

A collection of graph algorithms built on top of GraphBLAS

Scott McMillan, Carnegie Mellon University / Software Engineering Institute

Tim Mattson, Intel Corporation

With contributions from:

Gabor Szarnyas, CWI Amsterdam Tim Davis, Texas A&M University Erick Welch , Anaconda James Kitchen, Anaconda David Bader, NJIT Roi Lipman, RedisLabs

GraphBLAS Minisymposium @ SIAM CSE21 1 March 2021



Copyright 2021 Carnegie Mellon University and Timothy G. Mattson

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

DM21-0172

- Sep. 2013: GraphBLAS "position paper" at IEEE HPEC
- Jun. 2015: GraphBLAS Forum kickoff

Standards for Graph Algorithm Primitives

Tim Mattson (Intel Corporation), David Bader (Georgia Institute of Technology), Jon Berry (Sandia National Laboratory), Aydin Buluc (Lawrence Berkeley National Laboratory), Jack Dongarra (University of Tennessee), Christos Faloutsos (Carnegie Melon University), John Feo (Pacific Northwest National Laboratory), John Gilbert (University of California at Santa Barbara), Joseph Gonzalez (University of California at Berkeley), Bruce Hendrickson (Sandia National Laboratory), Jeremy Kepner (Massachusetts Institute of Technology), Charles Leiserson (Massachusetts Institute of Technology), Andrew Lumsdaine (Indiana University), David Padua (University of Illinois at Urbana-Champaign), Stephen Poole (Oak Ridge National Laboratory), Steve Reinhardt (Cray Corporation), Mike Stonebraker (Massachusetts Institute of Technology), Steve Wallach (Convey Corporation), Andrew Yoo (Lawrence Livermore National Laboratory)

- Sep. 2013: GraphBLAS "position paper" at IEEE HPEC
- Jun. 2015: GraphBLAS Forum kickoff
- Dec. 2015: Formation of API Working Group
- May 2017: GraphBLAS C API Specification v1.0 released ("provisional")
- Nov. 2017: SuiteSparse GraphBLAS v1.0 released
- May 2018: IBM GraphBLAS released, "provisional" removed from C API spec.

Standards for Graph Algorithm Primitives

Tim Mattson (Intel Corporation), David Bader (Georgia Institute of Technology), Jon Berry (Sandia National Laboratory), Aydin Buluc (Lawrence Berkeley National Laboratory), Jack Dongarra (University of Tennessee), Christos Faloutsos (Carnegie Melon University), John Feo (Pacific Northwest National Laboratory), John Gilbert (University of California at Santa Barbara), Joseph Gonzalez (University of California at Berkeley), Bruce Hendrickson (Sandia National Laboratory), Jeremy Kepner (Massachusetts Institute of Technology), Charles Leiserson (Massachusetts Institute of Technology), Andrew Lumsdaine (Indiana University), David Padua (University of Illinois at Urbana-Champaign), Stephen Poole (Oak Ridge National Laboratory), Steve Reinhardt (Cray Corporation), Mike Stonebraker (Massachusetts Institute of Technology), Steve Wallach (Convey Corporation), Andrew Yoo (Lawrence Livermore National Laboratory)

- Sep. 2013: GraphBLAS "position paper" at IEEE HPEC
- Jun. 2015: GraphBLAS Forum kickoff

- Dec. 2015: Formation of API Working Group
- May 2017: GraphBLAS C API Specification v1.0 released ("provisional")
- Nov. 2017: SuiteSparse GraphBLAS v1.0 released
- May 2018: IBM GraphBLAS released, "provisional" removed from C API spec.

... But the GraphBLAS are really low level. They are for algorithm developers and researchers, not algorithm users.

What can we do to reach graph algorithm users?

