q_ν be the set of queries $\boldsymbol{q}$ such that $\boldsymbol{q} = \boldsymbol{q}_{\nu'}$ for a path ν' in $\mathfrak{P} \times \mathfrak{N}$ using the same sequence of vertices as ν. For $\boldsymbol{q}_1 = \rho_0^1 \wedge \Diamond(\rho_1^1 \wedge \cdots \wedge \Diamond \rho_{\ell'}^1)$ and $\boldsymbol{q}_2 = \rho_0^2 \wedge \Diamond(\rho_1^2 \wedge \cdots \wedge \Diamond \rho_{\ell'}^2)$ in Q_ν, we define $\boldsymbol{q}_1 \cap \boldsymbol{q}_2 = \rho_0^1 \cap \rho_0^2 \wedge \Diamond(\rho_1^1 \cap \rho_1^2 \wedge \cdots \wedge \Diamond \rho_{\ell'}^1 \cap \rho_{\ell'}^2)$. We observe that

$$\boldsymbol{q}' \models \boldsymbol{q}_\pi \text{ for each } \boldsymbol{q}' \in Q_\nu \text{ iff } \bigcap Q_\nu \models \boldsymbol{q}_\pi. \tag{6}$$

It remains to note that $\bigcap Q_\nu = \boldsymbol{q}_\nu$, while we have $\boldsymbol{q}' \models \boldsymbol{q}_\pi$ for each $\boldsymbol{q}' \in Q_\nu$ from Lemma 35.

($\Leftarrow$) Let $\pi = -\mathbf{1} \hookrightarrow_{\rho_0} (\boldsymbol{n}_0, \boldsymbol{m}_0) \cdots \hookrightarrow_{\rho_\ell} (\boldsymbol{n}_\ell, \boldsymbol{m}_\ell)$ be the separating path in $\mathfrak{P} \otimes \mathfrak{N}$ such that $\boldsymbol{q}_\nu \models \boldsymbol{q}_\pi$ for each separating path ν in $\mathfrak{P} \otimes \mathfrak{N}$. Clearly, $\pi = -\mathbf{1} \to_{\rho_0} (\boldsymbol{n}_0, \boldsymbol{m}_0) \cdots \to_{\rho_\ell} (\boldsymbol{n}_\ell, \boldsymbol{m}_\ell)$ is a separating path in $\mathfrak{P} \otimes \mathfrak{N}$. It remains to show that $\boldsymbol{q}_\nu \models \boldsymbol{q}_\pi$ for any separating path ν in $\mathfrak{P} \otimes \mathfrak{N}$, then the result would follow by Lemma 35. Take an arbitrary separating $\nu = -\mathbf{1} \to_{\rho'_0} (\boldsymbol{n}'_0, \boldsymbol{m}'_0) \cdots \to_{\rho'_{\ell'}} (\boldsymbol{n}'_{\ell'}, \boldsymbol{m}'_{\ell'})$ in $\mathfrak{P} \times \mathfrak{N}$. Let $\boldsymbol{q} = \bigcap Q_\nu$. Observe that by our construction, $-\mathbf{1} \Rightarrow_{\rho_0} (\boldsymbol{n}'_0, \boldsymbol{m}'_0), \ldots, (\boldsymbol{n}'_{\ell'-1}, \boldsymbol{m}'_{\ell'-1}) \Rightarrow_{\rho_{\ell'}} (\boldsymbol{n}'_{\ell'}, \boldsymbol{m}'_{\ell'})$ is in $\mathfrak{P} \otimes \mathfrak{N}$, where $\Rightarrow_{\rho_i} \in \{\hookrightarrow_{\rho_i}, \rightsquigarrow_{\rho_i}\}$. Thus, $\boldsymbol{q} \models \boldsymbol{q}_\pi$ and by (6), $\boldsymbol{q}_\nu \models \boldsymbol{q}_\pi$. $\dashv$

Now we explain how, for given $(\boldsymbol{n}, \boldsymbol{m})$ and $(\boldsymbol{n}', \boldsymbol{m}')$ in $\mathfrak{P} \otimes \mathfrak{N}$, to compute ρ such that $(\boldsymbol{n}, \boldsymbol{m}) \hookrightarrow_\rho (\boldsymbol{n}', \boldsymbol{m}')$ (respectively, $(\boldsymbol{n}, \boldsymbol{m}) \rightsquigarrow_\rho (\boldsymbol{n}', \boldsymbol{m}')$) or decide that one does not exist, in time polynomial in $|\sigma| |\mathfrak{P}| |\mathfrak{N}|$. Let $X_{\boldsymbol{n},\boldsymbol{m}} = \boldsymbol{l}(\boldsymbol{n}) \cap \boldsymbol{l}(\boldsymbol{m})$ for all $\boldsymbol{n}$ in $\mathfrak{P}$ and $\boldsymbol{m}$ in $\mathfrak{N}$. Clearly, $(\boldsymbol{n}, \boldsymbol{m}) \to_\rho (\boldsymbol{n}', \boldsymbol{m}')$ for some ρ iff $X_{\boldsymbol{n}',\boldsymbol{m}'} \setminus X_{\boldsymbol{n}'',\boldsymbol{m}''} \neq \emptyset$ for all $\boldsymbol{n}'' \in \nabla_{\boldsymbol{n},\boldsymbol{n}'}$ and $\boldsymbol{m}'' \in \nabla_{\boldsymbol{m},\boldsymbol{m}'}$. It is easy to see that

