High-Performance Graph Analysis

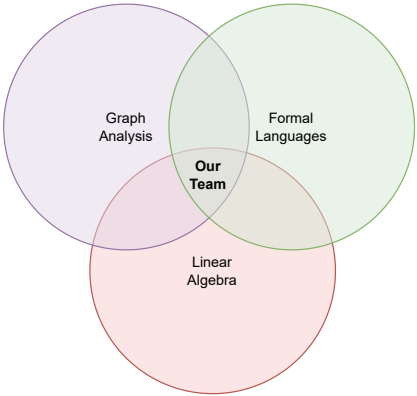
Semyon Grigorev

Saint Petersburg State University

March 22, 2024

Research Landscape

- Formal language constrained path querying[a]
  - Regular path querying (RPQ): partially supported in OpenCypher, part of GQL
  - Context-free path querying (CFPQ, CFL-r): proposal to OpenCypher
- Linear algebra based graph analysis
  - Specialized libraries for linear algebra (including GPGPU)
  - Specialized optimizations
  - New linear algebra based algorithms



[a] For given language $\mathcal{L}$ and edge-labelled graph $\mathcal{G}$ find $R = \{\pi \mid \pi - \text{path in } \mathcal{G}, \omega(\pi) \in \mathcal{L}\}$,
$\omega(v_0 \xrightarrow{l_0} v_1 \xrightarrow{l_1} v_2 \ldots) = l_0 l_1 \ldots$

Context-Free Path Querying (CFPQ)

- Multiple source CFPQ for RedisGraph
  - The **first full-stack CFPQ support**: OpenCypher and execution plan extension, evaluation algorithm
- CFPQ for Neo4j
  - GLL-based (non-matrix) CFPQ algorithm
  - **Computes all paths**, not reachability facts only
- Generic CFPQ solver
  - Optimization of the Context-Free Language Reachability Matrix-Based Algorithm
  - **Outperforms state-of-the-art generic CFPQ solvers and some specialized solutions**

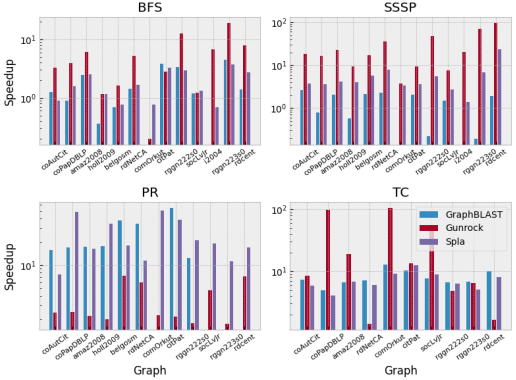
Regular Path Querying (RPQ)

- Linear algebra based multiple source RPQ
- Work in progress

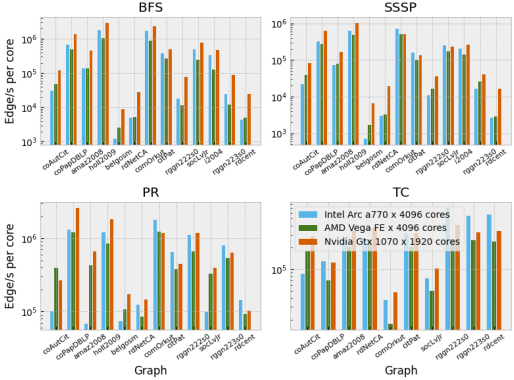
Spla[1]

Generalized sparse linear algebra library with vendor-agnostic GPUs accelerated computations

Performance comparison



Portability



[1] https://github.com/SparseLinearAlgebra/spla

Research Directions

- Graph analysis algorithms research and development
  - CFPQ and RPQ algorithms evaluation, optimization, etc.
  - Dataset collection (for CFPQ, RPQ)
  - New graph algorithms development
- Specialized software and hardware for high-performance linear algebra
  - Generic sparse linear algebra library with GPGPU support (Spla)
  - High-level optimizations for linear algebra (kernel fusion)
  - Specialized hardware for linear algebra: lambda-processors