A Worst-Case Optimal Multi-Round Algorithm for Parallel Computation of Conjunctive Queries on Graphs

Bas Ketsman*
Databases and Theoretical Computer Science Research Group,
Hasselt University

Abstract

This talk presents the results in [2] about the worst-case optimal communication cost for parallel evaluation of conjunctive queries. Particularly, we describe a multi-round algorithm that computes any conjunctive query with load $m/p^{1/\rho^*(\mathcal{Q})}$ per server, in the case where all input relations are binary. Here, $\rho^*(\mathcal{Q})$ denotes the fractional edge covering number of the query hypergraph, m is the maximum size of relations, and p is the number of servers.

In this setting two main prior results where known: First, for one-round algorithms over skew-free data, the optimal communication cost per server is $m/p^{1/\tau^*(\mathcal{Q})}$ [1], where m is the size of the largest input relation, and $\tau^*(\mathcal{Q})$ is the fractional vertex covering number of the query Hypergraph. Our algorithm exploits this result to obtain the mentioned bound when no skew is present. Second, it has been proven recently that for multi-round algorithms and unrestricted database instances, any algorithm requires at least $m/p^{1/\rho^*(\mathcal{Q})}$ communication cost per server [3]. Hence, our algorithm is essentially optimal.

References

[1] Paul Beame, Paraschos Koutris, and Dan Suciu. Skew in parallel query processing. In *Proceedings of PODS 2014*, pages 212–223, 2014.

[2] Bas Ketsman and Dan Suciu. A worst-case optimal multi-round algorithm for parallel computation of conjunctive queries. Accepted for publication at PODS, 2017.

[3] Paraschos Koutris, Paul Beame, and Dan Suciu. Worst-case optimal algorithms for parallel query processing. In *ICDT 2016*, pages 8:1–8:18, 2016.

*PhD Fellow of the Research Foundation Flanders - FWO.