Bounded Treewidth and the Infinite Core Chase: Complications and Workarounds toward Decidable Querying

Jean-François Baget, Marie-Laure Mugnier, Sebastian Rudolph

To cite this version:

HAL Id: lirmm-04320944
https://hal-lirmm.ccsd.cnrs.fr/lirmm-04320944
Submitted on 4 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Bounded Treewidth and the Infinite Core Chase

Complications and Workarounds toward Decidable Querying

Jean-François Baget
Marie-Laure Mugnier
LIRMM, Inria, University of Montpellier, CNRS
Montpellier, France

Sebastian Rudolph
Computational Logic Group, TU Dresden
Dresden, Germany

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.

CCS CONCEPTS

• Theory of computation → Automated reasoning; • Computing methodologies → Knowledge representation and reasoning.

KEYWORDS

existential rules, tuple-generating dependencies, chase, treewidth, universal models
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.

2 PRELIMINARIES

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

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

A substitution of a set of variables \( Y \subseteq \Delta_V \) is a mapping \( \sigma \) from \( Y \) to \( \Delta_T \). For an atom \( at = p(t_1, \ldots, t_k) \) and a substitution \( \sigma \) of \( Y \), let \( \sigma(at) = p(\sigma(t_1), \ldots, \sigma(t_k)) \) where \( \sigma(t_i) = t_i \) whenever \( i \in Y \) and \( \sigma(t_i) = t_i \) otherwise. If \( A \) is an atomset, then \( \sigma(A) = \{ \sigma(at) \mid at \in A \} \). For two substitutions \( \sigma \) and \( \sigma' \) of variable sets \( Y \) and \( Y' \), respectively, we let \( \sigma' \circ \sigma \) denote the substitution of \( Y' \cup Y \) defined by \( Y 
\Rightarrow \sigma'(\sigma(Y)) \cup \sigma(Y) \). 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 \( \pi \) with \( \pi(A) \subseteq B \). Given such a homomorphism \( \pi \), we also say that \( \pi \) maps \( A \) to \( B \), or that \( A \) maps to \( B \) (via \( \pi \)). An isomorphism from \( A \) to \( B \) is a bijective homomorphism \( \pi \) such that \( \pi^{-1} \) is an isomorphism 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 \( \pi \) where the restriction of \( \pi \) to \( \text{terms}(\pi(A)) \) (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; if we want to emphasize this aspect, we also refer to them as instances. A (Boolean) conjunctive query (CQ) is a

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

\[\text{\textbullet terminating cc = finite universal model} \]

\[\text{\textbullet inflating elevator = finite universal model} \]

\[\text{\textbullet steepening staircase = tw-finite universal model} \]

\[\text{\textbullet tw-bounded cc} \]

\[\text{\textbullet tw-bounded rc} \]

\[\text{\textbullet steepening staircase} \]
We now introduce a convenient notion of derivation to define these variables.\footnote{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).} 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 \rightharpoonup 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 ∀\(\forall\)∃ Y \land B[\bar{X}, \bar{Y}] \rightarrow ∃\bar{Z} \land H[\bar{X}, \bar{Z}]\) where \(\bar{X}, \bar{Y}, \bar{Z}\) 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 \rightharpoonup H, a trigger for I is a pair \(tr = (B \rightharpoonup H, \pi)\) such that \(\pi\) maps B to I; tr is satisfied in I if \(\pi\) can be extended to a homomorphism from B ∪ H to I. Note that an instance I is a model of a rule R if it satisfies every trigger for I of the form (R, \(\pi\)). Given a rule R = B \rightharpoonup H, an instance I and a trigger \(tr = (R, \pi)\) for I, the application of tr on I produces the instance \(\alpha(I, tr) = I \cup \pi_{\text{safe}}(H)\), where \(\pi_{\text{safe}}\) maps every frontier-variable X of R to \(\pi(X)\) and any existential variable in \(\text{vars}(H)\) to a fresh variable (usually called a labeled null).\footnote{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).}

**Universal Models.** A knowledge base (KB) is a pair \(\mathcal{K} = (F, \Sigma)\), where F is a finite instance and \(\Sigma\) 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 \(\Sigma\). 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.

### 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, \(\mathcal{\Sigma}\) denotes either the set \(\mathbb{N}\) of natural integers (for infinite derivations) or the interval \{0, \ldots, k\} ⋐ \mathbb{N} (for finite ones).

