Algebraic Path Problems and GraphBLAS: a Way to High-Performance Network Analysis

Semyon Grigorev

Saint Petersburg State University

April 28, 2022

Agenda

- Algebraic Path Problems
- GraphBLAS API
- Our team

Algebraic Path Problems

- Semiring-like structures to specify constraints on paths
  - Reachability — boolean semiring
  - Shortest paths — tropical semiring
  - ...

Algebraic Path Problems

- Semiring-like structures to specify constraints on paths
  - Reachability — boolean semiring
  - Shortest paths — tropical semiring
  - ...
- Linear algebra friendly algorithms
  - Transitive closure using matrix-matrix multiplication
  - APSP using matrix-matrix multiplication
  - BFS-like traversals using matrix-vector multiplication
  - ...

Algebraic Path Problems

- Semiring-like structures to specify constraints on paths
  - Reachability — boolean semiring
  - Shortest paths — tropical semiring
  - ...
- Linear algebra friendly algorithms
  - Transitive closure using matrix-matrix multiplication
  - APSP using matrix-matrix multiplication
  - BFS-like traversals using matrix-vector multiplication
  - ...
- Compositionality
  - Having two semirings one can create a new one
  - Single solution for similar problems
    - Generic solution
    - Configurable solution

Expressivity of the Framework

- Semiring with typical associativity and distributivity laws
  - Path Problems in Networks[1]
    - Most reliable path
    - *k*-shortest paths
    - Reachability under edge failures
    - ...
- Less restrictive structures
  - Without associativity and/or distributivity laws
  - Nonassociative, nonmonotonic, partially ordered, not antisymmetric
  - Negative cycles
    - Unstructured path problems and the making of semirings[2]
    - Efficient Algorithms for Path Problems with Gernal Cost Citeria[3]

[1] https://www.morganclaypool.com/doi/abs/10.2200/S00245ED1V01Y201001CNT003
[2] https://link.springer.com/chapter/10.1007/BFb0028261
[3] https://www.semanticscholar.org/paper/Efficient-Algorithms-for-Path-Problems-with-Gernal-Lengauer-Theune/3fd320d97db0a581952d2919587b112f8df57c0b

Why Associativity Matters

$$f(M_{n\times n}) = M^n = \underbrace{M \cdot M \ldots \cdot M}_{n}$$

Repeated squaring

$$\underbrace{(\underbrace{(M \cdot \ldots)}_{\frac{n}{4}} \cdot \underbrace{(M \cdot \ldots)}_{\frac{n}{4}})}_{\frac{n}{2}} \cdot \underbrace{(\underbrace{(M \cdot \ldots)}_{\frac{n}{4}} \cdot \underbrace{(M \cdot \ldots)}_{\frac{n}{4}})}_{\frac{n}{2}}$$

$$O(n^3)\log n$$

No ways to optimize

$$(\ldots((M \cdot M) \cdot M) \ldots \cdot M)$$

$$O(n^4)$$

GraphBLAS API

- Graph-matrix duality
- Operations over matrices and vectors
  - Parametrized by semiring-like structures
  - Based on sparse data structures
  - Highly parallel

[1] https://github.com/DrTimothyAldenDavis/GraphBLAS
[2] https://github.com/gunrock/graphblast
[3] https://gitee.com/CSL-ALP/graphblas

GraphBLAS API

- Graph-matrix duality
- Operations over matrices and vectors
  - Parametrized by semiring-like structures
  - Based on sparse data structures
  - Highly parallel

- High-performance implementations
  - SuiteSparse:GraphBLAS[1]: pure C
  - GraphBLAST[2]: GPGPU, Cuda C
  - Huawei's GraphBLAS[3]: C++
  - ...

