Bounded Treewidth and the Infinite Core Chase

Complications and Workarounds toward Decidable Querying

ABSTRACT

The *core chase*, a popular algorithm for answering conjunctive queries (CQs) over existential rules, is guaranteed to terminate and compute a finite universal model whenever one exists, leading to the equivalence of the universal-model-based and the chase-based definitions of *finite expansion sets* (fes) – a class of rulesets featuring decidable CQ entailment. In case of non-termination, however, it is non-trivial to define a "result" of the core chase, due to its non-monotonicity. This causes complications when dealing with advanced decidability criteria based on the existence of (universal) models of finite *treewidth*. For these, sufficient chase-based conditions have only been established for weaker, monotonic chase variants.

This paper investigates the – prima facie plausible – hypothesis that the existence of a treewidth-bounded universal model and the existence of a treewidth-bounded core-chase sequence coincide – which would conveniently entail decidable CQ entailment whenever the latter holds. Perhaps surprisingly, carefully crafted examples show that both directions of this hypothesized correspondence fail. On a positive note, we are still able to define an aggregation scheme for the infinite core chase that preserves treewidth bounds and produces a *finitely universal* model, i.e., one that satisfies exactly the entailed CQs. This allows us to prove that the existence of a treewidth-bounded core-chase sequence *does* warrant decidability of CQ entailment (yet, on other grounds than expected). Hence, for the first time, we are able to define a chase-based notion of *bounded treewidth sets* of rules that subsumes fes.

1 INTRODUCTION

The chase is a fundamental tool for the popular formalism of *existential rules*, also known as *tuple-generating dependencies*. Given a knowledge base (KB) composed of a finite set F of facts (the *database*) and a set Σ of (existential) rules, the chase repeatedly applies rules, giving rise to a sequence $F{=}F_0, F_1, F_2, \ldots$ If, in the course of this, a fixpoint is reached after a finite number of steps, one speaks of *chase termination*. Then, the final fact set obtained, seen as a structure, constitutes a finite model of the given KB, which is also *universal*, meaning that it can be homomorphically mapped to any model of the KB. This pleasant property allows one to consider this single model (instead of all models) to answer all queries preserved under homomorphisms, ranging from conjunctive queries (CQs) to datalog and other second-order queries.

In fact, there are different chase variants with differing behavior regarding redundancy treatment and termination. The simplest, most lavish, known as the *oblivious chase*, performs all possible rule applications, without checking for any redundancies [6]. The most frugal, known as the *core chase*, prunes all redundancies at each step, retaining a minimal set of atoms, which is called a *core* [9]. Between these two extremes, the *semi-oblivious* (aka skolem) and *restricted* (aka standard) chase avoid the creation of some redundancies, but not all [10, 17]. The core chase is the only chase variant that terminates exactly when the KB has a finite universal model, and produces the unique (up to isomorphism) smallest such model. Thus, the core chase is the best choice for a decision procedure that aims at chase termination. This motivates the definition of the *fes* (finite expansion sets) class containing all rule sets Σ for which the core chase for $\mathcal{K} = (F, \Sigma)$ terminates for all F [3]. For such Σ, the entailment $\mathcal{K} \models Q$ for any CQ Q can be decided by computing the core chase and evaluating Q against the resulting structure.

Yet, finite universal models may not exist. In such cases, no chase reaches a fixpoint, and there is no last chase sequence element to pick as a result. As a remedy, one may define the "result" of the chase as the infinite union over all the fact sets of the infinite sequence, obtaining an infinite structure. This will still yield a universal model for *monotonic* chase variants, where $F_i \subseteq F_{i+1}$ holds for all i, such as the oblivious, semi-oblivious and restricted chases. However, this does not work well for non-monotonic chase variants such as the core chase, where one cannot even be certain to obtain a model.

One could argue that these issues are of theoretical interest only, given that the non-terminating chase cannot actually be computed and cannot serve as a decision procedure. However, fortunately, decidability of CQ entailment can be established by other means, even when the chase does not terminate. In particular, it is ensured whenever an infinite universal model exists that is still reasonably "structurally well-behaved" by virtue of having a *bounded treewidth* [1, 7]. This insight gave rise to many existential rule fragments of high practical relevance, mostly based on varying notions of *guardedness*, which impose syntactic restrictions ensuring treewidth-boundedness for all chase sequences [1, 2, 7, 16]. Yet, these classes all have in common that the existence of a treewidth-bounded universal model can be established only via chase variants that are necessarily *monotonic*: the union over all F_i in a monotonic chase sequence is known to inherit the treewidth bound. Regrettably, for the core chase, which produces "smaller" intermediate structures and hence ensures treewidth-boundedness of the produced facts more often, no adequate model-producing "aggregation" strategy is known, let alone a treewidth-preserving one.

To overcome this issue, we provide a decidability guarantee, but also bring some unpleasant truths to light. We propose a treewidth-preserving aggregation scheme for the core chase that produces a model, but not a universal one. Luckily, we can still guarantee that the resulting model is *finitely universal* (that is, any of its finite substructures is universal) and thus sufficient for our purpose of decidable CQ entailment. Also, we show that the failure to construct a treewidth-bounded universal model out of a treewidth-bounded chase sequence is not a flaw of our approach, but unavoidable, by exhibiting the *steepening staircase example*: a uniformly treewidth-bounded core-chase sequence for a KB whose every universal model has infinite treewidth. Conversely, the *inflating elevator example* presents a KB with a universal model of finite treewidth, yet each of its core-chase sequences consists of structures of ever-growing treewidth, refuting the plausible hypothesis that any universal model of bounded treewidth can be obtained from a treewidth-bounded core-chase sequence. Figure 1 summarizes our findings.

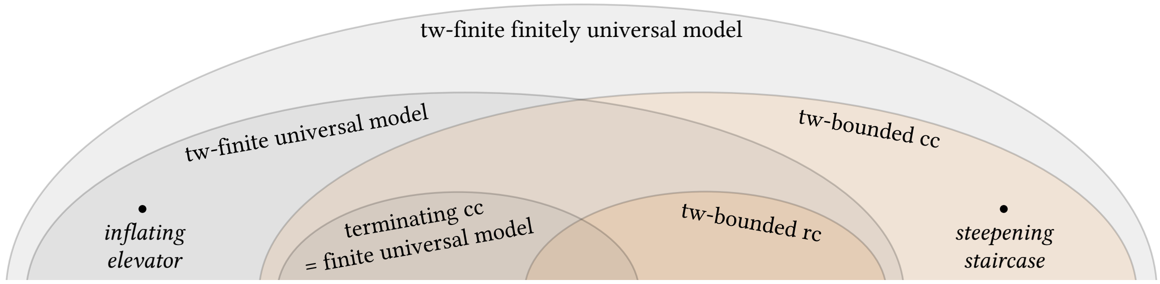


Figure 1: Venn diagram displaying the (non-)inclusion of decidable classes of existential rule sets discussed in the paper. We abbreviate treewidth by *tw*, and restricted and core chase by *rc* and *cc*, respectively. The rulesets entitled "steepening staircase" and "inflating elevator" demonstrate that existence of treewidth-finite universal models and treewidth-bounded core-chase sequences are independent properties. The tw-bounded cc class actually comes in two flavors, referred to as *uniform* and *recurring* boundedness. The latter is more general, but the distinction is irrelevant for this overview.

2 PRELIMINARIES

We use countably infinite disjoint sets Δ_V of *variables* (denoted by uppercase letters) and Δ_C of *constants* (denoted by lowercase letters). A *schema* $\mathcal{S}$ is a finite set of relation symbols (or predicates); each $\mathtt{p} \in \mathcal{S}$ is given an *arity* $ar(\mathtt{p}) \geq 0$. The set of *terms* is $\Delta_T = \Delta_C \cup \Delta_V$. A list $t_1, \dots, t_k$ of terms is also denoted by $\vec{t}$ with $|\vec{t}| = k$.

Atomsets and Homomorphisms. An *atom* over a schema $\mathcal{S}$ is an expression of the form $\mathtt{p}(\vec{t})$, $\mathtt{p} \in \mathcal{S}$ and $\vec{t} \in (\Delta_T)^k$ with $k = ar(\mathtt{p})$. An *atomset* over $\mathcal{S}$ is a countable set of atoms over $\mathcal{S}$. For an atom or atomset A, we let $terms(A)$ and $vars(A)$ denote the set of terms and variables in A, respectively.

A *substitution* of a set of variables $\mathcal{Y} \subseteq \Delta_V$ is a mapping σ from $\mathcal{Y}$ to Δ_T. For an atom $at = \mathtt{p}(t_1, \dots, t_k)$ and a substitution σ of $\mathcal{Y}$, let $\sigma(at) = \mathtt{p}(\sigma^+(t_1), \dots, \sigma^+(t_k))$ where $\sigma^+(t_i) = \sigma(t_i)$ whenever $t_i \in \mathcal{Y}$ and $\sigma^+(t_i) = t_i$ otherwise. If A is an atomset, then $\sigma(A) = \{\sigma(at) \mid at \in A\}$. For two substitutions σ and σ' of variable sets $\mathcal{Y}$ and $\mathcal{Y}'$, respectively, we let $\sigma' \circ \sigma$ denote the substitution of $\mathcal{Y}' \cup \mathcal{Y}$ defined by $Y \mapsto \sigma'^+(\sigma^+(Y))$. Two substitutions are *compatible* if they map the same variables to the same terms.

A *homomorphism* from an atomset A to an atomset B is a substitution π with $\pi(A) \subseteq B$. Given such a homomorphism π, we also say that π *maps* A to B, or that A *maps* to B (via π). An *isomorphism* from A to B is a bijective homomorphism π such that π^{-1} is a homomorphism from B to A (then A and B are called *isomorphic*). An *endomorphism* (*automorphism*) of A is a homomorphism (isomorphism) from A to itself. A *retraction* of A is an endomorphism π where the restriction of π to $terms(\pi(A))$ (the *retract*) is the identity. Note that the classes of homomorphisms, endomorphisms, isomorphisms, and retractions are all closed under composition. A finite atomset A is called a *core* if every retraction of A is the identity. Any finite atomset A admits a retract that is a core; this retract is unique up to isomorphism and called the *core of* A.

We identify an atomset with the (possibly infinite) formula obtained from the existential closure of the conjunction of its atoms. Finite or infinite atomsets also naturally correspond to first-order interpretations;[1] if we want to emphasize this aspect, we also refer to them as *instances*. A *(Boolean) conjunctive query (CQ)* is a finite atomset. Note that we conflate *labeled nulls* usually used in instances with *variables* usually used in queries, as they correspond to the same logical notion. We rely on the standard notions of model and semantic entailment, denoted by $\models$. An instance I is a model of a (possibly infinite) atomset A iff A maps to I; for A and B two (possibly infinite) atomsets, $A \models B$ iff B maps to A.

Existential Rules. An *(existential) rule* R is of the form $B \rightarrow H$, where the *body* $B = body(R)$ and the *head* $H = head(R)$ are nonempty finite atomsets. The variables in B are called *universal*, those both in B and H are called *frontier*, and those only in H are called *existential*. We identify a rule with the first-order sentence $\forall \vec{X}\vec{Y}. \bigwedge B[\vec{X}, \vec{Y}] \rightarrow \exists \vec{Z}. \bigwedge H[\vec{X}, \vec{Z}]$ where $\vec{X}, \vec{Y}, \vec{Z}$ are the frontier, non-frontier universal, and existential variables of R, respectively. In examples, we use the logical notation but omit universal quantifiers.