**Definition 1 (Derivation).** A derivation from a KB \(\mathcal{K} = (F, \Sigma)\) is a (possibly infinite) sequence \(\mathcal{D} = ((tr, \sigma_i, F_i))_{i \in \mathcal{\Sigma}}\), where the \(tr_i\) are triggers (except \(tr_0 = \emptyset\), the \(\sigma_i\) are rejections called simplifications, and the \(F_i\) are finite instances such that \(F_0 = \sigma_0(F)\); and, for all \(i \in \mathcal{\Sigma}\setminus\{0\}\), \(F_i = \sigma_i(\alpha(F_{i-1}, tr_i))\), 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 \(\mathcal{D} = (F_i)_{i \in \mathcal{\Sigma}}\), leaving the \(tr_i\) and \(\sigma_i\) implicit. A derivation is called monotonic if \(F_{i-1} \subseteq F_i\) holds for all \(i \in \mathcal{\Sigma}\setminus\{0\}\). In a monotonic derivation, the restriction of \(\sigma_i\) 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 \(\mathcal{D} = (F_i)_{i \in \mathcal{\Sigma}}\), the "result" of \(\mathcal{D}\) is usually defined as the (infinite) union of all instances along \(\mathcal{D}\). We denote this union by \(\mathcal{D}^* = \bigcup_{i \in \mathcal{\Sigma}} F_i\) 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 \((R, \pi)\) for some \(F_i\) may not remain a trigger for some \(F_j\) with \(j > i\): this is because \(\pi(\text{body}(R))\) 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 \(\mathcal{D} = (F_i)_{i \in \mathcal{\Sigma}}\) be a derivation, and X be a variable occurring in some \(F_i\). For any \(j \in \mathcal{\Sigma}\) with \(j \geq i\), we define \(\sigma_i^j(X) = X\) and \(\sigma_i^j(F_j) = \sigma_j \circ \cdots \circ \sigma_{i+1}(X)\) when \(j > i\).

It is immediate that \(\sigma_i^j\) (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, \(\sigma_i^j\) is the identity for any \(j\). In the following, if \(tr = (R, \pi)\) is a trigger for A and \(\sigma\) is a substitution, we note \(\sigma(tr) = (R, \sigma \circ \pi)\) the trigger for \(\sigma(A)\).

**Definition 3 (Fair Derivation).** A derivation \(\mathcal{D} = (F_i)_{i \in \mathcal{\Sigma}}\) is fair if, for any \(i \in \mathcal{\Sigma}\) and trigger \(tr\) for \(F_i\), there is some \(j \in \mathcal{\Sigma}\) with \(j \geq i\), such that \(\sigma_i^j(tr)\) 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 \(((tr, \sigma_i, F_i))_{i \in \mathcal{\Sigma}}\) in which \(\sigma_i\) 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 \(\sigma_i\) 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 TREETWIDTH TO THE PICTURE

We now recall the popular notion of the treewidth of a 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 ⊆ 2^terms(A) and edges E ⊆ V × V, where:
- for each at ∈ A exists some v ∈ V with terms(at) ⊆ v;
- for each t ∈ terms(A), letting V_t = {v ∈ V | t ∈ 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, denoted by tw(A), is the minimal width among all its tree decompositions.

Fact 1. A ⊆ B implies tw(A) ≤ tw(B).

Definition 5. Given a natural number n, we say that an atomset A contains an n × n-grid, if terms(A) contains n^2 distinct terms, denoted if for i, j ∈ {1, ..., n}, such that for all k ∈ {1, ..., n − 1} and t ∈ {1, ..., n},
- there is some at ∈ A with {t_{x_i^k}, t_{y_j^k+1}} ⊆ terms(at), and
- there is some at' ∈ A with {t_{x_i^k}, t_{y_j^k+1}} ⊆ terms(at').

Fact 2. If A contains an n × n-grid then tw(A) ≥ 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 ∈ N and a restricted chase sequence (F_i)_{i∈Σ} such that tw(F_i) ≤ b for all i ∈ Σ.

Proposition 2. CQ entailment for bts is decidable.

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 → |I|. An instance I is then called \( \mu \)-finite, if \( \mu(I) \neq \infty \).

Moreover, we say that a sequence \((F_i)_{i∈Σ}\) of atomsets is uniformly \( \mu \)-bounded, if there exists some k ∈ N such that \( \mu(F_i) ≤ k \) for all i ∈ Σ. \((F_i)_{i∈Σ}\) will be called recursively \( \mu \)-bounded if there exists some k ∈ N such that, for any j ∈ Σ there exists some i ≥ j from Σ, for which \( \mu(F_i) ≤ k \) holds. It is easy to see that uniform \( \mu \)-boundedness implies recursive \( \mu \)-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 \( \mu \), if the existence of a (uniformly or recursively) \( \mu \)-bounded chase sequence for a KB is a necessary and/or sufficient condition for the existence of a \( \mu \)-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 if and only if 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 recurrently) 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 \( \mathcal{K}^b = (P^b, \Sigma^b) \) where \( \Sigma^b = (R^b_1, R^b_2, R^b_3, E^b_1, \ldots, E^b_4) \), as given in Figure 2.