$$\rho_{\boldsymbol{n},\boldsymbol{m},\boldsymbol{n}',\boldsymbol{m}'} = \bigcup_{\substack{\boldsymbol{n}'' \in \nabla_{\boldsymbol{n},\boldsymbol{n}'}, \boldsymbol{m}'' \in \nabla_{\boldsymbol{m},\boldsymbol{m}'}, \\ |X_{\boldsymbol{n}',\boldsymbol{m}'} \setminus X_{\boldsymbol{n}'',\boldsymbol{m}''}| = 1}} (X_{\boldsymbol{n}',\boldsymbol{m}'} \setminus X_{\boldsymbol{n}'',\boldsymbol{m}''}).$$

It follows that $(\boldsymbol{n}, \boldsymbol{m}) \hookrightarrow_\rho (\boldsymbol{n}', \boldsymbol{m}')$ iff $\rho = \rho_{\boldsymbol{n}, \boldsymbol{m}, \boldsymbol{n}', \boldsymbol{m}'}$ and $(X_{\boldsymbol{n}', \boldsymbol{m}'} \setminus X_{\boldsymbol{n}'', \boldsymbol{m}''}) \cap \rho_{\boldsymbol{n}, \boldsymbol{m}, \boldsymbol{n}', \boldsymbol{m}'} \neq \emptyset$ for all $\boldsymbol{n}'' \in \nabla_{\boldsymbol{n}, \boldsymbol{n}'}$ and $\boldsymbol{m}'' \in \nabla_{\boldsymbol{m}, \boldsymbol{m}'}$. Then, $(\boldsymbol{n}, \boldsymbol{m}) \rightsquigarrow_\rho (\boldsymbol{n}', \boldsymbol{m}')$ iff $\rho = \rho_{\boldsymbol{n}, \boldsymbol{m}, \boldsymbol{n}', \boldsymbol{m}'}$, $(\boldsymbol{n}, \boldsymbol{m}) \rightarrow_\rho (\boldsymbol{n}', \boldsymbol{m}')$ for some ρ, and $(X_{\boldsymbol{n}', \boldsymbol{m}'} \setminus X_{\boldsymbol{n}'', \boldsymbol{m}''}) \cap \rho_{\boldsymbol{n}, \boldsymbol{m}, \boldsymbol{n}', \boldsymbol{m}'} = \emptyset$ for some $\boldsymbol{n}'' \in \nabla_{\boldsymbol{n}, \boldsymbol{n}'}$ and $\boldsymbol{m}'' \in \nabla_{\boldsymbol{m}, \boldsymbol{m}'}$. Thus, we have an algorithm to construct $\mathfrak{P} \otimes \mathfrak{N}$ in time polynomial in $|\sigma||\mathfrak{P}||\mathfrak{N}|$.

