Graph Library: Graph Container Interface

Document #: **P3130r1**Date: 2024-02-05

Project: Programming Language C++

Audience: Library Evolution

SG19 Machine Learning

SG14 Game, Embedded, Low Latency

SG6 Numerics

Revises: P3130r0

Reply-to: Phil Ratzloff (SAS Institute)

phil.ratzloff@sas.com Andrew Lumsdaine lumsdaine@gmail.com

Contributors: Kevin Deweese

Muhammad Osama (AMD, Inc)

 ${\rm Jesun\ Firoz}$

Michael Wong (Intel)

Jens Maurer

Richard Dosselmann (University of Regina)

Matthew Galati (Amazon)

Guy Davidson (Creative Assembly)

©ISO/IEC $extbf{P3130r1}$

1 Getting Started

This paper is one of several interrelated papers for a proposed Graph Library for the Standard C++ Library. The Table 1 describes all the related papers.

Paper	Status	Description
P1709	Inactive	Original proposal, now separated into the following papers.
P3126	Active	Overview, describes the big picture of what we are proposing.
P3127	Active	Background and Terminology provides the motivation, theoretical background, and
		terminology used across the other documents.
P3128	Active	Algorithms covers the initial algorithms as well as the ones we'd like to see in the future.
P3129	Active	Views has helpful views for traversing a graph.
P3130	Active	Graph Container Interface is the core interface used for uniformly accessing graph data
		structures by views and algorithms. It is also designed to easily adapt to existing graph data
		structures.
P3131	Active	Graph Containers describes a proposed high-performance compressed_graph container. It
		also discusses how to use containers in the standard library to define a graph, and how to
		adapt existing graph data structures.

Table 1: Graph Library Papers

Reading them in order will give the best overall picture. If you're limited on time, you can use the following guide to focus on the papers that are most relevant to your needs.

Reading Guide

- If you're **new to the Graph Library**, we recommend starting with the *Overview* paper (P3126) to understand the focus and scope of our proposals.
- If you want to **understand the theoretical background** that underpins what we're doing, you should read the *Background and Terminology* paper (P3127).
- If you want to **use the algorithms**, you should read the *Algorithms* paper (P3128) and *Graph Containers* paper (P3131).
- If you want to **write new algorithms**, you should read the *Views* paper (P3129), *Graph Container Interface* paper (P3130), and *Graph Containers* paper (P3131). You'll also want to review existing implementations in the reference library for examples of how to write the algorithms.
- If you want to **use your own graph container**, you should read the *Graph Container Interface* paper (P3130) and *Graph Containers* paper (P3131).

2 Revision History

P3130r0

- Split from P1709r5. Added Getting Started section.
- Add default implementation for target_id(g,uv) when the graph type matches the pattern random_access_range<forward_range<integral>> or random_access_range<forward_range<tuple<integral,...>>>; vertex_id_t<G> also defaults to the integral type given.
- Revised concept definitions, adding sourced_targeted_edge and target_edge_range, and replaced summary table with code for clarity. Also assured that all combinations of adjacency list concepts for basic, sourced and index exist.
- Move text for graph data structures created from std containers from Graph Container Interface to Container Implementation paper.

 $\S 2.0$

— Identify all concept definitions as "For exposition only" until we have consensus of whether they belong in the standard or not.

P3130r1

- Add num_edges(g) and has_edge(g) functions. Split function table into 3 tables for graph, vertex and edge functions because it was getting too big.
- Removed the Load Graph Data section with it's load functions from P3130 Graph Container Interface because it unnecessarily complicates the interface with constructors for graph data structures. To complement this, constructors have been added for compressed_graph in P3131 Graph Containers.
- Revised partition functions after implementation in compressed_graph to reflect usage, including: renaming partition_count(g) to num_partitions(g) to match other names used, changed partition_id(g,u) to partition_id(g,uid) because vertices may not exist when the function is called, and removing edges(g,u, pid) because it can easily be implemented as a filter using ranges functionality when target vertices can be in different partitions.

§2.0

3 Naming Conventions

Table 2 shows the naming conventions used throughout the Graph Library documents.