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

Definition 8. We define \( I^b \) as the infinite instance using the terms \( \text{terms}(I^b) = \{X^b_i | (i, j) ∈ \mathbb{N}^2, i + 1 ≥ j \} \) and consisting of the atoms
\[
\begin{align*}
\mathcal{f}(X^b_0) & h(X^b_i, X^b_{i+1}) & v(X^b_i, X^b_{i+1}), \\
& & c(X^b_i) \quad \text{for } i ≥ j ≥ 1 \quad \text{h}(X^b_i), \text{j}(X^b_i) \quad \text{for } i ≤ j.
\end{align*}
\]

The instance \( I^b \) is depicted in Figure 2; the names \( X^b_i \) of the variables of \( I^b \) are in correspondence to their cartesian coordinates (i, j) in the picture. We now consider some particular subsets of \( \text{terms}(I^b) \). For any k ∈ \( \mathbb{N} \), let \( P_k = (X^b_{2k} \cup \Sigma^b) \subseteq \mathcal{K}^b \), \( \Sigma^b_k = (X^b_{2k+1}) \subseteq \mathcal{K}^b \), and \( S_k = \Sigma^b \cup \Sigma^b_{k+1} \cup \{X^b_{2k+1}\} \). Let \( P^b_k \) (resp. \( C^b_k \), \( S^b_k \)) denote the subset of \( I^b \) induced by \( P_k \) (resp. \( C_k \), \( S_k \)). Intuitively, \( P^b_k \) is the finite part until column \( k \), \( C^b_k \) is the k-th column of \( I^b \) (minus its top element) and \( S^b_k \) is a step - a rectangle containing the two columns \( C^b_k \) and \( C^b_{k+1} \).

We first point out that there is a sequence of rule applications from any \( C^b_k \) producing \( S^b_k \). Indeed, we can apply \( R^b_1 \) on the top of \( C^b_k \) to "complete" \( C^b_k \) and obtain the two highest variables of \( C^b_{k+1} \). Then we apply \( R^b_2 \) k times (from top to bottom) to obtain the other variables of \( C^b_{k+1} \). Once \( X^b_{2k+1} \) has been generated, we can apply \( R^b_3 \) to generate the h-loop on \( X^b_{2k+1} \), then \( k \) successive applications of \( R^b_4 \) propagate the loops on \( C^b_{k+1} \), from bottom to top. There is thus a monotonic infinite derivation \( D^b = (F_i)_{i∈\mathbb{N}} \) from \( \mathcal{K}^b \), the natural aggregation of which yields \( I^b \). We successively apply \( R^b_2 \), \( R^b_3 \), and \( R^b_4 \) on \( I^b \) to obtain \( S^b_0 = I^b \). Since \( C^b_k \subseteq S^b_k \), we apply the rules on \( C^b_k \) as seen previously to obtain \( R^b_1 \) and thus \( S^b_k \), and so on. The infinite union of all atomsets along this derivation is \( D^b \).

Proposition 3. \( I^b \) is a result of the restricted chase on \( \mathcal{K}^b \).
**Sketch of proof.** The derivation $D_k$ given above is a restricted chase sequence. Clearly, no $c(F, R, π)$ in $D_k$ retracts to $F_f$ so it remains to check that $D_k$ is fair. Indeed, if $(R, π)$ is a trigger for some $F_i$, then it is a trigger w.r.t some $R_h^c$ that is necessarily satisfied (at most in $h_{k+1}^c$). Thus $(R, π)$ is satisfied in some $F_j \supseteq h_j^c$. \(\Box\)

As a result of the restricted chase, $\hat{h}$ is a universal model of $K_h$. 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 $D_k$ to build a derivation $D_k$ that relies upon those retractions. $D_k$ starts out like $D_k$, but as soon as $S_k^h$ is obtained, we retract it to its core $C_{k}^h$. Then, following $D_k$’s course, $D_k$ proceeds to build $S_k^h$ that retracts to its core $C_{k}^h$. As for $D_k$, we note that $D_k$ is fair. Moreover, each retraction to a core is done a finite number of rule applications after the previous one: $D_k$ is thus a core chase sequence. Finally, we point out that every atomset in $D_k$ is a subset of some $S_k^h$, and has thus treewidth at most 2.

**Proposition 4.** There is a core chase sequence for $K_h$ that is uniformly treewidth-bounded by 2.

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

**Proposition 5.** No universal model of $K_h$ has finite treewidth.

**Sketch of proof.** Any universal model $U$ of $K_h$ is homomorphically equivalent to $\hat{h}$. This allows to show that, for any $n \geq 1$, $U$ contains an $n \times n$-grid, hence $tw(U) \geq n$. \(\Box\)

