RBGL: R interface to boost graph library

VJ Carey `stvjc@channing.harvard.edu`

July 13, 2003

Summary. A very preliminary implementation of an interface from R to the Boost Graph Library (BGL, an alternative to STL programming for mathematical graph objects) is presented. *This 2003 update employs the* graph *class of Bioconductor.*

Contents

1 Working with the Bioconductor `graph` class

An example object representing file dependencies is included, as shown in Figure 1.

```
> library(RBGL)

Loading required package: graph

Attaching package 'RBGL':


        The following object(s) are masked _by_ package:graph :

         dfs
```

```
> data(FileDep)
> print(FileDep)

A graph with  directed  edges
Number of Nodes = 15
Number of Edges = 19

> require(Rgraphviz) || stop("This vignette requires Rgraphviz to be installed")

Loading required package: Rgraphviz
Creating a new generic function for "lines"
Creating a new generic function for "plot"
[1] TRUE
```

2 Algorithms supported by RBGL

2.1 Topological sort

The `tsort` function will return the indices of vertices in topological sort order:

```
> ts <- tsort(FileDep)
> print(nodes(FileDep)[ts + 1])

 [1] "zow_h"       "boz_h"       "zig_cpp"     "zig_o"       "dax_h"
 [6] "yow_h"       "zag_cpp"     "zag_o"       "bar_cpp"     "bar_o"
[11] "foo_cpp"     "foo_o"       "libfoobar_a" "libzigzag_a" "killerapp"
```

Note that if the input graph is not a DAG, BGL `topological_sort` will check this and throw 'not a dag'. This is crudely captured in the interface (a message is written to the console and zeroes are returned).

```
#FD2 <- FileDep
# now introduce a cycle
#FD2@edgeL[["bar_cpp"]]$edges <- c(8,1)
#tsort(FD2)
```

2.2 Kruskal's minimum spanning tree

Function `mstree.kruskal` just returns a list of edges, weights and nodes determining the minimum spanning tree (MST) by Kruskal's algorithm.

```
> km <- fromGXL(file(system.file("GXL/kmstEx.gxl", package = "graph")))
> print(mstree.kruskal(km))
```

```
> z <- plot(FileDep)
```