Template		Variable	
Parameter	Type Alias	Names	Description
G			Graph
	<pre>graph_reference_t<g></g></pre>	g	Graph reference
GV		val	Graph Value, value or reference
V	vertex_t <g></g>		Vertex
	vertex_reference_t <g></g>	u,v,x,y	Vertex reference. u is the source (or only)
			vertex. v is the target vertex.
VId	vertex_id_t <g></g>	uid, vid, seed	Vertex id. uid is the source (or only) vertex
			id. vid is the target vertex id.
VV	vertex_value_t <g></g>	val	Vertex Value, value or reference. This can be
			either the user-defined value on a vertex, or a
			value returned by a function object (e.g. VVF)
			that is related to the vertex.
VR	vertex_range_t <g></g>	ur, vr	Vertex Range
VI	<pre>vertex_iterator_t<g></g></pre>	ui,vi	Vertex Iterator. ui is the source (or only)
			vertex.
		first,last	vi is the target vertex.
VVF		vvf	Vertex Value Function: $vvf(u) \rightarrow vertex value$,
			or $vvf(uid) \rightarrow vertex value$, depending on re-
			quirements of the consume algorithm or view.
VProj		vproj	Vertex descriptor projection function: vproj(x
) vertex_descriptor <vid,vv>.</vid,vv>
	partition_id_t <g></g>	pid	Partition id.
		P	Number of partitions.
PVR	partition_vertex_range_t <g></g>	pur,pvr	Partition vertex range.
E	edge_t <g></g>		Edge
	edge_reference_t <g></g>	uv,vw	Edge reference. uv is an edge from vertices u
			to v . vw is an edge from vertices v to w .
EId	edge_id_t <g></g>	eid,uvid	Edge id, a pair of vertex_ids.
EV	edge_value_t <g></g>	val	Edge Value, value or reference. This can be
			either the user-defined value on an edge, or a
			value returned by a function object (e.g. EVF)
77			that is related to the edge.
ER	vertex_edge_range_t <g></g>		Edge Range for edges of a vertex
EI	vertex_edge_iterator_t <g></g>	uvi,vwi	Edge Iterator for an edge of a vertex. uvi is
			an iterator for an edge from vertices u to v.
			vwi is an iterator for an edge from vertices v
PWP		6	to w .
EVF		evf	Edge Value Function: $evf(uv) \rightarrow edge value$,
			or $evf(eid) \rightarrow edge$ value, depending on the
			requirements of the consuming algorithm or
ED			view.
EProj		eproj	Edge descriptor projection function: eproj(x)
			$ ightarrow$ edge_descriptor <vid,sourced,ev> .</vid,sourced,ev>

Table 2: Naming Conventions for Types and Variables

§3.0 4

4 Graph Container Interface

The Graph Container Interface defines the primitive concepts, traits, types and functions used to define and access an adacency graph, no matter its internal design and organization. Thus, it is designed to reflect all forms of adjacency graphs including a vector of lists, CSR-based graph and adjacency matrix, whether they are in the standard or external to the standard.

All algorithms in this proposal require that vertices are stored in random access containers and that vertex_id_t<G>
is integral, and it is assumed that all future algorithm proposals will also have the same requirements.

The Graph Container Interface is designed to support a wider scope of graph containers than required by the views and algorithms in this proposal. This enables for future growth of the graph data model (e.g. incoming edges on a vertex), or as a framework for graph implementations outside of the standard. For instance, existing implementations may have requirements that cause them to define features with looser constraints, such as sparse vertex_ids, non-integral vertex_ids, or storing vertices in associative bi-directional containers (e.g. std::map or std::unordered_map). Such features require specialized implementations for views and algorithms. The performance for such algorithms will be sub-optimal, but is preferrable to run them on the existing container rather than loading the graph into a high-performance graph container and then running the algorithm on it, where the loading time can far outweigh the time to run the sub-optimal algorithm. To achieve this, care has been taken to make sure that the use of concepts chosen is appropriate for algorithm, view and container.

5 Core Concepts

This section describes the core concepts to describe the adjacency lists used for graphs in the Graph Library. There are a number of qualifiers concept names that are are used.