### 7 THE INFLATING ELEVATOR

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

**Definition 9 (The Inflating Elevator $\mathcal{K}^*$).** We let $\mathcal{K}^* = (F^*, \Sigma^*)$ where $\Sigma^* = \{R_1^c, R_2^c, R_3^c, R_4^c, R_5^c, R_6^c\}$ and $F^*$ 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($\hat{h}$) = $\{X_{ij} | (i, j) \in \mathbb{N}, 1 \leq i, j \leq 2l\}$. Then $\mathcal{I}^*$ consists of the following atoms for all $i, j$ where all mentioned nulls are in terms($\mathcal{I}^*$):

\[
d(X_{ij}) \quad h(X_{ij}, X_{ij+1}) \quad v(X_{ij}, X_{ij+1}) \\
t(X_{ij}) \quad h(X_{ij}, X_{ij+1}) \quad v(X_{ij}, X_{ij+1}) \quad \text{for } i \leq j \\
c(X_{ij}) \quad h(X_{ij}, X_{ij+1}) \quad v(X_{ij}, X_{ij+1})
\]

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

**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 $i \geq 1$, $X_{2l-1}^i$ and $X_{2l}^i$ are introduced as instances of $Y'$ and $Y''$ through an application of Rule $R_1^c$ with $X \mapsto X_{2l-1}^i$ and $Y \mapsto X_{2l}^i$.
- for every $i \geq 1$, $X_{ij}^i$ is introduced as instance of $Y'$ through an application of Rule $R_2^c$ with $X \mapsto X_{ij}^i$, and $Y \mapsto X_{ij}^i$.
- every remaining $X_{ij}^i \in$ terms($\mathcal{I}^*$) with $i \geq 1$ is introduced as instance of $Y'$ through an application of Rule $R_3^c$ with $X \mapsto X_{ij}^i$, $Y \mapsto X_{ij}^i-1$, and $Y \mapsto X_{ij}^i-1$.

Fairness follows from the fact that $\mathcal{I}^*$ satisfies all its triggers, as can be checked easily. \(\Box\)

As a result of the restricted chase, $\mathcal{I}^*$ is a universal model of $\mathcal{K}^*$. As it turns out, it even contains another universal model of finite...
We next describe a sequence \( I^v_0, I^v_1, \ldots \) of subsets of \( I^v \) that exhibit increasing treewidths and will later be shown to occur as substructures in any core chase sequence of \( \mathcal{K}^v \). Figure 4 depicts the first elements of that sequence.

**Definition 12.** We define the sequence \( (I^v_n)_{n \in \mathbb{N}} \) of atomsets by letting
\[
I^v_0 = F^v \quad \text{and, for any } n > 0, \quad I^v_n = \text{the substructure of } I^v \text{ induced by terms } X_i^v \quad \text{for } \quad V = \{ X_i^v \mid i \leq \frac{n}{3} \} \cup \{ X_{i'}^v \mid i \leq n+1 \text{ and } j \geq n \} \text{ removing all atoms } v(X_i^v) \text{ and } f(X_i^v) \text{ with } j > n \text{ as well as all atoms } h(X_{i'}^v, X_{i'+1}^v) \text{ with } k > j \text{ and } k > n.
\]

**Proposition 8.** The following hold:
1. Every \( I^v_n \) is a core.
2. \( I^v_n \) has a treewidth of at least \([n/3] + 1\).
3. For every core chase sequence \( (F_i)_{i \in \mathbb{N}} \) for \( \mathcal{K}^v \), there is an unbounded monotonic function \( f : \mathbb{N} \to \mathbb{N} \) such that, for every \( n \in \mathbb{N}, \) \( \Gamma_n^{(n)} \) is isomorphic to a subset of \( F_n \).
4. For every core chase sequence \( (F_i)_{i \in \mathbb{N}} \) for \( \mathcal{K}^v \) and any \( m, n \in \mathbb{N} \) there is 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 \( \mathcal{K}^v \) is recurringly or uniformly treewidth-bounded.
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 recurring-treewidth-bounded core chase sequence (forthcoming Theorem 2). 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 recurring-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 1). In the following, we detail the employed notions and arguments laid out above.

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

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

Proof. ($\Rightarrow$) Let $\sigma$ 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 $K$ by some $\sigma'$, thus $\sigma' \circ \sigma$ maps $Q$ to $I$. ($\Leftarrow$) Since $K \models Q$ and $M$ is a model of $K$, $M \models Q$. $\square$

Defining Robust Aggregations. For non-monotonic derivations $D = (F_i)_{i \in \mathbb{Z}}$, it may happen that an atom $a_i$ at some later step $j$. Yet, at will still belong to the natural aggregation $D^+$. Intuitively, the natural aggregation generates 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 $\sigma_j$ 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 $X^I_0$ to $X^I_1$, then $X^I_1$ to $X^I_2$, etc., and there is no way we can define the ultimate image of $X^I_0$ 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 $X$ with $N$, and use the total ordering $<_X$ on $X$ defined by $X <_X Y$ iff rank($X$) < rank($Y$).