Given an instance I and a rule $B \rightarrow H$, a *trigger* for I is a pair $tr = (B \rightarrow H, \pi)$ such that π maps B to I; tr is *satisfied* in I if π can be extended to a homomorphism from $B \cup H$ to I. Note that an instance I is a model of a rule R iff it satisfies every trigger for I of the form (R, π). Given a rule $R = B \rightarrow H$, an instance I and a trigger $tr = (R, \pi)$ for I, the *application* of tr on I produces the instance $\alpha(I, tr) = I \cup \pi^{\mathrm{safe}}(H)$, where π^{safe} maps every frontier-variable X of R to $\pi(X)$ and any existential variable in $vars(H)$ to a fresh variable (usually called a *labeled null*).[2]

Universal Models. A *knowledge base (KB)* is a pair $\mathcal{K} = (F, \Sigma)$, where F is a *finite* instance and Σ is a finite set of rules. An instance I is a *model* of $\mathcal{K}$ if it is a model of F and of each rule in Σ. An instance I is *universal* for $\mathcal{K}$ if it (homomorphically) maps to every model of $\mathcal{K}$; note that this does not necessarily mean that I is a model of $\mathcal{K}$. An instance I is a *universal model* of $\mathcal{K}$ if it is a model of $\mathcal{K}$ and is universal for $\mathcal{K}$. We consider the following *CQ entailment* problem: given a KB $\mathcal{K}$ and a Boolean CQ Q, does $\mathcal{K} \models Q$ hold? For any universal model I of $\mathcal{K}$, $\mathcal{K} \models Q$ holds iff Q (homomorphically) maps to I, hence, a universal model of $\mathcal{K}$ is sufficient to decide CQ entailment.

[1] Note that we operate under the *unique name assumption.*

[2] The notion of *fresh variable* refers to the underlying assumption that the referred variable is not already present in F, but also, that it has not occurred at any potential previous computation step (which is particularly relevant when rule applications are iterated and/or intertwined with other operations).

3 DERIVATIONS AND THEIR RESULTS

In this paper, we focus on the restricted and the core chase variants. We now introduce a convenient notion of derivation to define these two variants. Actually, it would allow to define other variants that fall between these two variants in terms of redundancy removal, like e.g., the frugal chase [15]. Our type of derivation is not only a sequence of rule applications, but also incorporates a retraction that removes (some) redundancies after each rule application. In the following, $\mathfrak{I}$ denotes either the set $\mathbb{N}$ of natural integers (for infinite derivations) or the interval $\{0, \ldots, k\} \subseteq \mathbb{N}$ (for finite ones).

Definition 1 (Derivation). *A* derivation *from a KB $\mathcal{K} = (F, \Sigma)$ is a (possibly infinite) sequence $\mathcal{D} = ((tr_i, \sigma_i, F_i))_{i\in\mathfrak{I}}$, where the tr_i are triggers (except $tr_0 = \emptyset$), the σ_i are retractions called* simplifications*, and the F_i are finite instances such that: $F_0 = \sigma_0(F)$; and, for all $i \in \mathfrak{I}\backslash\{0\}$, $F_i = \sigma_i(\alpha(F_{i-1}, tr_i))$, where $tr_i = (R_i, \pi_i)$ with $R_i \in \Sigma$ is a trigger for F_{i-1} not satisfied in F_{i-1}.*

For the sake of brevity, we often denote a derivation simply by $\mathcal{D} = (F_i)_{i\in\mathfrak{I}}$, leaving the tr_i and σ_i implicit. A derivation is called *monotonic* if $F_{i-1} \subseteq F_i$ holds for all $i \in \mathfrak{I}\backslash\{0\}$. In a monotonic derivation, the restriction of σ_i to the terms of F_{i-1} is the identity.

When a derivation $\mathcal{D} = (F_i)_{0\leq i\leq k}$ is finite, its result can be defined by just taking its last instance: $\mathcal{D}^+ = F_k$. However, for infinite derivations of the form $\mathcal{D} = (F_i)_{i\in\mathbb{N}}$, the "result" of $\mathcal{D}$ is usually defined as the (infinite) union of all instances along $\mathcal{D}$. We denote this union by $\mathcal{D}^* = \bigcup_{i\in\mathfrak{I}} F_i$ and call it the *natural aggregation* of $\mathcal{D}$ (to distinguish it from the *robust aggregation* defined in Section 8). Note that if $\mathcal{D}$ is a finite monotonic derivation, then $\mathcal{D}^* = \mathcal{D}^+$.

As stated in the next proposition, $\mathcal{D}^*$ is universal for $\mathcal{K}$. Yet, to ensure that a *model* of $\mathcal{K}$ is obtained, we need to require *fairness*, which intuitively means that every trigger for some F_i has to be satisfied in some F_j with $j \geq i$. To formalize this notion, a difficulty with our derivation notion (which arises for any non-monotonic type of chase) is that a trigger (R, π) for some F_i may not remain a trigger for some F_j with $j > i$: this is because $\pi(body(R))$ may be "transformed away" by successive simplifications. To address this issue, we need to "trace" how a set of atoms is transformed along a derivation.

Definition 2. *Let $\mathcal{D} = (F_i)_{i\in\mathfrak{I}}$ be a derivation, and X be a variable occurring in some F_i. For any $j \in \mathfrak{I}$ with $j \geq i$, we define $\tilde{\sigma}_i^i(X) = X$ and $\tilde{\sigma}_i^j(X) = \sigma_j \circ \cdots \circ \sigma_{i+1}(X)$ when $j > i$.*

It is immediate that $\tilde{\sigma}_i^j$ (which is either the identity when $i = j$ or $\sigma_j \circ \cdots \circ \sigma_{i+1}$ otherwise) is a homomorphism from F_i to F_j. Note also that for a monotonic derivation, $\tilde{\sigma}_i^j$ is the identity for any j. In the following, if $tr = (R, \pi)$ is a trigger for A and σ is a substitution, we note $\sigma(tr) = (R, \sigma \circ \pi)$ the trigger for $\sigma(A)$.

Definition 3 (Fair derivation). *A derivation $\mathcal{D} = (F_i)_{i\in\mathfrak{I}}$ is* fair *if, for any $i \in \mathfrak{I}$ and trigger tr for F_i, there is some $j \in \mathfrak{I}$ with $j \geq i$, such that $\tilde{\sigma}_i^j(tr)$ is a satisfied trigger for F_j.*

In classical chase procedures, only *active* triggers (a notion specific to each chase variant) are applied. In the restricted chase, a trigger for F_i is active if it is not already satisfied in F_i. The core chase furthermore computes a retraction to a core after each (or a finite number of) rule application. For both variants, the classical definition of a chase sequence coincides with our notion of a fair derivation. A *restricted chase sequence* can be seen as a fair derivation $((tr_i, \sigma_i, F_i))_{i\in\mathfrak{I}}$ in which σ_i is the identity. Since this derivation is monotonic, it allows for a simpler expression of fairness: any trigger for an F_i has to be satisfied in some F_j, with $i \leq j$. A *core chase sequence* can be seen as a (non-monotonic) fair derivation in which each σ_i produces a core.

Finally, we adapt to our general framework some well-known properties of these chase variants [9, 10, 14]. Albeit $\mathcal{D}^*$ is not always a model, modelhood is guaranteed for monotonic derivations, as already known for the restricted chase.

Proposition 1. *Let $\mathcal{D}$ be a derivation from $\mathcal{K}$. Then:*

(1) $\mathcal{D}^*$ is universal for $\mathcal{K}$;
(2) if $\mathcal{D}$ is monotonic and fair, $\mathcal{D}^*$ is a model of $\mathcal{K}$;
(3) if $\mathcal{D}$ is fair, for all CQ Q, $\mathcal{K} \models Q$ iff $\mathcal{D}^* \models Q$.

4 ADDING TREEWIDTH TO THE PICTURE

We now recall the popular notion of the treewidth of an atomset as well as some well-known facts about it, which will be useful later.

Definition 4. *Given an atomset A, a* tree decomposition *of A is a (possibly infinite) tree $T = (V, E)$, with vertices $V \subseteq 2^{terms(A)}$ and edges $E \in V \times V$, where:*

- *for each $at \in A$ exists some $v \in V$ with $terms(at) \subseteq v$;*
- *for each $t \in terms(A)$, letting $V_t = \{v \in V \mid t \in v\}$, the subgraph of T induced by V_t is connected.*

The width *of $T = (V, E)$ is the size of its largest vertex, minus 1. The* treewidth *of an atomset A, denoted by $tw(A)$, is the minimal width among all its tree decompositions.*

Fact 1. *$A \subseteq B$ implies $tw(A) \leq tw(B)$.*

Definition 5. *Given a natural number n, we say that an atomset A* contains an $n \times n$-grid*, if $terms(A)$ contains n^2 distinct terms, denoted t_j^i for $i, j \in \{1, \ldots, n\}$, such that for all $k \in \{1, \ldots, n-1\}$ and $\ell \in \{1, \ldots, n\}$:*

- *there is some $at \in A$ with $\{t_\ell^k, t_\ell^{k+1}\} \subseteq terms(at)$, and*
- *there is some $at' \in A$ with $\{t_k^\ell, t_{k+1}^\ell\} \subseteq terms(at')$.*

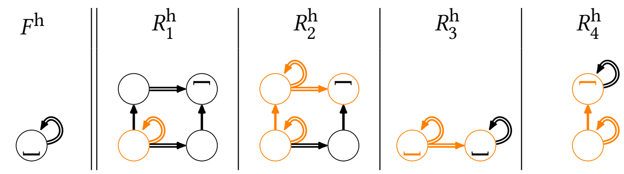
Fact 2. *If A contains an $n \times n$-grid then $tw(A) \geq n$.*

Treewidth is an important notion in the context of existential rules, as the existence of universal models with finite treewidth implies decidability of CQ entailment [3, 7]. In fact, many concrete and practically relevant classes of existential rule sets enjoy this property. One generic way to guarantee the existence of such models is by imposing conditions on the corresponding derivations. This approach underlies all definitions of so-called *bounded treewidth sets* of rules from the literature, but there is a certain disagreement and diversity as to certain details and the type of chase employed (cf. Footnote 4). Here, we will provide the most general such definition that is known to guarantee finite-treewidth universal models along the lines of previously established proofs.

Definition 6. *A ruleset Σ is called a* bounded treewidth set *(bts) if for any finite instance F, there exist some $b \in \mathbb{N}$ and a restricted chase sequence $(F_i)_{i\in\mathfrak{I}}$ such that $tw(F_i) \leq b$ for all $i \in \mathfrak{I}$.*

Proposition 2. *CQ entailment for bts is decidable.*

$$
\begin{array}{lr}
(R_1^{\mathrm{h}}) & \mathtt{h}(X,X) \to \exists X'YY'.\mathtt{h}(X,Y) \wedge \mathtt{v}(X,X') \wedge \mathtt{h}(X',Y') \wedge \mathtt{v}(Y,Y') \wedge \mathtt{c}(Y') \\
(R_2^{\mathrm{h}}) & \mathtt{h}(X,X) \wedge \mathtt{v}(X,X') \wedge \mathtt{h}(X',X') \wedge \mathtt{h}(X',Y') \to \exists Y.\mathtt{c}(Y') \wedge \mathtt{h}(X,Y) \wedge \mathtt{v}(Y,Y') \\
(R_3^{\mathrm{h}}) & \mathtt{f}(X) \wedge \mathtt{h}(X,X) \wedge \mathtt{h}(X,Y) \to \mathtt{f}(Y) \wedge \mathtt{h}(Y,Y) \\
(R_4^{\mathrm{h}}) & \mathtt{h}(X,X) \wedge \mathtt{v}(X,X') \wedge \mathtt{c}(X') \to \mathtt{h}(X',X') \\
(F^{\mathrm{h}}) & \{\mathtt{f}(X_0^0), \mathtt{h}(X_0^0, X_0^0)\}
\end{array}
$$



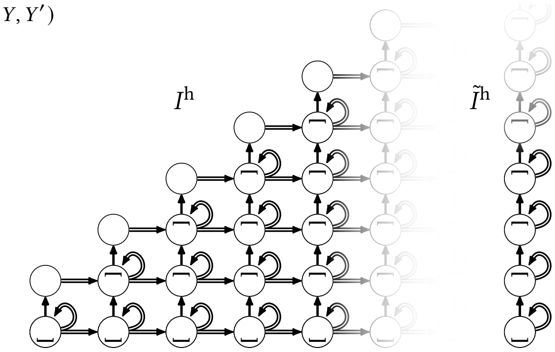


Figure 2: Left: rules of Σ^{h}, fact set F^{h}, and a graphical representation thereof. Orange (grey) elements represent the rule body, black elements the rule head. Visualization of atoms: $\Longrightarrow$ denotes $\mathtt{h}$ ("horizontal") and $\longrightarrow$ denotes $\mathtt{v}$ ("vertical"); we write $\overline{}$ for $\mathtt{c}$ ("ceiling") and $\underline{}$ for $\mathtt{f}$ ("floor"). Right: Atomset I^{h} from Definition 8 – an infinite universal model of $\mathcal{K}^{\mathrm{h}}$. Atomset $\tilde{I}^{\mathrm{h}}$ at the very right is another infinite model of $\mathcal{K}^{\mathrm{h}}$, which is not universal but satisfies exactly the same CQs.

