Assignment 4
The Circumnavigations of Denver Long

Prof. Darrell Long
CSE 13S – Spring 2021

Due: May 2nd at 11:59 pm

1 Introduction

I wonder why it is that when I plan a route too carefully, it goes to pieces, whereas if I blunder along in blissful ignorance aimed in a fancied direction I get through with no trouble.

—John Steinbeck, *Travels with Charley: In Search of Amerca*

Denver Long decides to augment his income during his retirement years by selling the fine products produced by the Shinola Corporation. He enjoys driving his Lincoln, and so it's the life of a traveling salesman for him. Having accidentally taken a wrong turn near Chula Vista and winding up in Mexico, he asks his son to have his class create a computer program that will provide an optimal route to all the cities along his route and then return him to his home in scenic Clearlake.

2 Directed Graphs

A graph is a data structure $G = \langle V, E \rangle$ where $V = \{v_0, \ldots, v_n\}$ is the set of vertices (or nodes) and $E = \{\langle v_i, v_j \rangle, \ldots\}$ is the set of edges that connect the vertices. For example, you might have a set of cities $V = \{\text{El Cajon}, \text{La Mesa}, \text{San Diego}, \ldots, \text{La Jolla}\}$ as the vertices and write "El Cajon" → "Lakeside" to indicate that there is a path (Los Coches Road) from El Cajon to Lakeside. If there is a path from El Cajon to Lakeside, as well as a path from Lakeside to El Cajon, then the edge connecting El Cajon and Lakeside is *undirected.*

Such a simple graph representation simply tells you how vertices are connected and provides the idea of one-way roads. But it really does not help Denver, since it does not provide any notion of distance. We solve this problem by associating a weight with each edge and so we might write "Santee → El Cajon, 2" to indicate that there is a path of two miles long from Santee to El Cajon. Given a set of edges and weights, then we can then find the shortest path from any vertex to any other vertex (the answer may be that there is no such path). There are elegant (and quick) algorithms for computing the shortest path, but that is not exactly what we want to do. We want to find a path through all of the vertices, visiting each *exactly once*, such that there is a direct (single step) connection from the last vertex to the first. This is called a *Hamiltonian path*. This will address Denver's need to get home, but won't necessarily be the shortest such path. So, we need to go through every possible Hamiltonian path given the list of cities Denver is traveling through and pick the shortest path. Note: if you find yourself on a path that is longer than the current best found Hamiltonian path, then you can preemptively reject your current path.

3 Representing Graphs

Nothing can be more limiting to the imagination than only writing about what you know.

—John W. Gardner

Perhaps the simplest way to represent a graph is with an *adjacency matrix*. Consider an $n \times n$ adjacency matrix M, where n is the number of vertices in the graph. If $M_{i,j} = k$, where $1 \le i \le j \le n$, then we say that there exists a *directed edge* from vertex i to vertex j with weight k. Traveling through wormholes is considered hazardous, so any valid edge weight k must be non-zero and positive.

$$\begin{array}{c|cccccccc|}
 & 0 & 1 & 2 & 3 & 4 & 5 & \cdots & 25 \\
0 & 0 & 10 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 2 & 5 & 0 & 0 & 0 & 0 \\
2 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 5 \\
3 & 0 & 0 & 0 & 0 & 21 & 0 & 0 & 0 \\
4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}$$

Each edge will be represented as a triplet $\langle i, j, k\rangle$. The set of edges in the adjacency matrix above is

$$E = \{\langle 0,1,10\rangle, \langle 1,2,2\rangle, \langle 1,3,5\rangle, \langle 2,5,3\rangle, \langle 2,25,5\rangle, \langle 3,4,21\rangle\}.$$

If the above adjacency matrix were made to be *undirected*, it would be reflected along the diagonal.

$$\begin{array}{c|cccccccc}
 & 0 & 1 & 2 & 3 & 4 & 5 & \cdots & 25 \\
\hline
0 & 0 & 10 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 10 & 0 & 2 & 5 & 0 & 0 & 0 & 0 \\
2 & 0 & 2 & 0 & 0 & 0 & 3 & 0 & 5 \\
3 & 0 & 5 & 0 & 0 & 21 & 0 & 0 & 0 \\
4 & 0 & 0 & 0 & 21 & 0 & 0 & 0 & 0 \\
5 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\
\vdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
25 & 0 & 0 & 5 & 0 & 0 & 0 & 0 & 0
\end{array}$$

