Bounded Treewidth and the Infinite Core Chase: Complications and Workarounds toward Decidable Querying (extended abstract of paper published at PODS 2023)

Jean-François Baget[1] , **Marie-Laure Mugnier**[1] , **Sebastian Rudolph**[2]
[1]LIRMM, Inria, University of Montpellier, CNRS [2]Computational Logic Group, TU Dresden

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, ... If a fixpoint is reached after a finite number of steps (one speaks of *chase termination*), the obtained structure constitutes a finite model of the given KB, which is also *universal* – it maps homomorphically into every model – allowing one to consider just this single model to answer a wide range of queries, including conjunctive queries (CQs).

Among the different chase variants, the *core chase* is the only one that terminates exactly when the KB has a finite universal model, and produces the unique smallest such model. Thus, the core chase is the best choice for a decision procedure that aims at chase termination, motivating 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. 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 do not always exist. In such cases, no chase arrives at a fixpoint. As a remedy, one may define the "result" of the chase as the infinite union over all the intermediate structures of the chase 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. However, the core chase is non-monotonic, in which case one cannot even be certain to obtain a model.

CQ entailment may be decidable even if there is no finite universal model: in particular, this is the case whenever an infinite universal model exists that is still reasonably "structurally well-behaved" by virtue of having a *bounded treewidth*. 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. 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 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" (called *robust aggregation*) for the core chase that produces a model, but not a universal one. Still, thanks to being *finitely universal* (each of its finite substructures maps into every model) the obtained model is a perfect proxy for detecting CQ entailment, thus sufficient for our purpose of showing decidability. Also, we show that the inability to construct a treewidth-bounded universal model out of a treewidth-bounded core chase sequence is unavoidable, by exhibiting the *steepening staircase example*: a uniformly treewidth-bounded core-chase sequence for a KB whose every universal model has infinite treewidth (Fig. 2). Conversely, the *inflating elevator example* (Fig. 3) presents a KB with a universal model of finite treewidth, yet each of its core-chase sequences consists of structures of ever-growing treewidth (cf. Fig. 4), 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.

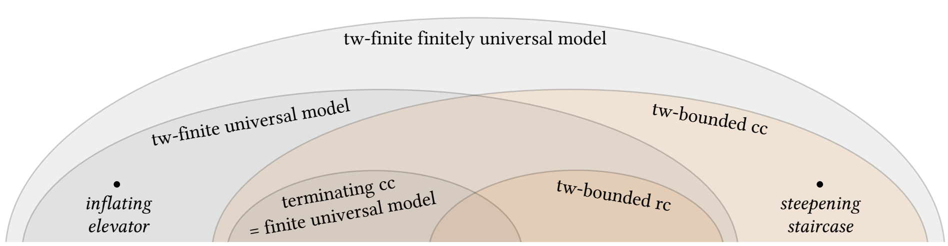


Figure 1: Set diagram displaying (non-)inclusion of decidable classes of existential rule sets from the paper. We abbreviate treewidth by *tw*, and restricted and core chase by *rc* and *cc*, respectively. The rulesets entitled "steepening staircase" and "inflating elevator" demonstrate that existence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

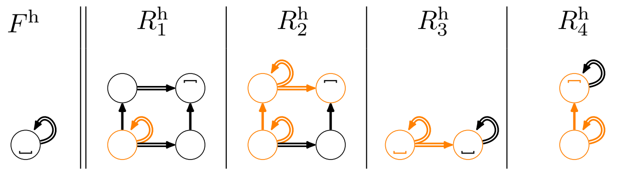
(R_1^{h}) $\mathtt{h}(X,X) \to \exists X'YY'.\mathtt{h}(X,Y) \wedge \mathtt{v}(X,X') \wedge \mathtt{h}(X',Y') \wedge \mathtt{v}(Y,Y') \wedge \mathtt{c}(Y')$

(R_2^{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^{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^{h}) $\mathtt{h}(X,X) \wedge \mathtt{v}(X,X') \wedge \mathtt{c}(X') \to \mathtt{h}(X',X')$

(F^{h}) $\{\mathtt{f}(X_0^0), \mathtt{h}(X_0^0, X_0^0)\}$



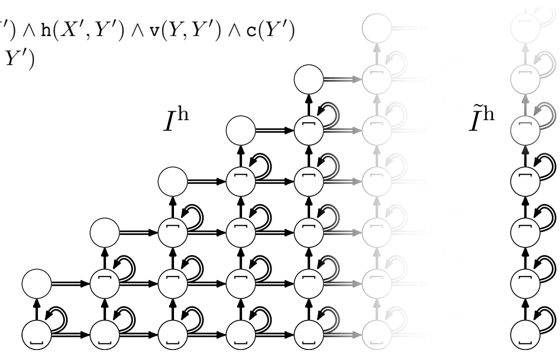


Figure 2: The *steepening staircase* example $\mathcal{K}^{\mathrm{h}} = (\Sigma^{\mathrm{h}}, F^{\mathrm{h}})$. Left: rules of Σ^{h}, fact set F^{h}, and a graphical representation thereof. Orange (grey) elements represent the rule body, black elements the rule head. Visualization of atoms: ⟹ denotes $\mathtt{h}$ ("horizontal") and ⟶ denotes $\mathtt{v}$ ("vertical"); we write ⎺ for $\mathtt{c}$ ("ceiling") and ⎵ for $\mathtt{f}$ ("floor"). Right: an infinite universal model I^{h} of $\mathcal{K}^{\mathrm{h}}$ (having infinite treewidth). Even more right: infinite model $\tilde{I}^{\mathrm{h}}$ of $\mathcal{K}^{\mathrm{h}}$, which is not universal (note that it does not have a homomorphism to I^{h}) but finitely universal, i.e., it satisfies exactly those CQs entailed by $\mathcal{K}^{\mathrm{h}}$. $\tilde{I}^{\mathrm{h}}$ can be obtained via robust aggregation (details not in this abstract).

