Algolab BGL Introduction

Andreas Bärtschi

October 12, 2016¹

¹Updated on October 14, 2016



Boost Graph Library

A generic C++ library of graph data structures and algorithms.

BGL docs – your new best friend:

http://www.boost.org/doc/libs/1_58_0/libs/graph/doc

Moodle: There's a brief copy & paste manual.

Algolab VM & General: There's a technical instructions page for all things Algolab.

BGL: A generic library

Genericity type	STL	BGL
Algorithm /	Decoupling of algorithms	Decoupling of graph algorithms
Data-Structure	and data-structures	and graph representations
Interoperability	Key ingredients: iterators	Vertex iterators, edge iterators,
		adjacency iterators
Parameterization	Element type	Vertex and edge property
	parameterization	multi-parametrization
		Associate <i>multiple</i> properties
		Accessible via property maps
Extensions	through function objects	through a <i>visitor object</i> , event points and methods depend on particular algorithm

BGL: Graph Representations / Data Structures

Structure	Representation	Advantages	Do
Graph classes	Adjacency list	Swiss army knife:	use this!
		Directed/undirected graphs,	
		allow/disallow parallel-edges,	
		efficient insertion, fast adja-	
		cency structure exploitation	
	Adjacency matrix	Dense graphs	use at your
Adaptors	Edge list	Simplicity	own risk!
	External adaptation	Convert existing graph struc-	Not covered
		tures (LEDA etc.) to BGL	in Algolab.

BGL: adjacency_list

Example without any vertex or edge properties:

Defines a *directed* Graph where the vertices are stored in a vector (VertexList vecS) and the outgoing edges in each vertex are stored in a vector (OutEdgeList vecS). (Also see *Useful stuff: Options for adjacency_list, page 40.*)

BGL: adjacency_list

Example with vertex property and multiple edge properties:

Interior properties are stored with the graph. Property Maps allow us to access the interior properties of the graph. Think of these as a map (object with operator []). Also see *Useful stuff: Interior property maps, pages 44–45*.

BGL: Graph Algorithms

Area	Topic	Details
Basics	Distances	Dijkstra shortest paths
		Prim minimum spanning tree
		Kruskal minimum spanning tree
	Components	Connected, biconnected &
		strongly connected components
	General Matchings	General unweighted matching
Flows	Maximum Flow	Graph setup (residual graph)
		Edmonds-Karp and Push-Relabel
	Disjoint paths	Vertex- / Edge-disjoint s-t paths
Advanced Flows	Minimum Cut	Maxflow-Mincut Theorem
	Bipartite Matchings	Vertex Cover & Independent Set
	Mincost Maxflow	Bipartite weighted matching & more

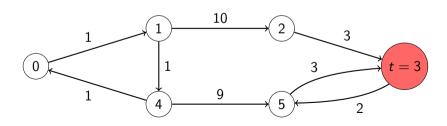
Many more (not in Algolab 2016): planarity testing, sparse matrix ordering, ... **Prerequisites**: Theory, BFS, DFS, topological sorting, Eulerian tours, Union-Find...

Tutorial problem: statement & example

Input A directed graph G with positive weights on edges and a vertex t, $|V(G)| \le 10^5$, $|E(G)| \le 2 \cdot 10^5$.

Definition We call a vertex *u universal* if all vertices in *G* can be reached from it.

Output Length of a shortest path $u \to t$ that starts in some universal vertex u. If such a path does not exist, output NO.

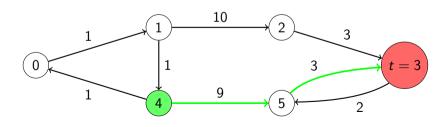


Tutorial problem: statement & example

Input A directed graph G with positive weights on edges and a vertex t, $|V(G)| \le 10^5$, $|E(G)| \le 2 \cdot 10^5$.

Definition We call a vertex *u universal* if all vertices in *G* can be reached from it.

Output Length of a shortest path $u \to t$ that starts in some universal vertex u. If such a path does not exist, output NO.



Tutorial problem: how to start?

Time's short, so hurry up!

