

# Project Report



2023

## Design and Analysis of Algorithms

### **Made By:**

Minam Faisal (21i-1901)

Momenah Saif (21i-1909)

Submission Date: 5<sup>th</sup> December, 2023

## **Codes Explanation:-**

### **1. Normalization:**

- **Data Structure Definition:**

Defines a structure named Data that has four members: column1, column2, column3, and column4.

- **File Reading Function:**

Declares a function called Filereading that reads data from a file specified by the filename parameter and stores it in a vector of Data structures. Opens the file using an input file stream and checks if it's successfully opened. Reads each line from the file, extracts four values using sscanf, creates a Data structure, and adds it to the vector.

- **Printing Function:**

Defines a function called print that prints the contents of a vector of Data structures.

- **Execution Time Measurement Function:**

Declares a function named ExecutionTimeCalculation that measures the execution time of a given function using the <chrono> library. Records the start time, executes the provided function, records the end time, and calculates the duration.

- **Main Function:**

Declares a vector named dataVector to store the data read from the file. Measures the execution time of the file reading function using the ExecutionTimeCalculation function and prints the time taken. Prints the contents of the vector using the print function.

```
~/Desktop/algo_project/Normalization.cpp - Mousepad
File Edit Search View Document Help
1 #include <iostream>
2 #include <fstream>
3 #include <vector>
4 #include <chrono>
5 #include <functional>
6 using namespace std;
7
8 //This representing different columns of data
9 struct Data
10 {
11     int col1;
12     int col2;
13     int col3;
14     long long col4;
15 };
16
17
18 //This function reads a file and processes its contents it takes a filename and a reference to a vector of Data as parameters.
19 void FileReading(const string filename, vector<Data>& datavector)
20 {
21     //Attempts to open the file using an input filestream
22     ifstream file(filename);
23     if (!file.is_open())
24     {
25         cout << "Error opening the file: " << filename << endl;
26         return;
27     }
28     //Reads each line from the file, extracts four values using sscanf, creates a Data structure, and adds it to the vector.
29     string line;
30     while (getline(file, line))
31     {
32         Data data;
33         sscanf(line.c_str(), "%i%ld%ld%ld", &data.col1, &data.col2, &data.col3, &data.col4);
34         datavector.push_back(data);
35     }
36     file.close();
37 }
38
39 //This function measures the execution time of a given function using the chrono library
40 void print(const vector<Data>& datavector)
41 {
42     for (const auto& data : datavector)
43     {
44         cout << data.col1 << " " << data.col2 << " " << data.col3 << " " << data.col4 << endl;
45     }
46 }
47
48 double ExecutionTimeCalculation(function<void()> function)
49 {
50     auto start = chrono::high_resolution_clock::now();
51     function();
52     auto end = chrono::high_resolution_clock::now();
53     chrono::duration<double> duration = end - start;
54     return duration.count();
55 }
56
57 int main()
58 {
59     //vector of Data named datavector is declared.
60     vector<Data> datavector;
61 }
```

## Output:

- Time taken to read and process data in csv file = 0.00654087

```
kali@kali: ~/Desktop/algo_project
File Actions Edit View Help
(kali@kali)~/Desktop/algo_project
$ g++ Normalization.cpp -o n
(kali@kali)~/Desktop/algo_project
$ ./n
Time taken to read and process data: 0.00654087 seconds
1 0 0 0
2 0 0 0
3 0 0 0
0 0 0 0
3 0 0 0
2 0 0 0
4 0 0 0
4 0 0 0
0 0 0 0
0 0 0 0
4 0 0 0
3 0 0 0
1 0 0 0
0 0 0 0
1 0 0 0
5 0 0 0
4 0 0 0
2 0 0 0
4 0 0 0
0 0 0 0
8 0 0 0
6 0 0 0
1 0 0 0
3 0 0 0
2 0 0 0
1 0 0 0
7 0 0 0
0 0 0 0
4 0 0 0
0 0 0 0
2 0 0 0
1 0 0 0
```

## 2. Graph Representation:

- **File Reading:**

The Filereading function reads data from a CSV file and stores it in a vector of structures Data. It checks if the file can be opened and reads each line of the file. If any line cannot be read correctly, it catches the error and continues processing the rest of the file.

- **Data Structure:**

The program uses a structure called Data to represent each record in the file. Each record has information about the source, destination, weight, and timestamp of a connection.

- **Sorting:**

The MergeSorting function uses a stable merge sort algorithm to sort the data based on the timestamp. This ensures that records with equal timestamps maintain their original order.