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

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

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

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

Note that $(G_i)$ is not a derivation, since the $\tau_i$ 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 $(G_i)_{i \in \mathbb{Z}}$ be an associated robust sequence. For $i, j \in \mathbb{Z}$ with $i < j$, let $\tau_{j-i} = \tau_j \circ \cdots \circ \tau_{i+1}$ denote the composition of all $\tau_i$ between $G_i$ and $G_j$. Then, for any $X \in \text{vars}(G_i)$, there is $j \in \mathbb{Z}$ with $j > i$ such that $\tau_{j-i}(X) = Y \in \text{terms}(G_j)$ and for all $k \in \mathbb{Z}$ with $k > j$, $\tau_{j-k}(Y) = Y$ (i.e., $Y$ is stable from $G_j$ on). We let $\tau(X) = Y$.

Proof. Let $X \in \text{vars}(G_i)$, then $\tau_{i+1}(X) = \tau_{i+1}'(X) \leq_X X$. Consider some arbitrary $j \in \mathbb{Z}$ with $j > i$. Among the homomorphisms $\tau_{i+1}$ that $\tau_{j-i}$ is composed of, there can be at most rank$_k(X)$ many of them that are effectively decreasing (causing $\tau_{j-i}^{-1}(X) <_X \tau_{j-i}(X)$).

We now use the $\tau(G_i)$ to define the robust aggregation. Note that, contrary to $(F_i)$ or $(G_i)$, the sequence $(\tau(G_i))$ is monotononic.

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

Semantic Properties of Robust Aggregations. The steepening staircase shows that the robust aggregation of a derivation is not always universal. Indeed, consider the KB $\mathbb{K}^0$ (from Definition 7) and let $<_X$ be an order on the variables with $j < k \Rightarrow X_j <_X X_k$. The core chase on $\mathbb{K}^0$ begins building the first step $S_0^0$ of $\mathbb{K}^0$, and all simplifications are the identity until done. Now, the first proper retraction maps $S_0^0$ to $S_1^0$ and $X_0^0$ to $X_1^0$, so the robust renaming generates $G_{i=1}$, which is isomorphic to the column $C_{i=1}$, but its variables are
named (from bottom to top) $X^\circ_0$ and $X^\circ_1$. Likewise, from successive proper retraction steps, we obtain $G_i$, isomorphic to $G_j$ but with variables named $X^\circ_0, X^\circ_1, X^\circ_2, \ldots, X^\circ_{j+1}$. Note that $\tau(G_i)$ is $G_i$ holds: every variable is stable since subsequent re-mappings would have to be within the same row, yet all variables therein are $<_X$-greater. Then, the robust aggregation $D^R$ is isomorphic to the infinite column $\bar{F}$, with variables named $X^\circ_0, X^\circ_1, X^\circ_2, \ldots, X^\circ_{j+1}, \ldots$, which is not universal, but is a finitely universal model, as stated below.

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

To prove this proposition, we rely on the next lemma, which states that any finite part of $D^R$ is “stably present” from a certain element on in the robust sequence associated with $D$.

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

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

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

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

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 2.

**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 $D$ that is recurringly treewidth-bounded by some integer $k$, the following hold:

1. $D$’s natural aggregation $D^N$ has treewidth $\leq k$, if $D$ is monotonic.
2. $D$’s robust aggregation $D^R$ has treewidth $\leq k$.

**Proof.** Let $\mathcal{K}$ be a finite subset of $D^R$ (for proof of (1)) or $D^R$ (for proof of (2)). There is some $p \in \mathcal{I}$ such that, for every $\tau(\bar{F}_s)$, we can exhibit some $I_r$ isomorphic to $F_r$ with $I \subseteq I_r$. To prove (1), $D$ being monotonic, we can define $I_r = F_r$. To prove (2), we rely upon Lemma 1 and define $I_r = G_r$. Since $D$ is recurringly treewidth-bounded, there is some $\tau(\bar{F}_s) = k$ such that $tw(\bar{F}_s) \leq k$. Thus $tw(I_s) = tw(F_s) \leq k$, and we conclude, thanks to compactness of treewidth, that $D^R$ 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. We obtain this result via a mild generalization of respective statements for universal models [3, 7, 11].

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

**Sketch of proof.** $\mathcal{K} \models Q$ can be detected in finite time due to the completeness of first-order logic. $\mathcal{K} \models \neg Q$ 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 1:

**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 2 – warrants decidable CQ entailment and properly subsumes other classes with that property.4

**Definition 17.** A ruleset $\Sigma$ is called core-bts, if for every finite atomset $F$, there exists a core chase sequence for the KB $(F, \Sigma)$, 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.

---

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

2. 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.
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 \( K \) allowing 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 \( K' \) 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 recurring treewidth-bounded, the aggregated atomset will be of finite treewidth. Together, these findings establish decidability of CQ entailment for all knowledge bases with a recurring treewidth-bounded core chase. Abstracting from concrete databases, this yields a novel, very general abstract class of recurring treewidth-bounded rule sets, 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.