It remains to show how to check the condition on $\mathfrak{P} \otimes \mathfrak{N}$ in Lemma 36. We will use an algorithm similar to the one for unique most specific separators in Lemma 26, marking each $(\boldsymbol{n}, \boldsymbol{m})$ in $\mathfrak{P} \otimes \mathfrak{N}$ by either a query $\boldsymbol{q}_\pi$ for some $\hookrightarrow$-path π (from $-\mathbf{1}$) or with ε. Intuitively, $\boldsymbol{q}_\pi$ marks $(\boldsymbol{n}, \boldsymbol{m})$ if $\boldsymbol{q}_\nu \models \boldsymbol{q}_\pi$ for each path ν to $(\boldsymbol{n}, \boldsymbol{m})$ and otherwise it is marked by ε. We start by marking all the states $(\bar{0}, \boldsymbol{m})$ such that $-\mathbf{1} \hookrightarrow_\rho (\bar{0}, \boldsymbol{m})$ with ρ and marking the states with $-\mathbf{1} \rightsquigarrow_\rho (\bar{0}, \boldsymbol{m})$ with ε. Now, assume that all immediate predecessors of $(\boldsymbol{n}, \boldsymbol{m})$ have already been marked. Let $\boldsymbol{q}_{\pi_i}$, $i = 1, \ldots, k$, be all of their marks different from ε. Consider the extended paths $\pi_i' = \pi_i \hookrightarrow_{\rho_i} (\boldsymbol{n}, \boldsymbol{m})$ and the set Π of all paths to $(\boldsymbol{n}, \boldsymbol{m})$. We check whether there is a containment witness for $\boldsymbol{q}_\pi$ and $\boldsymbol{q}_{\pi_1'}$, for every $\pi \in \Pi$, in which case we mark $(\boldsymbol{n}, \boldsymbol{m})$ by $\boldsymbol{q}_{\pi_1'}$. Otherwise, we do the same for $\boldsymbol{q}_{\pi_2'}$, and so on. If $(\boldsymbol{n}, \boldsymbol{m})$ has no mark after k iterations, we mark it by ε. Although Π can be of exponential size, we show that marking each node can be done in P. Indeed, consider the following problem: given $\boldsymbol{q}_{\pi_1'}$ and $(\boldsymbol{n}, \boldsymbol{m})$, decide if there exists $\pi \in \Pi$ such that there is no containment witness for $\boldsymbol{q}_\pi$ and $\boldsymbol{q}_{\pi_1'}$. This problem is in NL because such π, if exists, can be computed by a non-deterministic algorithm guessing π of the form (2) node-by-node. At each step $i \in \{0, \ldots, n\}$ we only need to remember the maximal $i' \in \{0, \ldots, n' - 1\}$, for the length n' of π_1', such that there exists a containment witness for $\boldsymbol{q}_{-\mathbf{1} \Rightarrow_{\rho_0} (\boldsymbol{n}_0, \boldsymbol{m}_0) \cdots \Rightarrow_{\rho_i} (\boldsymbol{n}_i, \boldsymbol{m}_i)}$ and the prefix of π_1' up to (and including) i'-th item. If, after the final step n, we have $i' \neq n' - 1$, we obtained $\pi \in \Pi$ such that there is no containment witness for $\boldsymbol{q}_\pi$ and $\boldsymbol{q}_{\pi_1'}$. Because NL = CONL and the latter problem is in NL, we conclude that checking whether there is a containment witness for $\boldsymbol{q}_\pi$ and $\boldsymbol{q}_{\pi_1'}$, for every $\pi \in \Pi$, is in NL. After all the nodes are marked, we consider all nodes $(\boldsymbol{n}_i, \bar{\infty})$, $i = 1, \ldots, l$, marked with $\boldsymbol{q}_{\pi_i}$ for $\pi_i = \pi_i' \hookrightarrow_{\rho_i} (\boldsymbol{n}_i, \bar{\infty})$ with $\rho_i \neq \emptyset$. If there are no such, there is no unique most general separator. Otherwise, take the set Π of all paths leading to the $(\boldsymbol{n}_i, \bar{\infty})$. We check whether there is a containment witness for $\boldsymbol{q}_\pi$ and $\boldsymbol{q}_{\pi_1}$, for every $\pi \in \Pi$, in which case $\boldsymbol{q}_{\pi_1}$ is returned as a unique most specific separator. Otherwise, we do the same for $\boldsymbol{q}_2$, and so on.

