Efficient Regular Path Query Evaluation in PGX*

Xuming Meng[1,2] Hassan Chafi[2] George Fletcher[1] Sungpack Hong[2]
Oskar van Rest[2]

[1]Eindhoven University of Technology
[2]Oracle Labs

In response to the wide availability of massive graph structured data sets (e.g., social, biological, and citation networks), there has been increasing interest in the design and engineering of graph management systems. Many open research challenges accompany these investigations. A fundamental topic is the engineering of graph query languages.

In our work, we are focusing on efficient evaluation of a basic query class, namely the *regular path queries* (RPQ), on property graphs. RPQs, which identify pairs of vertices connected via paths conforming to a regular expression specified over the edge labels of the graph instance, lie at the core of practical graph query languages such as PGQL, OpenCypher, and SPARQL 1.1. Different from previous graph data models such as RDF, property graphs are an emerging practical data model capturing edge and vertex-labeled directed graphs, where both edges and vertices carry properties as arbitrary key-values. Our study is undertaken in the context of *PGX*, a state-of-the-art parallel in-memory graph analytics framework developed at Oracle Labs.[1]

To efficiently process RPQs in a main-memory setting, we identify a practical approach based on state-of-the art transitive closure solutions, which we adapt to property graphs by introducing a graph summary structure called the *reachability graph*. Two particular implementation strategies are used interchangeably according to the estimated size of the reachability graph. To support effective query plan generation, we also study scalable solutions for transitive closure size estimation. In this talk, we will give an overview of our solutions and initial experimental studies which demonstrate their effectiveness and directions for further research.

*Contact author: Xuming Meng (jonas.meng.jm@gmail.com)

[1]http://www.oracle.com/technetwork/oracle-labs/parallel-graph-analytics/