Standards for Graph Algorithm Primitives

Tim Mattson (Intel Corporation), David Bader (Georgia Institute of Technology), Jon Berry (Sandia National Laboratory), Aydin Buluc (Lawrence Berkeley National Laboratory), Jack Dongarra (University of Tennessee), Christos Faloutsos (Carnegie Melon University), John Feo (Pacific Northwest National Laboratory), John Gilbert (University of California at Santa Barbara), Joseph Gonzalez (University of California at Berkeley), Bruce Hendrickson (Sandia National Laboratory), Jeremy Kepner (Massachusetts Institute of Technology), Charles Leiserson (Massachusetts Institute of Technology), Andrew Lumsdaine (Indiana University), David Padua (University of Illinois at Urbana-Champaign), Stephen Poole (Oak Ridge National Laboratory), Steve Reinhardt (Cray Corporation), Mike Stonebraker (Massachusetts Institute of Technology), Steve Wallach (Convey Corporation), Andrew Yoo (Lawrence Livermore National Laboratory)

- Sep. 2013: GraphBLAS "position paper" at IEEE HPEC
- Jun. 2015: GraphBLAS Forum kickoff
- Dec. 2015: Formation of API Working Group
- May 2017: GraphBLAS C API Specification v1.0 released ("provisional")
- Nov. 2017: SuiteSparse GraphBLAS v1.0 released
- May 2018: IBM GraphBLAS released, "provisional" removed from C API spec.
- May 2019: LAGraph "position paper" at IEEE IPDPSW
- Mar 2020: Formation of LAGraph Working Group
- May 2021: Release of LAGraph v1.0 (planned)

Standards for Graph Algorithm Primitives

Tim Mattson (Intel Corporation), David Bader (Georgia Institute of Technology), Jon Berry (Sandia National Laboratory), Aydin Buluc (Lawrence Berkeley National Laboratory), Jack Dongarra (University of Tennessee), Christos Faloutsos (Carnegie Melon University), John Feo (Pacific Northwest National Laboratory), John Gilbert (University of California at Santa Barbara), Joseph Gonzalez (University of California at Berkeley), Bruce Hendrickson (Sandia National Laboratory), Jeremy Kepner (Massachusetts Institute of Technology), Charles Leiserson (Massachusetts Institute of Technology), Andrew Lumsdaine (Indiana University), David Padua (University of Illinois at Urbana-Champaign), Stephen Poole (Oak Ridge National Laboratory), Steve Reinhardt (Cray Corporation), Mike Stonebraker (Massachusetts Institute of Technology), Steve Wallach (Convey Corporation), Andrew Yoo (Lawrence Livermore National Laboratory)

LAGraph: A Community Effort to Collect Graph Algorithms Built on Top of the GraphBLAS

Tim Mattson[‡], Timothy A. Davis[⋄], Manoj Kumar[¶], Aydın Buluç[†], Scott McMillan[§], José Moreira[¶], Carl Yang*,[†]

[†]Intel Corporation [†]Computational Research Division, Lawrence Berkeley National Laboratory

*Texas A&M University ¶IBM Corporation §Software Engineering Institute, Carnegie Mellon University

*Electrical and Computer Engineering Department, University of California, Davis

Sep. 2013: GraphBLAS "position paper" at **IEEE HPEC**

Jun. 2015: GraphBLAS Forum kickoff

Standards for Graph Algorithm Primitives

Tim Mattson (Intel Corporation), David Bader (Georgia Institute of Technology), Jon Berry (Sandia National Laboratory), Aydin Buluc (Lawrence Berkeley National Laboratory), Jack Dongarra (University of Tennessee), Christos Faloutsos (Carnegie Melon University), John Feo (Pacific Northwest National Laboratory), John Gilbert (University of California at Santa Barbara), Joseph Gonzalez (University of California at Berkeley), Bruce Hendrickson (Sandia National Laboratory), Jeremy Kepner tts Institute of Technology), Charles Leiserson (Massachusetts Institute of Technology) na University), David Padua (University of Illinois at Urbana-Champaign) aboratory), Steve Reinhardt (Cray Corporation), Mike Stonel teve Wallach (Convey Corporation).