We next prove Theorem 23 (a) for $\mathcal{Q}_p[\bigcirc\Diamond]$-separators. Let $E^+ = \{\mathcal{D}_1, \ldots, \mathcal{D}_k\}$. As in the previous part of the proof, we denote by $\boldsymbol{n}$ any vector $(n_1, \ldots, n_k)$ with $0 \leq n_j \leq \max \mathcal{D}_j$, $1 \leq j \leq k$. The vertices of $\mathfrak{P}$ are tuples $\boldsymbol{n}^t = (n_1, \ldots, n_k)^t$, where $(n_1, \ldots, n_k)$ is as above and $0 \leq t \leq \min\{\max \mathcal{D}_j - n_j \mid 1 \leq j \leq k\}$. Define $\boldsymbol{n} + t = (n_1 + t, \ldots, n_k + t)$ and $\boldsymbol{n} < \boldsymbol{n}'$ if $n_j < n_j'$ for all $1 \leq j \leq k$. $\mathfrak{P}$ has edges $\boldsymbol{n}^t \rightarrow \boldsymbol{n}_1^s$ for all vertices $\boldsymbol{n}^t, \boldsymbol{n}_1^s$ in $\mathfrak{P}$ such that $\boldsymbol{n} + t < \boldsymbol{n}_1$. Let $t_+(\boldsymbol{n}) = \bigcap_{0 \leq j \leq k} t_{\mathcal{D}_j}(n_j)$. We set the label $L(\boldsymbol{n}^t)$ of each $\boldsymbol{n}^t$ in $\mathfrak{P}$ to be a $t+1$-component vector $(t_+(\boldsymbol{n}), t_+(\boldsymbol{n}+1), \ldots, t_+(\boldsymbol{n}+t))$. It will be convenient to set $L(\boldsymbol{n}) = t_+(\boldsymbol{n})$. Similarly, given $E^- = \{\mathcal{E}_1, \ldots, \mathcal{E}_l\}$, we denote by $\boldsymbol{m}$ any vector $(m_1, \ldots, m_l)$ with $m_j \in [0, \max \mathcal{E}_j] \cup \{\infty\}$, $1 \leq j \leq l$. Let $m_j + t = \infty$ if $m_j + t > \max \mathcal{E}_j$ and $\infty + t = \infty$ for $t \geq 0$. We denote $(\infty, \ldots, \infty)$ by $\bar{\infty}$. The vertices of $\mathfrak{N}$ are tuples $\boldsymbol{m}^t = (m_1, \ldots, m_k)^t$, where m_j is as above and $0 \leq t \leq \min\{\max \mathcal{D}_j \mid 1 \leq j \leq k\}$. $\mathfrak{N}$ has edges $\boldsymbol{m}^t \rightarrow \boldsymbol{m}'^s$ for all vertices such that $\boldsymbol{m} + t < \boldsymbol{m}'$. Set $t_{\mathcal{E}_j}(\infty) = \sigma$ and $t_+(\boldsymbol{m}) = \bigcap_{1 \leq j \leq l} t_{\mathcal{E}_j}(m_j)$. We set the label $M(\boldsymbol{m}^t)$ of each $\boldsymbol{m}^t$ in $\mathfrak{N}$ to be a $t+1$-component vector $(t_+(\boldsymbol{m}), t_+(\boldsymbol{m}+1), \ldots, t_+(\boldsymbol{m}+t))$. Let $(\rho_0, \ldots, \rho_t) \subseteq (\rho_0', \ldots, \rho_t')$ if $\rho_i \subseteq \rho_i'$ for all i.

We take a product-like structure $\mathfrak{P} \otimes \mathfrak{N}$ containing the vertices $(\bar{0}^t, \boldsymbol{m}^t)$ such that $m_j = 0$ if $(t_{\mathcal{D}_j}(0), \ldots, t_{\mathcal{D}_j}(t)) \subseteq (t_{\mathcal{E}_j}(0), \ldots, t_{\mathcal{E}_j}(t))$, and $m_j = \infty$ otherwise, for all $1 \leq j \leq l$. Also, it contains vertices $(\boldsymbol{n}^t, \boldsymbol{m}^t)$ with $\boldsymbol{n}^t \neq \bar{0}^t$ from $\mathfrak{P}$ and with $\boldsymbol{m}^t$ from $\mathfrak{N}$ such that $L(\boldsymbol{n}^t) \subseteq M(\boldsymbol{m}^t)$. The edges of $\mathfrak{P} \times \mathfrak{N}$ are $(\boldsymbol{n}^t, \boldsymbol{m}^t) \rightarrow (\boldsymbol{n}'^s, \boldsymbol{m}'^s)$ such that $\boldsymbol{n}^t \rightarrow \boldsymbol{n}'^s$, $\boldsymbol{m}^t \rightarrow \boldsymbol{m}'^s$ and there exists no $\mathfrak{P} \times \mathfrak{N}$-vertex $(\boldsymbol{n}'^s, \boldsymbol{m}''^s)$ satisfying $\boldsymbol{m}^t \rightarrow \boldsymbol{m}''^s$, $m_i'' < m_i'$ and $m_i'' \neq \infty$ for some $1 \leq i \leq l$. The latter *minimality* condition on $\boldsymbol{m}'$ ensures that for given $(\boldsymbol{n}^t, \boldsymbol{m}^t)$ and $\boldsymbol{n}'^s \leftarrow \boldsymbol{n}^t$ there is a unique $\boldsymbol{m}'^s$ such that $(\boldsymbol{n}^t, \boldsymbol{m}^t) \rightarrow (\boldsymbol{n}'^s, \boldsymbol{m}'^s)$.

Given a path in $\mathfrak{P} \otimes \mathfrak{N}$

$$\pi = (\boldsymbol{n}_0^{t_0}, \boldsymbol{m}_0^{t_0}), \ldots, (\boldsymbol{n}_\ell^{t_\ell}, \boldsymbol{m}_\ell^{t_\ell}) \tag{7}$$

with $\boldsymbol{n}_0 = \bar{0}$, let $\boldsymbol{q}_\pi$ be the $\mathcal{Q}_p[\bigcirc\Diamond]$-query of the form (1) with $n = t_0 + \cdots + t_\ell + \ell$ such that:

