No Cliques Allowed: The Next Step Towards BDD/FC Conjecture (Extended Abstract)

Lucas Larroque[1], Piotr Ostropolski-Nalewaja[2] and Michaël Thomazo[1]

[1]Inria, DI ENS, ENS, CNRS, PSL University, France
[2]University of Wroclaw, Poland

Abstract

This paper addresses one of the fundamental open questions in the realm of existential rules: the conjecture on the finite controllability of bounded derivation depth rule sets (bdd ⇒fc). We take a step toward a positive resolution of this conjecture by demonstrating that universal models generated by bdd rule sets cannot contain arbitrarily large tournaments (arbitrarily directed cliques) without entailing a loop query, $\exists x\, E(x, x)$. This simple yet elegant result narrows the space of potential counterexamples to the (bdd ⇒fc) conjecture.

Keywords

Chase, Existential Rules, Finite unification sets, Bounded Derivation Depth Property, Tuple Generating Dependencies, TGDs, BDD, FUS, Finite Controllability, BDD/FC Conjecture, FO-Rewritability

The last decade has witnessed a fruitful interplay between the database and the knowledge representation communities, aiming to facilitate the integration and querying of legacy databases. The general idea is to allow a user to formulate queries using a logical formalism, whose predicates' meaning is constrained through first-order formulas. This leads to a core reasoning problem, called *ontology based query answering* (OBQA), that takes as input a database $\mathcal{I}$, a rule set $\mathcal{R}$ and a Boolean conjunctive query q, and asks whether $\mathcal{I}, \mathcal{R} \models q$, where $\models$ denotes the entailment relation of first-order logic. In other words, is that true that any model (finite or infinite) of $\mathcal{I}$ and $\mathcal{R}$ is a model of q? Rules are often expressed as *existential rules*[1] which are formulas of the shape $\forall \bar{x}, \bar{y}\ B(\bar{x}, \bar{y}) \rightarrow \exists \bar{z}\ H(\bar{y}, \bar{z})$, where B and H are conjunctions of atoms respectively called the *body* and the *head* of the rule. It is well known that when no further constraints are put on rules, OBQA is an undecidable problem.

Decidability of OBQA A significant research effort has thus been devoted to design conditions on $\mathcal{R}$ that ensure the decidability and sometimes the tractability of Boolean conjunctive query answering under $\mathcal{R}$. An important tool to understand these conditions is the *chase* [5]. In a nutshell, the chase is an algorithm that adds fresh terms and new atoms to ensure that each mapping of a rule body can be extended into a mapping of a rule head. The result of this possibly infinite process is a specific model of $\mathcal{I}$ and $\mathcal{R}$, which has a universal model property [6]: it homomorphically maps to any model of $\mathcal{I}$ and $\mathcal{R}$. When the chase is finite (which is for instance the case when the rule set fulfills some kind of acyclicity [7, 8, 9, 10, 11]), this provides a decision procedure for query answering. It is also the case when the chase is not finite, but is a structure of bounded treewidth — this is typically the case for guarded rules [12].

Finally, and of great interest for our work, is the case of UCQ-rewritability. A rule set $\mathcal{R}$ is said to be UCQ-rewritable if conjunctive queries (CQs) can always be rewritten. Specifically, for every CQ q, there exists a union of conjunctive queries (UCQ) $Q_\mathcal{R}$ such that, for any database $\mathcal{I}$:

$$\mathcal{I}, \mathcal{R} \models q \iff \mathcal{I} \models Q_\mathcal{R}.$$

DL 2025: 38th International Workshop on Description Logics, September 3–6, 2025, Opole, Poland
lucas.larroque@inria.fr (L. Larroque); postropolski@cs.uni.wroc.pl (P. Ostropolski-Nalewaja); michael.thomazo@inria.fr (M. Thomazo)
0009-0007-2351-2681 (L. Larroque); 0000-0002-8021-1638 (P. Ostropolski-Nalewaja); 0000-0002-1437-6389 (M. Thomazo)

[1]Existential rules are also referred to as a *tuple-generating dependencies* [1], *conceptual graph rules* [2], Datalog$^\pm$ [3], and $\forall\exists$-*rules* [4] in the literature.