[1] https://github.com/DrTimothyAldenDavis/GraphBLAS
[2] https://github.com/gunrock/graphblast
[3] https://gitee.com/CSL-ALP/graphblas

GraphBLAS API

- Graph-matrix duality
- Operations over matrices and vectors
  - Parametrized by semiring-like structures
  - Based on sparse data structures
  - Highly parallel
- High-performance implementations
  - SuiteSparse:GraphBLAS[1]: pure C
  - GraphBLAST[2]: GPGPU, Cuda C
  - Huawei's GraphBLAS[3]: C++
  - ...
- More information on GraphBLAS
  - Home page: `https://graphblas.org/`
  - GraphBLAS-related resources: `https://graphblas.org/GraphBLAS-Pointers/`
  - Introduction to GraphBLAS: `http://mit.bme.hu/~szarnyas/grb/graphblas-introduction.pdf`

[1] https://github.com/DrTimothyAldenDavis/GraphBLAS
[2] https://github.com/gunrock/graphblast
[3] https://gitee.com/CSL-ALP/graphblas

GraphBLAS Applications

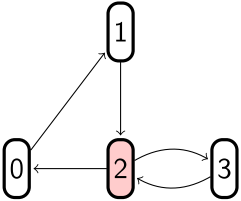
- BFS-like algorithms
  - BFS: levels, parents, multiple sources
  - SSSP
  - ...
- Graph clustering
- Transitive closure based algorithms
  - APSP
  - ...
- Triangle counting
- ...

GraphBLAS Applications

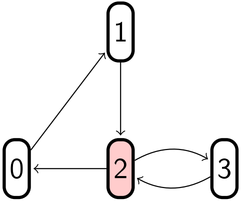
- BFS-like algorithms
  - BFS: levels, parents, multiple sources
  - SSSP
  - ...
- Graph clustering
- Transitive closure based algorithms
  - APSP
  - ...
- Triangle counting
- ...

- LAGraph: collection of GraphBLAS-based algorithms
  - GitHub: `https://github.com/GraphBLAS/LAGraph`
  - Latest report: `https://arxiv.org/pdf/2104.01661.pdf`

BFS-like Skeleton

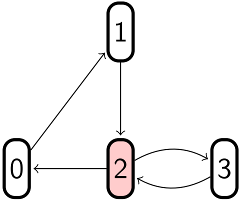


BFS-like Skeleton



$$\overbrace{\begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix}}^{\text{Current front}} \underset{\text{Semiring}}{\times} \overbrace{\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}}^{\text{Adjacency matrix}} = \overbrace{\begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix}}^{\text{New front}}$$

BFS-like Skeleton



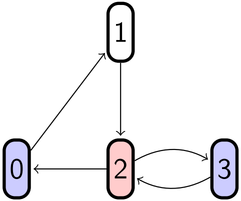
Adjacency matrix

Current front

New front

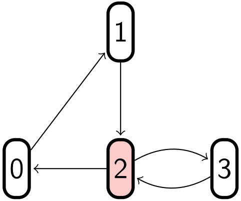
Semiring

$$\begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix} \times \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix}$$

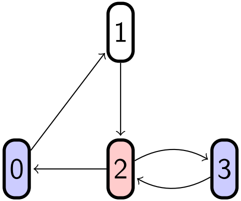


$$\begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix}$$

BFS-like Skeleton

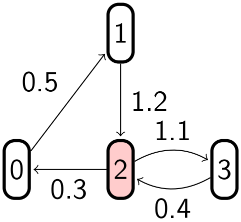


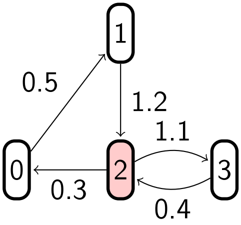
$$\overbrace{\begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix}}^{\text{Current front}} \underset{\text{Semiring}}{\times} \overbrace{\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}}^{\text{Adjacency matrix}} = \overbrace{\begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix}}^{\text{New front}}$$



$$\begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & \cancel{1} & 0 \end{pmatrix}$$

