

Name: Michael Rogers

ID: 105667404

CSCI 3104, Algorithms  
Problem Set 5b (48 points)

Profs. Hoenigman & Agrawal  
Fall 2019, CU-Boulder

---

**Instructions for submitting your solution:**

- The solutions **should be typed** and we cannot accept hand-written solutions. Here's a short intro to Latex.
  - You should submit your work through **Gradescope** only.
  - If you don't have an account on it, sign up for one using your CU email. You should have gotten an email to sign up. If your name based CU email doesn't work, try the identikey@colorado.edu version.
  - Gradescope will only accept **.pdf** files (except for code files that should be submitted separately on Gradescope if a problem set has them) and **try to fit your work in the box provided**.
  - You cannot submit a pdf which has less pages than what we provided you as Gradescope won't allow it.
  - Verbal reasoning is typically insufficient for full credit. Instead, write a logical argument, in the style of a mathematical proof.
  - For every problem in this class, you must justify your answer: show how you arrived at it and why it is correct. If there are assumptions you need to make along the way, state those clearly.
  - You may work with other students. However, **all solutions must be written independently and in your own words**. Referencing solutions of any sort is strictly prohibited. You must explicitly cite any sources, as well as any collaborators.
-

Name: Michael Rogers

ID: 105667404

CSCI 3104, Algorithms  
Problem Set 5b (48 points)

Profs. Hoenigman & Agrawal  
Fall 2019, CU-Boulder

---

1. (25 pts) For this question, you are going to implement Kruskal's algorithm and union-find to build an MST from supplied data. Refer to the python starter code **MST\_Q1\_starter\_code.py** on Canvas that generates a graph of US cities, where the cities are the vertices and the edges are the distances between them. The code requires **miles\_dat.txt.gz** file as the graph data source so keep it in the same folder as the code. Before you start writing any code, make sure you can build the code that's been supplied. The code uses the networkx library. You may need to install this library for the code to run.

**Read all instructions for this question carefully.**

- (a) (5 pts) Complete the code to find the edges that are part of the MST. You should add these edges in the list *kruskal\_selected\_edges*. Do not change the existing format of the edges. They are represented as a tuple of vertices and a vertex is represented like  $v = \text{"Waukegan, IL"}$ . Read the comments in the code for more information. You don't need to read/understand the *miles\_graph()* and *draw\_graph()* functions.
- (b) (10 pts) Implement the *union()* function to implement Kruskal's.
- (c) (10 pts) Modify your code slightly so that you can produce disconnected components. Let's call these components clusters. The "spacing" of any particular clustering (group of clusters) is defined as the smallest edge between vertices in any pair of different clusters. If we stop Kruskal's  $k$  iterations before the algorithm completes, what is the spacing value? Run your code for  $k = 2 \dots 10$  to generate spacing for all these  $k$  values. Your code needs to have this calculation for your answer to receive credit.
- (d) In the pdf that you submit for this assignment, please include the following:
  - i. One of the generated graphs **MST.png** that your code produces that shows the MST for that run. Note that on each run, you can get a different number of edges to begin with. Thus, you can expect a different answer each time you run.
  - ii. The spacing values for each  $k$  value that you use.
  - iii. Your .py file for this question needs to be submitted to Canvas.

