A probabilistic approach to complex logical query answering

Tamara Cucumides, Floris Geerts

University of Antwerp, Belgium
{tamara.cucumidesfaundez,floris.geerts}@uantwerpen.be

Query answering is a cornerstone of data management, yet it faces significant challenges when applied to knowledge graphs due to their inherent incompleteness. To overcome this, the machine learning community has introduced a new approach called *complex logical query answering* (CLQA) [1]. The goal in this approach is to predict answers that may not be explicitly found in the observable, often incomplete, knowledge graph, but rather in its inferred completion. While an intuitive solution might be to first complete the graph and then query it, this strategy remains underexplored. Instead, current prevailing methods emphasize combining symbolic query answering algorithms with neural link predictors. However, this approach faces limitations, including restrictions on the types of queries supported and a lack of clarity in the semantics of the query answering process.

In our ongoing work, we are exploring the intuitive approach of completing the knowledge graph before querying it, following two key steps.

1. Graph completion. Unlike traditional graph completion methods, our goal is to produce a completion optimized for subsequent querying, meaning that we prioritize obtaining correct answers to queries rather than generating an accurate graph completion. To achieve this, we experiment with link prediction methods and further calibrate them specifically for this task. Also, we explore leveraging the sparsity of data to improve resource efficiency.

2. Querying. We develop methods for querying the completed graph. We treat the completed graph as a probabilistic database and transform complex logical query answering into a probabilistic query answering problem, for which the *possible worlds* semantics is the standard [2]. Given that general query evaluation under this semantics is $\#P$-hard [3], we apply approximation techniques to balance accuracy with computational efficiency.

This work bridges machine learning and database systems by addressing challenges from both domains: graph completion and link prediction in machine learning, and probabilistic query semantics and evaluation in database systems. This interdisciplinary approach aims to provide both a formal and practical study of the problem of complex logical query answering.

References

[1] Hongyu Ren, Mikhail Galkin, Michael Cochez, Zhaocheng Zhu, and Jure Leskovec. Neural graph reasoning: Complex logical query answering meets graph databases. *arXiv preprint arXiv:2303.14617*, 2023.

[2] Dan Suciu, Dan Olteanu, Christopher Ré, and Christoph Koch. *Probabilistic databases*. Springer, 2022.

[3] Nilesh Dalvi and Dan Suciu. The dichotomy of conjunctive queries on probabilistic structures. In *Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 293–302, 2007.