We will create a graph ADT based on this `struct` definition:

```c
struct Graph {
  uint32_t vertices;                    // Number of vertices.
  bool undirected;                      // Undirected graph?
  bool visited[VERTICES];               // Where have we gone?
  uint32_t matrix[VERTICES][VERTICES];  // Adjacency matrix.
};
```

We elect to use an adjacency matrix with set maximum dimensions. This is both to simplify the abstraction and also due to the computational complexity of solving the Traveling Salesman Problem (TSP) with depth-first search (DFS), which is discussed in §5. The `VERTICES` macro will be defined and supplied to you in `vertices.h`. In this header file, there is another macro `START_VERTEX` which defines the origin vertex of the shortest Hamiltonian path we will be searching for. You may not modify this file. The `struct` definition of a graph *must* go in `graph.c`. The following subsections define the interface for the graph ADT.

vertices.h

```c
#ifndef __VERTICES_H__
#define __VERTICES_H__

#define START_VERTEX 0   // Starting (origin) vertex.
#define VERTICES     26  // Maximum vertices in graph.

#endif
```

3.1 `Graph *graph_create(uint32_t vertices, bool undirected)`

The constructor for a graph. It is through this constructor in which a graph can be specified to be undirected. Make sure each cell of the adjacency matrix, `matrix`, is set to zero. Also make sure that each index of the `visited` array is initialized as `false` to reflect that no vertex has been visited yet. The `vertices` field reflects the number of vertices in the graph.

3.2 `void graph_delete(Graph **G)`

The destructor for a graph. Remember to set the pointer `G` to `NULL`.

3.3 `uint32_t graph_vertices(Graph *G)`

Return the number of vertices in the graph.

3.4 `bool graph_add_edge(Graph *G, uint32_t i, uint32_t j, uint32_t k)`

Adds an edge of weight `k` from vertex `i` to vertex `j`. If the graph is undirected, add an edge, also with weight `k` from `j` to `i`. Return `true` if both vertices are within bounds and the edge(s) are successfully added and `false` otherwise.

3.5 `bool graph_has_edge(Graph *G, uint32_t i, uint32_t j)`

Return `true` if vertices `i` and `j` are within bounds and there exists an edge from `i` to `j`. Remember: an edge exists if it has a non-zero, positive weight. Return `false` otherwise.

3.6 `uint32_t graph_edge_weight(Graph *G, uint32_t i, uint32_t j)`

Return the weight of the edge from vertex `i` to vertex `j`. If either `i` or `j` aren't within bounds, or if an edge doesn't exist, return 0.

3.7 `bool graph_visited(Graph *G, uint32_t v)`

Return `true` if vertex `v` has been visited and `false` otherwise.

3.8 `void graph_mark_visited(Graph *G, uint32_t v)`

If vertex `v` is within bounds, mark `v` as visited.

3.9 `void graph_mark_unvisited(Graph *G, uint32_t v)`

If vertex `v` is within bounds, mark `v` as unvisited.

3.10 `void graph_print(Graph *G)`

A debug function you will want to write to make sure your graph ADT works as expected.

4 Depth-first Search

Again it might have been the American tendency in travel. One goes, not so much to see but to tell afterward.

—John Steinbeck, *Travels with Charley: In Search of Amerca*

We need a methodical procedure for searching through the graph. Once we have examined a vertex, we do not want to do so again—we don't want Denver going through cities where he has already been (he has been known to wear out his welcome: charming women and fighting men).

Depth-first search (DFS) first marks the vertex v as having been visited, then it iterates through all of the edges $\langle v, w \rangle$, recursively calling itself starting at w if w has not already been visited.

```
procedure DFS(G,v):
    label v as visited
    for all edges from v to w in G.adjacentEdges(v) do
       if vertex w is not labeled as visited then
          recursively call DFS(G,w)
    label v as unvisited
```

Finding a Hamiltonian path then reduces to:

1. Using DFS to find paths that pass through all vertices, and
2. There is an edge from the last vertex to the first. The solutions to the Traveling Salesman Problem are then the shortest found Hamiltonian paths.