Shortest Paths

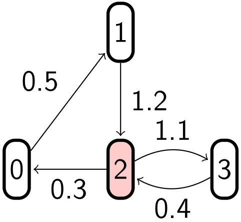




$$\begin{pmatrix} +\infty & +\infty & 0 & +\infty \end{pmatrix} \times \begin{pmatrix} 0 & 0.5 & +\infty & +\infty \\ +\infty & 0 & 1.2 & +\infty \\ 0.3 & +\infty & 0 & 1.1 \\ +\infty & +\infty & 0.4 & 0 \end{pmatrix} = \begin{pmatrix} 0.3 & +\infty & 0 & 1.1 \end{pmatrix}$$

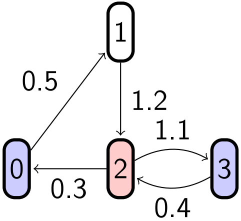
$\{min, +\}$

Shortest Paths



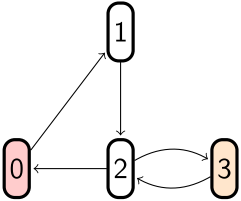
$$\begin{pmatrix} +\infty & +\infty & \colorbox{pink}{0} & +\infty \end{pmatrix} \times \begin{pmatrix} 0 & 0.5 & +\infty & +\infty \\ +\infty & 0 & 1.2 & +\infty \\ 0.3 & +\infty & 0 & 1.1 \\ +\infty & +\infty & 0.4 & 0 \end{pmatrix} = \begin{pmatrix} 0.3 & +\infty & 0 & 1.1 \end{pmatrix}$$

$\{min, +\}$

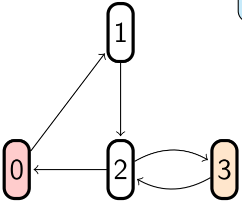


$$\begin{pmatrix} 0.3 & +\infty & 0 & 1.1 \end{pmatrix} \times \begin{pmatrix} +\infty & 0.5 & +\infty & +\infty \\ +\infty & +\infty & 1.2 & +\infty \\ 0.3 & +\infty & +\infty & 1.1 \\ +\infty & +\infty & 0.4 & +\infty \end{pmatrix} = \begin{pmatrix} 0.3 & 0.8 & 0 & 1.1 \end{pmatrix}$$

Multiple Sources Traversion Skeleton



Multiple Sources Traversion Skeleton

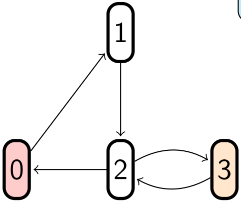


Current fronts for independent tracking

New indepen- dent fronts

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

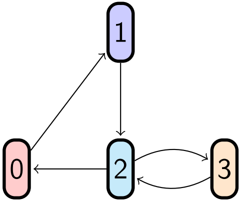
Multiple Sources Traversion Skeleton



Current fronts for independent tracking

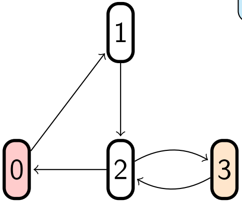
New independent fronts

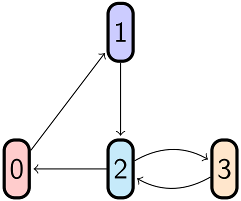
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$



$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \times \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

Multiple Sources Traversion Skeleton





Current fronts for independent tracking

New independent fronts

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \times \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & \cancel{1} \end{pmatrix}$$

Team

- Semyon Grigorev (Lead)
  - PhD (2016)
  - Associate professor (2016, SPbSU)
  - s.v.grigoriev@spbu.ru
  - High-performance graph analysis
  - Graph databases
  - dblp: https://dblp.org/pid/181/9903.html
- Ekaterina Shemetova
  - PhD student
  - Path problems with constraints
  - Fine-grained complexity
  - Dynamic graph problems
- Rustam Azimov
  - PhD student
  - Linear algebra based graph analysis
  - GraphBLAS API
  - Algebraic path problem