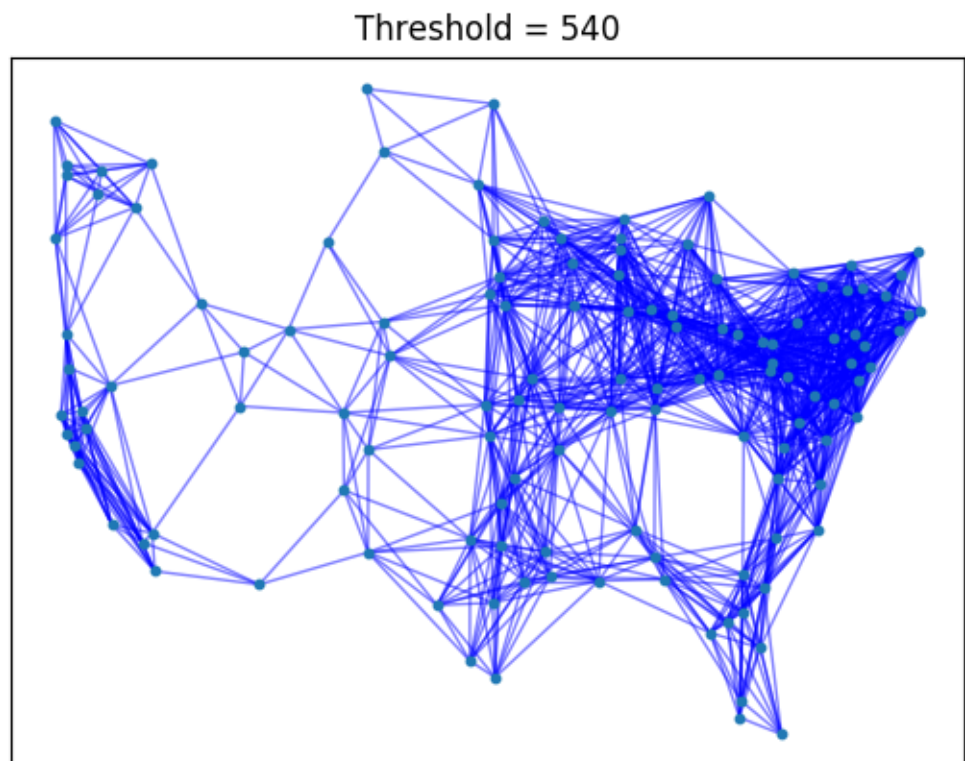
Name: Michael Rogers

ID: 105667404

CSCI 3104, Algorithms  
Problem Set 5b (48 points)

Profs. Hoenigman & Agrawal  
Fall 2019, CU-Boulder

(Space for Q1 image and spacing values)



I don't understand what the k spacing is. I attempted to implement the code,  
but it doesn't work.

Name: Michael Rogers

ID: 105667404

CSCI 3104, Algorithms  
Problem Set 5b (48 points)

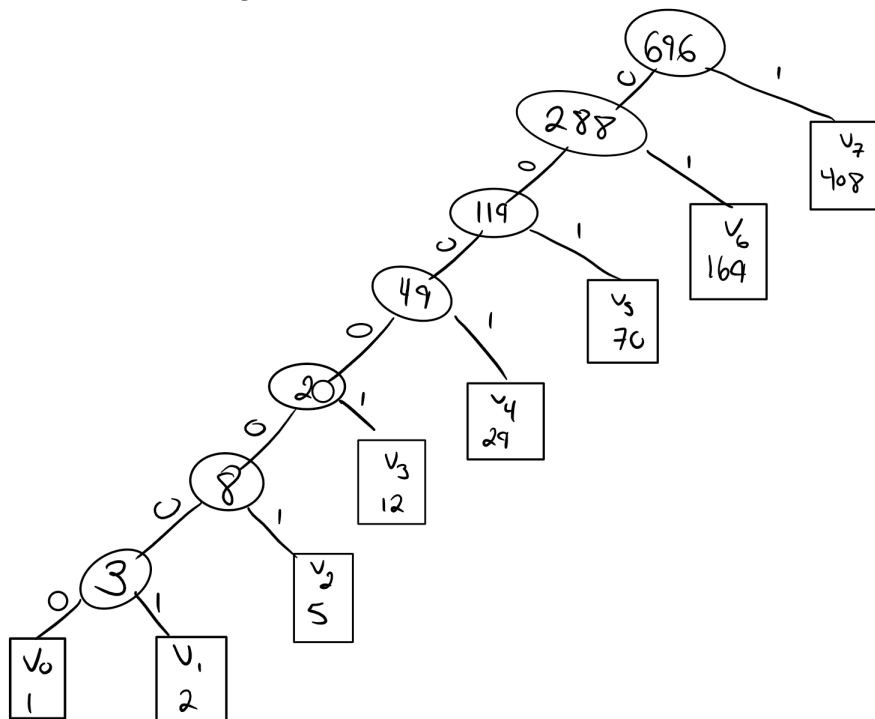
Profs. Hoenigman & Agrawal  
Fall 2019, CU-Boulder

2. (3 pts) How many disconnected components are there when you stop Kruskal's  $k$  round before you complete the MST? Justify your answer.