- "Check if there is a unique u with no in-edges, if yes output shortest path $u \rightarrow t$." (what if there is no such u?)
- "For each *u* check with DFS if *u* reaches all vertices, then..." (**too slow**)
- Start coding:

```
1 #include <iostream>
2 int main() {
3     // some random algorithm
4 }
```

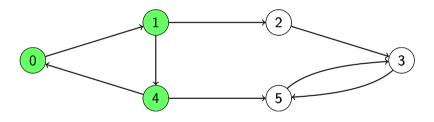
No! Take your time, model the problem, design the algorithm, understand why it should work, ⇒ then code.

Tutorial problem: how to start?

- Bad question: Why shouldn't it work? ("It is correct on all three examples I came up with", etc.)
- Good question: Why should it work? ("How would I prove it works?")
- Applies to Moodle forums as well!

Tutorial problem: example

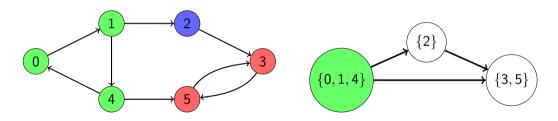
What are the universal vertices?



 \Rightarrow must be related to some sort of connected component concept in directed graphs!

Tutorial problem: strongly connected components (SCC) example

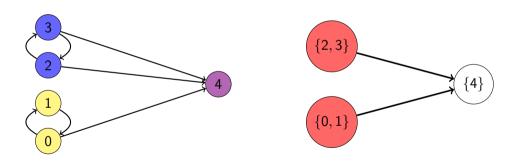
Strongly connected components:



Is there always a universal vertex?

Tutorial problem: strongly connected components "NO" example

No!



Tutorial problem: modeling

Let us call a strongly connected component with no in-edges in the SCC Directed Acyclic Graph a *minimal component*.

Fact

If there is more than one minimal component in G, then there is no universal u.

Lemma

If there is exactly one minimal component in G, then its vertices are exactly the universal vertices.

Tutorial problem: modeling

New formulation of the problem:

- II If there exists > 1 minimal strongly connected component in G, output NO.
- 2 Output the shortest distance $u \rightarrow t$ for best universal u in G.

Worst-case: Still $\Omega(|V| \cdot \text{Dijkstra's shortest paths}) = \Omega(|V|^2 \log |V| + |V||E|)!$ I.e. around $|V||E| \approx 10^5 \cdot 2 \cdot 10^5 = 20000000000$ operations.

Tutorial problem: modeling

New formulation of the problem:

- I If there exists > 1 maximal strongly connected component in G_T , output NO.
- 2 Output the shortest distance $t \to u$ for universal u in G_T .

Now we can work only with G_T and one single Dijkstra run! I.e. around $|V| \log |V| + |E| \approx 2 \cdot 10^5 = 200'000$ operations.

Tutorial problem: implementation

First and foremost, BGL docs:

- How to find the strong_components.
- How to check how many maximal components are there? topological_sort? Maybe there is a simple ad hoc?
- Compute shortest t u path on G_T with dijkstra_shortest_paths.

Tutorial problem: code – preamble

```
10 // STL includes
#include <iostream>
#include <vector>
13 #include <algorithm>
14 #include <climits>
15 // BGL includes
16 #include <boost/graph/adjacency_list.hpp>
17 #include <boost/graph/strong components.hpp>
18 #include <boost/graph/dijkstra_shortest_paths.hpp>
19 // Namespaces
20 using namespace std:
21 using namespace boost:
```

Tutorial problem: code – typedefs

Tutorial problem: code – reading the input

```
38 void testcases() {
      // Read and build graph
      int V, E, t; // 1st line: <vertex_no> <edge_no> <target>
40
      cin >> V >> E >> t:
41
     \mathsf{Graph}\ \mathsf{G}(\mathsf{V}); // Creates an empty graph on V vertices
42
     WeightMap weightmap = get(edge weight, G);
43
      for (int i = 0: i < E: ++i) {
44
          int u, v, w; // Each edge: <from> <to> <weight>
45
          cin >> u >> v >> w:
46
          Edge e; bool success; // Swapping u and v
47
          tie(e, success) = add_edge(v, u, G); // to create G_T!
48
         weightmap[e] = w;
49
50
```

Tutorial problem: code – strong components

Exterior property: strong_components assigns to each vertex the index of its strong component. This is a *property* of the vertex stored *outside* of the graph itself, namely in the vector sccmap. To access the vector, we turn it into an *exterior property map*.

