

Fine-Grained Complexity of Query Answering

Internship Proposal

Nofar Carmeli (<https://nofar.carme.li/>)

This research is in the field of database theory, with intersections with fine-grained complexity and enumeration algorithms. The long-term goal of this line of work is, given a user-specified query, to determine what is the most efficient algorithm that can evaluate this query.

In terms of time complexity, as the number of answers to the query may be much larger than the size of the input database, we cannot expect to have a linear-time algorithm for query evaluation (only printing the answers may require longer), so it is very natural to consider enumeration measures. In this sense, the ideal time guarantees are linear preprocessing and constant delay between successive answers. As an intermediate goal, we ask which queries can be solved within this time complexity. The goal is to identify a property that distinguishes the tractable queries from the intractable ones. For queries that are a *natural join*, we understand the complete picture, and the tractability depends on the structure of the hypergraph describing the query [2, 1].

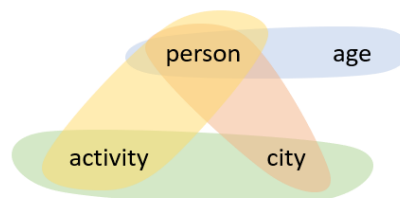
As an example, consider the following query and database. The query asks for a list of people who have a hobby that they can practice in the city where they live. You can also see the query answers and the hypergraph describing the query.

Q(person,activity,city,age):-Hobbies(person,activity),Residence(person,city),Availability(city,activity),Age(person,age)

Hobbies		Residence		Availability		Age	
person	activity	person	city	city	activity	person	age
Léa	climbing	Léa	Montpellier	Montpellier	climbing	Léa	22
Léa	hiking	Thomas	London	Montpellier	swimming	Thomas	24
Thomas	theatre			London	theatre		

Query Answers

person	age	activity	city
Léa	22	climbing	Montpellier
Thomas	24	theatre	London



As the hypergraph describing this query is cyclic, we know that there is no algorithm that can produce the answers to this query with linear preprocessing and constant delay, under some complexity assumptions. If the atom `Residence(person,city)` was removed from the query, it would have been acyclic and solvable efficiently.

In general, when we consider more expressive query classes, the classification is well-understood only in some cases [6, 5, 3, 4], and we can explore during the internship some of the other cases. This type of research usually includes coming up with algorithms, identifying the class of queries for which the algorithms can be applied, coming up with reductions for proving hardness, identifying the class of queries for which the reductions can be applied, identifying queries that are not covered by the algorithms and reductions, trying to extend the algorithms or reductions to cover these queries too, and repeating until we reach a complete classification.

If you find these topics interesting, you can also watch one of the recorded talks on my website to get a better sense of the area. Feel free to contact me so we can discuss more specifically a problem we can solve during the internship.

References

- [1] Guillaume Bagan, Arnaud Durand, and Etienne Grandjean. On acyclic conjunctive queries and constant delay enumeration. In *International Workshop on Computer Science Logic*, pages 208–222. Springer, 2007.
- [2] Johann Brault-Baron. *De la pertinence de l'énumération: complexité en logiques propositionnelle et du premier ordre*. PhD thesis, Université de Caen, 2013.
- [3] Karl Bringmann and Nofar Carmeli. Unbalanced triangle detection and enumeration hardness for unions of conjunctive queries. *Logical Methods in Computer Science*, 21, 2025.
- [4] Nofar Carmeli and Markus Kröll. Enumeration complexity of conjunctive queries with functional dependencies. *Theory of Computing Systems*, 64(5):828–860, 2020.
- [5] Nofar Carmeli and Markus Kröll. On the enumeration complexity of unions of conjunctive queries. *ACM Transactions on Database Systems (TODS)*, 46(2):1–41, 2021.
- [6] Nofar Carmeli and Luc Segoufin. Conjunctive queries with self-joins, towards a fine-grained enumeration complexity analysis. In *Proceedings of the 42nd ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, pages 277–289, 2023.