Constant delay enumeration for First Order queries over classes of graphs with local bounded expansion

Alexandre Vigny

Join work with:
Luc Segoufin

1 Introduction

Query evaluation is a fundamental task in databases and a vast literature is devoted to the complexity of this problem. Given a database $\mathbf{D}$ and a query q the goal is to compute the set $q(\mathbf{D})$ of all solutions for q over $\mathbf{D}$. Unfortunately, the set $q(\mathbf{D})$ might be way bigger than the database itself as the number of solutions could be exponential in the arity of the query. It can therefore be unrealistic to compute all solutions, even for small queries. One could imagine many scenarios to overcome this situation. We could for instance only want to compute the number of solutions or just compute the k most relevant solutions relative to some ranking function.

We consider here the complexity of the enumeration of the set $q(\mathbf{D})$, i.e. generating one by one all the solutions for q over $\mathbf{D}$. In this context two parameters play an important role. The first one is the *preprocessing time*, i.e. the time it takes to produce the first solution. The second one is the *delay*, i.e. the maximum time between the output of any two consecutive solutions. An enumeration algorithm is then said to be *efficient* if these two parameters are small. For the delay, the best we can hope for is constant time: depending only on the query and independent from the size of the database. For the preprocessing time an ideal goal would be linear time: linear in the size of the database with a constant factor depending on the query. When both are achieved we say that the query can be enumerated with constant delay after linear preprocessing.

2 Previous work

Constant delay enumeration after linear preprocessing cannot be achieved for all queries. However, for restricted classes of queries and databases several efficient enumeration algorithms have been obtained. This is the case for instance for first-order (FO) queries over graphs with bounded degree [2, 5], monadic second-order (MSO) queries over databases with bounded tree-width [1, 7] and FO queries over databases with bounded expansion [6]. Bounded expansion is a large class of databases as it contains in particular all structures excluding at least one minor (planarity, bounded tree-width etc.) and all structures of bounded degree [9].

In some scenarios only pseudo-linear preprocessing time has been achieved. A query can be enumerated with constant delay after a pseudo-linear preprocessing time if for all $\epsilon > 0$ there exists an enumeration procedure with constant delay and preprocessing time in $O(\|\mathbf{D}\|^{1+\epsilon})$. This is the case for FO queries over databases with low degree [3].

A special case of enumeration is when the query is boolean. In this case the preprocessing computes the answer to the query. In order to be able to enumerate queries of a given language efficiently, it is therefore necessary to be able to solve the boolean case efficiently.

It has been shown recently that boolean FO queries could be computed in pseudo-linear time over nowhere dense graphs [4]. Nowhere dense is an important class of graphs generalizing bounded expansion [9]. Amongst classes of graphs closed under sub-graphs, Nowhere dense is the largest possible class enjoying efficient evaluation for FO queries [8].

It's a major open problem to show that over nowhere dense graphs the boolean case can be extended to a constant delay enumeration for FO queries of higher arities.

We made one step towards solving this problem, extending the bounded expansion result to graphs having local bounded expansion. Local bounded expansion lies strictly between bounded expansion and nowhere dense. It contains for instance all graphs having local bounded tree-width, or excluding locally a minor.

3 Our results

Our main contributions, proved in [10] are the two following theorem.

Theorem 1. *The enumeration of first-order query over class of graphs with local bounded expansion can be done with constant delay, after pseudo-linear preprocessing. Moreover the output tuples are given in lexicographiacal order.*

Theorem 2. *Let $\mathcal{C}$ be a class of graphs with local bounded expansion and $q(\overline{x})$ be a first-order query. Then for all graph G in $\mathcal{C}$, we can compute $|q(G)|$ in pseudo-linear time.*

References

[1] Guillaume Bagan. MSO queries on tree decomposable structures are computable with linear delay. In *Computer Science Logic (CSL'06)*, 2006.

[2] Arnaud Durand and Etienne Grandjean. First-order queries on structures of bounded degree are computable with constant delay. *ACM Trans. Comput. Log.*, 8(4), 2007.

[3] Arnaud Durand, Nicole Schweikardt, and Luc Segoufin. Enumerating answers to first-order queries over databases of low degree. In *Symp. on Principles of Database Systems (PODS'14)*, 2014.

[4] Martin Grohe, Stephan Kreutzer, and Sebastian Siebertz. Deciding first-order properties of nowhere dense graphs. In *Symp. on Theory of Computing (STOC)*, 2014.

[5] Wojciech Kazana and Luc Segoufin. First-order query evaluation on structures of bounded degree. *Logical Methods in Computer Science*, 7(2), 2011.

[6] Wojciech Kazana and Luc Segoufin. Enumeration of first-order queries on classes of structures with bounded expansion. In *Symp. on Principles of Database Systems (PODS)*, 2013.

[7] Wojciech Kazana and Luc Segoufin. Enumeration of monadic second-order queries on trees. *ACM Trans. Comput. Log.*, 14(4), 2013.

[8] Stephan Kreutzer and Anuj Dawar. Parameterized complexity of first-order logic. *Electronic Colloquium on Computational Complexity (ECCC)*, 16:131, 2009.

[9] Jaroslav Nešetřil and Patrice Ossona de Mendez. On nowhere dense graphs. *Eur. J. Comb.*, 32(4):600–617, 2011.

[10] Luc Segoufin and Alexandre Vigny. Constant delay enumeration for FO queries over databases with local bounded expansion. In *20th International Conference on Database Theory, (ICDT'17)*, 2017.