- $(\rho_0, \ldots, \rho_{t_0}) = L(\boldsymbol{n}_0^{t_0})$ and $\boldsymbol{o}_i = \bigcirc$ for $i \in (0, t_0]$
- $(\rho_{t_0+1}, \ldots, \rho_{t_0+t_1+1}) = L(\boldsymbol{n}_1^{t_1})$, $\boldsymbol{o}_{t_0+1} = \Diamond$ and $\boldsymbol{o}_i = \bigcirc$ for $i \in (t_0 + 1, t_0 + t_1 + 1]$
- ...
- $(\rho_{t_0+\cdots+t_{\ell-1}+\ell}, \ldots, \rho_{t_0+\cdots+t_\ell+\ell}) = L(\boldsymbol{n}_\ell^{t_\ell})$, $\boldsymbol{o}_{t_0+\cdots+t_{\ell-1}+\ell} = \Diamond$ and $\boldsymbol{o}_i = \bigcirc$ for $i \in (t_0 + \cdots + t_{\ell-1} + \ell, t_0 + \cdots + t_\ell + \ell]$.

(Note that $\boldsymbol{q}_\pi$ is not necessarily in normal form.) If $L(\boldsymbol{n}_\ell + t_\ell) \neq \emptyset$ and $\boldsymbol{m}_\ell + t_\ell = \bar{\infty}$, we call π a *separating* path.

Lemma 37. *E is $\mathcal{Q}_p[\bigcirc\Diamond]$-separable iff there exists a separating path in $\mathfrak{P} \otimes \mathfrak{N}$.*

Proof. ($\Rightarrow$) Suppose $\boldsymbol{q} = \bar{\rho}_0 \wedge \Diamond(\bar{\rho}_1 \wedge (\ldots \Diamond \bar{\rho}_\ell) \ldots)$ (in the normal form) separates E, where each $\bar{\rho}_i = \rho_i^0 \wedge \bigcirc \rho_i^1 \wedge \cdots \wedge \bigcirc^{p_i} \rho_i^{p_i}$. If $\ell = 0$, we can further assume $\rho_0^{p_0} \neq \emptyset$. It follows $\bar{\rho}_0 \subseteq L(\bar{0}^{p_0})$ and $\bar{\rho}_0 \not\subseteq (t_{\mathcal{E}_j}(0), \ldots, t_{\mathcal{E}_j}(p_0))$ for all $1 \leq j \leq l$. Thus, $L(\bar{0} + p_0) \neq \emptyset$, and $(\bar{0}^{p_0}, \bar{\infty}^{p_0})$ (so, $\infty + p_0 = \bar{\infty}$) is in $\mathfrak{P} \otimes \mathfrak{N}$, which completes the proof.

Suppose $\ell > 0$, $\bar{\rho}_0 \wedge \Diamond(\bar{\rho}_1 \wedge (\ldots \Diamond \bar{\rho}_{\ell-1})\ldots)$ does not separate E, and $\rho_\ell^{p_\ell} \neq \emptyset$. We define $\boldsymbol{n}_0 = \bar{0}$. Assuming that $\boldsymbol{n}_i = (n_1, \ldots, n_k)$ has been defined, define $\boldsymbol{n}_{i+1} = (n'_1, \ldots, n'_k)$ where $n'_j = \min\{n \mid n > n_j + p_i, \bar{\rho}_{i+1} \subseteq (t_{\mathcal{D}_j}(n), t_{\mathcal{D}_j}(n + 1), \ldots, t_{\mathcal{D}_j}(n + p_{i+1}))\}$, for $0 \leq i < \ell$. Clearly, we have $\boldsymbol{n_0}^{p_0} \to \cdots \to \boldsymbol{n_\ell}^{p_\ell}$ and $L(\boldsymbol{n_\ell} + p_\ell) \neq \emptyset$. Then, we take $\boldsymbol{m}_0$ equal to $\boldsymbol{m}$ satisfying $m_j = 0$ if $(t_{\mathcal{D}_j}(0), \ldots, t_{\mathcal{D}_j}(p_0)) \subseteq (t_{\mathcal{E}_j}(0), \ldots, t_{\mathcal{E}_j}(p_0))$, and $m_j = \infty$ otherwise, for all $1 \leq j \leq l$. We set $\boldsymbol{m}_{i+1}$ to be $\boldsymbol{m}$ such that $(\boldsymbol{n}_i^{p_i}, \boldsymbol{m}_i^{p_i}) \to (\boldsymbol{n}_{i+1}^{p_{i+1}}, \boldsymbol{m}^{p_{i+1}})$. We claim that the path $(\boldsymbol{n}_0^{p_0}, \boldsymbol{m}_0^{p_0}), \ldots, (\boldsymbol{n}_\ell^{p_\ell}, \boldsymbol{m}_\ell^{p_\ell})$ is such that $\boldsymbol{m}_\ell + p_\ell = \bar{\infty}$. To prove the claim, set $d_{i,j} = \min\{t \mid t \geq i + p_0 + \cdots + p_i \text{ and } \mathcal{E}_j^{\leq t} \models \bar{\rho}_0 \wedge \Diamond(\bar{\rho}_1 \wedge (\ldots \Diamond \bar{\rho}_i)\ldots)\}$ or $d_{i,j} = \infty$ if t as above does not exist, for $0 \leq i \leq \ell$ and $1 \leq j \leq l$. Let $\boldsymbol{d}_i = (d_{i,1}, \ldots, d_{i,l})$. It follows that $\boldsymbol{d}_\ell = \bar{\infty}$. It is easily shown by induction on i that $\boldsymbol{d}_i = \boldsymbol{m}_i + p_i$ which completes the proof.

