Counting the Solutions to a Query

Marcelo Arenas ✉

Pontificia Universidad Católica de Chile, Santiago, Chile
Millennium Institute Foundational Research on Data, Santiago, Chile

Abstract

In this talk, we consider the problem of counting the solutions to a query. Our first motivating scenario is the use of regular expressions to extract paths from a graph database. More specifically, given a graph database D, a regular expression r and a natural number n, consider the problem of counting the number of paths p in D such that p conforms to r and the length of p is n. This problem is known to be hard, namely #P-complete. In this talk, we show that this problem admits a fully polynomial-time randomized approximation scheme (FPRAS). Remarkably, the key idea to prove this result is to show that the fundamental problem #NFA admits an FPRAS, where #NFA is the problem of counting the number of strings of length n accepted by a non-deterministic finite automaton (NFA). While this problem is known to be #P-complete and, more precisely, SpanL-complete, it was open whether this problem admits an FPRAS. In this work, we solve this open problem and obtain as a welcome corollary that every function in SpanL admits an FPRAS.

As a second motivating scenario, we consider the widely used class of conjunctive queries over relational databases. More specifically, for every class $\mathcal{C}$ of conjunctive queries with bounded treewidth, we introduce the first FPRAS for counting the answers to a query in $\mathcal{C}$. In fact, our FPRAS is more general, and also applies to conjunctive queries with bounded hypertree width, as well as unions of such queries. As for the case of graph databases, the key ingredient in our proof is the resolution of a fundamental counting problem from automata theory. Specifically, we show that the problem #TA admits an FPRAS, where #TA is the problem of counting the number of trees of size n accepted by a tree automaton (TA).

This talk is based on the results presented in [1, 2].

References

1 Marcelo Arenas, Luis Alberto Croquevielle, Rajesh Jayaram, and Cristian Riveros. #NFA admits an FPRAS: efficient enumeration, counting, and uniform generation for logspace classes. *J. ACM*, 68(6):48:1–48:40, 2021.

2 Marcelo Arenas, Luis Alberto Croquevielle, Rajesh Jayaram, and Cristian Riveros. When is approximate counting for conjunctive queries tractable? In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1015–1027, 2021.