Spectra of Cardinality Queries over Description Logic Knowledge Bases (Extended Abstract)

Quentin Manière[1,2], Marcin Przybyłko[1]

[1]Department of Computer Science, Leipzig University, Germany
[2]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden/Leipzig, Germany

Keywords
Description Logics, Counting queries, Spectrum, Complexity of reasoning

1. Introduction

A recent line of research has explored ways of leveraging ontology-mediated query answering (OMQA) to support counting queries, a class of aggregate queries that allows to perform analytics on data. Several semantics for such queries have been investigated, differing on how the possibility of multiple models can be taken into account [1, 2, 3].

In this paper we adopt the semantic proposed in [2], extending [4], that defines a counting query as a conjunctive query (CQ) in which some variables have been designated as *counting variables*. Such a query is evaluated in every model of the knowledge base (KB) by considering every possible homomorphism of the query into the model and by then returning the number of obtained assignments for the counting variables. A certain answer of the counting query over the KB is then defined as an interval $[\![m, M]\!]$ that contains all the possible answers across all possible models, *i.e.* a uniform lower bound m and a uniform upper bound M. The complexity of deciding whether a given interval is a certain answer under this semantics is now well-understood for a variety of DLs [5, 6].

In the present paper, rather than providing uniform bounds, we aim to compute (a representation of) the set of possible answers, which is a subset (of tuples) of natural numbers with infinity, *i.e.* a subset of $\mathbb{N}^\infty := \{0, 1, 2, \ldots, \infty\}$. We call this subset the *spectrum of the counting query*, inspired by the notion of spectrum of a formula, that is the set of the possible cardinalities of its models [7, 8]. To do so, we first investigate the possible shapes of spectra for *counting conjunctive queries (CCQs)* and ontologies expressed in the $\mathcal{ALCIF}$. Traditional CQ answering is well-understood in this expressive DL [9, 10] that supports functionality constraints whose interactions with counting queries have never been studied to the best of our knowledge (those proposed in [5] and denoted $\mathcal{N}^-$ being more restricted).

One of the challenges encountered in our work is to clarify how to represent spectra. Indeed, the set of possible answers of a CCQ across models of a KB might, *a priori*, be an arbitrarily complex set of natural numbers, and thus hard to describe by means other than providing the CCQ-KB couple. We aim to identify classes of ontology-mediated queries (OMQs) whose spectra admit an *effective representation*. By effective, we intend that (i) the representation is finite, ideally with a size that can be bounded by the size of the OMQ, (ii) independent of the specific description logic, and (iii) spectrum membership can be efficiently tested, *i.e.* membership can be tested in polynomial time with respect to the size of the integer and of the representation. Finding such a representation, whenever it exists, can be viewed as a precomputation allowing for further analytics.

DL 2024: 37th International Workshop on Description Logics, June 18–21, 2024, Bergen, Norway
quentin.maniere@uni-leipzig.de (Q. Manière); marcin.przybylko@uni-leipzig.de (M. Przybyłko)
0000-0001-9618-8359 (Q. Manière); 0000-0003-1859-7055 (M. Przybyłko)