($\Leftarrow$) Let (7) be a separating path. We will show that $\boldsymbol{q}_\pi$ separates E. Indeed, clearly $\mathcal{D} \models \boldsymbol{q}_\pi$ for each $\mathcal{D} \in E^+$. By induction on i, we show that $\boldsymbol{d}_i = \boldsymbol{m}_i + t_i$. It follows that $\mathcal{E} \not\models \boldsymbol{q}_\pi$ for all $\mathcal{E} \in E^-$ which completes the proof. $\dashv$

The existence of a separating path in $\mathfrak{P} \otimes \mathfrak{N}$ can be checked in time $O(t_+^{c_+} t_-^{c_-})$. If c_+ and c_- are bounded, given $(\boldsymbol{n}^t, \boldsymbol{m}^t)$ and $(\boldsymbol{n}'^s, \boldsymbol{m}'^s)$, we can check in logspace whether $(\boldsymbol{n}^t, \boldsymbol{m}^t)$ is the root of $\mathfrak{P} \times \mathfrak{N}$ and $(\boldsymbol{n}^t, \boldsymbol{m}^t) \to (\boldsymbol{n}'^s, \boldsymbol{m}'^s)$. So the existence of a separating path can be decided in NL.

We now prove Theorem 23 (b) for $\mathcal{Q}_p[\bigcirc\Diamond]$-separators. As it was already done in the proof of Lemma 37, we assume that the $\mathcal{Q}_p[\bigcirc\Diamond]$-queries are

$$\boldsymbol{q} = \bar{\rho}_0 \wedge \Diamond(\bar{\rho}_1 \wedge \cdots \wedge \Diamond \bar{\rho}_\ell), \tag{8}$$

where each $\bar{\rho}_i = \rho_i^0 \wedge \bigcirc \rho_i^1 \wedge \cdots \wedge \bigcirc^{p_i} \rho_i^{p_i}$. Furthermore, we treat each query $\bar{\rho}_i$ as a vector $(\rho_i^0, \ldots, \rho_i^{p_i})$ when convenient. The proof of the following statement is a straightforward modification of the proof of Lemma 38 (in the same way as the proof of Lemma 37 modifies the proof of Lemma 24):

Lemma 38. *If $\boldsymbol{q}$ of the form* (8) *separates E, then there exists a separating path* (7) *such that $t_i = p_i$ and $\bar{\rho}_i \subseteq L(\boldsymbol{n}_i^{t_i})$, for all $i \leq \ell$.*

In essence, we can now use any existing algorithm for computing a shortest/longest path in the directed acyclic graph[2] $\mathfrak{P} \times \mathfrak{N}$, where the edges $(\boldsymbol{n}^t, \boldsymbol{m}^t) \to (\boldsymbol{n}'^s, \boldsymbol{m}'^s)$ have weights $s + 1$ (we also need to add a new root and extra edges leading to the root $(\bar{0}^s, \boldsymbol{m}^s)$ of $\mathfrak{P} \otimes \mathfrak{N}$). Some straightforward algorithms are provided below.

For the shortest separator, we compute the closures $\mathcal{C}^\ell$ containing all the tuples $(\boldsymbol{n}'^t, \boldsymbol{m}'^t, w)$ such that $(\boldsymbol{n}'^t, \boldsymbol{m}'^t)$ is reachable in $\mathfrak{P} \times \mathfrak{N}$ from some $(\bar{0}^s, \boldsymbol{m}^s)$ on a path of length $\leq \ell$, while

$$w = \min\{\ell' + t_1 + \cdots + t_{\ell'} \mid (\bar{0}^{t_1}, \boldsymbol{m}^{t_1}), \ldots, (\boldsymbol{n}'^{t_{\ell'}}, \boldsymbol{m}'^{t_{\ell'}})$$

$$\text{is a path from some } (\bar{0}^s, \boldsymbol{m}^s) \text{ to } (\boldsymbol{n}'^t, \boldsymbol{m}'^t)\}. \quad (9)$$