Alternative: Define your own custom interior vertex property vertex_component_t. See Useful stuff: Custom properties, page 47. Then create an (interior) property map and call the algorithm with this map (hence without make_iterator_property_map).

Tutorial problem: code – maximal SCCs

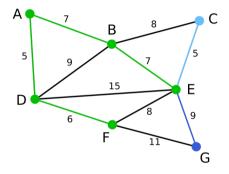
```
56 void testcases() {
57
      // Find universal strong component (if any)
      // Why does this approach work? Exercise.
59
      vector<int> is_max(nscc, 1);
60
      Edgelt ebeg, eend;
61
      // Iterate over all edges.
62
      for (tie(ebeg, eend) = edges(G); ebeg != eend; ++ebeg) {
63
          // ebeg is an iterator, *ebeg is a descriptor.
64
          Vertex u = source(*ebeg, G), v = target(*ebeg, G);
65
          if (sccmap[u] != sccmap[v]) is _max[sccmap[u]] = 0;
66
67
      int max_count = count(is_max.begin(), is_max.end(), true);
68
      if (max_count != 1) {
69
          cout << "NO" << endl;
70
          return:
71
72
```

Tutorial problem: code – Dijkstra

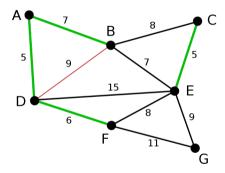
```
72 void testcases() {
73
     // Compute shortest t-u path in G T
74
     vector <int> distmap(V); // We must use at least one of these
75
     vector < Vertex > predmap(V); // vectors as an exterior property map.
76
     dijkstra shortest paths(G, t,
77
         78
                predmap.begin(), get(vertex_index, G))).
79
         distance map(make iterator property map( // concatenated by .
80
                distmap.begin(), get(vertex index, G))));
81
     int res = INT MAX:
82
     for (int u = 0; u < V; ++u)
83
         // Minimum of distances to 'maximal' universal vertices
84
         if (is_max[sccmap[u]])
85
             res = min(res, distmap[u]);
86
     cout << res << endl:
87
```

Tutorial problem: code – main

Minimum spanning trees



Intermediate step of Prim's algorithm to compute a Minimum Spanning Tree.



Intermediate step of Kruskal's algorithm to compute a Minimum Spanning Tree.

Minimum spanning tree algorithms

Algorithm	starts with	next	Time
Prim MST	Arbitrary start vertex	Adds connection (if possible) to the closest neighbour of all so far discovered vertices.	$O(E \log V)$
Kruskal	Edge of minimum weight	Adds next smallest edge, if this is possible without creating a cycle.	$O(E \log E)$

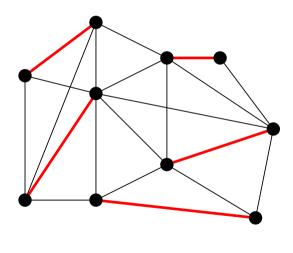
We need to provide a predecessor vector (as an exterior property map) to Prim (maps nodes to their parents in MST), and an edge vector (to store MST edges) to Kruskal.

Minimum spanning tree implementations

Prim's algorithm

Kruskal's algorithm

Maximum matching: general unweighted version



- G = (V, E)
- $M \subseteq E$ is a matching if and only if no two edges of M are adjacent.
- In an unweighted graph, a maximum matching is a matching of maximum cardinality.
- In a weighted graph, a maximum matching is a matching such that the weight sum over the included edges is maximum.
- BGL does not provide weighted matching algorithms.

Maximum matching: invoking algorithm

```
1 // Compute Matching
vector < Vertex > matemap(V);
3 // Use the vector as an Exterior Property Map: Vertex -> Matched mate
4 edmonds_maximum_cardinality_matching(G, make_iterator_property_map(
          matemap.begin(), get(vertex index, G)));
7 // Look at the matching
8 // Matching size
9 int matchingsize = matching size(G, make iterator property map(
          matemap.begin(), get(vertex index, G)));
10
12 // unmatched vertices get the NULL VERTEX as mate.
13 const Vertex NULL_VERTEX = graph_traits < Graph >:: null_vertex ();
14 for (int i = 0: i < V: ++i) {
      if (matemap[i] != NULL VERTEX && i < matemap[i]) {</pre>
16
           . . .
```