2. Contributions

Our main contributions are the following. First, we introduce the notion of a spectrum of a CCQ and formalize the problem of computing effective representation thereof, if it exists. We show that connected and individual-free CCQs evaluated on $\mathcal{ALCIF}$ KBs always admit effectively representable spectra, as those must be finitely generated subsets of $\mathbb{N}^\infty$. This further motivates a focus on cardinality queries, *i.e.* Boolean atomic CCQs, introduced in [11], which admit effective representation. We fully characterize the possible shapes of spectra for concept cardinality queries on $\mathcal{ALCIF}$ KBs, in particular showing that every finitely generated subset of $\mathbb{N}^\infty$ can be realized. We also study several sublogics of $\mathcal{ALCIF}$, from $\mathcal{EL}$ and DL-Lite$_{\mathsf{core}}$, for most of which we obtain full characterizations of possible shapes of spectra. For some, only simpler shapes, such as $[\![m, +\infty]\!]$, are possible (see Table 1). For the $\mathcal{ELIF}_\perp$ DL, corresponding to the Horn fragment of $\mathcal{ALCIF}$, we notably use tailored variations of the cycle-reversion techniques introduced to tackle finite model reasoning in such DLs [12, 13, 14]. Our work also features a wealth of examples to prove whether a given shape can indeed be obtained. We further investigate the case of role cardinality queries which feature challenging shapes of spectra already for $\mathcal{EL}_\perp$ KBs. Through connections with the concept case and refinements of the corresponding constructions, we are able to show that, also in the case of role cardinality queries, the possible shapes for $\mathcal{ELIF}_\perp$ KBs are better-behaved than for general $\mathcal{ALCIF}$ KBs. We conclude with bounds regarding the data complexity of computing some of those effective representations, notably relying on existing results on DLs augmented with closed predicates.

The rest of this extended abstract highlights preliminaries regarding spectra, shows a full characterization for $\mathcal{ALCIF}$ and concept cardinality queries, and discusses our tailored version of the cycle-reversion technique.

3. Effective representations of spectra

We assume familiarity with the DL $\mathcal{ALCIF}$, its sublogics and semantics [15]. We assume two disjoint sets of *variables* and *counting variables.* A *counting conjunctive query (CCQ)* takes the form $q(\bar{x}) = \exists \bar{y}\ \exists \bar{z}\ \psi(\bar{x}, \bar{y}, \bar{z})$, where $\bar{x}$ and $\bar{y}$ are tuples of distinct variables, $\bar{z}$ is a tuple of distinct counting variables and ψ is a conjunction of concept and role atoms whose terms are drawn from $\bar{x} \cup \bar{y} \cup \bar{z}$ or individual names. For a given tuple $\bar{a}$ of individuals, and a given model $\mathcal{I}$ of a KB $\mathcal{K}$, we define: $\#q(\bar{a})^{\mathcal{I}} := \#\{\pi_{|\bar{z}} \mid \pi : q \to \mathcal{I}$ homomorphism such that $\pi(\bar{x}) = \bar{a}\}$. The spectrum of q and $\bar{a}$ over $\mathcal{K}$ is further defined as: $\mathsf{Sp}_{\mathcal{K}}(q(\bar{a})) := \{\#q(\bar{a})^{\mathcal{I}} \mid \mathcal{I} \models \mathcal{K}\}$.

While we do not know whether all spectra can be effectively represented, in the sense explained in the introduction, we identify a class of CCQs, namely connected and individual-free CCQs, whose spectra admit such a representation based on the following property:

Lemma 1. *If q is connected and individual-free, then $\mathsf{Sp}_{\mathcal{K}}(q)$ is closed under addition.*

It is indeed well-known that every subset of $\mathbb{N}^\infty$ closed under addition is finitely generated (see *e.g.* [16], Chapter 2, Proposition 4.1). In particular, for every spectrum $\mathsf{Sp}_{\mathcal{K}}(q)$ of a satisfiable connected and individual-free CCQ q, there exist a finite subset S of $\mathbb{N}^\infty$ and two numbers $M, n \in \mathbb{N}^\infty$ such that $\mathsf{Sp}_{\mathcal{K}}(q) = S \cup \{M + k \cdot n \mid k \in \mathbb{N}\}$. Therefore, the problem of computing $\mathsf{Sp}_{\mathcal{K}}(q)$ for such a pair $(\mathcal{K}, q)$ can be properly defined as the task of computing S, M and n.