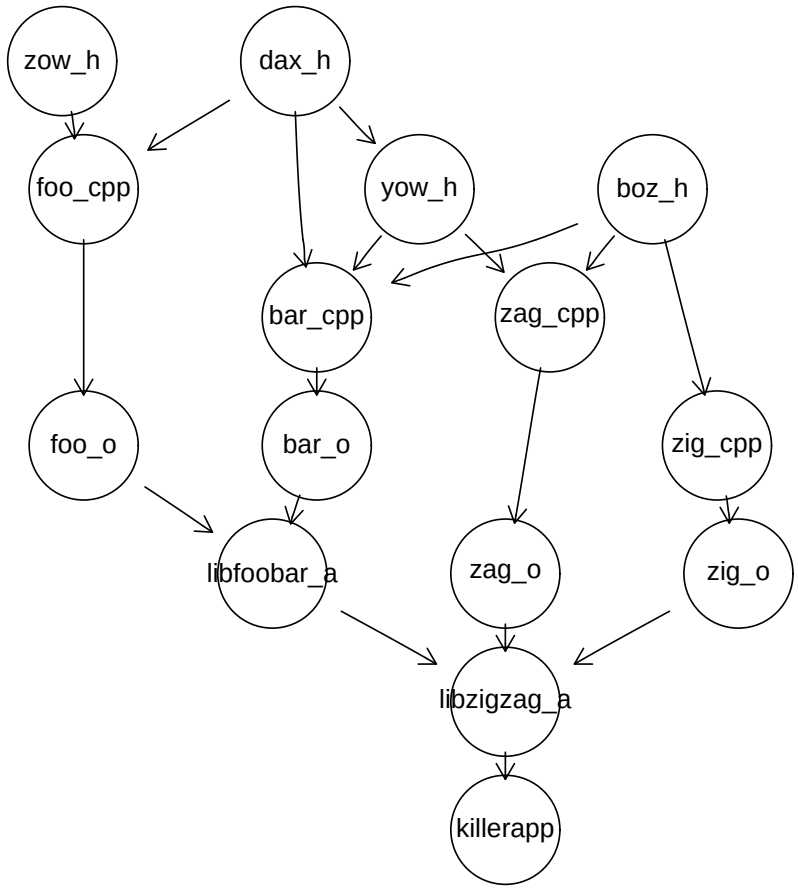


Figure 1: File dependency digraph example from Boost library.

```
warning: directed graph supplied; directions ignored.
$edgeList
     [,1] [,2] [,3] [,4]
[1,]    1    4    5    2
[2,]    3    5    1    4

$weights
     [,1] [,2] [,3] [,4]
[1,]    1    1    1    1

$nodes
[1] "A" "B" "C" "D" "E"
```

2.3 Depth first search

The `dfs` function returns a list of node indices by discovery and finish order.

```
> df <- fromGXL(file(system.file("XML/dfsex.gxl", package = "RBGL")))
> print(o <- dfs(df))

u v w x y z
1 1 2 1 1 2
```

Here is the list of nodes in DFS discovery order.

```
> print(nodes(df)[o$discovered])

character(0)
```

2.4 Breadth first search

The `bfs` function returns a vector of node indices for a breadth-first search (BFS) starting at the node indexed by `init.ind`.

```
> bf <- fromGXL(file(system.file("XML/bfsex.gxl", package = "RBGL")))
> bf@edgemode <- "undirected"
> print(o <- bfs(bf, init.ind = 2))

[1] 2 6 1 3 7 5 4 8

> z <- plot(bf)
```

```
> z <- plot(km)
```