- **basic** where there is only a vertex id but no vertex reference.
- **index** where the vertex range is random-access and the vertex id is integral.
- **sourced** where an edge has a source id.

While we belive the use of concepts is appropriate for graphs as a range-of-ranges, we are marking them as "For exposition only" until we have consensus of whether they belong in the standard or not.

5.1 Edge Concepts

The types of edges that can occur in a graph are described with the edges concepts. The E edge template parameter allows for different types of edges, such as incoming and outgoing edge types.

§5.1 5

There are two edge range concepts.

 $[Phil: edges(g,u) isn't being used anywhere. We could use basic_adjacency_list in the View and algorithm definitions.]$

5.2 Vertex Concepts

The vertex_range concept is the general definition used for adjacency lists while index_vertex_range is used for high performance graphs where vertices typically stored in a vector.

5.3 Adjacency List Concepts

The basic adjacency lists add concepts where there is no vertex object, only a vertex id.

§5.3

The adjacency list concepts bring together the vertex and edge concepts used for core graph concepts. All algorithms initially proposed for the Graph Library use the index_adjacency_list .

```
// For exposition only
template <class G>
concept adjacency_list = vertex_range<G> && //
                      targeted_edge_range<G> && //
                      targeted_edge<G, edge_t<G>>;
template <class G>
concept index_adjacency_list = index_vertex_range<G> && //
                            targeted_edge_range<G> && //
                            targeted_edge<G, edge_t<G>>>;
template <class G>
concept sourced_adjacency_list = vertex_range<G> && //
                              targeted_edge_range<G> && //
                              sourced_targeted_edge<G, edge_t<G>>;
template <class G>
concept sourced_index_adjacency_list = index_vertex_range<G> && //
                                   targeted_edge_range<G> && //
                                   sourced_targeted_edge<G, edge_t<G>>;
```

6 Traits

Table 3 summarizes the type traits in the Graph Container Interface, allowing views and algorithms to query the graph's characteristics.

7 Types

Table 4 summarizes the type aliases in the Graph Container Interface. These are the types used to define the objects in a graph container, no matter its internal design and organization. Thus, it is designed to be able to reflect all forms of adjacency graphs including a vector of lists, compressed graph and adjacency matrix.

The type aliases are defined by either a function specialization for the underlying graph container, or a refinement of one of those types (e.g. an iterator of a range). Table 5 describes the functions in more detail.

graph_value(g), vertex_value(g,u) and edge_value(g,uv) can be optionally implemented, depending on whether the graph container supports values on the graph, vertex and edge types.

There is no contiguous requirement for vertex_id from one partition to the next, though in practice they will often be assigned contiguously. Gaps in vertex_id s between partitions should be allowed.

§8.0 7

Trait	Type	Comment
has_degree <g></g>	concept	Is the degree(g,u) function available?
has_find_vertex <g></g>	concept	Are the find_vertex(g,_) functions available?
has_find_vertex_edge <g></g>	concept	Are the find_vertex_edge(g,_) functions available?
has_contains_edge <g></g>	concept	Is the contains_edge(g,uid,vid) function available?
<pre>define_unordered_edge<g,e> : false_type</g,e></pre>	struct	Specialize for edge implementation to derive from true_type for unordered edges
is_unordered_edge <g,e></g,e>	struct	<pre>conjunction<define_unordered_edge<e>, is_sourced_edge<g, e="">></g,></define_unordered_edge<e></pre>
is_unordered_edge_v <g,e></g,e>	type alias	
unordered_edge <g,e></g,e>	concept	
is_ordered_edge <g,e></g,e>	struct	negation <is_unordered_edge<g,e>></is_unordered_edge<g,e>
is_ordered_edge_v <g,e></g,e>	type alias	
ordered_edge <g,e></g,e>	concept	
<pre>define_adjacency_matrix<g> : false_type</g></pre>	struct	Specialize for graph implementation to derive from true_type for edges stored as a square 2-dimensional array
is_adjacency_matrix <g></g>	struct	
<pre>is_adjacency_matrix_v<g></g></pre>	type alias	
adjacency_matrix <g></g>	$\operatorname{concept}$	