oratory)

Dec. 2015: Formation of APJ

LAG

- May 2017:
- Nov. 2017:
- May 2018: IBI
- May 2019: **IPDPSW**
- Formation of LAGraph Working Group Mar 2020:

May 2021: Release of LAGraph v1.0 (planned)

THIS IS A WORK IN PROGESS (some details could change)

LAGraph: A Community Effort to Collect Graph Algorithms Built on Top of the GraphBLAS

Tim Mattson[‡], Timothy A. Davis^o, Manoj Kumar[¶], Aydın Buluc[†], Scott McMillan[§], José Moreira[¶], Carl Yang^{*,†}

‡Intel Corporation †Computational Research Division, Lawrence Berkeley National Laboratory *Texas A&M University ¶IBM Corporation \{\sigma Software Engineering Institute, Carnegie Mellon University *Electrical and Computer Engineering Department, University of California, Davis

LAGraph Working Group

- Members (meeting ~ weekly since March 2020):
 - Tim Davis, Gabor Szarnyas, Tim Mattson, Scott McMillan, Jim Kitchen, Erik Welch, David Bader, Roi Lipman, and others
- The purpose of the LAGraph effort:
 - Provide a repository for researchers to share graph algorithms based on GraphBLAS.
 - Create a library of "commercial-grade" algorithms for data scientists and other users.
- Supporting goals
 - Portability across different implementations of GraphBLAS and different platforms
 - Test GraphBLAS implementations, expose weaknesses, and push new features
 - Create a transparent process to new algorithms/utilities to the LAGraph user library
 - Support development of other language bindings for LAGraph+GraphBLAS

LAGraph Target Audience

- We serve two communities
 - Developers of graph algorithms (new and/or improved implementations)
 - Users of graph algorithms (application writers, data scientists, etc)
- Our users come in two flavors:
 - Basic users:
 - Want things to "just work"
 - Modest graphs, exploratory exercises
 - Ease of use over performance
 - Advanced users:
 - HUGE graphs
 - Parallel applications
 - Need more control on memory use, etc...

Elements of the LAGraph library Design

- Graph data structure
- Basic vs. Advanced API
- Signature conventions
- Error handling

The LAGraph "graph" data structure: LAGraph_Graph

- Transparent struct
- Contains two types of data:
 - Primary components
 - "Cached" properties
- New members added as need arises

 Construction "moves" matrix into LAGraph object

 Utilities provided for explicit computation of cached properties

```
struct LAGraph Graph
   // Primary components (REQUIRED)
   GrB Matrix
                          // adjacency matrix
  LAGraph Kind kind;
                          // directed, undirected, etc
   // Cached properties (OPTIONAL)
                          // transposed matrix
   GrB Matrix
   GrB Vector
                row degree;
                col degree;
   GrB Vector
  LAGraph BooleanProperty
                A pattern is symmetric; // T/F/Unk
   // etc...
};
// Matrix ownership transfers on construction
GrB Matrix M;
                 //...construction of M omitted
LAGraph Graph G;
LAGraph New(&G, &M, LAGRAPH DIRECTED, msg);
// explicit computation of cached properties
LAGraph Property AT(G, msg);
```

Basic vs. Advanced Interfaces

Basic Interfaces

- Limited options
 - Likely only one function for a given algorithm
- May inspect input and compute expensive cached properties as needed
 - E.g. compute vertex degrees and sort prior to computation

Advanced Interfaces

- Multiple implementations (algorithms) for the same computation
 - E.g., push vs. pull, or batched modes
- Stricter requirements on inputs
 - E.g., will return an error before computing needed cached properties
- May include lower level entry points into existing algorithms (e.g. single-hop BFS)