5 CORE CHASE & STRUCTURAL MEASURES

In what follows, we will use the term *structural measure* to generically denote any function μ that maps instances to elements of $\mathbb{N} \cup \{\infty\}$. An easy example would be the *size* of an instance defined by $size : I \mapsto |I|$. An instance I is then called *μ-finite*, if $\mu(I) \neq \infty$. Moreover, we say that a sequence $(F_i)_{i \in \mathfrak{I}}$ of atomsets is *uniformly μ-bounded*, if there exists some $k \in \mathbb{N}$ such that $\mu(F_i) \leq k$ for all $i \in \mathfrak{I}$. $(F_i)_{i \in \mathfrak{I}}$ will be called *recurringly μ-bounded* if there exists some $k \in \mathbb{N}$ such that for any $j \in \mathfrak{I}$ there exists some $i \geq j$ from $\mathfrak{I}$ for which $\mu(F_i) \leq k$ holds. It is easy to see that uniform μ-boundedness implies recurring μ-boundedness, but not vice versa.

Since – on an intuitive level – universal models can be seen as "limits" of appropriate chase sequences, it is a natural question to ask to what extent this limit process preserves structural measures. More specifically, one may ask oneself, given a particular type of chase and structural measure μ, if the existence of a (uniformly or recurringly) μ-bounded chase sequence for a KB is a necessary and/or sufficient condition for the existence of a μ-finite universal model. As mentioned before, for the structural measure of size, this question can be answered positively: A knowledge base $\mathcal{K}$ has a (size-)finite universal model iff it has a size-bounded core chase sequence [9].

Turning to the structural measure of treewidth, however, we found that, surprisingly, both directions fail, witnessed by counterexamples for either direction: The "steepening staircase" KB (Section 6) allows for a (even uniformly) treewidth-bounded chase sequence while lacking a treewidth-finite universal model, whereas the "inflating elevator" KB (Section 7) has a universal model of finite treewidth while not exhibiting a (even just recurringly) treewidth-bounded core-chase sequence.

Irrespective of the fact that our presentation focuses on treewidth as the arguably most prominent structural measure, it should be noted that our counterexamples are based on grid structures and therefore also immediately work for other measures, such as clique-width [11] or (generalized) hypertreewidth [13].

6 THE STEEPENING STAIRCASE

For the KB below, the core chase sequence is uniformly treewidth-bounded by 2, but none of its universal models has finite treewidth.

Definition 7 (The Steepening Staircase KB). *We let $\mathcal{K}^{\mathrm{h}} = (F^{\mathrm{h}}, \Sigma^{\mathrm{h}})$ where $\Sigma^{\mathrm{h}} = \{R_1^{\mathrm{h}}, R_2^{\mathrm{h}}, R_3^{\mathrm{h}}, R_4^{\mathrm{h}}\}$, as given in Figure 2.*

We now describe the instance I^{h}, which is a universal model of $\mathcal{K}^{\mathrm{h}}$ that we can obtain via both the restricted and the core chase.

Definition 8. *We define I^{h} as the infinite instance using the terms $terms(I^{\mathrm{h}}) = \{X_j^i \mid (i,j) \in \mathbb{N}^2, i+1 \geq j\}$ and consisting of the atoms*

$$
\begin{array}{lll}
\mathtt{f}(X_0^i) & \mathtt{h}(X_j^i, X_j^{i+1}) & \mathtt{v}(X_j^i, X_{j+1}^i) \\
\mathtt{c}(X_j^i) \quad \textit{for } i \geq j \geq 1 & \mathtt{h}(X_j^i, X_j^i) \quad \textit{for } i \leq j. &
\end{array}
$$

The instance I^{h} is depicted in Figure 2; the names X_j^i of the variables of I^{h} are in correspondence to their cartesian coordinates (i,j) in the picture. We now consider some particular subsets of $terms(I^{\mathrm{h}})$. For any $k \in \mathbb{N}$, let $P_k = \{X_j^i\}_{i \leq k}$, $C_k = \{X_j^k\}_{j \leq k}$, and $S_k = C_k \cup C_{k+1} \cup \{X_{k+1}^k\}$. Let P_k^{h} (resp. C_k^{h}, S_k^{h}) denote the subset of I^{h} induced by P_k (resp. C_k, S_k). Intuitively, P_k^{h} is the finite *part* until column k, C_k^{h} is the k^{th} *column* of I^{h} (minus its top element) and S_k^{h} is a *step* – a rectangle containing the two columns C_k^{h} and C_{k+1}^{h}.

We first point out that there is a sequence of rule applications from any C_k^{h} producing S_k^{h}. Indeed, we can apply R_1^{h} on the top of C_k^{h} to "complete" C_k^{h} and obtain the two highest variables of C_{k+1}^{h}. Then we apply R_2^{h} k times (from top to bottom) to obtain the other variables of C_{k+1}^{h}. Once X_0^{k+1} has been generated, we can apply R_3^{h} to generate the $\mathtt{h}$-loop on X_0^{k+1}, then k successive applications of R_4^{h} propagate the loops on C_{k+1}^{h}, from bottom to top. There is thus a monotonic infinite derivation $\mathcal{D}_{\mathrm{r}} = (F_i)_{i \in \mathbb{N}}$ from $\mathcal{K}^{\mathrm{h}}$, the natural aggregation of which yields I^{h}. We successively apply R_1^{h}, R_3^{h}, and R_4^{h} on F^{h} to obtain $S_0^{\mathrm{h}} = P_1^{\mathrm{h}}$. Since $C_1^{\mathrm{h}} \subseteq S_0^{\mathrm{h}}$, we apply the rules on C_1^{h} as seen previously to obtain R_1^{h} and thus S_2^{h}, and so on. The infinite union of all atomsets along this derivation is $\mathcal{D}_{\mathrm{r}}^* = I^{\mathrm{h}}$.

Proposition 3. *I^{h} is a result of the restricted chase on $\mathcal{K}^{\mathrm{h}}$.*

Sketch of proof. The derivation $\mathcal{D}_{\mathrm{r}}$ given above is a restricted chase sequence. Clearly, no $\alpha(F, R, \pi)$ in $\mathcal{D}_{\mathrm{r}}$ retracts to F, so it remains to check that $\mathcal{D}_{\mathrm{r}}$ is *fair*. Indeed, if (R, π) is a trigger for some F_i, then it is a trigger wrt some R_k^{h} that is necessarily satisfied (at most in P_{k+2}^{h}). Thus (R, π) is satisfied in some $F_j \supseteq P_{k+2}^{\mathrm{h}}$. □

As a result of the restricted chase, I^{h} is a universal model of $\mathcal{K}^{\mathrm{h}}$. Now, we point out that for any k, C_{k+1}^{h} is a retract of S_k^{h} that is a core. Then we can use $\mathcal{D}_{\mathrm{r}}$ to build a derivation $\mathcal{D}_{\mathrm{c}}$ that relies upon those retractions. $\mathcal{D}_{\mathrm{c}}$ starts out like $\mathcal{D}_{\mathrm{r}}$, but as soon as S_0^{h} is obtained, we retract it to its core C_1^{h}. Then, following $\mathcal{D}_{\mathrm{r}}$'s course, $\mathcal{D}_{\mathrm{c}}$ proceeds to build S_1^{h} that retracts to its core C_2^{h} … As for $\mathcal{D}_{\mathrm{r}}$, we note that $\mathcal{D}_{\mathrm{c}}$ is fair. Moreover, each retraction to a core is done a finite number of rule applications after the previous one: $\mathcal{D}_{\mathrm{c}}$ is thus a core chase sequence. Finally, we point out that every atomset in $\mathcal{D}_{\mathrm{c}}$ is a subset of some S_k^{h}, and has thus treewidth at most 2.

Proposition 4. *There is a core chase sequence for $\mathcal{K}^{\mathrm{h}}$ that is uniformly treewidth-bounded by* 2.

However, all the core computations done in $\mathcal{D}_{\mathrm{c}}$ with the goal of producing a "leaner" result turn out to be futile when it comes to the aggregation: $\mathcal{D}_{\mathrm{c}}^* = \mathcal{D}_{\mathrm{r}}^* = I^{\mathrm{h}}$ contains an $n{\times}n$ grid for any n, and has thus unbounded treewidth. The next proposition even shows that $\mathcal{K}^{\mathrm{h}}$ admits no universal model of finite treewidth. For instance, the atomset $\tilde{I}^{\mathrm{h}}$ pictured in Figure 2 is a model of $\mathcal{K}^{\mathrm{h}}$ but it is not universal: it does not map to I^{h}, since it features an infinite v-path, while all v-paths contained in I^{h} are of finite length.

Proposition 5. *No universal model of $\mathcal{K}^{\mathrm{h}}$ has finite treewidth.*

Sketch of proof. Any universal model U of $\mathcal{K}^{\mathrm{h}}$ is homomorphically equivalent to I^{h}. This allows to show that, for any $n \geq 1$, U contains an $n \times n$-grid, hence $tw(U) \geq n$. □

7 THE INFLATING ELEVATOR

We now present a knowledge base $\mathcal{K}^{\mathrm{v}}$ which does have a universal model with a treewidth of 1, while any (fair) core chase sequence for $\mathcal{K}^{\mathrm{v}}$ contains atomsets whose associated treewidths grow monotonically beyond any given bound.

Definition 9 (The Inflating Elevator KB). *We let $\mathcal{K}^{\mathrm{v}} = (F^{\mathrm{v}}, \Sigma^{\mathrm{v}})$ where $\Sigma^{\mathrm{v}} = \{R_1^{\mathrm{v}}, R_2^{\mathrm{v}}, R_3^{\mathrm{v}}, R_4^{\mathrm{v}}, R_5^{\mathrm{v}}, R_6^{\mathrm{v}}, R_7^{\mathrm{v}}\}$ and F^{v} are as given in the upper part of Figure 3.*

We describe an atomset (shown on the left in Figure 4) representing a universal model that can be obtained via the natural aggregation over the restricted chase or a core chase. We use the same naming convention for nulls as before.

Definition 10. *Let $terms(I^{\mathrm{h}}) = \{X_j^i \mid (i,j) \in \mathbb{N}, i-1 \leq j \leq 2i\}$. Then I^{v} consists of the following atoms for all i, j where all mentioned nulls are in $terms(I^{\mathrm{v}})$:*

$\mathtt{d}(X_j^i)$	$\mathtt{h}(X_j^i, X_j^{i+1})$	$\mathtt{v}(X_j^i, X_{j+1}^i)$
$\mathtt{f}(X_j^i)$	$\mathtt{h}(X_{2i}^i, X_{2i+1}^{i+1})$	$\mathtt{v}(X_j^i, X_j^i)$ *for $i \leq j$*
$\mathtt{c}(X_{2i}^i)$	$\mathtt{h}(X_{2i}^i, X_{2i+2}^{i+1})$	

Proposition 6. *I^{v} is a result of the restricted chase on $\mathcal{K}^{\mathrm{v}}$.*

Sketch of proof. The claim can be shown inductively by assuming that rules without existential variables are prioritized and new nulls are created according to the following scheme:

- for every $i \geq 1$, X_{2i-1}^i and X_{2i}^i are introduced as instances of Y' and Y'' through an application of Rule R_1^{v} with $X \mapsto X_{2i-2}^{i-1}$ and $Y \mapsto X_{2i-2}^i$.
- for every $i \geq 1$, X_i^{i+1} is introduced as instance of Y' through an application of Rule R_2^{v} with $X \mapsto X_{i-1}^i$ and $X' \mapsto X_i^i$.
- every remaining $X_j^i \in terms(I^{\mathrm{v}})$ with $i \geq 1$ is introduced as instance of Y' through an application of Rule R_3^{v} with $X \mapsto X_{j-1}^{i-1}$, $X' \mapsto X_j^{i-1}$, and $Y \mapsto X_{j-1}^i$.