Team: Master Students

- Alexandra Istomina
  - Master student
  - Fine-grained complexity
  - Path problems with constraints
  - Algebraic path problem
- Egor Orachev
  - Master student
  - Linear algebra based graph analysis
  - GraphBLAS API
  - GPGPU programming
- Vladimir Kutuev
  - Master student
  - Linear algebra based graph analysis
  - GraphBLAS API
  - Parallel programming
- Julia Susanina
  - Master student
  - Linear algebra based graph analysis
  - Probabilistic graph analysis
  - GPGPU programming

Research areas

- Linear algebra based algorithms for graph analysis
  - GraphBLAS-based algorithms design, implementation and evaluation
  - Portable multi-GPGPU implementation of GraphBALS-like API
  - GraphBLAS API analysis

Research areas

- Linear algebra based algorithms for graph analysis
  - GraphBLAS-based algorithms design, implementation and evaluation
  - Portable multi-GPGPU implementation of GraphBALS-like API
  - GraphBLAS API analysis
- Path problems with constraints
  - Formal Language Constrained Path Querying
    - New algorithms development
    - Complexity analysis
    - New classes of languages investigation
    - High performance algorithms implementation and evaluation

Formal Language Constrained Path Querying

- Particular case of algebraic path problem
  - Multiplication is not associative
  - Multiplication is not commutative
  - ...

Formal Language Constrained Path Querying

- Particular case of algebraic path problem
  - Multiplication is not associative
  - Multiplication is not commutative
  - ...
- Examples
  - Regular path querying (RPQ)
  - Context-free path querying (CFPQ)

Formal Language Constrained Path Querying

- Particular case of algebraic path problem
  - Multiplication is not associative
  - Multiplication is not commutative
  - ...
- Examples
  - Regular path querying (RPQ)
  - Context-free path querying (CFPQ)
- Applications
  - Graph analysis
  - Interprocedural static code analysis
  - Graph database querying

Our Results

- Tools
  - Spla: sparse linear algebra framework for multi-GPU computations based on OpenCL
  - SPbLA: library of GPGPU-powered sparse boolean linear algebra operations

Our Results

- Tools
  - Spla: sparse linear algebra framework for multi-GPU computations based on OpenCL
  - SPbLA: library of GPGPU-powered sparse boolean linear algebra operations
  - CFPQ_PyAlgo: set of GraphBLAS-based FLPQ algorithms
  - LDBC Graphalytics extension for evaluation of formal language constrained path querying

Our Results

- Tools
  - Spla: sparse linear algebra framework for multi-GPU computations based on OpenCL
  - SPbLA: library of GPGPU-powered sparse boolean linear algebra operations
  - CFPQ_PyAlgo: set of GraphBLAS-based FLPQ algorithms
  - LDBC Graphalytics extension for evaluation of formal language constrained path querying
  - GLL4Graph: CFPQ for Neo4j
  - CFPQ for RedisGraph

Our Results

- Tools
  - Spla: sparse linear algebra framework for multi-GPU computations based on OpenCL
  - SPbLA: library of GPGPU-powered sparse boolean linear algebra operations
  - CFPQ_PyAlgo: set of GraphBLAS-based FLPQ algorithms
  - LDBC Graphalytics extension for evaluation of formal language constrained path querying
  - GLL4Graph: CFPQ for Neo4j
  - CFPQ for RedisGraph
- Papers (> 10)
  - SPbLA: The Library of GPGPU-Powered Sparse Boolean Linear Algebra Operations (GrAPL@IPDPS)
  - Evaluation of the context-free path querying algorithm based on matrix multiplication (GRADES-NDA@SIGMOD)
  - Multiple-Source Context-Free Path Querying in Terms of Linear Algebra (EDBT, Core A)
  - Context-free path querying by matrix multiplication (GRADES-NDA@SIGMOD)

Possible Ways for Collaboration

- Algebraic Path Problem framework applicability for network analysis
  - Which constraints can be specified in terms of semirings?
    - Length minimality
    - Nodes to visit
    - ...
  - Is it flexible enough?
- High-performance network analysis
  - GraphBLAS-based solution
  - Algorithms development and analysis
  - Algorithms implementation and evaluation