5 Computational Complexity

Many a trip continues long after movement in time and space have ceased. I remember a man in Salinas who in his middle years traveled to Honolulu and back, and that journey continued for the rest of his life. We could watch him in his rocking chair on his front porch, his eyes squinted, half-closed, endlessly traveling to Honolulu.

—John Steinbeck, *Travels with Charley: In Search of Amerca*

How long will this take? The answer is, it will take a very long time if there are a large number of vertices. The running time of the simplest algorithm is $O(n!)$ and there is no known algorithm that runs in less than $O(2^n)$ time. In fact, the TSP has been shown to be NP-hard, which means that it is as difficult as any problem in the class NP (you will learn more about this in CSE 104: Computability and Computational Complexity). Basically, it means that it can be solved in polynomial time if you have a magical computer that at each if-statement is takes both branches every time (creating a copy of the computer for each such branch).

6 Representing Paths

We find after years of struggle that we do not take a trip; a trip takes us.

—John Steinbeck, *Travels with Charley: In Search of Amerca*

Given that vertices are added to and removed from the traveled path in a stack-like manner, we decide to abstract a path as follows:

```c
struct Path {
  Stack *vertices; // The vertices comprising the path.
  uint32_t length; // The total length of the path.
};
```

The following subsections define for the interface for the path ADT.

6.1 `Path *path_create(void)`

The constructor for a path. Set `vertices` as a freshly created stack that can hold up to `VERTICES` number of vertices. Initialize `length` to be 0. The `length` field will track the length of the path. In other words, it holds the sum of the edge weights between consecutive vertices in the `vertices` stack.

6.2 `void path_delete(Path **p)`

The destructor for a path. Remember to set the pointer `p` to `NULL`.

6.3 `bool path_push_vertex(Path *p, uint32_t v, Graph *G)`

Pushes vertex `v` onto path `p`. The `length` of the path is *increased* by the edge weight connecting the vertex at the top of the stack and `v`. Return `true` if the vertex was successfully pushed and `false` otherwise.

6.4 `bool path_pop_vertex(Path *p, uint32_t *v, Graph *G)`

Pops the `vertices` stack, passing the popped vertex back through the pointer `v`. The `length` of the path is *decreased* by the edge weight connecting the vertex at the top of the stack and the popped vertex. Return `true` if the vertex was successfully popped and `false` otherwise.

6.5 `uint32_t path_vertices(Path *p)`

Returns the number of vertices in the path.

6.6 `uint32_t path_length(Path *p)`

Returns the length of the path.

6.7 `void path_copy(Path *dst, Path *src)`

Assuming that the destination path `dst` is properly initialized, makes `dst` a copy of the source path `src`. This will require making a copy of the `vertices` stack as well as the `length` of the source path.

6.8 `void path_print(Path *p, FILE *outfile, char *cities[])`

Prints out a path to `outfile` using `fprintf()`. Requires a call to `stack_print()`, as defined in §7.10, in order to print out the contents of the `vertices` stack.

7 Stacks, Revisited

If you have some respect for people as they are, you can be more effective in helping them to become better than they are.

—John W. Gardner

You will use the stack that you implemented for assignment 3 with slight modifications. If there were any problems with your stack for that assignment, make sure to fix them for this assignment. Here is the modified stack interface for this assignment.

```c
struct Stack {
    uint32_t top;
    uint32_t capacity;
    uint32_t *items;
};
```

7.1 `Stack *stack_create(uint32_t capacity)`

The constructor function for a `Stack`. The `top` of a stack should be initialized to 0. The capacity of a stack is set to the specified capacity. The specified capacity also indicates the number of items to allocate memory for, the items in which are held in the dynamically allocated array `items`.

7.2 `void stack_delete(Stack **s)`

The destructor function for a stack. Remember to set the pointer `s` to `NULL`.

7.3 `bool stack_empty(Stack *s)`

Return `true` if the stack is empty and `false` otherwise.

7.4 `bool stack_full(Stack *s)`

Return `true` if the stack is full and `false` otherwise.

7.5 `uint32_t stack_size(Stack *s)`

Return the number of items in the stack.