*Solution.*  $k-1$ . There will be at most  $k-1$  disconnected components at the  $k^{\text{th}}$  stop because you will have just merged two clusters at  $k-2$ , yet they won't all be merged at  $k$ .

3. (5 pts) Consider the recurrence  $F_n = 2F_{n-1} + F_{n-2}$ , with the base cases  $F_0 = 1$  and  $F_1 = 2$ . Suppose we have letters  $v_0, \dots, v_7$ ; where for  $i \in \{0, \dots, 7\}$ , the frequency of  $v_i$  is given by  $F_i$ . Draw a Huffman tree for  $v_0, \dots, v_7$ .

*Solution.* Drawing of Huffman Tree:



Name: Michael Rogers

ID: 105667404

**CSCI 3104, Algorithms**  
**Problem Set 5b (48 points)**

**Profs. Hoenigman & Agrawal**  
**Fall 2019, CU-Boulder**

---

4. (5 pts) Assume you run your Huffman tree algorithm and you produce the following pre-fix codes. Describe why there must be an error in your algorithm.

S = 00  
c = 01  
i = 001  
e = 011  
n = 101

*Solution.* This cannot be a valid Huffman code because Huffman codes are prefix codes. The algorithm must have an error in it because no code can be a prefix for another, and S is a prefix for c, i, and e and c is a prefix for e, etc.

Name: Michael Rogers

ID: 105667404

**CSCI 3104, Algorithms**  
**Problem Set 5b (48 points)**

**Profs. Hoenigman & Agrawal**  
**Fall 2019, CU-Boulder**

---

5. (10 pts) Assume you're given an integer matrix that represents a plot of land, where the value at that location in the matrix represents the height above sea level. A value of zero indicates water. A pond is a region of water connected vertically, horizontally, or diagonally. The size of the pond is the total number of connected water cells. Write an algorithm to compute the sizes of all ponds in the matrix.

Example:

```
0 2 1 0
0 1 0 1
1 1 0 1
0 1 0 1
```

would output 1, 2, 4.

- (a) (3 pts) Describe the graph data structure that your algorithm will use for this problem.

*Solution.* For this problem, the graph will be an undirected graph where all vertices have edges to all of the vertices that are adjacent to it. For example, if the 0 in the top left hand corner is  $s$ , it would have edges from  $s - 0$ ,  $s - 2$ ,  $s - 1$ .

Name: Michael Rogers

ID: 105667404

**CSCI 3104, Algorithms**  
**Problem Set 5b (48 points)**

**Profs. Hoenigman & Agrawal**  
**Fall 2019, CU-Boulder**

---

- (b) (2 pts) Provide a 3-4 sentence description of how your algorithm works, including how the matrix is converted to the graph, how adjacent vertices are identified, and how the algorithm traverses the graph to identify connected vertices.

*Solution.* This algorithm will go through each vertex in the graph and check if the value of the vertex is 0. If not, it will move on to the next. For each vertex equal to zero, it will iterate through its adjacent vertices and select edges where the adjacent vertex is 0 and increment the count of the size of that pond. It will then mark the vertex as visited and move on to the next vertex with value 0, or move on to the next.

- (c) (5 pts) Write an algorithm to solve this problem.

*Solution.* Pseudocode:

```
def selectPonds(G):
    initialize all vertices as unvisited
    lengths = []
    length = 0
    for each vertex in G:
        for adjVertex in vertex.adj :
            if adjVertex == 0 and adjVertex == unvisited:
                perform a depth first search to find other 0's adjacent.
                // In the DFS, increment length
                lengths.append(length)
    return lengths
```