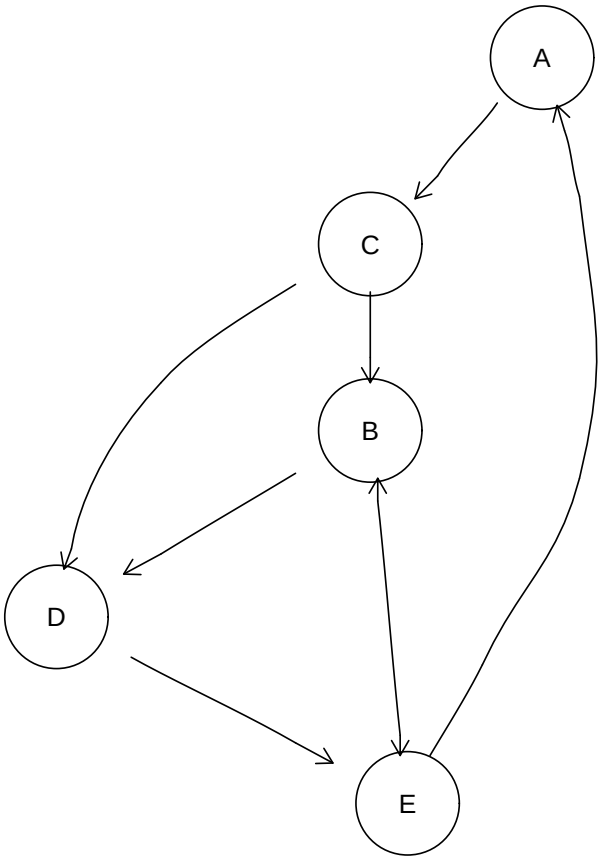


Figure 2: Kruskal MST example from Boost library.

```
> z <- plot(df)
```

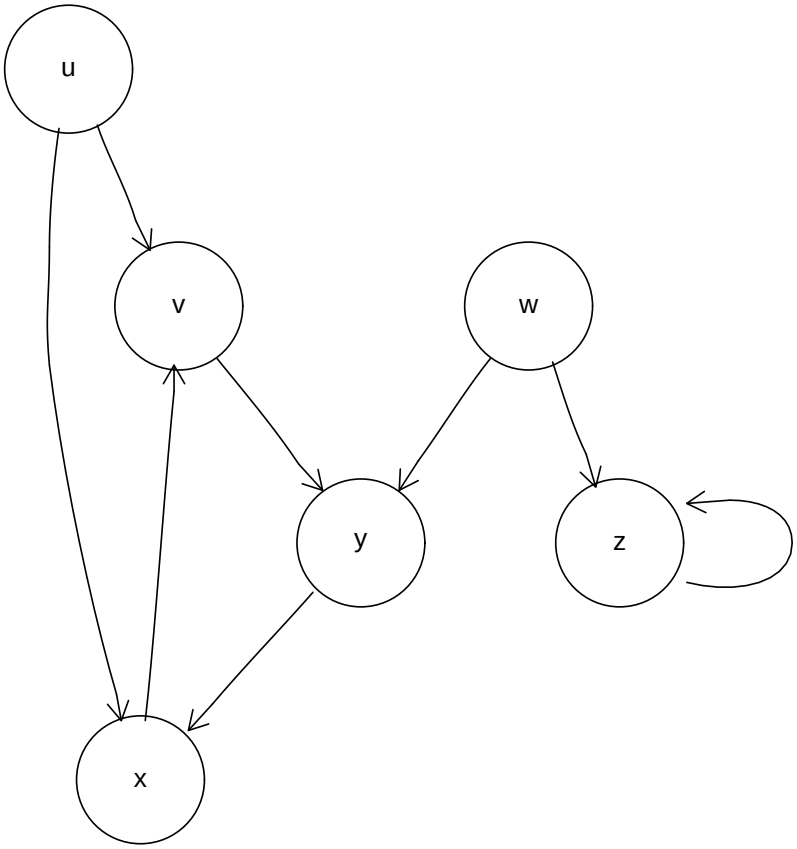
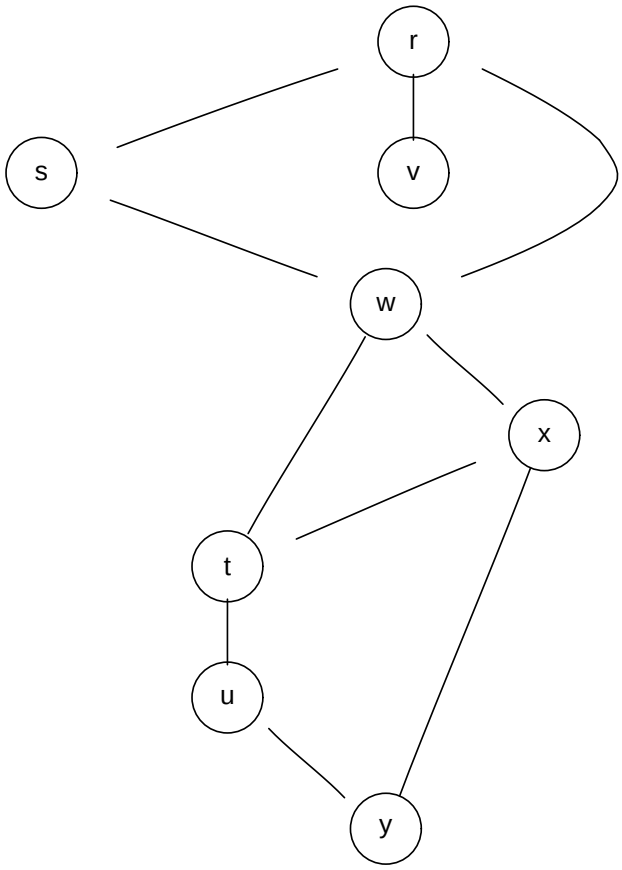


Figure 3: DFS example from Boost library.



The nodes in BFS order starting with the second node are

```
> print(nodes(bf)[o])
```

```
[1] "s" "w" "r" "t" "x" "v" "u" "y"
```

2.5 Dijkstra's shortest paths

```
> dd <- fromGXL(file(system.file("XML/dijkex.gxl", package = "RBGL")))
> print(dijkstra.sp(dd))
```

```
$distances
[1] 0 6 1 4 5

$penult
[1] 1 5 1 3 4

$start
[1] 1
```