7.6 `bool stack_push(Stack *s, uint32_t x)`

If the stack is full prior to pushing the item `x`, return `false` to indicate failure. Otherwise, push the item and return `true` to indicate success.

7.7 `bool stack_peek(Stack *s, uint32_t *x)`

Peeking into a stack is synonymous with querying a stack about the element at the top of the stack. If the stack is empty prior to peeking into it, return `false` to indicate failure.

7.8 `bool stack_pop(Stack *s, uint32_t *x)`

If the stack is empty prior to popping it, return `false` to indicate failure. Otherwise, pop the item, set the value in the memory `x` is pointing to as the popped item, and return `true` to indicate success.

7.9 `void stack_copy(Stack *dst, Stack *src)`

Assuming that the destination stack `dst` is properly initialized, make `dst` a copy of the source stack `src`. This means making the *contents* of `dst->items` the same as `src->items`. The top of `dst` should also match the top of `src`.

7.10 `void stack_print(Stack *s, FILE *outfile, char *cities[])`

Prints out the contents of the stack to `outfile` using `fprintf()`. Working through each vertex in the stack starting from the *bottom*, print out the name of the city each vertex corresponds to. This function will be given to aid you.

```c
void stack_print(Stack *s, FILE *outfile, char *cities[]) {
    for (uint32_t i = 0; i < s->top; i += 1) {
        fprintf(outfile, "%s", cities[s->items[i]]);
        if (i + 1 != s->top) {
            fprintf(outfile, " -> ");
        }
    }
    fprintf(outfile, "\n");
}
```

8 Command-line Options

Attitude is a choice. Happiness is a choice. Optimism is a choice. Kindness is a choice. Giving is a choice. Respect is a choice. Whatever choice you make makes you. Choose wisely.

—Roy T. Bennett, *The Light in the Heart*

Your program must support any combination of the following command-line options.

- `-h`: Prints out a help message describing the purpose of the graph and the command-line options it accepts, exiting the program afterwards. Refer to the reference program in the resources repo for an idea of what to print.

- `-v`: Enables verbose printing. If enabled, the program prints out *all* Hamiltonian paths found as well as the total number of recursive calls to `dfs()`.
- `-u`: Specifies the graph to be undirected.
- `-i infile`: Specify the input file path containing the cities and edges of a graph. If not specified, the default input should be set as `stdin`.
- `-o outfile`: Specify the output file path to print to. If not specified, the default output should be set as `stdout`.

9 Reading An Input Graph

It behooves a man who wants to see wonders sometimes to go out of his way.

—John Mandeville, *The Travels of Sir John Mandeville*

We will be storing graphs in specially formatted files. Here is an example:

```
$ cat mythical.graph
4
Asgard
Elysium
Olympus
Shangri-La
0 3 5
3 2 4
2 1 10
1 0 2
```

The first line of a graph file is the number of vertices, or cities, in the graph. Assuming n is the number of vertices, the next n lines of the file are the names of the cities. Each line after that is an edge. It is to be scanned in as a triplet $\langle i, j, k \rangle$ and interpreted as an edge from vertex i to vertex j with weight k.

10 Specifics

The gladdest moment in human life, methinks, is a departure into unknown lands.

—Sir Richard Burton

Here are the specifics for your program implementation.

1. Parse command-line options with looped calls to `getopt()`. This should be familiar from assignments 2 and 3.

2. Scan in the first line from the input. This will be the number of vertices, or cities, that will be in the graph. Print an error if the number specified is greater than `VERTICES`, the macro defined in `vertices.h`.

3. Assuming the number of specified vertices is n, read the next n lines from the input using `fgets()`. Each line is the name of a city. Save the name of each city to an array. You will want to either make use of `strdup()` from `<string.h>` or implement your own `strdup()` function. If the line is malformed, print an error and exit the program. Note: using `fgets()` will leave in the newline character at the end, so you will manually have to change it to the null character to remove it.

4. Create a new graph G, making it undirected if specified.

5. Scan the input line by line using `fscanf()` until the end-of-file is reached. Add each edge to G. If the line is malformed, print an error and exit the program.

6. Create two paths. One will be for tracking the current traveled path and the other for tracking the shortest found path.