Signature Conventions

- Naming conventions
 - LAGraph "namespace"
 - Category of algorithm or utility
 - Algorithm name
- Examples
 - LAGraph Community Louvain
 - LAGraph Community Markov
 - LAGraph_Community_LabelPropagation
 - LAGraph_Property_AT
 - LAGraph_Property_RowDegree
 - LAGraph_Property_ColDegree

```
int LAGraph <Category> <Algorithm>
   // outputs
                    // allocated internally
   Type1
           *out1,
   Type2
           *out2,
  // input/output
   Type3
            inout, // allocated by user
   // inputs
   Type4
            input1,
   Type5
            input2,
   char *msg
```

Signature Conventions

Outputs

- Pass by reference, pointer created by user
- Allocated by algorithm
- Pass **NULL** for optional outputs
- Input/Outputs
 - Pass by value
 - Allocated before call (by user or other functions)
 - Pass graph object here if properties are computed
- Inputs
 - Pass by value (not modified)

```
int LAGraph <Category> <Algorithm>
   // outputs
                     // allocated internally
   Type1
           *out1,
   Type2
           *out2,
   // input/output
   Type3
            inout,
                    // allocated by user
   // inputs
   Type4
            input1,
   Type5
            input2,
   char *msg
```

Signature Conventions

- Return Value, signed integer
 - $= 0 \rightarrow success$
 - > 0 → algorithm specific warning code
 - < 0 → algorithm specific error code
- Error messages, msg
 - User-allocated char buffer of size LAGRAPH_MSG_LEN
 - Holds algorithm-specific error or warning messages

```
int LAGraph <Category> <Algorithm>
   // outputs
                    // allocated internally
   Type1
           *out1,
           *out2,
  Type2
  // input/output
                   // allocated by user
  Type3
            inout,
   // inputs
  Type4
            input1,
  Type5
            input2,
   char *msg
```

Error Handling

- All GraphBLAS and LAGraph methods can return an error.
- Errors should be checked before proceeding.
- Resources should be released after an unrecoverable error.
- Exception handling in C:

```
#define LAGraph_TRY(LAGraph_method)
{
   int LAGraph_status = LAGraph_method;
   if (LAGraph_status < 0) {
       LAGraph_CATCH(LAGraph_status);
   }
}</pre>
```

- Developers define LAGraph_CATCH before a function using LAGraph_TRY
 - Ensure proper freeing of memory
 - Perform any other necessary tasks
- Similar mechanism provided for GraphBLAS calls.

LAGraph: Pushing the state of the art in algorithms using GraphBLAS

- Libraries are important, but we also want to drive research on algorithms that use the GraphBLAS.
- We found that it was hard to share algorithms given everyone used different notations.
- Hence ... we are working to define a consensus notation for expressing graph algorithms using linear algebra