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

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

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

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

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

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

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

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

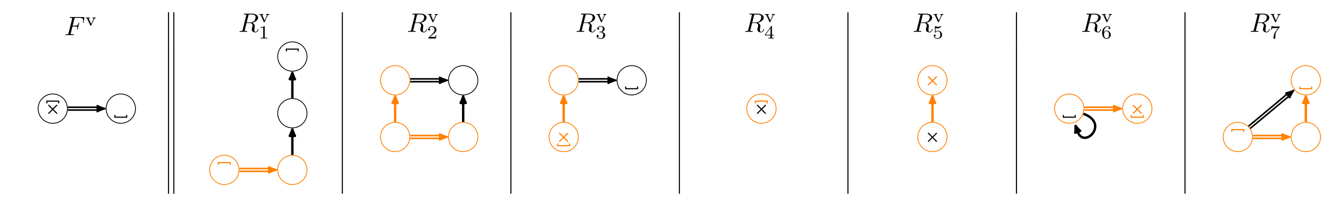


Figure 3: The *inflating elevator* example $\mathcal{K}^{\mathrm{v}} = (\Sigma^{\mathrm{v}}, F^{\mathrm{v}})$. Fact set F^{v} and rules of Σ^{v} (top) and their graphical depictions (bottom). Orange (grey) elements represent the rule body and black elements the rule head. Atoms are encoded as follows: ⟹ denotes $\mathtt{h}$ ("horizontal") and ⟶ denotes $\mathtt{v}$ ("vertical"); we write ⎺ for $\mathtt{c}$ ("ceiling"), ⎵ for $\mathtt{f}$ ("floor"), and × for $\mathtt{d}$ ("done").

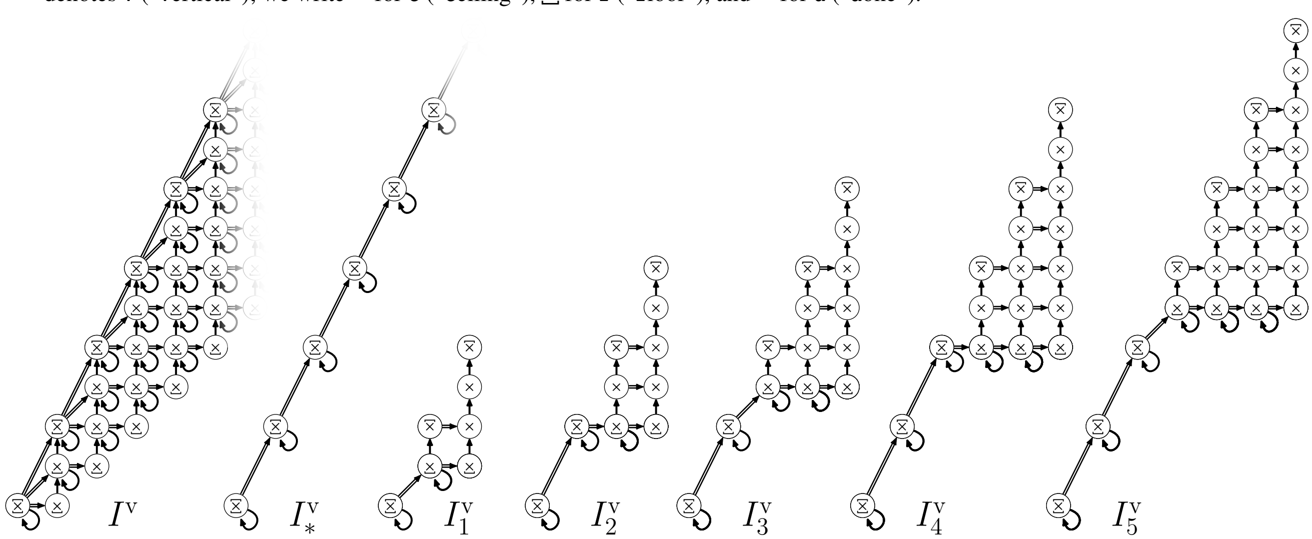


Figure 4: Two infinite universal models of $\mathcal{K}^{\mathrm{v}}$ (I^{v} having infinite treewidth and I_*^{v} having a treewidth of 1), and intermediate (non-consecutive) "snapshots" I_1^{v} – I_5^{v} of a core chase sequence for $\mathcal{K}^{\mathrm{v}}$, demonstrating that the treewidth grows monotonically beyond any bound.