Fairness follows from the fact that I^{v} satisfies all its triggers, as can be checked easily. □

As a result of the restricted chase, I^{v} is a universal model of $\mathcal{K}^{\mathrm{v}}$. As it turns out, it even contains another universal model of finite treewidth. This second universal model I_*^{v}, also shown in Figure 4, is given in the next definition.

Definition 11. *We define the atomset I_*^{v} as the set of those atoms from I^{v} only containing variables of the form X_{2i}^i.*

Proposition 7. *I_*^{v} is a universal model of $\mathcal{K}^{\mathrm{v}}$.*

Proof. I_*^{v} is a model of $\mathcal{K}^{\mathrm{v}}$: it receives a homomorphism from F^{v} and satisfies all rules from Σ^{v}. It is universal, since the identity is a homomorphism from I_*^{v} to I^{v} which is itself a universal model. □

This implies that no finite universal model of $\mathcal{K}^{\mathrm{v}}$ can exist (as any such model would receive a homomorphism from I_*^{v} and thus contain a h-cycle, thus not be homomorphically equivalent to I_*^{v}).

We next describe a sequence $I_0^{\mathrm{v}}, I_1^{\mathrm{v}}, \ldots$ of subsets of I^{v} that exhibit increasing treewidths and will later be shown to occur as substructures in any core chase sequence of $\mathcal{K}^{\mathrm{v}}$. Figure 4 depicts the first elements of that sequence.

Definition 12. *We define the sequence $(I_n^{\mathrm{v}})_{n \in \mathbb{N}}$ of atomsets by letting $I_0^{\mathrm{v}} = F^{\mathrm{v}}$ and, for any $n > 0$, obtaining I_n^{v} as the substructure of I^{v} induced by $terms(I_n^{\mathrm{v}}) = \{X_{2i}^i \mid i \leq \frac{n}{2}\} \cup \{X_j^i \mid i \leq n{+}1 \text{ and } j \geq n\}$ removing all atoms $\mathtt{v}(X_j^i, X_j^i)$ and $\mathtt{f}(X_j^i)$ with $j > n$ as well as all atoms $\mathtt{h}(X_j^i, X_k^{i+1})$ with $k > j$ and $k > n$.*

Proposition 8. *The following hold:*

(1) Every I_n^{v} is a core.
(2) I_n^{v} has a treewidth of at least $\lceil n/3 \rceil + 1$.
(3) For every core chase sequence $(F_i)_{i \in \mathbb{N}}$ for $\mathcal{K}^{\mathrm{v}}$, there is an unbounded monotonic function $f : \mathbb{N} \to \mathbb{N}$ such that, for every $n \in \mathbb{N}$, $I_{f(n)}^{\mathrm{v}}$ is isomorphic to a subset of F_n.
(4) For every core chase sequence $(F_i)_{i \in \mathbb{N}}$ for $\mathcal{K}^{\mathrm{v}}$ and any $m \in \mathbb{N}$ exists a $k \in \mathbb{N}$ such that $tw(F_i) \geq m$ for all $i \geq k$.

From these technical insights, we obtain the strong guarantee regarding the growth of the treewidth:

Corollary 1. *No core chase sequence for $\mathcal{K}^{\mathrm{v}}$ is recurringly or uniformly treewidth-bounded.*

$$(F^{\mathrm{v}}) \qquad \{\mathtt{c}(X_0^0), \mathtt{d}(X_0^0), \mathtt{h}(X_0^0, X_0^1), \mathtt{f}(X_0^1)\}$$

$$(R_1^{\mathrm{v}}) \qquad \mathtt{c}(X) \wedge \mathtt{h}(X, Y) \to \exists Y'Y''.\mathtt{v}(Y, Y') \wedge \mathtt{v}(Y', Y'') \wedge \mathtt{c}(Y'')$$

$$(R_3^{\mathrm{v}}) \qquad \mathtt{v}(X, X') \wedge \mathtt{h}(X, Y) \to \exists Y'.\mathtt{v}(Y, Y') \wedge \mathtt{h}(X', Y')$$

$$(R_2^{\mathrm{v}}) \qquad \mathtt{d}(X) \wedge \mathtt{f}(X) \wedge \mathtt{v}(X, X') \to \exists Y'.\mathtt{h}(X', Y') \wedge \mathtt{f}(Y')$$

$$(R_4^{\mathrm{v}}) \qquad \mathtt{c}(X) \to \mathtt{d}(X)$$

$$(R_5^{\mathrm{v}}) \qquad \mathtt{v}(X, X') \wedge \mathtt{d}(X') \to \mathtt{d}(X)$$

$$(R_6^{\mathrm{v}}) \qquad \mathtt{h}(X, Y) \wedge \mathtt{d}(Y) \wedge \mathtt{f}(Y) \to \mathtt{f}(X) \wedge \mathtt{v}(X, X)$$

$$(R_7^{\mathrm{v}}) \qquad \mathtt{c}(X) \wedge \mathtt{h}(X, Y) \wedge \mathtt{v}(Y, Y') \wedge \mathtt{f}(Y') \to \mathtt{h}(X, Y')$$

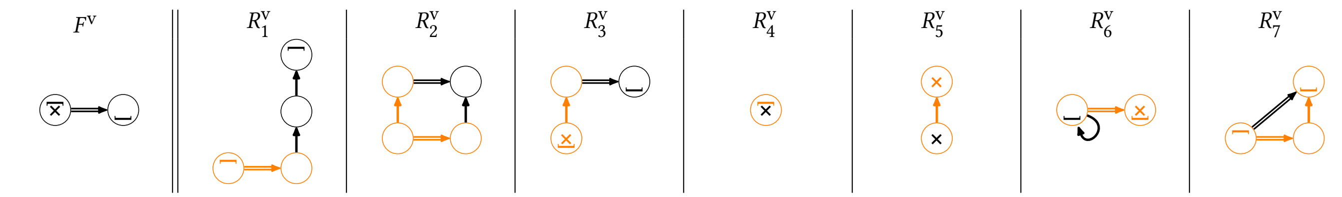


Figure 3: F^{v} and rules of Σ^{v} (top) and their graphical depictions (bottom). Orange (grey) elements represent the rule body and black elements the rule head. Atoms are encoded as follows: $\Longrightarrow$ denotes h ("horizontal") and $\longrightarrow$ denotes v ("vertical"); we write ‾ for c ("ceiling"), _ for f ("floor"), and × for d ("done").

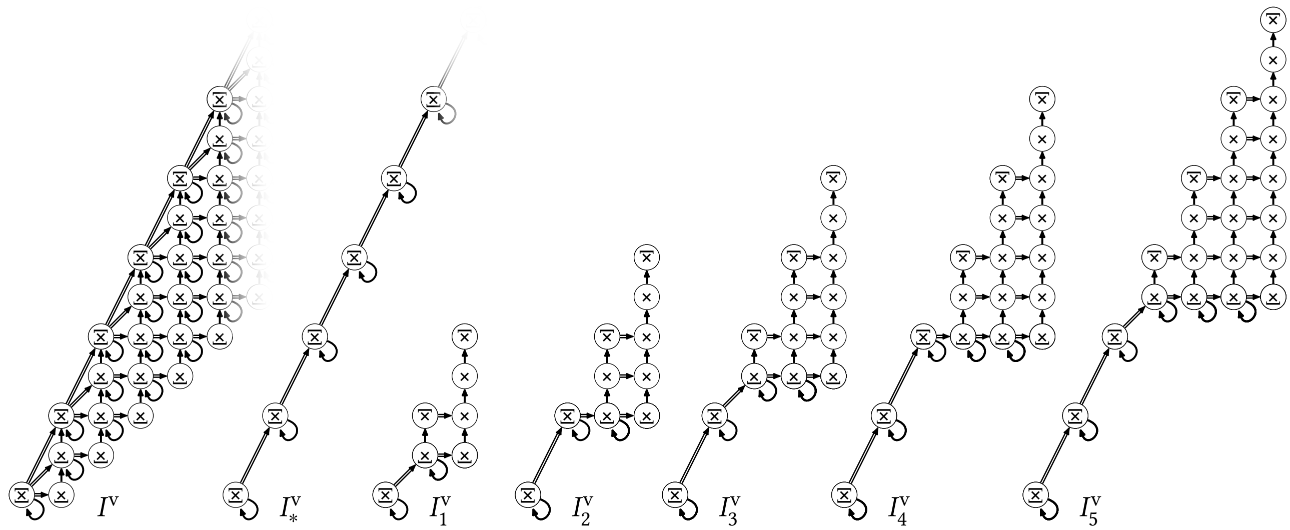


Figure 4: Two infinite universal models of $\mathcal{K}^{\mathrm{v}}$ (I^{v} from Definition 10 and I_*^{v} from Definition 11), and finite atomsets $I_1^{\mathrm{v}} - I_5^{\mathrm{v}}$ of the sequence $(I_n^{\mathrm{v}})_{n \in \mathbb{N}}$ from Definition 12 (recall that $I_0^{\mathrm{v}} = F^{\mathrm{v}}$).

8 ROBUST AGGREGATIONS SAVE THE DAY

Recall that the steepening staircase example demonstrates that a bounded-treewidth chase sequence does not warrant the existence of a universal model of finite treewidth. This blocks the traditional approach for showing decidability of CQ entailment. However, we are still able to establish this desired result by other means, as demonstrated in the course of the next two sections.

More specifically, we show that *CQ entailment is decidable for the class of KBs having a recurringly treewidth-bounded core chase sequence* (forthcoming Theorem 1). To do so, we go through the following steps. Firstly, we resort to a weaker notion than universality, namely *finite universality* (Definition 13). We show that finitely universal models can play the same role as universal models when it comes to CQ entailment (Proposition 9). Secondly, we define a novel way to compute the result of a derivation, namely the *robust aggregation* of a derivation, and show that the robust aggregation of any fair derivation is a finitely universal model (Proposition 11). Finally, in Section 9, we show that the robust aggregation of a derivation having recurringly bounded treewidth has finite treewidth (Proposition 12), and conclude by adapting Courcelle's theorem to show that CQ entailment is decidable for KBs admitting a finitely universal model of finite treewidth (Theorem 2). In the following, we detail the employed notions and arguments laid out above.

Definition 13 (Finite universality). *An atomset I is* finitely universal *for $\mathcal{K}$ if each finite subset of I is universal for $\mathcal{K}$.*

Proposition 9. *Let M be a finitely universal model of a KB $\mathcal{K}$, and let Q be a CQ. Then $\mathcal{K} \models Q$ iff $M \models Q$.*

Proof. ($\Leftarrow$) Let σ be a homomorphism from Q to M. As M is finitely universal, the finite subset $\sigma(Q)$ of M maps to any model I of $\mathcal{K}$ by some σ', thus $\sigma' \circ \sigma$ maps Q to I. ($\Rightarrow$) Since $\mathcal{K} \models Q$ and M is a model of $\mathcal{K}$, $M \models Q$. □

Defining Robust Aggregations. For non-monotonic derivations $\mathcal{D} = (F_i)_{i \in \mathfrak{I}}$, it may happen that an atom *at* is in some F_i, but "disappears" at some later step j. Yet, *at* will still belong to the natural aggregation $\mathcal{D}^*$. Intuitively, the natural aggregation generates

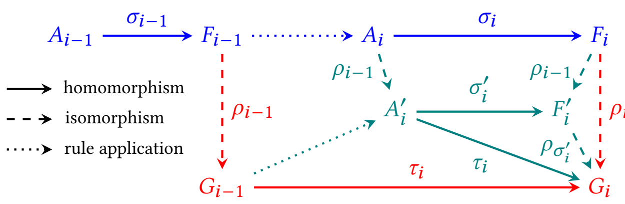


Figure 5: Building the robust sequence associated with $\mathcal{D}$.

atomsets that are "too big" (this is why they may not be models). We thus introduce a new type of aggregation, called *robust aggregation*, that, instead of merely combining all atomsets F_i along the derivation, combines their collapsed versions obtained via preemptive applications of future simplifications σ_j along the derivation. Defining this result is not immediate, however, since a variable could be indefinitely re-mapped through simplifications along a derivation. Observe that, in the staircase example, the core chase maps X_0^0 to X_0^1, then X_0^1 to X_0^2, etc., and there is no way we can define the ultimate image of X_0^0 unless we can force the simplification to stabilize at some point. This is the goal of the *robust renaming*, for which we assume a bijection *rank* of the variables $\mathcal{X}$ with $\mathbb{N}$, and use the total ordering $<_{\mathcal{X}}$ on $\mathcal{X}$ defined by $X <_{\mathcal{X}} Y$ iff $rank(X) < rank(Y)$.