We compute all $\mathcal{C}^1, \ldots, \mathcal{C}^\ell$, where ℓ is the length of the longest path in $\mathfrak{P} \otimes \mathfrak{N}$. Let $\mathcal{C}_\infty^i$ be the restriction of $\mathcal{C}^i$ to tuples $(\boldsymbol{n}'^t, \boldsymbol{m}'^t, w)$ with $\boldsymbol{m}' + t = \bar{\infty}$ and $L(\boldsymbol{n}' + t) \neq \emptyset$. We select i^* and w^* such that $(\boldsymbol{n}'^t, \boldsymbol{m}'^t, w^*) \in \mathcal{C}_\infty^{i^*}$ and $w^* = \min\{w \mid (\boldsymbol{n}'^t, \boldsymbol{m}'^t, w) \in \mathcal{C}_\infty^i, 1 \leq i \leq \ell\}$. In essence, w^* is the length of the shortest query separating E. To construct a $\boldsymbol{q}$ with length w^*, we use the sets $\mathcal{C}^1, \ldots, \mathcal{C}^{i^*-1}, \mathcal{C}_\infty^{i^*}$. $\varkappa$ will be $\bar{\rho}_1 \wedge \Diamond(\bar{\rho}_2 \wedge (\ldots \Diamond \bar{\rho}_i)\ldots)$ where each $\bar{\rho}_i = \rho_i^0 \wedge \bigcirc \rho_i^1 \wedge \cdots \wedge \bigcirc^{p_i} \rho_i^{p_i}$. Take an $\boldsymbol{n}_{i^*}^{t_{i^*}}$ such that $(\boldsymbol{n}_{i^*}^{t_{i^*}}, \boldsymbol{m}'^{t_{i^*}}, w^*) \in \mathcal{C}_\infty^{i^*}$. We set $p_{i^*} = t_{i^*}$ and $(\rho_{i^*}^0, \ldots, \rho_{i^*}^{p_{i^*}}) = L(\boldsymbol{n}_{i^*}^{t_{i^*}})$. Now, take an $\boldsymbol{n}_{i^*-1}^{t_{i^*-1}}$ such that $(\boldsymbol{n}_{i^*-1}^{t_{i^*-1}}, \boldsymbol{m}^{t_{i^*-1}}, w_{i^*-1}) \in \mathcal{C}^{i^*-1}$, $(\boldsymbol{n}_{i^*-1}^{t_{i^*-1}}, \boldsymbol{m}^{t_{i^*-1}}) \to (\boldsymbol{n}_{i^*}^{t_{i^*}}, \boldsymbol{m}'^{t_{i^*}})$, and $w^* = w_{i^*-1} + t_{i^*} + 1$. We set $p_{i^*-1} = t_{i^*-1}$ and $(\rho_{i^*-1}^0, \ldots, \rho_{i^*-1}^{p_{i^*-1}}) = L(\boldsymbol{n}_{i^*-1}^{t_{i^*-1}})$. We continue in this fashion to obtain t_1 such that $(\bar{0}^{t_1}, \boldsymbol{m}^{t_1}, t_1 + 1) \in \mathcal{C}^1$ and $w_2 = t_1 + t_2 + 2$. We define $p_1 = t_1$ and $(\rho_1^0, \ldots, \rho_1^{p_1}) = L(\bar{0}^{t_1})$. This completes the definition of a shortest $\boldsymbol{q}$ separating E.

To compute a longest separator, we compute the closures $\mathcal{C}^i$ as in the case of the shortest separator except for using $\max$ instead of $\min$ in (9). We proceed with computing the separator $\boldsymbol{q}$ as in the case above setting w^* with $\max$ instead of $\min$.

Finally, we prove Theorem 23 (c) for unique most specific $\mathcal{Q}_p[\bigcirc\Diamond]$-separators. First, we observe that the proof of Lemma 26 is modified in a straightforward way to demonstrate that this lemma holds for $\mathcal{Q}_p[\bigcirc\Diamond]$-queries and separating paths (7) in $\mathfrak{P} \otimes \mathfrak{N}$. Secondly, we note that checking the existence of a separating path π such that $\boldsymbol{q}_\pi \models \boldsymbol{q}_\nu$ for every separating path ν in $\mathfrak{P} \otimes \mathfrak{N}$ can be checked in P by a very similar algorithm as for $\mathcal{Q}_p[\Diamond]$-queries. To see that the marking of the nodes can still be done in P for $\mathcal{Q}_p[\bigcirc\Diamond]$, we only need to note that checking $\boldsymbol{q} \models \boldsymbol{q}'$ for $\boldsymbol{q}, \boldsymbol{q}' \in \mathcal{Q}_p[\bigcirc\Diamond]$ remains in P (see Theorem 2). Moreover, the problem of checking whether there is a containment witness for a given $\boldsymbol{q}_{\pi'_1}$ and $\boldsymbol{q}_\pi$ for every π in the set Π of path (7) in $\mathfrak{P} \otimes \mathfrak{N}$ ending at a given $(\boldsymbol{n}^t, \boldsymbol{m}^t)$, is in NL.