Given that the chase can produce an infinite structure, the UCQ-rewritability of a rule set ensures decidable OBQA, as the rewriting can be computed and subsequently used to query the database. Unsurprisingly, UCQ-rewritability has been the subject of extensive research, leading to the introduction of numerous subclasses over the past few decades, including: *linear theories*, which permit at most a single atom in rule bodies [13]; *guarded* `bdd` *theories*, which generalize linear theories [14, 15]; *sticky theories*, where strong restrictions are imposed on joins [16].

UCQ-rewritability is also studied independently under a number of different names — highlighting the fundamental nature of the property — with notable cases of finite unification sets class (`fus`) and bounded derivation depth property (`bdd`). The `fus` class is closely tied to the concept of the backward chaining procedure, which is essentially a process of "chasing in reverse." In this approach, to answer the entailment question, one starts with a query and attempts to derive a substructure of the database by applying rules in reverse. The `bdd` class has a much older origin, dating back to the 1980s. It stems from the classical notion of the boundedness of a Datalog program [17].

Definition 1. *A rule set has* bounded derivation depth (bdd) *if for every CQ q, there is some $k \geq 0$ such that for all instances $\mathcal{I}$, we have $\langle \mathcal{I}, \mathcal{R} \rangle \models q$ if and only if q is already entailed at step k of the chase from $\mathcal{I}$ and $\mathcal{R}$. We use* `bdd` *to denote the class of rule sets with bdd.*

OBQA in the finite Databases, however, are finite structures, and under the understanding that $\mathcal{I}$ is a partial description of the real world, and that $\mathcal{R}$ are constraints that should hold on the finite real world, the classical notion of entailment provided by first-order logic does not correspond to the desired notion of entailment. Indeed, one should be more interested in knowing whether "q holds in any *finite* database that contains $\mathcal{I}$ and is a model of $\mathcal{R}$". This semantics is clearly related with the previous one, as a consequence of $\mathcal{I}$ and $\mathcal{R}$ in the unrestricted semantics is also one in the finite one. The converse is however not true, as the following prototypical example witnesses.

Example 2. *Let us consider $\mathcal{I} = \{\mathtt{E}(a, b)\}$, and $\mathcal{R}$ containing the two rules $\forall x \forall y\ \mathtt{E}(x, y) \rightarrow \exists z\ \mathtt{E}(y, z)$ and $\forall x \forall y \forall z\ \mathtt{E}(x, y) \wedge \mathtt{E}(y, z) \rightarrow \mathtt{E}(x, z)$. The query $\exists x\ \mathtt{E}(x, x)$ is not a consequence of $\mathcal{I}, \mathcal{R}$ under the unrestricted semantics, as witnessed by the chase. In any finite model however, there must be a cycle, and hence a loop because of the transitivity enforced by the second rule.*

Finite Controllability Finite reasoning provides us with a semi-decision procedure for non-entailment (enumerating finite interpretations, and checking whether there exists one that is not a model of Q), but entailment is not semi-decidable anymore, as a finite universal model may not exist. An interesting question, already raised by Rosati [18], is, given a rule set $\mathcal{R}$, whether for any database $\mathcal{I}$ and any query Q the unrestricted and the finite semantics coincide. If so, the rule set $\mathcal{R}$ is said to be *finitely controllable* (`fc`). Note that whenever a rule set is `fc`, query answering becomes decidable. Indeed, the classical mathematical tools for semi-deciding first-order entailment are still available, while non-entailment can be witnessed by finite structures. Rosati showed that inclusion dependencies, a very restricted form of existential rules, enjoy finite controllability [18]. This has been generalized in two ways, for guarded rules [19], as well as for sticky rules [20]. Note that both classes generalize inclusion dependencies in various ways, guarded rules enjoying the bounded treewidth property, while sticky rules enjoy the bounded derivation depth property.

