Graphflow: An Active Graph Database
Chathura Kankanamge, Siddhartha Sahu, Amine Mhedhbi, Jeremy Chen, Semih Salihoglu
Data Systems Group, University of Waterloo, Canada

Graph Databases - Overview
- Data management system for querying graph data.

Motivation
- Existing GDBs are passive (i.e., one-time queries only).
- Some apps need continuous queries, i.e., triggers.

Graphflow Highlights
- Active GDB supporting cont. subgraph queries/triggers.
- Worst case optimal join algorithms.
  - Generic Join - One-time subgraph queries.
  - Delta Generic Join - Continuous subgraph queries.

One-time Subgraph Queries: Generic Join vs Binary Join
- Generic Join: Column-at-a-time plan
  - Pick a column ordering: e.g., a, b, c
- Binary Join: Table-at-a-time plan
  - Pick a table order. i.e., join tree

Continuous Subgraph Queries: Delta Generic Join
- IVM on an n-relation join can be decomposed into n delta queries.
  - e.g. - for the 3-relation triangle query
- Graphflow internally stores;
  - all stored as adjacency lists

Theorem: Under insertion-only workloads, cumulative runtime of Delta-GJ on t batches ≤ O(GJ on R₁ U R₂ U ... U Rₙ)

Graphflow Architecture
- Cypher++
  - Updates
  - Subgraph
  - Shortest Path
  - Continuous Subgraph
- One-time Query Processor
  - One-time Subgraph
  - Generic Join
- Continuous Query Processor
  - Continuous Subgraph
  - Delta Generic Join
- Notify updates
- In-memory Graph Store

Evaluations
- One-time query (Diamond) runtime - Neo4j vs GF
- Cont. query (Diamond) runtime - PostgreSQL vs GF
- One-time query (4 - clique) GF Generic J. vs GF Binary J.