Notice that $\mathsf{Sp}_{\mathcal{K}}(q) = \emptyset$ if and only if $\mathcal{K}$ is unsatisfiable; and, similarly, $\mathsf{Sp}_{\mathcal{K}}(q) = \{0\}$ if and only if $\mathcal{K}$ is satisfiable but q is unsatisfiable with respect to $\mathcal{K}$. Spectra non-closed under addition are easily found by dropping the above restriction to connected and individual-free CCQs, as shown by the following:

Example 1. *Consider the empty KB $\mathcal{K}$ and the two Boolean CCQs $q_1 := \exists z_1\ \exists z_2\ \mathrm{C}(z_1) \wedge \mathrm{C}(z_2)$ and $q_2 := \exists z_1\ \exists z_2\ \mathrm{r}(\mathsf{a}, z_1) \wedge \mathrm{r}(\mathsf{a}, z_2)$, where z_1 and z_2 are counting variables. It is easily verified that* $\mathsf{Sp}_{\mathcal{K}}(q_1) = \mathsf{Sp}_{\mathcal{K}}(q_2) = \{n^2 \mid n \in \mathbb{N}\} \cup \{\infty\}$.

Table 1
Possible shapes of spectra for some description logics, here $m \in \mathbb{N}$ and V is any subsemigroup of $\mathbb{N}$. $\star$ indicates no other shape is possible.

$\mathcal{L}$		$[\![m,\infty]\!]$	$\emptyset$	$\{0\}$	$\{0\} \cup [\![m,\infty]\!]$	$\{\infty\}$	$\{0,\infty\}$	$V \cup \{\infty\}$
$\mathcal{ALCIF}$	$\star$	✓	✓	✓	✓	✓	✓	✓
$\mathcal{ELIF}_\bot$	·	✓	✓	✓	✓	✓	✓	·
DL-Lite$_\mathcal{F}$	$\star$	✓	✓	✓	✓	✓	✓	·
DL-Lite$_{\text{core}}$, $\mathcal{ALCI}$, $\mathcal{ALCF}$	$\star$	✓	✓	✓	✓	·	·	·
$\mathcal{ELIF}$	$\star$	✓	✓	·	·	✓	·	·
$\mathcal{EL}$	$\star$	✓	·	·	·	·	·	·

4. Spectrum of a concept cardinality query

We now present two results regarding the query $q_\mathrm{C} := \exists z\ \mathrm{C}(z)$, where C is a concept name and z a counting variable. Computing the spectrum of q_C over a KB $\mathcal{K}$ thus corresponds to the natural task of deciding the possible values of $\left|\mathrm{C}^\mathcal{I}\right|$ across the models $\mathcal{I}$ of $\mathcal{K}$. Naturally, q_C satisfies preconditions of Lemma 1 and, thus, its spectrum is finitely generated. Conversely, one can ask which sets are spectra of concept cardinality queries. For a DL $\mathcal{L}$, we say that a set V is *$\mathcal{L}$-concept realizable* if there is a concept C and an $\mathcal{L}$ KB $\mathcal{K}$ such that $\mathsf{Sp}_\mathcal{K}(q_\mathrm{C}) = V$. For $\mathcal{ALCIF}$ KBs, we have the following complete characterization.

Theorem 1. *A subset of $\mathbb{N}^\infty$ is $\mathcal{ALCIF}$-concept realizable iff it is $\emptyset$, $\{0\}$, or any subsemigroup of $\mathbb{N}^\infty$ containing ∞.*

The "only-if" direction is essentially a consequence of Lemma 1, while the "if" direction is a generalization of the following example in which $\mathsf{Sp}_\mathcal{K}(q_\mathrm{C}) = 2\mathbb{N} \cup \{\infty\}$.

Example 2. *Consider the KB $\mathcal{K}$ with the empty ABox and the following $\mathcal{ALCIF}$ TBox enforcing that r is a bijection between A and B:*

$$\mathrm{C} \equiv \mathrm{A} \sqcup \mathrm{B} \quad \mathrm{A} \sqcap \mathrm{B} \sqsubseteq \bot \quad \mathrm{A} \sqsubseteq \exists \mathrm{r}.\mathrm{B} \quad \mathrm{B} \sqsubseteq \exists \mathrm{r}^-.\mathrm{A} \quad \top \sqsubseteq \leq 1\mathrm{r}.\top \quad \top \sqsubseteq \leq 1\mathrm{r}^-.\top$$

We now turn to $\mathcal{ELIF}_\bot$ KBs, for which we are not able to obtain a full characterization of the realizable spectrum. However, we prove that for a set to be realizable, it must have $\alpha = 1$, that is the possible shapes simplify to $S \cup [\![m,\infty]\!]$ for some $m \in \mathbb{N}^\infty$ and $S \subseteq [\![0,m]\!]$.