There is a well-known conjecture by Gogacz and Marcinkowski [21] stating that rule sets with the bounded derivation depth (`bdd`) property are finitely controllable (`bdd` $\Rightarrow$`fc`). Moreover, they demonstrated that the conjecture holds in a specific restricted setting where `bdd` rule sets are defined over a binary signature and have heads containing only a single atom [21]. However, this result applies to a highly limited case, and little progress has been made towards resolving the conjecture in its general form. The broader question of whether finite controllability extends to all `bdd` rule sets remains an important open problem.

Contributions Our main contribution is to show that `bdd` rule sets enjoy certain model-theoretic property. We write $\textsc{Tournaments}_{\mathtt{E}}$ to denote the query: *for all integers k, there is a set $\{ x_1, \ldots, x_k \}$*

of elements in the given instance such that $\mathtt{E}(x_i, x_j)$ *or* $\mathtt{E}(x_j, x_i)$ *holds in the instance for all* $i \neq j$, and write $\textsc{Loop}_{\mathtt{E}}$ to denote the query $\exists x\ \mathtt{E}(x, x)$. Then, the following is our main result:

Theorem 3. *For every* bdd *rule set* $\mathcal{R}$ *and every instance* $\mathcal{I}$*:*

$$(\mathcal{I}, \mathcal{R}) \models \textsc{Tournaments}_{\mathtt{E}} \quad \Rightarrow \quad (\mathcal{I}, \mathcal{R}) \models \textsc{Loop}_{\mathtt{E}}.$$

In addition to providing insights into the structural properties of universal models of bdd rule sets, this result serves as an important step towards proving the (bdd $\Rightarrow$fc) conjecture. Specifically, it narrows the space of potential counterexamples to the conjecture by eliminating the most natural ones. We briefly explain why this is the case. First note that any model that entails $\textsc{Tournaments}_{\mathtt{E}}$ and not $\textsc{Loop}_{\mathtt{E}}$ is infinite, as the terms $x_1, \ldots, x_k$ in $\textsc{Tournaments}_{\mathtt{E}}$ must be distinct for $\textsc{Loop}_{\mathtt{E}}$ not to be entailed. Thus, $\textsc{Tournaments}_{\mathtt{E}}$ does imply $\textsc{Loop}_{\mathtt{E}}$ in the finite setting, and any bdd rule set for which this implication does not hold (in the unrestricted setting) would disprove (bdd $\Rightarrow$fc).

Note however that Example 2 does not constitute such a counterexample, as its rule set is simply not bdd, since the transitivity rule has to be applied at least as many times as the distance between a and b to entail $\mathtt{E}(a, b)$, which cannot be bounded independently of the size of the database. Expanding on that example, one could try to mimic the behavior using bdd rules by replacing the transitivity rule with $\forall x \forall x' \forall y \forall y'\ \mathtt{E}(x, x') \wedge \mathtt{E}(y, y') \rightarrow \mathtt{E}(x, y')$, which entails it. With this change, the rule set is finally bdd, and the chase entails $\textsc{Tournaments}_{\mathtt{E}}$. However, this new rule triggers the entailment of $\exists x\ \mathtt{E}(x, x)$ as soon as $\exists x \exists y\ \mathtt{E}(x, y)$ is entailed, as expected due to Theorem 3.

While Theorem 3 does not entail the (bdd $\Rightarrow$fc) conjecture, as a different type of counterexample could in principle exist, we are confident that the insights and carefully curated toolkit developed to tackle this case represent an important milestone in the effort to settle the (bdd $\Rightarrow$fc) conjecture in the general setting.