Definition 14 (Robust renaming). *Let A be an atomset and let σ be a retraction of A. The* robust renaming associated with σ *is the substitution ρ_σ of $vars(\sigma(A))$ that maps any variable X of $\sigma(A)$ to the $<_{\mathcal{X}}$-smallest variable of $\sigma^{-1}(X)$. We let $\tau_\sigma = \rho_\sigma \circ \sigma$.*

It is immediate that ρ_σ is an isomorphism from $\sigma(A)$ to $\tau_\sigma(A)$, and, for any variable X in A, $\tau_\sigma(X)$ is a constant or $\rho_\sigma(X) \leq_{\mathcal{X}} X$. Let us now inductively apply those robust renamings along a derivation.

Definition 15 (Robust sequence). *Let $\mathcal{D} = (F_i)_{i\in\mathfrak{I}}$ be a derivation. The* robust sequence associated with $\mathcal{D}$ *is the sequence of atomsets $(G_i)_{i\in\mathfrak{I}}$ defined inductively by (see Figure 5 for induction step):*

- *With $A_0 = F$, $F_0 = \sigma_0(A_0)$, and $\rho_0 = \rho_{\sigma_0}$, we define $G_0 = \rho_0(F_0)$;*
- *$\forall i \in \mathfrak{I}$ with $i > 0$, if $F_{i-1} = \sigma_{i-1}(A_{i-1})$, $A_i = \alpha(F_{i-1}, tr)$, $F_i = \sigma_i(A_i)$ and $G_{i-1} = \rho_{i-1}(F_{i-1})$ (F_{i-1} and G_{i-1} being isomorphic), we build G_i and an isomorphism ρ_i from F_i to G_i as follows:*
  - *let $A'_i = \rho_{i-1}(A_i)$ (see that $A'_i = \alpha(G_{i-1}, \rho_{i-1}(tr))$), with the same fresh variables as in $\alpha(F_{i-1}, tr)$) and $F'_i = \rho_{i-1}(F_i)$;*
  - *then $\sigma'_i = \rho_{i-1} \circ \sigma_i \circ \rho_{i-1}^{-1}$ is a retraction such that $\sigma'_i(A'_i) = F'_i$;*
  - *we define $G_i = \rho_{\sigma'_i}(F'_i)$, with $\rho_{\sigma'_i}$ the robust renaming associated with σ'_i and $\rho_i = \rho_{\sigma'_i} \circ \rho_{i-1}$ an isomorphism from F_i to G_i;*
  - *furthermore, we denote by $\tau_i = \tau_{\sigma'_i} = \rho_{\sigma'_i} \circ \sigma'_i$ the homomorphism from A'_i to G_i. See that τ_i also maps $G_{i-1} \subseteq A'_i$ to G_i.*

Note that (G_i) is not a derivation, since the τ_i from A'_i to G_i are not endomorphisms. However, every G_i is isomorphic to F_i, and we show that variables are finitely renamed along this sequence.

Proposition 10. *Let $(G_i)_{i\in\mathfrak{I}}$ be an associated robust sequence. For $i, j \in \mathfrak{I}$ with $i < j$, let $\bar{\tau}_i^j = \tau_j \circ \cdots \circ \tau_{i+1}$ denote the composition of all τ_ℓ between G_i and G_j. Then, for any $X \in vars(G_i)$, there is $j \in \mathfrak{I}$ with $j > i$ such that $\bar{\tau}_i^j(X) = Y \in terms(G_j)$ and for all $k \in \mathfrak{I}$ with $k > j$, $\bar{\tau}_j^k(Y) = Y$ (i.e., Y is stable from G_j on). We let $\bar{\tau}(X) = Y$.*

Proof. Let $X \in vars(G_i)$, then $\tau_{i+1}(X) = \tau_{\sigma'_{i+1}}(X) \leq_{\mathcal{X}} X$. Consider some arbitrary $j \in \mathfrak{I}$ with $j > i$. Among the homomorphisms τ_ℓ that $\bar{\tau}_i^j$ is composed of, there can be at most $rank_{\mathcal{X}}(X)$ many of them that are effectively decreasing (causing $\bar{\tau}_i^{\ell-1}(X) <_{\mathcal{X}} \bar{\tau}_i^\ell(X)$). □

We now use the $\bar{\tau}(G_i)$ to define the *robust aggregation*. Note that, contrary to (F_i) or (G_i), the sequence $(\bar{\tau}(G_i))$ is monotonic.

Definition 16 (Robust aggregation). *Given a derivation $\mathcal{D} = (F_i)_{i\in\mathfrak{I}}$ and its associated robust sequence $(G_i)_{i\in\mathfrak{I}}$, the* robust aggregation *of $\mathcal{D}$ is the (possibly infinite) atomset $\mathcal{D}^{\circledast} = \bigcup_{i\in\mathfrak{I}} \bar{\tau}(G_i)$.*

Semantic Properties of Robust Aggregations. The steepening staircase shows that the robust aggregation of a derivation is not always universal. Indeed, consider the KB $\mathcal{K}^{\mathrm{h}}$ (from Definition 7) and let $<_{\mathcal{X}}$ be an order on the variables with $j < k \Rightarrow X_j^i <_{\mathcal{X}} X_k^i$. The core chase on $\mathcal{K}^{\mathrm{h}}$ begins building the first step S_0^{h} of I^{h}, and all simplifications are the identity until done. Now, the first proper retraction maps X_0^0 to X_0^1 and X_1^0 to X_1^1, so the robust renaming generates G_{i_1}, which is isomorphic to the column C_1^{h}, but its variables are named (from bottom to top) X_0^0 and X_1^0. Likewise, from successive proper retraction steps, we obtain G_{i_j} isomorphic to C_j^{h} but with variables named $X_0^0, X_1^0, X_2^1, \ldots, X_{j+1}^j$. Note that $\bar{\tau}(G_{i_j}) = G_{i_j}$ holds: every variable is stable since subsequent re-mappings would have to be within the same row, yet all variables therein are $<_{\mathcal{X}}$-greater. Then, the robust aggregation $\mathcal{D}^{\circledast}$ is isomorphic to the infinite column $\tilde{I}^{\mathrm{h}}$, with variables named $X_0^0, X_1^0, X_2^1, \ldots, X_{j+1}^j, \ldots$, which is not universal, but is a finitely universal model, as stated below.

Proposition 11. *Let $\mathcal{D}$ be a derivation from $\mathcal{K}$. Then (1) $\mathcal{D}^{\circledast}$ is finitely universal for $\mathcal{K}$; and (2) if $\mathcal{D}$ is fair, $\mathcal{D}^{\circledast}$ is a model of $\mathcal{K}$.*

To prove this proposition, we rely on the next lemma, which states that any finite part of $\mathcal{D}^{\circledast}$ is "stably present" from a certain element on in the robust sequence associated with $\mathcal{D}$.

Lemma 1. *Let $\mathcal{D}$ be a derivation and let $(G_i)_{i\in\mathfrak{I}}$ be the robust sequence associated with $\mathcal{D}$. For any finite subset A of $\mathcal{D}^{\circledast}$, there is some $k \in \mathfrak{I}$ such that $A \subseteq G_r$ for every $r \in \mathfrak{I}$ with $r \geq k$.*

Sketch of proof. See that *(i)* the $\bar{\tau}(G_i)$ form a monotonic sequence and then, thanks to Proposition 10, that *(ii)* for every $\bar{\tau}(G_i)$, there exists $k \in \mathfrak{I}$ such that $\bar{\tau}(G_i) \subseteq G_r$ for every $r \geq k$. Thanks to *(i)*, there is some i with $A \subseteq \bar{\tau}(G_i)$ and we conclude with *(ii)*. □

Proof of Proposition 11. *(1)* Let M be an arbitrary model of $\mathcal{K}$, and let I be be any finite subset of $\mathcal{D}^{\circledast}$. By Lemma 1, there is some k such that $I \subseteq G_k$. Now G_k is isomorphic to F_k, which is universal (from Proposition 1), so G_k (hence also I) maps to M.

(2) Let $\mathcal{D} = (F_i)_{i\in\mathfrak{I}}$ be a fair derivation from (F, Σ) and $(G_i)_{i\in\mathfrak{I}}$ be its associated robust sequence. Since τ_0 maps F to G_0, $\bar{\tau} \circ \tau_0$ maps F to $\mathcal{D}^{\circledast}$, thus $\mathcal{D}^{\circledast}$ is a model of F. Consider now any trigger tr for $\mathcal{D}^{\circledast}$. By Lemma 1, there exists some $j \in \mathfrak{I}$ such that tr is a trigger for G_r for any $r \in \mathfrak{I}$ with $r \geq j$. Since ρ_r is an isomorphism from F_r to G_r, we obtain that $\rho_r^{-1}(tr)$ is a trigger for F_r. Since $\mathcal{D}$ is fair, there exists some $s \in \mathfrak{I}$ with $s \geq r$ such that the trigger $\tilde{\sigma}_r^s \circ \rho_r^{-1}(tr)$ for F_s is satisfied in F_s. Now since ρ_s is an isomorphism from F_s to G_s, it follows that $\rho_s \circ \tilde{\sigma}_r^s \circ \rho_r^{-1}(tr)$ is a satisfied trigger for G_s. We first see that $\sigma_{r+1} \circ \rho_r^{-1} = \rho_{r+1}^{-1} \circ \tau_{r+1}$. By applying this property iteratively, we show that $\rho_s \circ \tilde{\sigma}_r^s \circ \rho_r^{-1} = \bar{\tau}_r^s$. Then $\rho_s \circ \tilde{\sigma}_r^s \circ \rho_r^{-1}(tr) = \bar{\tau}_r^s(tr) = tr$ is a trigger for G_s satisfied in G_s, and thus satisfied in $\mathcal{D}^{\circledast}$. □

Hence, both natural and robust aggregations indicate whether a CQ is entailed by a KB. Yet, natural aggregation provides an instance that is universal but not always a model, while the more complex robust aggregation provides a model which might be only finitely universal. We show next how the latter case can still be utilized towards proving Theorem 1.

9 DECIDABILITY THROUGH TREEWIDTH

The steepening staircase example shows that the natural aggregation of the core chase may have infinite treewidth even if the chase sequence is uniformly treewidth-bounded. The next proposition provides two results: Firstly, the natural aggregation is indeed treewidth-preserving for *monotonic* derivations, generalizing a result by Baget et al. [3] for the restricted chase. Secondly (and more importantly), robust aggregation is superior to natural aggregation in that treewidth preservation can be shown to hold even for non-monotonic chases. Both results rely upon the compactness of treewidth [18]: if F is an atomset where $tw(F') \leq k$ holds for every finite subset $F' \subseteq F$, then $tw(F) \leq k$.

Proposition 12. *For any derivation $\mathcal{D}$ that is recurringly treewidth-bounded by some integer k, the following hold:*
(1) $\mathcal{D}$'s natural aggregation $\mathcal{D}^*$ has treewidth $\leq k$, if $\mathcal{D}$ is monotonic.
(2) $\mathcal{D}$'s robust aggregation $\mathcal{D}^{\circledast}$ has treewidth $\leq k$.

Proof. Let I be a finite subset of $\mathcal{D}^*$ (for proof of *(1)*) or $\mathcal{D}^{\circledast}$ (for proof of *(2)*). There is some $p \in \mathfrak{I}$ such that, $\forall r \geq p \in \mathfrak{I}$, we can exhibit some I_r isomorphic to F_r with $I \subseteq I_r$. To prove *(1)*, $\mathcal{D}$ being monotonic, we can define $I_r = F_r$. To prove *(2)*, we rely upon Lemma 1 and define $I_r = G_r$. Since $\mathcal{D}$ is recurrently treewidth-bounded, there is some $s \geq p \in \mathfrak{I}$ such that $tw(F_s) \leq k$. Thus $tw(I) \leq tw(I_s) = tw(F_s) \leq k$, and we conclude, thanks to compactness of treewidth, that $\mathcal{D}^*$ or $\mathcal{D}^{\circledast}$ has treewidth $\leq k$. ◻

The last missing insight is that the existence of treewidth-bounded *finitely universal* models suffices to establish decidability of CQ entailment.[3] We obtain this result via a mild generalization of respective statements for *universal* models [3, 7, 11].

