Ketsman, Bas; Suciu, Dan; Tao, Yufei

Summary: We present a constant-round algorithm in the massively parallel computation (MPC) model for evaluating a natural join where every input relation has two attributes. Our algorithm achieves a load of $\tilde{O}(m/p^{1/\rho})$ where $m$ is the total size of the input relations, $p$ is the number of machines, $\rho$ is the join’s fractional edge covering number, and $\tilde{O}(\cdot)$ hides a polylogarithmic factor. The load matches a known lower bound up to a polylogarithmic factor. At the core of the proposed algorithm is a new theorem (which we name the isolated cartesian product theorem) that provides fresh insight into the problem’s mathematical structure. Our result implies that the subgraph enumeration problem, where the goal is to report all the occurrences of a constant-sized subgraph pattern, can be settled optimally (up to a polylogarithmic factor) in the MPC model.

MSC: 03B70 Logic in computer science 68-XX Computer science

Keywords: joins; conjunctive queries; parallel computing; database theory

Full Text: arXiv Link

References:

Edited by FIZ Karlsruhe, the European Mathematical Society and the Heidelberg Academy of Sciences and Humanities © 2022 FIZ Karlsruhe GmbH


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.