Table 3: Graph Container Interface Type Traits

Type Alias	Definition	Comment	
<pre>graph_reference_t<g></g></pre>	add_lvalue_reference <g></g>		
<pre>graph_value_t<g></g></pre>	<pre>decltype(graph_value(g))</pre>	optional	
vertex_range_t <g></g>	<pre>decltype(vertices(g))</pre>		
<pre>vertex_iterator_t<g></g></pre>	iterator_t <vertex_range_t<g>></vertex_range_t<g>		
vertex_t <g></g>	range_value_t <vertex_range_t<g>></vertex_range_t<g>		
<pre>vertex_reference_t<g></g></pre>	<pre>range_reference_t<vertex_range_t<g>></vertex_range_t<g></pre>		
vertex_id_t <g></g>	<pre>decltype(vertex_id(g,ui))</pre>		
vertex_value_t <g></g>	<pre>decltype(vertex_value(g,u))</pre>	optional	
vertex_edge_range_t <g></g>	<pre>decltype(edges(g,u))</pre>		
<pre>vertex_edge_iterator_t<g></g></pre>	<pre>iterator_t<vertex_edge_range_t<g>></vertex_edge_range_t<g></pre>		
edge_t <g></g>	<pre>range_value_t<vertex_edge_range_t<g>>></vertex_edge_range_t<g></pre>		
edge_reference_t <g></g>	<pre>range_reference_t<vertex_edge_range_t<g>></vertex_edge_range_t<g></pre>		
edge_value_t <g></g>	<pre>decltype(edge_value(g,uv))</pre>	optional	
The following is only available when the optional source_id(g,uv) is defined for the edge			
edge_id_t <g></g>	edge_descriptor <vertex_id_t<g>,true,void,void></vertex_id_t<g>		
partition_id_t <g></g>	<pre>decltype(partition_id(g,uid))</pre>	optional	
_partition_vertex_range_t <g></g>	vertices(g,pid)	optional	

Table 4: Graph Container Interface Type Aliases

8 Classes and Structs

The <code>graph_error</code> exception class is available, inherited from <code>runtime_error</code>. While any function may use it, it is only anticipated to be used by the <code>load</code> functions at this time. No additional functionality is added beyond that

§8.0

provided by runtime_error.

While we belive the use of concepts is appropriate for graphs as a range-of-ranges, we are marking them as "For exposition only" until we have consensus of whether they belong in the standard or not.

9 Functions

Tables 5, 6 and 7 summarize the primitive functions in the Graph Container Interface. used to access an adacency graph, no matter its internal design and organization. Thus, it is designed to be able to reflect all forms of adjacency graphs including a vector of lists, CSR-based graph and adjacency matrix.

Function	Return Type	Complexity	Default Implementation
graph_value(g)	graph_value_t <g></g>	constant	n/a, optional
vertices(g)	vertex_range_t <g></g>	constant	<pre>g if random_access_range<g> , n/a otherwise</g></pre>
<pre>num_vertices(g)</pre>	integral	constant	<pre>size(vertices(g))</pre>
<pre>num_edges(g)</pre>	integral	$ \mathrm{E} $	<pre>n=0; for(u: vertices(g))n+=distance(</pre>
has_edge(g)	bool	V	<pre>edges(g,u)); return n; for(u: vertices(g))if !empty(edges(g, u))return true; return false;</pre>
num_partitions(g)	integral	constant	1
<pre>vertices(g,pid)</pre>	<pre>partition_vertex_range_t<g></g></pre>	constant	<pre>vertices(g)</pre>
<pre>num_vertices(g,pid)</pre>	integral	constant	<pre>size(vertices(g))</pre>

Table 5: Graph Functions

The complexity shown above for num_edges(g) and has_edge(g) is for the default implementation. Specific graph implementations may have better characteristics.

The complexity shown above for vertices(g,pid) and num_vertices(g,pid) is for the default implementation. Specific graph implementations may have different characteristics.