Theorem 1. *Let $\mathfrak{C}$ be a class of knowledge bases for which every $\mathcal{K} = (F, \Sigma) \in \mathfrak{C}$ has a model I that is finitely universal for $\mathcal{K}$ and that satisfies $tw(I) \in \mathbb{N}$. Then CQ entailment for $\mathfrak{C}$ is decidable.*

Sketch of proof. $\mathcal{K} \models Q$ can be detected in finite time due to the completeness of first-order logic. $\mathcal{K} \not\models Q$ can be detected by incrementing k stepwise and checking if $\mathcal{K} \wedge (\neg Q)$ has a model of treewidth k, which is decidable. ◻

We finally obtain our main result, which follows from Propositions 11 and 12, and Theorem 2:

Theorem 2. *CQ entailment is decidable for the class of KBs having a recurringly treewidth-bounded core chase sequence.*

We end this section by using this decidability result to define a new class of rulesets and discussing its relationship with existing abstract decidable classes. As usual in the existential rule setting, the considered property can be abstracted from the underlying database, obtaining a new fragment of existential rules that – thanks to Theorem 1 – warrants decidable CQ entailment and properly subsumes and reconciles other classes with that property.[4]

[3]However, no upper complexity bounds are entailed. This holds even for the more restricted class of KBs with finite, "properly" universal models [5].

Definition 17. *A ruleset Σ is called* core-bts*, if for every finite atomset F, there exists a core chase sequence for the KB (F, Σ), whose treewidth is recurringly bounded by some $k \in \mathbb{N}$.*

Proposition 13. *CQ entailment is decidable for any ruleset that is core-bts. Moreover, core-bts subsumes both finite expansion sets (fes) and bounded treewidth sets (bts), which are mutually incomparable.*

10 CONCLUSION AND FUTURE WORK

In this paper, we have investigated ways of exploiting properties of the core chase in non-terminating settings, with the main goal of ensuring decidability of CQ entailment based on treewidth guarantees for the atomsets occurring in chase sequence.

On the negative side, we found that, contrary to plausible expectations, the existence of a treewidth-bounded core-chase sequence does not coincide with the existence of a treewidth-bounded universal model, nor is there a subsumption in one of the two directions: On one hand, we exhibited a KB $\mathcal{K}^{\mathrm{h}}$ admitting a core-chase sequence the treewidth of which is uniformly bounded by 2, while all its universal models are of unbounded treewidth. On the other hand, we described a KB $\mathcal{K}^{\mathrm{v}}$ admitting an infinite universal model of treewidth 1, while all corresponding core chase sequences consist of structures of ever increasing treewidth.

On the positive side, we showed how a given core chase sequence can be robustly aggregated into a (potentially infinite) atomset that is a model of the underlying knowledge base, while satisfying exactly those CQs entailed by it. We also showed that for any such core chase sequence that is recurringly treewidth-bounded, the aggregated atomset will be of finite treewidth. Together, these findings establish decidability of CQ entailment for all knowledge bases with a recurringly treewidth-bounded core chase. Abstracting from concrete databases, this yields a novel, very general abstract class of recurringly treewidth-bounded rulesets, ensuring decidability of CQ entailment and subsuming the two previously known incomparable classes fes and bts.

Future work on the topic will clarify under what circumstances the robust aggregation produces cores (according to some of the many existing non-equivalent definitions of cores in the infinite [4]). Also, we will investigate the relationship of our approach to the *stable chase* introduced by Carral et al. [8], which also produces (not necessarily universal) models satisfying exactly the entailed CQs. Note that the stable chase is quite elaborate and not subsumed by our current generic definition of derivation: the computation occasionally "jumps back" to earlier sequence elements and starts rebuilding the sequence from there.

[4]Notably, this corrects inaccurate statements in prior work by Baget et al. [3], where bts was claimed to subsume fes. The reason for this misconception was a definition of bts using cores, whereas the proof of decidability of CQ entailment for this class was flawed, as it erroneously assumed that the natural aggregation over a (treewidth-bounded) core chase sequence produces a (treewidth-bounded) universal model. The current paper also corrects this earlier work, showing that the decidability claim made therein can be salvaged by other means.

REFERENCES

[1] Jean-François Baget, Michel Leclère, and Marie-Laure Mugnier. 2010. Walking the Decidability Line for Rules with Existential Variables. In *Proceedings of the 12th International Conference on Principles of Knowledge Representation and Reasoning (KR'10)*, Fangzhen Lin, Ulrike Sattler, and Miroslaw Truszczynski (Eds.). AAAI Press. http://aaai.org/ocs/index.php/KR/KR2010/paper/view/1216

[2] Jean-François Baget, Marie-Laure Mugnier, Sebastian Rudolph, and Michaël Thomazo. 2011. Walking the Complexity Lines for Generalized Guarded Existential Rules. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11)*, Toby Walsh (Ed.). IJCAI/AAAI, 712–717. https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-126

[3] Jean-François Baget, Michel Leclère, Marie-Laure Mugnier, and Eric Salvat. 2011. On rules with existential variables: Walking the decidability line. *Artificial Intelligence* 175, 9 (2011), 1620–1654. https://doi.org/10.1016/j.artint.2011.03.002

[4] Bruce L. Bauslaugh. 1995. Core-like properties of infinite graphs and structures. *Discrete Mathematics* 138, 1-3 (1995), 101–111.

[5] Camille Bourgaux, David Carral, Markus Krötzsch, Sebastian Rudolph, and Michaël Thomazo. 2021. Capturing Homomorphism-Closed Decidable Queries with Existential Rules. In *Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning, KR 2021*, Meghyn Bienvenu, Gerhard Lakemeyer, and Esra Erdem (Eds.). 141–150. https://doi.org/10.24963/kr.2021/14

[6] A. Calì, G. Gottlob, and M. Kifer. 2008. Taming the Infinite Chase: Query Answering under Expressive Relational Constraints. In *Proceedings of the 11th International Conference on Principles of Knowledge Representation and Reasoning (KR'08)*. AAAI Press, 70–80.

[7] Andrea Calì, Georg Gottlob, and Michael Kifer. 2013. Taming the Infinite Chase: Query Answering under Expressive Relational Constraints. *Journal of Artificial Intelligence Research* 48 (2013), 115–174. https://doi.org/10.1613/jair.3873

[8] David Carral, Markus Krötzsch, Maximilian Marx, Ana Ozaki, and Sebastian Rudolph. 2018. Preserving Constraints with the Stable Chase. In *Proceedings of the 21st International Conference on Database Theory (ICDT'18) (LIPIcs, Vol. 98)*, Benny Kimelfeld and Yael Amsterdamer (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 12:1–12:19. https://doi.org/10.4230/LIPIcs.ICDT.2018.12

[9] Alin Deutsch, Alan Nash, and Jeffrey B. Remmel. 2008. The chase revisited. In *Proceedings of the 27th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS'08)*, Maurizio Lenzerini and Domenico Lembo (Eds.). ACM, 149–158. https://doi.org/10.1145/1376916.1376938

[10] R. Fagin, P. G. Kolaitis, R. J. Miller, and L. Popa. 2005. Data Exchange: Semantics and Query Answering. *Theoretical Computer Science* 336, 1 (2005), 89–124.

[11] Thomas Feller, Tim S. Lyon, Piotr Ostropolski-Nalewaja, and Sebastian Rudolph. 2023. Finite-Cliquewidth Sets of Existential Rules: Toward a General Criterion for Decidable yet Highly Expressive Querying. In *Proceedings of the 26th International Conference on Database Theory (ICDT 2023) (LIPIcs)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik. To appear. Preprint available via https://arxiv.org/abs/2209.02464.

[12] Kurt Gödel. 1929. *Über die Vollständigkeit des Logikkalküls.* Ph. D. Dissertation. Universität Wien.

[13] Georg Gottlob, Nicola Leone, and Francesco Scarcello. 2003. Robbers, Marshals, and Guards: Game Theoretic and Logical Characterizations of Hypertree Width. *J. Comput. Syst. Sci.* 66, 4 (jun 2003), 775–808. https://doi.org/10.1016/S0022-0000(03)00030-8

[14] Gösta Grahne and Adrian Onet. 2018. Anatomy of the Chase. *Fundamenta Informaticae* 157, 3 (2018), 221–270. https://doi.org/10.3233/FI-2018-1627

[15] George Konstantinidis and José Luis Ambite. 2014. Optimizing the Chase: Scalable Data Integration under Constraints. *Proc. VLDB Endow.* 7, 14 (2014), 1869–1880. https://doi.org/10.14778/2733085.2733093

[16] Markus Krötzsch and Sebastian Rudolph. 2011. Extending Decidable Existential Rules by Joining Acyclicity and Guardedness. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11)*, Toby Walsh (Ed.). IJCAI/AAAI, 963–968. https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-166

[17] B. Marnette. 2009. Generalized schema-mappings: from termination to tractability. In *Proceedings of the 28th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS'09)*, J. Paredaens and J. Su (Eds.). ACM, 13–22.

[18] Robin Thomas. 1988. The Tree-Width Compactness Theorem for Hypergraphs. Available via https://people.math.gatech.edu/~thomas/PAP/twcpt.pdf.

PROOFS OF SECTION 3

Fact 3. *If tr is a trigger for F, μ maps F to I and I satisfies $\mu(tr)$, then there is μ' (compatible with μ) that maps $\alpha(F, tr)$ to I.*

Lemma 2. *For every fair derivation $\mathcal{D} = (F_i)_{i \in \mathfrak{I}}$, there exists a fair monotonic derivation $\mathcal{D}_{mon} = (G_i)_{i \in \mathfrak{I}}$ such that for every $i \in \mathfrak{I}$, there is a retraction from G_i to F_i.*

Proof. From $\mathcal{D} = (F_i)_{i \in \mathfrak{I}}$, let us first build inductively a derivation $\mathcal{D}_{mon} = (G_i)_{i \in \mathfrak{I}}$ such that $G_0 = A_0 = F$, $F_0 = \sigma_0(F_0)$ and $\forall i > 0 \in \mathfrak{I}$, $F_i = \sigma_i(A_i)$ with $A_i = \alpha(F_{i-1}, tr_i)$, we can define $G_i = \alpha(G_{i-1}, tr_i)$. See that σ_0 is a retraction from G_0 to F_0, and so the trigger tr_1 is also a trigger in G_0, allowing us to build G_1. Now we claim that σ_0 is a retraction from G_1 to A_1, and thus $\sigma_1 \circ \sigma_0$ is a retraction from G_1 to F_1. An induction based upon these remarks shows that for $i \in \mathfrak{I}$, $\tilde{\sigma}^i = \sigma_i \circ \cdots \circ \sigma_0$ is a retraction from G_i to F_i that allows us to build G_{i+1}. The derivation $\mathcal{D}_{mon}$ we obtain is monotonic, but it remains to check that it is fair. Given any trigger tr for some G_i, $\tilde{\sigma}^i(tr)$ is a trigger for F_i and thus (fairness of $\mathcal{D}$) there exists $j \in \mathfrak{I}$ such that $\tilde{\sigma}^j(tr)$ is a trigger for F_j satisfied in F_j, and the trigger tr for G_j for which $\tilde{\sigma}^j$ is a retraction into F_j is satisfied in G_j. □

Proposition 1 *(Extended version)* Let $\mathcal{D}$ be a derivation from $\mathcal{K}$. Then:

(1) $\mathcal{D}^*$ is universal for $\mathcal{K}$;
(2) if $\mathcal{D}$ is finite, $\mathcal{D}^+$ is universal for $\mathcal{K}$;
(3) if $\mathcal{D}$ is monotonic and fair, $\mathcal{D}^*$ is a model of $\mathcal{K}$;
(4) if $\mathcal{D}$ is finite and fair, $\mathcal{D}^+$ is a model of $\mathcal{K}$;
(5) if $\mathcal{D}$ is fair and Q is a CQ, $\mathcal{K} \models Q$ iff $\mathcal{D}^* \models Q$.

