

# Efficient Query Processing in Probabilistic-Temporal Databases

Maximilian Dylla<sup>1</sup>, Iris Miliaraki<sup>1</sup>, and Martin Theobald<sup>2</sup>  
{mdylla, miliaraki, mtb}@mpii.de

<sup>1</sup>Max Planck Institute for Informatics, Campus E1.4, 66123 Saarbrücken, Germany

<sup>2</sup>University of Antwerp, Middelheimlaan 1, 2020 Antwerp, Belgium

## Abstract

Managing uncertain data via probabilistic databases (PDBs) has evolved as an established field of research which has found a plethora of applications, ranging from scientific data management, sensor networks, data integration, to information extraction and knowledge management systems. Despite the polynomial runtime complexity for the data computation step involved in finding probabilistic answer candidates, confidence computations for these answers are known to be  $\#\mathcal{P}$ -hard already for fairly simple select-project-join (SPJ) queries [1]. Thus, efficient strategies for confidence computations and early pruning of low-confidence query answers remain a key challenge for the scalable management of uncertain data. In our current work [2, 4], we specifically focus on efficient query answering in *PDBs with non-materialized views*. We are aiming to identify the top- $k$  query answers, based on their marginal probabilities, before all input tuples that would be needed to compute the query answers in an exhaustive way have been seen by the query processor. Moreover, in contrast to all lineage models known to us which consider lineage as purely propositional formulas, where each formula represents a single query answer, we more generally introduce *first-order lineage formulas*, where each formula may represent an entire set of query answers. Our main observation is that each intermediate step of query processing can unambiguously be captured by such a first-order lineage formula, which is our key for combining data and confidence computations in a tuple-independent PDB setting. In a further step, we describe how the developed query processing techniques for PDBs can seamlessly be applied to a temporal database setting [3, 5, 6]. Also here, lineage serves as a key for the efficient alignment of time-intervals, known as *sequenced semantics* in temporal databases, thus resulting in a *closed and complete* probabilistic-temporal database model.

## References

1. Nilesch Dalvi and Dan Suciu. The dichotomy of conjunctive queries on probabilistic structures. In *PODS*, pages 293–302, 2007.
2. Maximilian Dylla, Iris Miliaraki, and Martin Theobald. Top- $k$  Query Processing in Probabilistic Databases with Non-Materialized Views. To appear in ICDE 2013.
3. Maximilian Dylla, Mauro Sozio, and Martin Theobald. Resolving Temporal Conflicts in Inconsistent RDF Knowledge Bases. In *BTW*, pages 474–493, 2011.
4. Timm Meiser, Maximilian Dylla, and Martin Theobald. Interactive Reasoning in Uncertain RDF Knowledge Bases. In *CIKM*, pages 2557–2560, 2011.
5. Anish Das Sarma, Martin Theobald, and Jennifer Widom. LIVE: A Lineage-Supported Versioned DBMS. In *SSDBM*, pages 416–433, 2010.
6. Yafang Wang, Mohamed Yahya, and Martin Theobald. Time-aware Reasoning in Uncertain Knowledge Bases. In *MUD*, pages 51–65, 2010.