ACKNOWLEDGMENTS

The authors thank the anonymous reviewers for their helpful comments. Jean-François Baget and Marie-Laure Mugnier were partially supported by the ANR project CQFD (ANR-18-CE23-0003). Sebastian Rudolph has received funding from the European Research Council (Consolidator Grant Agreement no. 771779, DeciGUT).

REFERENCES


A PROOFS OF SECTION 3

The following appendices are devoted to the complete proofs that are missing or only sketched in the paper.

FACT 3. If \( t \) is a trigger for \( F \), \( \mu \) maps \( F \) to \( I \) and \( I \) satisfies \( \mu(\tau) \), then there is \( \mu' \) (compatible with \( \mu \)) that maps \( \alpha(F, \tau) \) to \( I \).

LEMMA 2. For every fair derivation \( D = (F_i)_{i \in \mathcal{I}} \), there exists a fair monotonic derivation \( D_{\text{mon}} = (G_i)_{i \in \mathcal{I}} \) such that for every \( i \in \mathcal{I} \), there is a retraction from \( G_i \) to \( F_i \).
monotonic, but it remains to check that it is fair. Given an trigger
for some \( G_i, \tilde{\sigma}^i(tr) \) is a trigger for \( F_i \) and thus (fairness of \( D \))
there exists \( j \in \mathcal{A} \) 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 \( G_j \).
\( \square \)

**Proposition 1 (Extended version)** Let \( D \) be a derivation from \( \mathcal{K} \).
Then:
1. \( D^* \) is universal for \( \mathcal{K} \).
2. If \( D \) is finite, \( D^* \) is universal for \( \mathcal{K} \).
3. If \( D \) is monotonic and fair, \( D^* \) is a model of \( \mathcal{K} \).
4. If \( D \) is finite and fair, \( D^* \) is a model of \( \mathcal{K} \).
5. If \( D \) is fair and \( Q \) is a CQ, \( \mathcal{K} \models Q \) iff \( D^* \models Q \).

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

(2) Hence, \( M \) is a model of every \( F_i \) in \( D \): each instance \( F_i \) is universal and, if \( D \) is finite, then the final result \( D^* = F_k \) is universal. \( \square \)

(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 \( \mu_i \) between successive atomsets implies global compatibility of all \( \mu_i \). We conclude by pointing out that \( \bigcup_{i \in \mathcal{A}} \mu_i \) is a homomorphism from \( D^* \) to \( M \), and thus that \( D^* \) is universal. \( \square \)

(4) The final result \( D^* = F_k \) of a finite derivation is a model of \( F \)
(\( \tilde{\sigma}^0_0 \circ \sigma_0 \) is a homomorphism from \( F \) to any \( F_i \) in the derivation)
and, by Definition 3, for any trigger \( tr \) for \( F_k \), there is some \( j \geq k \)
(thus \( j = k \)) such that \( \tilde{\sigma}^k_k(tr) = tr \) is a satisfied trigger for \( F_k \). \( \square \)

(3) In the case of an infinite fair derivation, we first point out that
\( D^* \) contains \( F_0 = \sigma_0(F) \), so it is a model of \( F \). Then consider any

**Table 1: Steepling staircase: from column to step.**

<table>
<thead>
<tr>
<th>Rule</th>
<th>Homomorphism</th>
<th>Atoms produced</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R^i_1 )</td>
<td>( X \rightarrow X^i_k )</td>
<td>( v(X^i_k, X^{i+1}_k), h(X^i_k, X^{i+1}_k), c(X^i_k, X^{i+1}_k) )</td>
</tr>
<tr>
<td>( R^i_2 )</td>
<td>( X \rightarrow X^i_k, X' \rightarrow X^{i+1}_k, Y \rightarrow X^{i+1}_k )</td>
<td>( h(X^i_k, X^{i+1}_k), v(X^{i+1}_k, X^{i+1}_k), c(X^{i+1}_k) )</td>
</tr>
<tr>
<td>( \cdots )</td>
<td>( \cdots )</td>
<td>( \cdots )</td>
</tr>
<tr>
<td>( R^i_3 )</td>
<td>( X \rightarrow X^i_k, X' \rightarrow X^{i+1}_k, Y \rightarrow X^{i+1}_k )</td>
<td>( h(X^i_k, X^{i+1}_k), v(X^{i+1}_k, X^{i+1}_k), c(X^{i+1}_k) )</td>
</tr>
<tr>
<td>( R^i_4 )</td>
<td>( X \rightarrow X^i_k, X' \rightarrow X^{i+1}_k )</td>
<td>( h(X^i_k, X^{i+1}_k), v(X^{i+1}_k, X^{i+1}_k), c(X^{i+1}_k) )</td>
</tr>
</tbody>
</table>

**B PROOFS OF SECTION 6**

We first prove the following claim (see the explanations before Proposition 5).