Theorem 2. *If a subset of $\mathbb{N}^\infty$ is $\mathcal{ELIF}_\bot$-concept realizable, then it has shape $\emptyset$, $\{0\}$, $\{\infty\}$, $\{0,\infty\}$, or $S \cup [\![m,\infty]\!]$ for some $m \in \mathbb{N}$ and $S \subseteq [\![0,m]\!]$.*

The key ingredient to prove the above is a construction of two (potentially infinite) models $\mathcal{I}$ and $\mathcal{J}$ of an $\mathcal{ELIF}_\bot$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ in which $\left|\mathrm{C}^\mathcal{J}\right| = \left|\mathrm{C}^\mathcal{I}\right| + 1 < \infty$. To this end, we refine a cycle-reversion technique which has been developed to study finite reasoning in $\mathcal{ELIF}_\bot$ [14]. More precisely, we tailor the notion of cycles to characterize under which conditions the extension of concept C may be finite, and then carefully manipulate the corresponding models to produce the above $\mathcal{I}$ and $\mathcal{J}$. Definition 1 below describes the cycles of interest for our study.

An *inverse functional path (IFP)* is a sequence $K_0, \mathrm{r}_1, K_1, \ldots, \mathrm{r}_n, K_n$ where $n \geq 1$, $K_0, \ldots, K_n$ are conjunctions of concept names and $\mathrm{r}_1, \ldots, \mathrm{r}_n$ are (potentially inverse) roles such that for all $0 \leq i < n$ we have $\mathcal{T} \models K_i \sqsubseteq \exists \mathrm{r}_{i+1}.K_{i+1}$ and $\mathcal{T} \models K_{i+1} \sqsubseteq \leq 1\mathrm{r}_{i+1}^-.K_i$.

The interesting cycles for a concept C are the IFPs looping on themselves and forcing the presence of (at least) one instance of C "per instance of the cycle", as follows:

Definition 1. *An IFP $K_0, \mathrm{r}_1, K_1, \ldots, \mathrm{r}_n, K_n$ is a* C-generating cycle *if $\mathcal{T} \models K_n \sqsubseteq K_0$ and there exists an IFP $L_0, \mathrm{s}_1, L_1, \ldots, \mathrm{s}_m, L_m$ such that $\mathcal{T} \models L_m \sqsubseteq \mathrm{C}$ and $\mathcal{T} \models K_i \sqsubseteq L_0$ for some $0 \leq i \leq n$.*

Reversing those cycles now means to consider the $\mathcal{ELIF}_\perp$ TBox $\mathcal{T}_{\mathrm{C}}$ obtained from $\mathcal{T}$ by adding, for each C-generating cycle $K_0, \mathrm{r}_1, K_1, \ldots, \mathrm{r}_n, K_n$ and each $0 \leq i < n$, the axioms $K_{i+1} \sqsubseteq \exists \mathrm{r}_i^-.K_i$ and $K_i \sqsubseteq \leq 1\mathrm{r}_{i+1}.K_{i+1}$. A key result towards the proof of Theorem 2 is now:

Lemma 2. *There is a model $\mathcal{I}$ of $\mathcal{K}$ such that $\left|\mathrm{C}^{\mathcal{I}}\right| < \infty$ if and only if the KB $(\mathcal{T}_{\mathrm{C}}, \mathcal{A})$ is satisfiable.*

5. Perspectives

A full characterization of spectra shapes for $\mathcal{ELIF}_\perp$ appears challenging as Theorem 2 suggests that those spectra may have the shapes of arbitrary *numerical semigroups*. Furthermore, while Theorem 1 offers a complete characterization for $\mathcal{ALCIF}$, how to compute the corresponding effective representations remains an open question.