At the final stage of the algorithm, we need to consider all the nodes of the form $(\boldsymbol{n}_i^{t_i}, \boldsymbol{m}_i^{t_i})$, $i = 1, \ldots, l$, with $L(\boldsymbol{n}_i + t_i) \neq \emptyset$, $\boldsymbol{m}_i + t_i = \bar{\infty}$ and marked with $\boldsymbol{q}_i \neq \varepsilon$. If there are no such, there is no unique most specific separator. Otherwise, take the set Π of all paths leading to the $(\boldsymbol{n}_i^{t_i}, \boldsymbol{m}_i^{t_i})$. We check whether there is a containment witness for $\boldsymbol{q}_1$ and every $\boldsymbol{q}_\pi$, $\pi \in \Pi$, in which case $\boldsymbol{q}_1$ is returned as a unique most specific separator. Otherwise, we do the same for $\boldsymbol{q}_2$, and so on. $\dashv$

---

[2]see e.g. Sedgewick, Robert; Wayne, Kevin Daniel (2011), Algorithms (4th ed.), Addison-Wesley Professional.

F Proofs for Section 8

We show that for $\mathcal{Q} \in \{\mathcal{Q}_p[\diamond], \mathcal{Q}_p[\bigcirc\diamond], \mathcal{Q}_{in}\}$, verifying longest and shortest separators is CONP-complete. The upper bound is trivial, and so we focus on the lower bound. We start with a general construction that shows how one can combine two example sets in a controlled way.

Let $\mathcal{Q} \in \{\mathcal{Q}_p[\diamond], \mathcal{Q}_p[\bigcirc\diamond], \mathcal{Q}_{in}\}$ be given and consider example set $E_1 = (E_1^+, E_1^-)$ and $E_2 = (E_2^+, E_2^-)$ with disjoint signatures such that all separating queries in $\mathsf{s}(E_1, \mathcal{Q})$ and $\mathsf{s}(E_2, \mathcal{Q})$ start with $\diamond$. Define E^+ by taking all

$$\mathcal{D}_1\mathcal{D}_2, \quad \mathcal{D}_2\mathcal{D}_1$$

with $\mathcal{D}_1 \in E_1^+$ and $\mathcal{D}_2 \in E_2^+$ and let $E^- = E_1^- \cup E_2^-$.

Lemma 39. $\mathsf{s}(E, \mathcal{Q}) = \mathsf{s}(E_1, \mathcal{Q}) \cup \mathsf{s}(E_2, \mathcal{Q})$.

Proof. $\mathsf{s}(E, \mathcal{Q}) \supseteq \mathsf{s}(E_1, \mathcal{Q}) \cup \mathsf{s}(E_2, \mathcal{Q})$ follows from the condition that queries in $\mathsf{s}(E_1, \mathcal{Q})$ and $\mathsf{s}(E_2, \mathcal{Q})$ start with $\diamond$.

To show that $\mathsf{s}(E, \mathcal{Q}) \subseteq \mathsf{s}(E_1, \mathcal{Q}) \cup \mathsf{s}(E_2, \mathcal{Q})$ observe that it follows from the definition of E^+ that every query in $\mathsf{s}(E, \mathcal{Q})$ either only uses symbols from E_1 or only symbols from E_2. Now the inclusion is trivial. ⊣

Theorem 40. *Let $\mathcal{Q} \in \{\mathcal{Q}_p[\diamond], \mathcal{Q}_p[\bigcirc\diamond], \mathcal{Q}_{in}\}$. Then verification of longest and shortest separating queries is CONP-hard.*

Proof. By reduction from SAT for the complement. Assume a propositional formula φ is given. Construct E_1 as in the proof of Point 2 of Theorem 21 such that there is a separating query in $\mathcal{Q}$ if φ is satisfiable. Assume all separating queries have length at least 2 and length at most n.

First let E_2 be any example set in a signature disjoint from the signature of E_1 and such that all separators have the form $\diamond\rho$ with ρ a set of atoms. Assume some such $\boldsymbol{q}$ separates E_2. Then $\boldsymbol{q}$ is not a longest separator of E (as defined above) iff φ is satisfiable.

Now let E_2 be any example set in a signature disjoint from the signature of E_1 and such that all separators start with $\diamond$ and have length $n + 1$. Assume some such $\boldsymbol{q}$ separates E_2. Then $\boldsymbol{q}$ is not a shortest separator of E (as defined above) iff φ is satisfiable. ⊣