Getting started: BGL installation

- Pre-installed in ETH computer rooms and the Algolab Virtualbox Image.
 Most likely also already installed on your system if you installed CGAL last week.
- On "standard" Linux distributions try getting a package from the repository. On macOS package from Homebrew.
- Comments on the versions:
 - 1.58: This version is recommended (current Ubuntu LTS, Algolab VM).
 - 1.55+: These versions have Mincost-maxflow, should be fine.
 - 1.54: Prim MST bug (unless Ubuntu)

Getting started: BGL without installing

- BGL is a Header-only library.
- Download recent version from: http://www.boost.org/users/download/.
- Just unpack the .tar.bz2 file, no installation required, see Section 3 here: http://www.boost.org/doc/libs/1_58_0/more/getting_started/unix-variants.html.
- To build using this version of boost use this command: g++ -02 -std=c++11 -I path/to/boost_1_58_0 test.cpp -o test
- Explanation: The '-I' flag tells the compile to include all the files from this directory, so that it can find header files like 'boost/graph/adjacency_list.hpp'

Getting started: compilation problems

Error messages can be terrible.

- Consider re-compiling the code after every line after it is first written. This will help to identify the problem quickly.
- Especially after the typedefs, and again after building the graph, before you do anything else!
- There will be confusing typedefs, nested types, iterators etc. Come up with a naming pattern and stick to it.

Getting started: runtime problems

- Isolate the smallest possible example where the program misbehaves.
- Watch out for invalidated iterators.
- Print a graph and see if it looks as expected. In particular, check if the number of vertices didn't increase due to mistakes in your edge insertion.
- More on the slides of the next (and also of the last) Section of today.

Getting started: Using the forums

Some post	Good post
I tried to solve this question as	My code to Problem xy gets a timelimit on the
mentioned in the lecture slides,	last test set and I don't know why. My approach
I got timelimit, I did not yet ap-	was the following:
ply the	Spoiler »
Spoiler» sort to make it fast «,	«
but in the slides it is mentioned	I can argue that my solution is correct, because
that without	Spoiler »
Spoiler» sort «, it is still fast	«
enough, I will be grateful if	The overall running time of my solution is
you could mention the prob-	Spoiler»
lem with my code that makes	«
it slow, thanks	
	Attached you can find my (reasonably com-
	mented) submission.

Getting started: Problem of the week

As usual, on Monday. Don't miss it! Be advised it doesn't have to be BGL. Anything already covered in the course can be used.

Conclusion



BGL THE BOOST GRAPH LIBRARY

REDUCE GIVEN SETTING TO A WELL- KNOWN GRAPH PROBLEM				
path	connected components	minimum spanning tree	maximum matching	
read setup suitable invoke specifications structures BGL				
USE BGL TO QUICKLY SOLVE THIS CORE PROBLEM				

Useful stuff: Algolab BGL documentation

For more information please have a look at the following provided files:

Tutorial slides A PDF of today's tutorial. Homework: Section Useful stuff.

Copy & paste A PDF manual containing code snippets and some detailed explanations of the concepts presented in all BGL tutorials.

Tutorial problem Code and Input file of today's tutorial problem.

Code snippets Self contained code demonstrating many useful code snippets. Some of it can also be found in the rest of this Section.

Useful stuff: Options for adjacency_list

adjacency_list is the class you almost always need.

```
2 typedef adjacency_list < vecS, vecS, undirectedS.</pre>
     no_property,
                               // nested vertex properties
property < edge_weight_t, int> // nested edge properties
                                          Graph:
     >
 OutEdgeList (1st vecS) — for each vertex, adjacency list kept in a vector.
              Choosing setS instead disallows parallel edges.
   VertexList (2nd vecS) — a list of all edges is kept in a vector. Use this!
   Directivity directed — directed graph.
             Other choices: undirectedS (undirected graph).
              Rarely needed: bidirectionalS (efficient access to incoming edges)
```

1 // Graph Type, OutEdgeList Type, VertexList Type, (un)directedS

Useful stuff: Building a graph

```
1 Graph G(n);  // Constructs empty graph with n vertices
2 ...
3 Edge e;
4 bool success;
5 tie(e, success) = add_edge(u, v, G);
```

- Adds edge from u to v in G.
- Caveat: if u or v don't exist in the graph, G is automatically extended.
- Returns an (Edge, bool) pair. First coordinate is an edge descriptor. If parallel edges are allowed, second coordinate is always true. Otherwise it is false in case of a failure (when the edge is a duplicate).