**Claim.** There is a sequence of rule applications from any column \( C^i_k \) producing step \( S^i_k \).

**Proof.** Let us consider \( C^i_k \) with variables named \( (X^i_k, X^{i+1}_k, \ldots, X^{i+k}_k) \),
from bottom to top. Let us apply rules as shown in Table 1. The obtained result is indeed \( S^i_k \).

**Proposition 5.** No universal model of \( \mathcal{R}^h \) has finite treewidth.

**Proof.** We call \( v \)-path (resp. \( h \)-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 \( v(X_i, X_{i+1}) \) (resp. \( h(X_i, X_{i+1}) \)). 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 \( \mathcal{R}^h \). We first point out that \( \mathcal{R}^h \) and \( U \) being both universal models, they homomorphically map to each other. We let \( h_1 \) denote the homomorphism from \( \mathcal{R}^h \) to \( U \) and let \( h_2 \) denote the homomorphism from \( U \) to \( \mathcal{R}^h \). Then \( h = h_2 \circ h_1 \) is an endomorphism on \( \mathcal{R}^h \), the properties of which we will now inspect further. We make use of the following notation: for \( h(X^i_j) = X^j_k \), we denote \( k \) by \( h_k(i, j) \) and \( t \) by \( h_t(i, j) \), that is, we
let \( h(X^i_j) = X^i_j \cap I^h \). We make the following observations (which hold for all endomorphisms on \( I^h \)):

1. \( h_y(i, 0) = 0 \) since \( f \) holds precisely for all nulls \( X^i_j \).

2. \( h_y(i, j) = j \), inductively with (1) as base case and the observation that \( h \) must preserve the length of incoming \( v \)-paths rooted in some \( f \).

3. \( h_x(i, j) = h_x(i, j+1) \), since this is the only way for \( h \) to preserve the \( v \)-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 \( X^i_j \) 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 \( h \)-atoms.

7. \( h_y(i+1, j) = h_y(i, j) \) since this is the only way for \( h \) to preserve the \( h \)-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 \( X^i_j \) with \( i > k \) is injective. Follows from (8), for the same \( k \), and (2).

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

\[
I^h_{\ell} \subseteq I^h \implies h_1(I^h_{\ell}) \subseteq h_1(I^h) \subseteq U \quad \text{Fact 1}
\]

\[
\therefore \text{tw}(h_1(I^h_{\ell})) \leq \text{tw}(U)
\]

\[
\therefore \text{tw}(I^h_{\ell}) \leq \text{tw}(U) \implies \text{tw}(U) \notin \mathbb{N}.
\]

\[\square\]

C PROOFS OF SECTION 7

Proposition 8. The following hold:

1. Every \( I^h_{\ell} \) is a core.

2. \( I^h_{\ell} \) has a treewidth of at least \( \lceil n/3 \rceil + 1 \).

3. For every core chase sequence \( (F_t)_{t \in \mathbb{N}} \) for \( \mathcal{K}^c \), there is an unbounded monotonic function \( f : \mathbb{N} \to \mathbb{N} \) such that, for every \( n \in \mathbb{N} \), \( I^h_{f(n)} \) is isomorphic to a subset of \( F_n \).

4. For every core chase sequence \( (F_t)_{t \in \mathbb{N}} \) for \( \mathcal{K}^c \) and any \( m \in \mathbb{N} \) exists a \( k \in \mathbb{N} \) such that \( \text{tw}(F_t) \geq m \) for all \( i \geq k \).

Proof. We show these claims consecutively.

1. It is straightforward to check that \( I^h_{\ell} \) is a core. To show that \( I^h_{\ell} \) is a core for every \( n > 0 \), pick an arbitrary retraction \( \sigma \) of \( I^h_{\ell} \).

2. Toward showing that \( \sigma \) is the identity, first note that it must be column-preserving (i.e., satisfy \( \sigma(X^i_j) = X^i_j \)), since for any two \( X^i_j, X^i_j \in \Delta^n \) hold:

- if they are connected by a \( v \)-path exactly if \( i = j \),
- if there is an \( h \) connection from the former to the latter, then \( i + 1 = j \).