7. Starting from the origin vertex, defined by the macro `START_VERTEX` in `vertices.h`, perform a depth-first search on G to find the shortest Hamiltonian path. Here is an example function prototype that you may use as the recursive depth-first function:

```c
void dfs(Graph *G, uint32_t v, Path *curr, Path *shortest
    , char *cities[], FILE *outfile);
```

   The parameter `v` is the vertex that you are currently on. The currently traversed path is maintained with `curr`. The shortest found path is tracked with `shortest`. The array of city names is `cities`. Finally, `outfile` is the output to print to.

8. After the search, print out the length of the shortest path found, the path itself (remember to return back to the origin), and the number of calls to `dfs()`.

```
$ ./tsp < mythical.graph
Path length: 21
Path: Asgard -> Shangri-La -> Olympus -> Elysium -> Asgard
Total recursive calls: 4
```

   If the verbose command-line option was enabled, print out *all* the Hamiltonian paths that were found as well. It is recommended that you print out the paths as you find them.

```
$ ./tsp -v < ucsc.graph
Path length: 7
Path: Cowell -> Stevenson -> Merrill -> Cowell
Path length: 6
Path: Cowell -> Merrill -> Stevenson -> Cowell
Path length: 6
Path: Cowell -> Merrill -> Stevenson -> Cowell
Total recursive calls: 5
```

11 Deliverables

Travel isn't always pretty. It isn't always comfortable. Sometimes it hurts, it even breaks your heart. But that's okay. The journey changes you; it should change you. It leaves marks on your memory, on your consciousness, on your heart, and on your body. You take something with you. Hopefully, you leave something good behind.

—Anthony Bourdain

1. Your program, called `tsp`, *must* have the following source and header files:
   - `vertices.h` defines macros regarding vertices.
   - `graph.h` specifies the interface to the graph ADT.
   - `graph.c` implements the graph ADT.
   - `stack.h` specifies the interface to the stack ADT.
   - `stack.c` implements the stack ADT.
   - `path.h` specifies the interface to the path ADT.
   - `path.c` implements the path ADT.
   - `tsp.c` contains `main()` and *may* contain any other functions necessary to complete the assignment.

   You can have other source and header files, but *do not try to be overly clever*. **You may not modify any of the supplied header files.**
2. `Makefile`: This is a file that will allow the grader to type `make` or `make all` to compile your program.
   - `CC=clang` must be specified.
   - `CFLAGS=-Wall -Wextra -Werror -Wpedantic` must be included.
   - `make` should build your program, as should `make all`.
   - `make clean` must remove all files that are compiler generated.
   - `make format` should format all your source code, including the header files.
3. Your code must pass `scan-build` *cleanly*.
4. `README.md`: This *must* be in *Markdown*. This must describe your program briefly, its usage, and how to build it using your `Makefile`.
5. `DESIGN.pdf`: This *must* be a PDF. The design document should contain answers to the pre-lab questions at the beginning and describe your design for your program with enough detail that a sufficiently knowledgeable programmer would be able to replicate your implementation. This does not mean copying your entire program in verbatim. You should instead describe how your program works with supporting pseudo-code.

12 Supplemental Readings

One of the reasons people stop learning is that they become less and less willing to risk failure. you learn, the more places you'll go.

—John W. Gardner

- *The C Programming Language* by Kernighan & Ritchie
  - Chapter 4 §4.10
  - Chapter 7 §7.4–7.8
- *Introduction to Algorithms* by T. Cormen, C. Leiserson, R. Rivest, & C. Stein
  - Chapter 10 §10.1
  - Chapter 35 §35.2

13 Submission

A man who procrastinates in his choosing will inevitably have his choice made for him by circumstance.

—Hunter S. Thompson, *The Proud Highway*

Refer back to `asgn0` for the steps on how to submit your assignment through `git`. Remember: *add, commit,* and *push*! Your assignment is turned in *only* after you have pushed *and* submitted the commit ID on Canvas. If you forget to push, you have not turned in your assignment and you will get a *zero*. "I forgot to push" is not a valid excuse. It is *highly* recommended to commit and push your changes *often*.

`Fhewhqccydw yd S yi byau wylydw q cedauo q sxqydiqm.`