Function	Return Type	ComplexityDefault Implementation		
<pre>find_vertex(g,uid)</pre>	vertex_iterator_t <g></g>	constant	begin(vertices(g))+ uid	
			<pre>if random_access_range<vertex_range_t<g>>></vertex_range_t<g></pre>	
<pre>vertex_id(g,ui)</pre>	vetex_id_t <g></g>	constant	<pre>(size_t)(ui - begin(vertices(g)))</pre>	
			Override to define a different <pre>vertex_id_t<g></g></pre>	
			type (e.g. int32_t).	
<pre>vertex_value(g,u)</pre>	vertex_value_t <g></g>	constant	n/a, optional	
<pre>vertex_value(g,uid)</pre>	vertex_value_t <g></g>	constant	<pre>vertex_value(g,*find_vertex(g,uid)) ,</pre>	
			optional	
degree(g,u)	integral	constant	size(edges(g,u)) if	
			<pre>sized_range<vertex_edge_range_t<g>></vertex_edge_range_t<g></pre>	
<pre>degree(g,uid)</pre>	integral	constant	<pre>size(edges(g,uid)) if</pre>	
			<pre>sized_range<vertex_edge_range_t<g>>></vertex_edge_range_t<g></pre>	
edges(g,u)	<pre>vertex_edge_range_t<g></g></pre>	constant	u if forward range <vertex_t<g>>> , n/a</vertex_t<g>	
			otherwise	
edges(g,uid)	<pre>vertex_edge_range_t<g></g></pre>	constant	edges(g,*find_vertex(g,uid))	
partition_id(g,uid)	partition_id_t <g></g>	constant	0	

Table 6: Vertex Functions

The default implementation for the degree functions assumes that vertex_edge_range_t<G> is a sized range to have constant complexity. If the underlying container has a non-linear size(R) function, the degree functions will also be non-linear. This is expected to be an uncommon case.

§9.0

Function	Return Type	ComplexityDefault Implementation		
target_id(g,uv)	vertex_id_t <g></g>	constant	(see below)	
<pre>target(g,uv)</pre>	vertex_t <g></g>	constant	*(begin(vertices(g))+ target_id(g, uv)) if	
			random_access_range <vertex_range_t<g>> &&</vertex_range_t<g>	
			<pre>integral<target_id(g,uv)></target_id(g,uv)></pre>	
edge_value(g,uv)	edge_value_t <g></g>	constant	uv if forward_range <vertex_t<g>>> , n/a</vertex_t<g>	
			otherwise, optional	
<pre>find_vertex_edge(g,u,vid)</pre>	vertex_edge_t <g></g>	linear	find(edges(g,u), [](uv)target_id(g,uv)==	
			<pre>vid;})</pre>	
<pre>find_vertex_edge(g,uid,vid)</pre>	vertex_edge_t <g></g>	linear	find_vertex_edge(g,*find_vertex(g,uid),vid	
)	
<pre>contains_edge(g,uid,vid)</pre>	bool	constant	uid < size(vertices(g))&& vid < size(
			<pre>vertices(g)) if is_adjacency_matrix_v<g> .</g></pre>	
		linear	<pre>find_vertex_edge(g,uid)!= end(edges(g,uid)</pre>	
) otherwise.	
The following are on	aly available when the	e optional so	ource_id(g,uv) is defined for the edge	
source_id(g,uv)	vertex_id_t <g></g>	constant	n/a, optional	
source(g,uv)	vertex_t <g></g>	constant	*(begin(vertices(g))+ source_id(g,uv)) if	
			random_access_range <vertex_range_t<g>>> &&</vertex_range_t<g>	
			<pre>integral<target_id(g,uv)></target_id(g,uv)></pre>	
edge_id(g,uv)	edge_id_t <g></g>	constant	edge_descriptor <vertex_id_t<g>,true,void,</vertex_id_t<g>	
			<pre>void>{source_id(g,uv),target_id(g,uv)}</pre>	

Table 7: Edge Functions

When the graph matches the pattern random_access_range<forward_range<integral>>> or random_access_range<forward_range<tuple<integral,...>>> , the default implementation for target_id(g,uv) will return the integral . Additionally, if the caller does not override vertex_id(g,ui) , the integral value will define the vertex_id_t<G> type.

Functions that have n/a for their Default Implementation must be defined by the author of a Graph Container implementation.