3. This is the case of the iterated application of \( h \) from “right to left”, starting from \( f(0) = 0 \). We proceed iteratively for larger \( n \). For any subsequent \( n \), we can assume that \( F_{n-1} \) contains some \( I^h_{\ell} \). Therefore, the only interesting case is if, upon producing \( F_n \), nulls of \( F_{n-1} \) are removed through the non-trivial retraction \( \sigma_n \). 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 \( \Gamma \)), removal of nulls will always simultaneously affect all nulls in a row, leaving behind only those of the form \( X^i_j \). Therefore, we obtain \( i = \lfloor j/2 \rfloor + 1 \). Also, by maximality of \( j \) and the fact that there are no row-decreasing \( v \)-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 \( \sigma_n \) to be a retraction, we require \( h(X^i_{j+1}) \in I^h_{\ell} \). Yet, as row-increasing \( h \)-edges can only be the consequence of a potentially iterated) prior application of \( R^c \), the atom \( h(X^i_{j+1}) \) must occur in some atomset preceding \( I^h_{\ell} \). Yet, this can only be the consequence of the iterated application of \( R^c \) propagating \( f \) from “right to left”, starting from \( f(X^i_{j+1}) \). The latter atom must, in turn have been created through iterated application of \( R^c ) \), propagating \( f \) “top-down” starting from \( f(X^i_{j+1}) \). 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 \( I^h_{\ell} \) must have previously existed in the derivation. Among those, the facts involving nulls \( X^i_j \) with \( \ell > j \), cannot have been removed by our maximality assumption. The remaining facts of \( I^h_{\ell} \) are indefinitely exempt from removal because the participating nulls are column-wise unique wrt carrying \( c \). We can therefore conclude that upon removal of \( X^i_j \) toward the creation of \( F_n \), the latter must contain \( I^h_{\ell} \).

4. Finally, we observe that, as an indirect consequence of fairness, every \( X^i_j \) with \( j \geq 2i \) is removed in some derivation step, leading to the consequence that ever growing elements \( I^h_{\ell} \) come into operation.

This claim is a direct consequence of Item 2 and Item 3, given monotonicity of treewidth (Fact 1).
D PROOFS OF SECTION 8

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

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

Then we prove (ii) for every $\tau(G_j)$, there exists some $k \geq j$ such that for every $r \geq k$, $\tau(G_j) \subseteq G_r$. For every variable $X$ in $G_j$, there is some $k_X \in \mathbb{N}$ such that $\tau_i^X(X)$ is stable in all atomsets after $G_{k_X}$ (Proposition 10). If we take $k = \max_{X \in \mathcal{V}(G_j)} k_X$, then for every $r \geq k$, $\tau_j^i = \tau_r^k \circ \tau_j^i$ is a homomorphism from $G_j$ to $G_r$, and thus $\tau(G_j) \subseteq G_r$.

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

E PROOFS OF SECTION 9

Theorem 1. CQ entailment is decidable for the class of KBs having a recursively treewidth-bounded core chase sequence.

Proof. Let $\mathcal{K}$ be the class of KBs having a recursively 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 $\mathcal{K} \in \mathcal{C}$ 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 $\text{rw}(I) \in \mathbb{N}$. From $I$ being finitely universal for $\mathcal{K}$ and $\mathcal{K} \models Q$, we can conclude $I \not\models Q$. But then we obtain $I \models F \land (\Sigma) \land (\neg Q)$ (assuming that $F$ and $Q$ are represented as first-order sentences and $\Sigma$ as a set of first-order sentences). This means, whenever $\mathcal{K} \not\models Q$, then there exists some $k$ (namely $\text{rw}(I)$) such that the first-order sentence $F \land (\Sigma) \land (\neg Q)$ 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 $F \land (\Sigma) \land (\neg Q)$ 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 $\mathcal{K} \not\models Q$. \hfill \Box

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 2. We successively 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 $(\tau(X, Y) \rightarrow \exists Z. \tau(Y, Z))$ is bts but not fes, whereas the singleton ruleset $(\tau(X, Y) \land \tau(Y, Z) \rightarrow \exists V. \tau(X, Y) \land \tau(X, Z) \land \tau(Z, V))$ 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 $\max_{i \in \mathbb{N}} |T(F_i)|$.

For the third bullet point, we observe that any treewidth-bounded restricted chase sequence $(F_i)_{i \in \mathbb{N}}$ can be transformed into a core-chase sequence $(F'_i)_{i \in \mathbb{N}}$ as follows: Let $\sigma'_0$ 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 $\sigma'_i$ where $\sigma'_i(F_i) = F'_i$ is a core, to define $\sigma'_i : F_i \rightarrow F_{i+1}$ such that $\sigma'_i(F_{i+1}) = F'_{i+1}$ is a core as follows: assuming $F_{i+1} = \alpha(F_i, (R, \pi))$, we let $\sigma'_i$ be an endomorphism of $\alpha(\sigma'_i(F_i), (R, \sigma'_i))$ 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 $\sigma'_i : F_{i+1} \rightarrow F_{i+1}$. Note that $(F'_i)_{i \in \mathbb{N}}$ 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 $(F_i)_{i \in \mathbb{N}}$, the same must hold for $(F'_i)_{i \in \mathbb{N}}$, given that $F'_i \subseteq F_i$ for all $i \in \mathbb{N}$. Thus $(F'_i)_{i \in \mathbb{N}}$ (and any pruned subsequence of it) is uniformly (and hence also recurrently) treewidth-bounded. \hfill \Box