

Ontology-Mediated Queries for Probabilistic Databases (Extended Abstract)^{*}

Stefan Borgwardt¹, İsmail İlkan Ceylan¹, and Thomas Lukasiewicz²

¹ Faculty of Computer Science, Technische Universität Dresden, Germany
`firstname.lastname@tu-dresden.de`

² Department of Computer Science, University of Oxford, UK
`thomas.lukasiewicz@cs.ox.ac.uk`

The semantics of large-scale knowledge bases like NELL and Google’s Knowledge Vault is founded on (tuple-independent) probabilistic databases (PDBs) [3]. As for ordinary databases, they employ the closed-world assumption, i.e., missing facts are treated as being false (having the probability 0), which leads to unintuitive results when querying PDBs. Recently, *open-world probabilistic databases (OpenPDBs)* were proposed to address this issue by allowing probabilities of facts that are not in the database, called *open tuples*, to take any value from a fixed probability interval [1]. In the resulting framework of OpenPDBs, query probabilities are given in terms of *upper* and *lower* probability values, which is more in line with an incomplete view of the world. However, OpenPDBs are still limited in the sense that they do not allow to distinguish queries that are intuitively different with respect to our commonsense knowledge. In order to model such knowledge, we extend OpenPDBs by ontology-mediated queries using Datalog[±] ontologies (which cover some Horn Description Logics). We show that the data complexity dichotomy between P and PP in (Open)PDBs [1, 2] can be lifted to the case of first-order rewritable positive ontologies (without negative constraints); and that the problem can become NP^{PP}-complete once negative constraints are allowed. This result demonstrates the difference between OpenPDBs and PDBs, as in the latter reasoning with ontologies remains in PP. We also propose an approximating semantics that circumvents the increase in complexity caused by negative constraints. Finally, we provide complexity results beyond the data complexity for ontology-mediated query evaluation relative to (tuple-independent) PDBs and OpenPDBs.

References

1. Ceylan, İ.İ., Darwiche, A., Van den Broeck, G.: Open-world probabilistic databases. In: Proc. of KR. pp. 339–348 (2016)
2. Dalvi, N., Suciu, D.: The dichotomy of probabilistic inference for unions of conjunctive queries. J. ACM 59(6), 1–87 (2012)
3. Suciu, D., Olteanu, D., Ré, C., Koch, C.: Probabilistic Databases. Morgan & Claypool (2011)

^{*} Full paper accepted at AAAI’17. This work was supported by DFG in the CRC 912 (HAEC) and the GRK 1907 (RoSI), and by the EPSRC grants EP/J008346/1, EP/L012138/1, EP/M025268/1, and EP/N510129/1.