GraphBLAS Notation is Evolving

| Operation | Notation | Operation | Notation |
|----------------|---|---------------------------------|--|
| mxm | $C\langle M \rangle \bigcirc = (A^T \oplus . \otimes B^T)$ | transpose | C ⟨ M ⟩ ⊙= A ^T |
| mxv vxm | $\mathbf{w}\langle \mathbf{m} \rangle \bigcirc = (\mathbf{A}^{T} \oplus . \otimes \mathbf{u})$ $\mathbf{w}^{T}\langle \mathbf{m} \rangle \bigcirc = (\mathbf{u}^{T} \oplus . \otimes \mathbf{A}^{T})$ | reduce (row) reduce (scalar) | $\mathbf{w}\langle \mathbf{m} \rangle \bigcirc = [\bigoplus_{j} \mathbf{A}^{T}(:,j)]$ $s \bigcirc = [\bigoplus_{i,j} \mathbf{A}^{T}(i,j)]$ $s \bigcirc = [\bigoplus_{i} \mathbf{u}(i)]$ |
| eWiseMult | $C\langle M \rangle$ $\bigcirc = (A^T \otimes B^T)$ $W\langle m \rangle$ $\bigcirc = (u \otimes v)$ | eWiseAdd | $C\langle M \rangle$ $\bigcirc = (A^T \oplus B^T)$ $w\langle m \rangle$ $\bigcirc = (u \oplus v)$ |
| extract | $C\langle M \rangle$ $\bigcirc = A^{T}(i,j)$ $w\langle m \rangle$ $\bigcirc = A^{T}(:,j)$ $w\langle m \rangle$ $\bigcirc = w(i)$ | assign | $\begin{array}{ll} \mathbf{C}\langle\mathbf{M}\rangle(i,j)\bigcirc=\mathbf{A}^T, & \mathbf{w}\langle\mathbf{m}\rangle(i)\bigcirc=\mathbf{u} \\ \mathbf{C}\langle\mathbf{M}\rangle(:,j)\bigcirc=\mathbf{u}, & \mathbf{C}\langle\mathbf{M}\rangle(i,:)\bigcirc=\mathbf{u}^T \\ \mathbf{C}\langle\mathbf{M}\rangle(i,j)\bigcirc=s, & \mathbf{w}\langle\mathbf{m}\rangle(i)\bigcirc=s \end{array}$ |
| apply | $\mathbf{C}\langle \mathbf{M} \rangle \bigcirc = f(\mathbf{A}^T, s)$ $\mathbf{w}\langle \mathbf{m} \rangle \bigcirc = f(\mathbf{u}, s)$ | select (NEW) | $\mathbf{C}\langle \mathbf{M} \rangle \bigcirc = \mathbf{A}^{T} \langle f(\mathbf{A}^{T}, s) \rangle$ $\mathbf{w}\langle \mathbf{m} \rangle \bigcirc = \mathbf{u} \langle f(\mathbf{u}, s) \rangle$ |
| | | | |
| build | $\mathbf{C} \leftarrow \{i, j, x\}$ $\mathbf{w} \leftarrow \{i, x\}$ | extractTuples | $\{i, j, x\} \leftarrow A$ $\{i, x\} \leftarrow u$ |
| extractElement | $s = \mathbf{A}(i, j)$ $s = \mathbf{u}(i)$ | setElement | $\mathbf{C}(i,j) = s$ $\mathbf{w}(i) = s$ |

Notation: i, j – indices, i, j – (bold) index arrays, x – (bold) scalar array, m – 1D mask, u, v, w – vectors (column), M – 2D mask, A, B, C – matrices, T – optional transpose,

^{¬ -} structural complement, r – clear output, ⊙, ⊕, or ⊗ monoid/binary function, ⊕.⊗ semiring, blue – optional parameters, red – optional modifiers

GraphBLAS Notation is Evolving

| Operation | Notation | Operation | Notation | | |
|--------------|---|--|--|--|--|
| mxm | $C\langle M \rangle \odot = (A^T \oplus . \otimes B^T)$ | transpose | $C\langle M \rangle$ $\bigcirc = A^T$ | | |
| mxv | $\mathbf{w}\langle\mathbf{m}\rangle \bigcirc = (\mathbf{A}^T \oplus . \otimes \mathbf{u})$ | reduce (row) | $\mathbf{w}\langle\mathbf{m}\rangle \bigcirc = \left[\bigoplus_{j} \mathbf{A}^{T}(:,j)\right]$ | | |
| vxm | Main points of contention among GraphBLAS researchers: | | | | |
| eWiseMult | Mask options: how to incorporate all the mask options (complement, structure-only, replace vs. merge, etc.) | | | | |
| extract | with an intuitive | with an intuitive notation: $\mathbf{C}(\neg \mathbf{s}(\mathbf{M}), \mathbf{r})$??? | | | |
| apply | C(M) | eMult: two options ⊙= (A ^T ⊗ B ^T) | $C\langle M \rangle \odot = (A^T \oplus B^T)$ | | |
| build | C(M) | $\bigcirc = (\mathbf{A}^{T} op_{\cap} \mathbf{B}^{T})$ | $\mathbf{C}\langle \mathbf{M} \rangle \bigcirc = (\mathbf{A}^{T} \text{ op}_{\cup} \mathbf{B}^{T})$ | | |
| extractEleme | $s = \mathbf{A}(i, j)$ $s = \mathbf{u}(i)$ | setElement | $\mathbf{C}(i,j) = s$ $\mathbf{w}(i) = s$ | | |