Useful stuff: Removing vertices and edges, Clearing a graph

Dangerous: Deletions of single vertices and edges.

Takes time, invalidates descriptors and iterators, might behave counterintuitively. Consult the docs. Not recommended.

```
remove_edge(u, v, G);
remove_edge(e, G);
clear_vertex(u, G);
clear_out_edges(u, G);
remove_vertex(u, G);
```

OK: Clearing a graph once it is no longer needed.

```
1 G.clear(); // Removes all edges and vertices.
2 G=Graph(n); // Destroys old graph; creates a new one with n vertices.
```

Useful stuff: Iterators

- edges(G) returns a pair of iterators which define a range of all edges.
- For undirected graphs each edge is visited once, with some orientation.

```
10 // Iterating over outgoing edges
11 OutEdgelt oeit, oeend;
12 for (tie(oeit, oeend) = out_edges(u, G); oeit != oeend; ++oeit) {
13     Vertex v = target(*oeit, G);
14     ...
```

■ source(*eit, G) is guaranteed to be u, even in an undirected graph.

Useful stuff: Interior property maps – vertices

Think of a **property map** as a map (i.e., object with operator []) indexed by vertices or edges. Property maps of vertices could be simulated with a vector, but maps of edges are very convenient.

- namemap is just a handle (pointer), copying it costs O(1).
- vertex_name_t is a predefined tag. It is purely conventional (you can create property<vertex_name_t, int> and store distances), but algorithms use them as default choices if not instructed otherwise.

Useful stuff: Interior property maps – edges

```
typedef adjacency_list<vecS, vecS, directedS,
no_property, // No vertex properties this time.
// Edge properties as fifth template argument.
property<edge_weight_t, int>> Graph;
typedef property_map<Graph, edge_weight_t>::type WeightMap;
...
WeightMap weightmap = get(edge_weight, G);
weightmap[e] = cost;
```

weightmap is used by many algorithms (Prim, Dijkstra, Kruskal, ...) as default choice for the edge weight.

Useful stuff: Predefined properties

Some *predefined* vertex and edge properties:

- vertex_name_t
- vertex_distance_t
- vertex_color_t
- vertex_degree_t
- edge_name_t
- edge_weight_t
- edge_weight2_t

Do not be misled into, e.g., thinking that vertex_degree_t will automatically keep track of the degree for you.

More: http://www.boost.org/doc/libs/1_58_0/boost/graph/properties.hpp

Useful stuff: Custom properties

Can be defined if you want to keep additional info associated with edges.

```
1 namespace boost {
    enum edge_info_t { edge_info = 219 }; // A unique ID.
     BOOST INSTALL PROPERTY(edge, info);
5 struct EdgeInfo {
  . . . .
7 };
9 typedef adjacency list < vecS, vecS, directedS,</pre>
      no property.
10
      property < edge info t, EdgeInfo > > Graph;
12 typedef property_map<Graph, edge_info_t >::type InfoMap;
13 . . .
14 InfoMap infomap = get(edge info, G);
infomap[e] = \dots
```

Useful stuff: Named parameters I

Using named parameters is a way to pass parameters (usually property maps) to functions (BGL algorithms) which is useful in two cases:

■ Many algorithms have a long list of parameters. Without named parameters, all of these must be provided in the correct order, even if only some are actually needed:

For e.g. Dijkstra calling the non-named parameter version is even worse!

Useful stuff: Named parameters II

Using named parameters is a way to pass parameters (usually property maps) to functions (BGL algorithms) which is useful in two cases:

Some algorithms can record additional information to exterior property maps if provided by named parameters.

Always concatenate named parameters by a . Do not pass them as separate parameters (i.e. separated by a .).

Useful stuff: Where to be careful

Be careful when you deviate from the provided instructions, in particular if. . .

...you use a pointer type as a property map (see e.g. here): Buggy for Dijkstra calls in combination with Strong components header.

- ...you use bundled properties instead of nested properties: Not well documented in the BGL examples; Buggy for MinCost flows.
- 3 ...you use named parameters for flow algorithms: Buggy for MinCost flows. Stick to the non-named versions, provide all property maps in correct order.