Value functions (graph_value(g) , vertex_value(g,u) and edge_value(g,uv)) can be optionally implemented, depending on whether the graph container supports values on the graph, vertex and edge types. They return a single value and can be scaler, struct, class, union, or tuple. These are abstract types used by the GVF, VVF and EVF function objects to retrieve values used by algorithms. As such it's valid to return the "enclosing" owning class (graph, vertex or edge), or some other embedded value in those objects.

find_vertex(g,uid) is constant complexity because all algorithms in this proposal require that vertex_range_t<G>
is a random access range.

If the concept requirements for the default implementation aren't met by the graph container the function will need to be overridden.

9.1 Determining the Vertex Id Type

To determine the type for vertex_id_t<G> the following steps are taken, in order, to determine its type.

- 1. Use the type returned by vertex_id(g,ui) when overridden for a graph.
- 2. When the graph matches the pattern random_access_range<forward_range<integral>> or random_access_range<forward_range<tuple<integral,...>>>, use the integral type specified.
- 3. Use size_t in all other cases.

vertex_id_t<G> is defined by the type returned by vertex_id(g) and it defaults to the difference_type of
the underlying container used for vertices (e.g int64_t for 64-bit systems). This is sufficient for all situations.

§9.1 10

However, there are often space and performance advantages if a smaller type is used, such as int32_t or even int16_t. It is recommended to consider overriding this function for optimal results, assuring that it is also large enough for the number of possible vertices and edges in the application. It will also need to be overridden if the implementation doesn't expose the vertices as a range.

10 Unipartite, Bipartite and Multipartite Graph Representation

num_partitions(g) returns the number of partitions, or partiteness, of the graph. It has a range of 1 to n, where 1 identifies a unipartite graph, 2 is a bipartite graph, and a value of 2 or more can be considered a multipartite graph.

If a graph data structure doesn't support partitions then it is unipartite with one partition and partite functions will reflect that. For instance, num_partitions(g) returns a value of 1, and vertices(g,0) (vertices in the first partition) will return a range that includes all vertices in the graph.

A partition identifies a type of a vertex, where the vertex value types are assumed to be uniform in each partition. This creates a dilemma because the existing vertex_value(g,u) returns a single type based template parameter for the vertex value type. Supporting multiple types can be addressed in different ways using C++ features. The key to remember is that the actual value used by algorithms is done by calling a function object that retrieves the value to be used. That function is specific to the graph data structure, using the partition to determine how to get the appropriate value.

- std::variant: The lambda returns the appropriate variant value based on the partition.
- Base class pointer: The lambda can call a member function to return the value based on the partition.
- void*: The lambda can cast the pointer to a concrete type based on the partition, and then return the appropriate value.

edges(g,uid,pid) and edges(g,ui,pid) filter the edges where the target is in the partition pid passed. This isn't needed for bipartite graphs.

11 Edgelists

An edgelist is a range of edge_descriptor values, used by some algorithms. Examples include the edgelist adjacency list view, or possibly a transform of an existing range that projects edge_descriptor values. There is no concrete data structure in this proposal like there is for an adjacency list.

12 Using Existing Graph Data Structures

Reasonable defaults have been defined for the GCI functions to minimize the amount of work needed to adapt an existing graph data structure to be used by the views and algorithms.

There are two cases supported. The first is for the use of standard containers to define the graph and the other is for a broader set of more complicated implementations.

This is described in more detail in the paper for Graph Library Container Implementations.

Acknowledgements

Phil Ratzloff's time was made possible by SAS Institute.

Portions of *Andrew Lumsdaine's* time was supported by NSF Award OAC-1716828 and by the Segmented Global Address Space (SGAS) LDRD under the Data Model Convergence (DMC) initiative at the U.S. Department of Energy's Pacific Northwest National Laboratory (PNNL). PNNL is operated by Battelle Memorial Institute under Contract DE-AC06-76RL01830.

§12.0 11

 $Michael\ Wong's$ work is made possible by Codeplay Software Ltd., ISOCPP Foundation, Khronos and the Standards Council of Canada.

Muhammad Osama's time was made possible by Advanced Micro Devices, Inc.

The authors thank the members of SG19 and SG14 study groups for their invaluable input.

§12.0 12