Organization of the proof The proof of Theorem 3 goes in two main steps. First, we introduce a series of rule set normalizations to constrain universal models of FUS rule sets in a more manageable structure. While some of these are well-known, the properties we require are tailored to the specific case at hand and need a separate formal proof. Then, we introduce the crucial notion of valley queries that are used to define the E predicate, and we show through a nice application of Ramsey theorem that such queries cannot define arbitrary tournaments.

An extended version of this work with fully detailed proofs can be found in [22].

Next steps A natural next step in this line of research is to establish that with UCQ-rewritable rule sets, one cannot define structures of arbitrarily high chromatic number without entailing the $\textsc{Loop}_{\mathtt{E}}$ query:

Conjecture 4. *For every signature* $\mathbb{S}$*, every UCQ-rewritable rule set* $\mathcal{R}$ *over* $\mathbb{S}$*, and every instance* $\mathcal{I}$*, we have:*

$$Ch(\mathcal{I}, \mathcal{R})|_{\mathtt{E}} \text{ cannot be colored with a finite number of colors} \quad \Rightarrow \quad Ch(\mathcal{I}, \mathcal{R}) \models \textsc{Loop}_{\mathtt{E}}.$$

We believe that this conjecture, if proven, would constitute an elegant model-theoretic property of UCQ-rewritable rule sets and represent a significant step toward proving the (bdd $\Rightarrow$fc) conjecture. Let us elaborate. Similar to the results presented in this paper, if Conjecture 4 holds, it would eliminate a substantial portion of the potential counterexample space: any structure that cannot be colored with a finite number of colors cannot be homomorphically embedded into a finite structure that does not entail $\textsc{Loop}_{\mathtt{E}}$.

Our hope is that the tools developed throughout this paper will serve as a foundation for the proof of Conjecture 4. However, we note that such a proof would not be a straightforward extension of the techniques developed in this paper. The current proof aims at showing existence of four-tournament witnessed by a single valley query. We know, however, that there exists structures of arbitrarily high chromatic numbers that do not contain the four-clique:

Theorem 5 (Erdős [23]). *There exist graphs with arbitrarily high girth and chromatic number.*

Acknowledgments

Piotr Ostropolski-Nalewaja was supported by grant 2022/45/B/ST6/00457 from the Polish National Science Centre (NCN).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] S. Abiteboul, R. Hull, V. Vianu, Foundations of Databases, Addison-Wesley, 1995. URL: http://webdam.inria.fr/Alice/.

[2] E. Salvat, M.-L. Mugnier, Sound and complete forward and backward chainings of graph rules, in: P. W. Eklund, G. Ellis, G. Mann (Eds.), Proceedings of the 4th International Conference on Conceptual Structures (ICCS'96), volume 1115 of *LNCS*, Springer, 1996, pp. 248–262. URL: https://doi.org/10.1007/3-540-61534-2_16. doi:10.1007/3-540-61534-2_16.

[3] G. Gottlob, Datalog+/-: A unified approach to ontologies and integrity constraints, in: V. D. Antonellis, S. Castano, B. Catania, G. Guerrini (Eds.), Proceedings of the 17th Italian Symposium on Advanced Database Systems, (SEBD'09), Edizioni Seneca, 2009, pp. 5–6.

[4] J.-F. Baget, M. Leclère, M.-L. Mugnier, E. Salvat, On rules with existential variables: Walking the decidability line, Artificial Intelligence 175 (2011) 1620–1654. URL: https://www.sciencedirect.com/science/article/pii/S0004370211000397. doi:https://doi.org/10.1016/j.artint.2011.03.002.

[5] D. Maier, A. O. Mendelzon, Y. Sagiv, Testing implications of data dependencies, ACM Transactions on Database Systems (TODS) 4 (1979). doi:10.1145/320107.320115.

[6] A. Deutsch, A. Nash, J. Remmel, The chase revisited, 2008. doi:10.1145/1376916.1376938.

[7] R. Fagin, P. G. Kolaitis, R. J. Miller, L. Popa, Data exchange: Semantics and query answering, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2572 (2003). doi:10.1007/3-540-36285-1_14.

