SuReQL: A Subgraph Match Based Relational Model For Large Graphs

Shijie Zhang, Jiong Yang, and Boya Sun
Case Western Reserve University
Motivation

Graphs are everywhere

- **Social networks**
  - Facebook, Google+, Linkedin, Twitter
- **Biological graphs**
  - Chemicals, protein interaction networks, protein structure graphs
Motivation

We need a sophisticated tool to query graph databases

- SQL for relational database
- Dremel for web-scale datasets
- ? for a graph of billions of nodes.
SuReQL for Graph Querying

- Subgraph match-based
- SQL like syntax
- Used for single large graph database
Graph Match Algebra

● **Selection**
  ○ Select matches that meet certain conditions.

● **Projection**
  ○ Select matches that are restricted to a vertex set.

● **Set Operators**
  ○ Operators on the matches of the same query graph

● **Natural Join**
  ○ Combined matches of two different graphs.

● **Theta Join**
  ○ Combined matches, meeting certain conditions, of two different graphs.
Basic Syntax

Similar to SQL

SELECT QUERY MATCHES
FROM MATCHES OF QUERY IN DATABASE
WHERE CONDITION
Access Method

- Subgraph isomorphism problem is NP-complete
- An index based approach is preferred
- The index structure should not grow exponentially to the graph size.
- May adopt Pregel or other distributed computing methods to build index.
Thank you!