Proof. Let M be an arbitrary model of $\mathcal{K}$. We first prove the existence of homomorphisms $F_i \to M$ by induction over i. The existence of some homomorphism $F_0 \to M$ is immediate by assumption. Then, if there is a homomorphism μ_j from some F_j of $\mathcal{D}$ to M, then there is a homomorphism μ_{j+1} of F_{j+1} to M such that μ_{j+1} is compatible with μ_j. We have $F_{j+1} = \sigma_{j+1}(\alpha(F_j, tr_{i+1}))$. See that $\mu_j(tr_{i+1})$ is a trigger for M, satisfied in M since it is a model of $\mathcal{K}$. Then (Fact 3) there is a homomorphism μ from $\alpha(F_j, tr_{j+1})$ to M compatible with μ_j and its restriction μ_{j+1} to the variables of $\sigma_{j+1}(\alpha(F_j, tr_{j+1}))$ is a homomorphism from F_{j+1} to M compatible with μ_j.

(2) Hence, M is a model of every F_i in $\mathcal{D}$: each instance F_i is universal and, if $\mathcal{D}$ is finite, then the final result $\mathcal{D}^+ = F_k$ is universal. □

(1) Now we claim that since a variable present both in F_i and F_j must appear in all atomsets between F_i and F_j (a consequence of the usage of fresh variables), the pairwise compatibility of the μ_i between succsessive atomsets implies global compatibility of all μ_i. We conclude by pointing out that $\bigcup_{i \in \mathfrak{I}} \mu_i$ is a homomorphism from $\mathcal{D}^*$ to M, and thus that $\mathcal{D}^*$ is universal. □

(4) If the derivation is finite, its final result $\mathcal{D}^+ = F_k$ is a model of F (as $\tilde{\sigma}_0^j \circ \sigma_0$ is a homomorphism from F to any F_j in the derivation) and, by Definition 3, for any trigger tr for F_k, there is some $j \geq k$ (which must be $j = k$) such that $\tilde{\sigma}_k^k(tr) = tr$ is a satisfied trigger for F_k. □

(3) In the case of an infinite fair derivation, we first point out that $\mathcal{D}^*$ contains $F_0 = \sigma_0(F)$, so it is a model of F. Then consider any trigger tr for $\mathcal{D}^*$: it is also a trigger for some F_i in $\mathcal{D}$. By Definition 3, there exists some $j \in \mathfrak{I}$ with $j \geq i$ such that $\tilde{\sigma}_i^j(tr)$ is a satisfied trigger for F_j. Since $\mathcal{D}$ is monotonic, σ_i^j is the identity and thus $\tilde{\sigma}_i^j(tr) = tr$ is satisfied in $\mathcal{D}^*$. □

$(5, \Leftarrow)$ Let π be a homomorphism from Q to $\mathcal{D}^*$. Since $\mathcal{D}^*$ is universal (by (1)), it maps to any model M of $\mathcal{K}$. Let τ_M be a homomorphism from $\mathcal{D}^*$ to M, then $\tau_M \circ \pi$ maps Q to M. □

$(5, \Rightarrow)$ Let us now consider the fair monotonic derivation $\mathcal{D}_{mon}$ from Lemma 2. We now that $\mathcal{D}^*_{mon}$ is a model of $\mathcal{K}$, and then if $\mathcal{K} \models Q$, then there is a homomorphism π from Q to $\mathcal{D}^*_{mon}$. Since $\pi(Q)$ is finite, there is some atomset G_i in $\mathcal{D}_{mon}$ such that $\pi(Q) \subseteq G_i$. We know there is a retract $\tilde{\sigma}^i$ from G_i to F_i, so $\tilde{\sigma}^i \circ \pi$ is a homomorphism from Q to F_i and so from Q to $\mathcal{D}^*$. □

PROOFS OF SECTION 6

There is a sequence of rule applications from any C_k^{h} producing S_k^{h}. Let us consider C_k^h with variables named, from bottom to top $(X_0^k, X_1^k, \ldots, X_k^k)$. Then we apply the rules in the following way:

Rule	Homomorphism	Atoms produced
R_1^h	$X \mapsto X_k^k$	$\mathtt{v}(X_k^k, X_{k+1}^k), \mathtt{h}(X_{k+1}^k, X_{k+1}^{k+1}), \mathtt{c}(X_{k+1}^{k+1}), \mathtt{h}(X_k^k, X_,^{k+1}k), \mathtt{v}(X_k^{k+1}, X_{k+1}^{k+1})$
R_2^h	$X \mapsto X_{k-1}^k, X' \mapsto X_k^k, Y' \mapsto X_k^{k+1}$	$\mathtt{h}(X_{k-1}^k, X_{k-1}^{k+1}), \mathtt{v}(X_{k-1}^{k+1}, X_k^{k+1}), \mathtt{c}(X_k^{k+1})$
$\vdots$	$\vdots$	$\vdots$
R_2^h	$X \mapsto X_0^k, \quad X' \mapsto X_1^k, Y' \mapsto X_1^{k+1}$	$\mathtt{h}(X_0^k, X_0^{k+1}), \mathtt{v}(X_0^{k+1}, X_1^{k+1}), \mathtt{c}(X_1^{k+1})$
R_3^h	$X \mapsto X_0^k Y \mapsto X_0^{k+1}$	$\mathtt{c}(X_0^{k+1}), \mathtt{h}(X_0^{k+1}, X_0^{k+1})$
R_4^h	$X \mapsto X_0^{k+1}, X' \mapsto X_1^{k+1}$	$\mathtt{h}(X_1^{k+1}, X_1^{k+1})$
$\vdots$	$\vdots$	$\vdots$
R_4^h	$X \mapsto X_k^{k+1}, X' \mapsto X_{k+1}^{k+1}$	$\mathtt{h}(X_{k+1}^{k+1}, X_{k+1}^{k+1})$

The obtained result is indeed S_k^h.

Proposition 5. *No universal model of $\mathcal{K}^{\mathrm{h}}$ has finite treewidth.*

Proof. We call $\mathtt{v}$-path (resp. $\mathtt{h}$-path) in an atomset a non-empty sequence of nulls such that, for any two consecutive nulls X_i and X_{i+1}, the atomset contains the atom $\mathtt{v}(X_i, X_{i+1})$ (resp. $\mathtt{h}(X_i, X_{i+1})$). By analogy to graphs, the length of a path is $n - 1$ if it is a sequence of n nulls.

Let U be an arbitrary universal model of $\mathcal{K}^{\mathrm{h}}$. We first point out that I^{h} and U being both universal models, they homomorphically map to each other. We let h_1 denote the homomorphism from I^{h} to U and let h_2 denote the homomorphism from U to I^{h}. Then $h = h_2 \circ h_1$ is an endomorphism on I^{h}, the properties of which we will now inspect further. We make use of the following notation: for $h(X_j^i) = X_\ell^k$, we denote k by $h_x(i, j)$ and ℓ by $h_y(i, j)$, that is, we let $h(X_j^i) = X_{h_y\langle i,j\rangle}^{h_x\langle i,j\rangle}$. We make the following observations (which hold for all endomorphisms on I^{h}):

(1) $h_y(i, 0) = 0$ since $\mathtt{f}$ holds precisely for all nulls X_0^i.
(2) $h_y(i, j) = j$, inductively with (1) as base case and the observation that h must preserve the length of incoming $\mathtt{v}$-paths rooted in some $\mathtt{f}$.
(3) $h_x(i, j) = h_x(i, j + 1)$, since this is the only way for h to preserve the $\mathtt{v}$-atoms.
(4) $h_x(i, j) = h_x(i, k)$, via iteration of (3).
(5) $h_x(i, j) \geq i$, due to (2) and the fact that X_j^i does not exist for $j > i + 1$.
(6) $h_x(i + 1, j) = h_x(i, j)$ or $h_x(i + 1, j) = h_x(i, j) + 1$, since this is the only way for h to preserve the $\mathtt{h}$-atoms.
(7) $h_y(i + 1, j) = h_y(i, j)$ since this is the only way for h to preserve the $\mathtt{h}$-atoms.
(8) There are $k, \ell \in \mathbb{N}$ such that $h_x(i, j) = i + \ell$ for all $i > k$. This is a consequence of (5) and (6).
(9) There is a $k \in \mathbb{N}$ such that the restriction of h to the X_j^i with $i > k$ is injective. Follows from (8), for the same k, and (2).

If we now let I_-^{h} be I^{h} restricted to terms X_j^i with $i > k$, we obtain that h is an isomorphism from I_-^{h} to $h(I_-^{\mathrm{h}})$, i.e., $I_-^{\mathrm{h}} \cong h(I_-^{\mathrm{h}})$. Since $h = h_2 \circ h_1$, this means that h_1 must be an isomorphism from I_-^{h} to $h_1(I_-^{\mathrm{h}})$ and h_2 must be an isomorphism from $h_1(I_-^{\mathrm{h}})$ to $h_2(h_1(I_-^{\mathrm{h}})) = h(I_-^{\mathrm{h}})$. Therefore, $tw(I_-^{\mathrm{h}}) = tw(h_1(I_-^{\mathrm{h}})) = tw(h_2(h_1(I_-^{\mathrm{h}})))$ (*). Now, for any given $n \in \mathbb{N}$ with $n > k$, take $\mathcal{T}_{n\times n} = \{X_j^i \mid n + 1 \leq i \leq 2n \text{ and } 0 \leq i \leq n - 1\} \subseteq terms(I^{\mathrm{h}})$. Consequently, $\mathcal{T}_{n\times n}$ witnesses that I_-^{h} contains a $n \times n$ grid. Yet, as n can be chosen arbitrarily large, I_-^{h} contains grids of arbitrary size and thus cannot have finite treewidth, i.e., $tw(I_-^{\mathrm{h}}) \notin \mathbb{N}$ (**). From these insights, we can conclude

$$I_-^{\mathrm{h}} \subseteq I^{\mathrm{h}} \Longrightarrow h_1(I_-^{\mathrm{h}}) \subseteq h_1(I^{\mathrm{h}}) \subseteq U \overset{\text{Fact 1}}{\Longrightarrow} tw(h_1(I_-^{\mathrm{h}})) \leq tw(U) \overset{(*)}{\Longrightarrow} tw(I_-^{\mathrm{h}}) \leq tw(U) \overset{(**)}{\Longrightarrow} tw(U) \notin \mathbb{N}.$$

□

PROOFS OF SECTION 7

Proposition 8. *The following hold:*

(1) Every I_n^{v} is a core.
(2) I_n^{v} has a treewidth of at least $\lceil n/3 \rceil + 1$.
(3) For every core chase sequence $(F_i)_{i\in\mathbb{N}}$ for $\mathcal{K}^{\mathrm{v}}$, there is an unbounded monotonic function $f : \mathbb{N} \to \mathbb{N}$ such that, for every $n \in \mathbb{N}$, $I_{f(n)}^{\mathrm{v}}$ is isomorphic to a subset of F_n.
(4) For every core chase sequence $(F_i)_{i\in\mathbb{N}}$ for $\mathcal{K}^{\mathrm{v}}$ and any $m \in \mathbb{N}$ exists a $k \in \mathbb{N}$ such that $tw(F_i) \geq m$ for all $i \geq k$.

Proof. We show these claims consecutively.

(1) It is straightforward to check that I_0^{v} is a core. To show that I_n^{v} is a core for every $n > 0$, pick an arbitrary retraction σ of I_n^{v}. Toward showing that σ is the identity, first note that it must be column-preserving (i.e., satisfy $\sigma(X_\ell^i) = X_{\ell'}^i$), since for any two $X_k^i, X_\ell^j \in \Delta_n^{\mathrm{v}}$ hold:
   - they are connected by a $\mathtt{v}$-path exactly if $i = j$,
   - if there is an $\mathtt{h}$ connection from the former to the latter, then $i + 1 = j$,

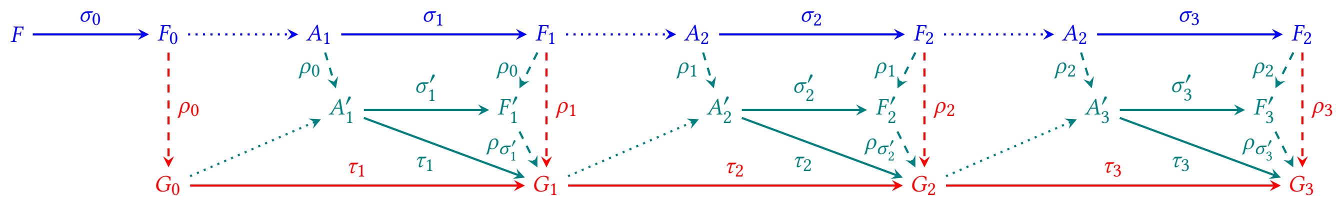


Figure 6: Depiction of the inductive definition of the robust associated sequence (Definition 12). Also useful to follow proof of Proposition 11.