[8] M. Krötzsch, S. Rudolph, Extending decidable existential rules by joining acyclicity and guardedness, 2011. doi:10.5591/978-1-57735-516-8/IJCAI11-166.

[9] B. Cuenca Grau, I. Horrocks, M. Krötzsch, C. Kupke, D. Magka, B. Motik, Z. Wang, Acyclicity notions for existential rules and their application to query answering in ontologies, Journal of Artificial Intelligence Resesearch (JAIR) 47 (2013) 741–808.

[10] L. Gerlach, D. Carral, General acyclicity and cyclicity notions for the disjunctive skolem chase, in: B. Williams, Y. Chen, J. Neville (Eds.), Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence, 2023, USA, AAAI Press, 2023, pp. 6372–6379. URL: https://ojs.aaai.org/index.php/AAAI/article/view/25784. doi:10.1609/aaai.v37i5.25784.

[11] D. Carral, I. Dragoste, M. Krötzsch, Restricted chase (non)termination for existential rules with disjunctions, in: C. Sierra (Ed.), Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017), Australia, ijcai.org, 2017, pp. 922–928.

[12] A. Calì, G. Gottlob, M. Kifer, Taming the infinite chase: Query answering under expressive relational constraints, in: Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning, KR'08, AAAI Press, 2008, p. 70–80.

[13] A. Calì, G. Gottlob, M. Kifer, Taming the infinite chase: query answering under expressive relational constraints, J. Artif. Int. Res. 48 (2013) 115–174.

[14] P. Barceló, G. Berger, C. Lutz, A. Pieris, First-order rewritability of frontier-guarded ontology-mediated queries, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden., 2018, pp. 1707–1713. URL: https://doi.org/10.24963/ijcai.2018/236. doi:10.24963/ijcai.2018/236.

[15] C. Civili, R. Rosati, On the first-order rewritability of conjunctive queries over binary guarded existential rules (extended abstract), 2015.

[16] A. Calì, G. Gottlob, A. Pieris, Advanced processing for ontological queries, Proceedings of the VLDB Endowment 3 (2010). doi:10.14778/1920841.1920912.

[17] H. Gaifman, H. Mairson, Y. Sagiv, M. Y. Vardi, Undecidable optimization problems for database logic programs, J. ACM 40 (1993) 683–713. URL: https://doi.org/10.1145/174130.174142. doi:10.1145/174130.174142.

[18] R. Rosati, On the finite controllability of conjunctive query answering in databases under open-world assumption, J. Comput. Syst. Sci. 77 (2011) 572–594. URL: https://doi.org/10.1016/j.jcss.2010.04.011. doi:10.1016/J.JCSS.2010.04.011.

[19] V. Bárány, G. Gottlob, M. Otto, Querying the guarded fragment, in: 2010 25th Annual IEEE Symposium on Logic in Computer Science, 2010, pp. 1–10. doi:10.1109/LICS.2010.26.

[20] T. Gogacz, J. Marcinkowski, Converging to the chase – a tool for finite controllability, Journal of Computer and System Sciences 83 (2017) 180–206. URL: https://www.sciencedirect.com/science/article/pii/S0022000016300617. doi:https://doi.org/10.1016/j.jcss.2016.08.001.

[21] T. Gogacz, J. Marcinkowski, On the bdd/fc conjecture, in: Proceedings of the 32nd ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS '13, Association for Computing Machinery, New York, NY, USA, 2013, p. 127–138. URL: https://doi.org/10.1145/2463664.2463668. doi:10.1145/2463664.2463668.

[22] L. Larroque, P. Ostropolski-Nalewaja, M. Thomazo, No cliques allowed: The next step towards bdd/fc conjecture, Proc. ACM Manag. Data 3 (2025). doi:10.1145/3725238.

[23] P. Erdös, Graph theory and probability, Canadian Journal of Mathematics 11 (1959) 34–38. doi:10.4153/CJM-1959-003-9.