We believe that it could also be interesting to study the impact of our results on the closely related problem of answering (Boolean atomic) queries under the bag semantics. While the semantics adopted in the present paper does not coincide with bag semantics, as discussed for example in [17, 5], considerations regarding the spectra and some of the corresponding techniques might be adapted to this setting.

Acknowledgments

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany and by the Sächsische Staatsministerium für Wissenschaft Kultur und Tourismus in the program Center of Excellence for AI-research “Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig”, project identification number: ScaDS.AI. Second author was supported by the DFG project LU 1417/3-1 QTEC.

References

[1] D. Calvanese, E. Kharlamov, W. Nutt, C. Thorne, Aggregate queries over ontologies, in: Proceedings of the 2nd International Workshop on Ontologies and Information Systems for the Semantic Web (ONISW), 2008, pp. 97–104.

[2] M. Bienvenu, Q. Manière, M. Thomazo, Answering counting queries over DL-Lite ontologies, in: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI), 2020, pp. 1608–1614.

[3] C. Feier, C. Lutz, M. Przybylko, Answer counting under guarded TGDs, in: Proceedings of the 24th International Conference on Database Theory (ICDT), 2021, pp. 11:1–11:22.

[4] E. V. Kostylev, J. L. Reutter, Complexity of answering counting aggregate queries over DL-Lite, Journal of Web Semantics (JWS) 33 (2015) 94–111.

[5] D. Calvanese, J. Corman, D. Lanti, S. Razniewski, Counting query answers over a DL-Lite knowledge base, in: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI), 2020, pp. 1658–1666.

[6] M. Bienvenu, Q. Manière, M. Thomazo, Counting queries over $\mathcal{ALCHI}$ ontologies, in: Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning (KR), 2022, pp. 53–62.

[7] R. Fagin, Generalized first-order spectra and polynomial-time recognizable sets, Complexity of computation 7 (1974) 43–73.

[8] A. Durand, R. Fagin, B. Loescher, Spectra with only unary function symbols, in: Proceedings of the 11th International Workshop on Computer Science Logic (CSL), 1997, pp. 189–202.

[9] B. Glimm, I. Horrocks, C. Lutz, U. Sattler, Conjunctive query answering for the description logic $\mathcal{SHIQ}$, Journal of Artificial Intelligence Research (JAIR) 31 (2008) 157–204.

[10] C. Lutz, The complexity of conjunctive query answering in expressive description logics, in: Proceedings of the 4th International Joint Conference on Automated Reasoning (IJCAR), 2008, pp. 179–193.
[11] M. Bienvenu, Q. Manière, M. Thomazo, Cardinality queries over DL-Lite ontologies, in: Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI), 2021, pp. 1801–1807.
[12] S. S. Cosmadakis, P. C. Kanellakis, M. Y. Vardi, Polynomial-time implication problems for unary inclusion dependencies, Journal of the ACM 37 (1990) 15–46.
[13] R. Rosati, Finite model reasoning in DL-Lite, in: Proceedings of the 5th European Semantic Web Conference (ESWC), 2008, pp. 215–229.
[14] Y. A. Ibáñez-García, C. Lutz, T. Schneider, Finite Model Reasoning in Horn Description Logics, in: Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR), 2014, pp. 490–509.
[15] F. Baader, I. Horrocks, C. Lutz, U. Sattler, An Introduction to Description Logic, Cambridge University Press, 2017.
[16] P. A. Grillet, Commutative Semigroups, Springer New York, NY, 2001.
[17] C. Nikolaou, E. V. Kostylev, G. Konstantinidis, M. Kaminski, B. Cuenca Grau, I. Horrocks, Foundations of ontology-based data access under bag semantics, Journal of Artificial Intelligence (AIJ) (2019) 91–132.