Notation: i, j – indices, i, j – (bold) index arrays, x – (bold) scalar array, m – 1D mask, u, v, w – vectors (column), M – 2D mask, A, B, C – matrices, T – optional transpose, T – structural complement, T – clear output, T , or T monoid/binary function, T semiring, blue – optional parameters, T – optional modifiers

Example of Notation: Triangle Counting

Inputs: $\mathbf{A} \in \mathbb{B}^{n \times n}$ (symmetric bool adjacency matrix)

Result: $t \in UINT64$

Function *TriangleCount*

p = permutation to sort by degree descending

$$A = A(p, p)$$

$$L = A\langle tril(A)\rangle$$

$$\mathbf{C}\langle s(\mathbf{L})\rangle = \mathbf{L} + . \& \mathbf{L}$$

 $\mathbf{t} = [+_{ii} \mathbf{C}(i, j)]$

```
int TriangleCount(uint64 t *t, GrB Matrix A, char *msg) {
 LAGraph TRY ( LAGraph SortByDegree (&P, A, descending, msg) );
 GrB TRY( GrB extract(A,...,A, P, n, P, n, NULL) );
 GrB TRY( GrB select(L,...,GrB TRIL, A, NULL) );
 GrB TRY( GrB mxm(C, L,..., GrB PLUS AND, L, L, NULL) );
 GrB TRY( GrB reduce(*t,..., GrB PLUS MONOID UINT64, C, NULL) );
 LAGRAPH FREE ALL;
 return 0:
```

Algorithms

- BFS: level and parent*
- SSSP: Bellman Ford, Delta-stepping*
- Connected Components: LACC, FastSV*, SCC-min-label, Boruvka's
- Minimum Spanning Forest: Boruvka's
- Vertex Betweenness Centrality*
- PageRank: GAP-PR*, Graphalytics-PR
- Triangle Counting* (6 methods)
- K-truss enumeration, All K-truss
- Community detection/clustering: label propagation, Louvain, markov, peer pressure
- Local clustering coefficient
- DNN inference

*GAP algorithms targeted for first release

Utilities

- Loaders: matrix market, binary
- Memory management: override malloc/free
- Cached property operations: verify, compute, query, delete
- Sorting: vectors, graphs
- Sampling: graph degrees
- Output: "pretty print" graphs

Next Steps

- Continue to refactor existing algorithms
- Developing the evaluation and testing processes to ensure high-quality implementations:
 - Correctness
 - Robustness
 - Performance

*** REQUEST FOR PARTICIPATION ***

- Contribute new/improved algorithms (in order of preference)
 - Fork the repo, develop code in Experimental, write tests, issue Pull Requests
 - Email code to the working group
 - Publish (we *might* read and incorporate)

Repository Information

- Location
 - https://github.com/GraphBLAS/LAGraph
- License for all contributed code: BSD 2-clause
 - https://github.com/GraphBLAS/LAGraph/blob/master/LICENSE
- Contents [FUTURE]
 - Doc: documentation
 - Include: contains the LAGraph.h header file
 - Source: <u>will</u> contain stable (curated) source code for the library (contains Algorithms and Utilities)
 - Experimental: draft, experimental, submissions, etc. (contains Algorithms and Utilities)
 - Test: programs to verify correctness of LAGraph code

Questions?

Presenters / POC

Scott McMillan, PhD Timothy G. Mattson, PhD

Principal Research Engineer Senior Principal Engineer

Carnegie Mellon University / SEI Intel Corp.

Email: smcmillan@sei.cmu.edu Email: timothy.g.mattson@intel.com

Repository: https://github.com/GraphBLAS/LAGraph

Minutes at: https://github.com/GraphBLAS/LAGraph-Working-Group/tree/master/minutes