- if $i+1=j$, then there are k' and ℓ' satisfying $\mathtt{h}(X^i_{k'}, X^j_{\ell'}) \in I^{\mathrm{v}}_n$.

  Yet then, for every $X^i_k \in \Delta^{\mathrm{v}}_n$, the corresponding column (the substructure of I^{v}_n induced by all X^j_ℓ with $j=i$) has an retraction obtained by restricting σ accordingly. Yet, each of these column-wise retractions must map the unique elements carrying $\mathtt{f}$ and $\mathtt{c}$ to themselves, which also forces all other elements (on the intermediate directed $\mathtt{v}$-path) to be identically mapped. Consequently, every row-wise retraction must be the identity function. Yet then, σ as a whole must be the identity as well.

(2) This claim is a consequence of Fact 2, since, for every n, the elements X^i_k with $\lfloor 2n/3 \rfloor + 1 \leq i \leq n+1$ and $n \leq k \leq \lceil 4n/3 \rceil$ witness that I^{v}_n contains a $(\lfloor n/3 \rfloor + 1) \times (\lfloor n/3 \rfloor + 1)$-grid.

(3) Without loss of generality, we assume the considered core chase employs the same naming scheme as I^{v}. Therefore, any intermediate atomset of the considered chase can be described by a subset of I^{v}. We first observe that $I^{\mathrm{v}}_0 = F^{\mathrm{v}}$, thus the claim is satisfied for $n=0$ once we set $f(0)=0$. We proceed iteratively for larger n. For any subsequent n, we can assume that F_{n-1} contains some I^{v}_m. Therefore, the only interesting case is if, upon producing F_n, nulls of F_{n-1} are removed through the non-trivial retraction σ_n. Among the nulls removed, let X^i_j be the one with maximal j and (among all these) the one with minimal i. By construction (observing I^{v}), removal of nulls will always simultaneously affect all nulls in a row, leaving behind only those of the form X^k_{2k}. Therefore, we obtain $i = \lfloor j/2 \rfloor + 1$. Also, by maximality of j and the fact that there are no row-decreasing $\mathtt{v}$-atoms, we know that $\sigma_n(X^i_j) = X^i_{j+1}$ (note that retractions must be column-preserving, as argued before). Then, for σ_n to be a retraction, we require $\mathtt{h}(X^{i-1}_{2\lfloor j/2 \rfloor}, X^i_{j+1}) \in I^{\mathrm{v}}_{n-1}$. Yet, as row-increasing $\mathtt{h}$-edges can only be the consequence of a (potentially iterated) prior application of R^{v}_7, the atom $\mathtt{f}(X^i_{j+1})$ must occur in some atomset preceding I^{v}_n. Yet, this can only be the consequence of the iterated application of R^{v}_7 propagating $\mathtt{f}$ from "right to left", starting from $\mathtt{f}(X^{j+2}_{j+1})$, $\mathtt{d}(X^{j+2}_{j+1})$. The latter atom must, in turn have been created through iterated application of R^{v}_5, propagating $\mathtt{d}$ "top-down" starting from $\mathtt{d}(X^{j+2}_{2j+4})$ which must have been created through application of R^{v}_4 to $\mathtt{c}(X^{j+2}_{2j+4})$. Yet, the only way to produce the latter is through R^{v}_1 following iterated application of R^{v}_2 preceded by an application of R^{v}_3 to $\mathtt{d}(X^{j+1}_j)$, $\mathtt{f}(X^{j+1}_j)$, and $\mathtt{v}(X^{j+1}_j, X^{j+1}_{j+1})$. This argument can then be repeated for columns further left, leading to the insight that removal of X^i_j requires that all facts from I^{v}_{j+1} must have previously existed in the derivation. Among those, the facts involving nulls X^k_ℓ with $\ell > j$, cannot have been removed by our maximality assumption. The remaining facts of I^{v}_{j+1} are indefinitely exempt from removal because the participating nulls are column-wise unique wrt carrying $\mathtt{c}$. We can therefore conclude that upon removal of X^i_j toward the creation of F_n, the latter must contain I^{v}_{j+1}.

    Finally, we observe that, as an indirect consequence of fairness, every X^i_j with $j \neq 2i$ will be removed in some derivation step, leading to the consequence that ever growing elements I^{v}_{j+1} will come into operation.

(4) This claim is a direct consequence of Item 2 and Item 3, given monotonicity of treewidth (Fact 1). □

PROOFS OF SECTION 8

Lemma 1. *Let $(G_i)_{i \in \mathfrak{I}}$ be the robust sequence associated with a derivation $\mathcal{D}$. If A is a finite subset of $\mathcal{D}^{\circledast}$, then there exists some $k \in \mathfrak{I}$ such that, for every $r \geq k \subseteq \mathfrak{I}$, $A \subseteq G_r$.*

Proof. We first prove *(i)* for every $i > 0 \in \mathfrak{I}$, $\bar{\tau}(G_{i-1}) \subseteq \bar{\tau}(G_i)$. Indeed, since τ_i is a homomorphism from G_{i-1} to G_i, then $\tau_i(G_{i-1}) \subseteq G_i$ and thus for any $j > i \in \mathfrak{I}$, $\bar{\tau}^j_i(\tau_i(G_{j-1})) \subseteq \bar{\tau}^j_i(G_i)$, meaning $\bar{\tau}(G_{i-1}) \subseteq \bar{\tau}(G_i)$.

Then we prove *(ii)* for every $\bar{\tau}(G_j)$, there exists some $k \geq j$ such that for every $r \geq k$, $\bar{\tau}(G_j) \subseteq G_r$. For every variable X in G_j, there is some $k_X \in \mathfrak{I}$ such that $\bar{\tau}(X) = \tau^{k_X}_j(X)$ is stable in all atomsets after G_{k_X} (Proposition 10). If we take $k = max_{X \in vars(G_j)} k_X$, then for every $r \geq k$, $\tau^r_j = \tau^r_k \circ \tau^k_j = \bar{\tau}$ is a homomorphism from G_j to G_r, and thus $\bar{\tau}(G_j) \subseteq G_r$.

Finally, since A is finite and the successive $\bar{\tau}(G_i)$ form a monotonic sequence (see *(i)*), there exists $j \in \mathfrak{I}$ such that $A \subseteq \bar{\tau}(G_j)$. Then *(ii)* there exists $k \geq j$ such that for every $r \geq k$, $\bar{\tau}(G_j) \subseteq G_r$ and thus $A \subseteq G_r$. □

PROOFS OF SECTION 9

Theorem 2. *CQ entailment is decidable for the class of KBs having a recurringly treewidth-bounded core chase sequence.*

Proof. Let $\mathfrak{C}$ be the class of KBs having a recurringly treewidth-bounded core chase sequence. The proof closely follows arguments from previous work [3, 7]. An algorithm deciding $\mathcal{K} \models Q$ for a given $\mathcal{K} \in \mathfrak{C}$ and conjunctive query Q can be devised from two semi-decision procedures (which, when executed in parallel give rise to a decision algorithm): one guaranteed to detect $\mathcal{K} \models Q$ after finite time and another detecting $\mathcal{K} \not\models Q$. For the former, we can evoke the fact that thanks to the completeness of first-order logic [12], the consequences of a first-order theory are recursively enumerable. So, the first part of the algorithm can just enumerate the consequences of $\mathcal{K}$ and terminate answering "yes" as soon as Q is found among the consequences. It remains to be shown that there is a semi-decision procedure detecting $\mathcal{K} \not\models Q$. By assumption, $\mathcal{K}$ has a finitely universal model I with $tw(I) \in \mathbb{N}$. From I being finitely universal for $\mathcal{K}$ and $\mathcal{K} \not\models Q$, we can conclude $I \not\models Q$. But then we obtain $I \models F \wedge (\bigwedge \Sigma) \wedge (\neg Q)$ (assuming that F and Q are represented as first-order sentences and Σ as a set of first-order sentences). This means, whenever $\mathcal{K} \not\models Q$, then there exists some k (namely $tw(I)$) such that the first-order sentence $F \wedge (\bigwedge \Sigma) \wedge (\neg Q)$ is satisfiable over the class of structures of treewidth k. Fortunately, as previously observed [3, 7], satisfiability of monadic second-order logic – and thus also of first-order logic – over classes of structures with a treewidth bounded by a given k is decidable. This allows to design a semi-decision procedure that increases k stepwise and in each step applies the decision procedure that checks if $F \wedge (\bigwedge \Sigma) \wedge (\neg Q)$ has a model of treewidth k. If so, the procedure terminates with the output "no", since we have shown that Q cannot be a consequence of $\mathcal{K}$. If not, we increment k and repeat. Clearly, thanks to the above assumption, this semi-decision procedure will output "no" and terminate exactly if $\mathcal{K} \not\models Q$. ◻

Proposition 13. *CQ entailment is decidable for any ruleset that is core-bts. Moreover, core-bts subsumes both finite expansion sets (fes) and bounded treewidth sets (bts), which are mutually incomparable.*

Proof. Decidability follows from Theorem 1. We sucessively prove the following items:

- fes and bts are incomparable,
- fes is subsumed by core-bts.
- bts is subsumed by core-bts.

For the first bullet point, note that the singleton ruleset $\{\mathtt{r}(X, Y) \rightarrow \exists Z.\mathtt{r}(Y, Z)\}$ is bts but not fes, whereas the singleton ruleset $\{\mathtt{r}(X, Y) \wedge \mathtt{r}(Y, Z) \rightarrow \exists V.\mathtt{r}(X, X) \wedge \mathtt{r}(X, Z) \wedge \mathtt{r}(Z, V)\}$ is fes but not bts.

For the second bullet point, recall that finite extension sets guarantee core-chase termination. Yet, for any finite sequence of finite structures one can find a uniform finite bound on the treewidth, it suffices to pick $\max_{i \in \mathfrak{I}} |\mathcal{T}(F_i)|$.

For the third bullet point, we observe that any treewidth-bounded restricted chase sequence $(F_i)_{i \in \mathfrak{I}}$ can be transformed into a core-chase sequence $(F'_i)_{i \in \mathfrak{I}}$ as follows: Let σ'_0 be an endomorphism turning F_0 into a core and let $F'_0 = \sigma'_0(F_0) = \sigma'_0(F)$. From this starting point, we can always use F_i, F'_i, and σ'_i where $\sigma'_i(F_i) = F'_i$ is a core, to define σ'_{i+1} and F'_{i+1} such that $\sigma'_{i+1}(F_{i+1}) = F'_{i+1}$ is a core as follows: assuming $F_{i+1} = \alpha(F_i, (R, \pi))$, we let $\widetilde{\sigma}_{i+1}$ be an endomorphism of $\alpha(\sigma'_i(F_i), (R, \sigma'_i \circ \pi))$ producing a core, which we choose as F'_{i+1}. Clearly then F'_{i+1} is also a core of $F_{i+1} = \alpha(F_i, (R, \pi))$ witnessed by the endomorphism $\sigma'_{i+1} = \widetilde{\sigma}_{i+1} \circ \sigma'_i$. Note that $(F'_i)_{i \in \mathfrak{I}}$ is indeed a core chase sequence, except for some elements being repeated, which can be removed. Now given that there exists a bound b greater than the treewidth of each element of $(F_i)_{i \in \mathfrak{I}}$, the same must hold for $(F'_i)_{i \in \mathfrak{I}}$, given that $F'_i \subseteq F_i$ for all $i \in \mathfrak{I}$. Thus $(F'_i)_{i \in \mathfrak{I}}$ (and any pruned subsequence of it) is uniformly (and hence also recurrently) treewidth-bounded. ◻