- **File Writing:**

The Filewriting function writes the sorted data back to a file in CSV format.

- **Graph Representation:**

The displayGraph function outputs a representation of the graph connections to the console. It shows the source and destination of each connection.

- **Timing Measurement:**

The Timemeasurement function measures the time it takes to execute a given function.

- **Main Function:**

In the main function, reads data from "data.csv" into the dataVector. Sorts the data based on timestamp using merge sort. Displays the graph representation on the terminal.

```
File Edit Search View Document Help
~/Desktop/algo_project/graph.cpp - Mousepad

1 #include <iostream>
2 #include <fstream>
3 #include <vector>
4 #include <string>
5 #include <algorithm>
6 #include <sstream>
7 #include <chrono>
8 #include <functional>
9 using namespace std;
10
11 struct Data
12 {
13     int source;
14     int destination;
15     int weight;
16     long long timestamp;
17 };
18
19 void FileReading(const string filename, vector<Data> &dataVector)
20 {
21     ifstream file(filename);
22     if(!file.is_open()) {
23         cout << "Error opening the file: " << filename << endl;
24         return;
25     }
26     string line;
27     int Counter = 1;
28     while (getline(file, line))
29     {
30         Counter++;
31         stringstream ss(line);
32         Data data;
33
34         string token;
35         // By using try and catch, the program can continue processing the remaining lines of the file even if an error occurs during the parsing of a specific line
36         try
37         {
38             if (getline(ss, token, ','))
39                 data.source = stoi(token);
40             if (getline(ss, token, ','))
41                 data.destination = stoi(token);
42             if (getline(ss, token, ','))
43                 data.weight = stoi(token);
44             if (getline(ss, token, ','))
45                 data.timestamp = stoi(token);
46             dataVector.push_back(data);
47         }
48         catch (const exception &e)
49         {
50             cout << "Error parsing line " << Counter << ": " << e.what() << endl;
51         }
52     }
53     file.close();
54 }
```

Output:

```
kali@kali: ~/Desktop/algo_project
File Actions Edit View Help

(kali@kali)~/Desktop/algo_project
$ g++ graph.cpp -o g
(kali@kali)~/Desktop/algo_project
$ ./g

Graph Representation:
Source: 3 → Destination: 12
Source: 3 → Destination: 24
Source: 32 → Destination: 41
Source: 32 → Destination: 41
Source: 37 → Destination: 45
Source: 38 → Destination: 48
Source: 21 → Destination: 48
Source: 62 → Destination: 79
Source: 23 → Destination: 17
Source: 76 → Destination: 23
Source: 76 → Destination: 23
Source: 78 → Destination: 24
Source: 1 → Destination: 2
Source: 1 → Destination: 2
Source: 1 → Destination: 2
Source: 7 → Destination: 2
Source: 3 → Destination: 1
Source: 3 → Destination: 1
Source: 3 → Destination: 1
Source: 8 → Destination: 17
Source: 0 → Destination: 9
Source: 0 → Destination: 9
Source: 1 → Destination: 0
Source: 1 → Destination: 0
Source: 1 → Destination: 0
Source: 0 → Destination: 4
Source: 0 → Destination: 4
Source: 0 → Destination: 4
Source: 6 → Destination: 1
Source: 2 → Destination: 3
Source: 2 → Destination: 3
Source: 2 → Destination: 3
Source: 3 → Destination: 7
Source: 2 → Destination: 11
Source: 5 → Destination: 14
Source: 5 → Destination: 22
Source: 34 → Destination: 38
Source: 34 → Destination: 38
```

### 3. Sorting:

- **Data Structure:**

The Data struct represents a connection for source, destination, weight, and timestamp.

- **File Reading:**

The Filereading function reads data from a CSV file into a vector of Data structures. It handles errors if the file cannot be opened or if there's an issue in reading the data.

- **Sorting Algorithms:**

Three sorting algorithms:-

Quick Sort: It's a fast sorting algorithm that uses a divide-and-conquer strategy.

Merge Sort: It's a stable sorting algorithm that also uses a divide-and-conquer approach.

Heap Sort: It's a comparison-based sorting algorithm that uses a binary heap data structure.

- **Sorting Functions:**

QuickSorting, MergeSorting, and HeapSorting functions sort the data vector using their respective sorting algorithms.

- **File Writing:**

The Filewriting function writes the sorted data to a new CSV file.

- **Display Function:**

The displayGraph function prints the data in a tabular format on the terminal.

- **Time Measurement:**

The Timemeasurement function measures the execution time of a provided function using the C++ <chrono> library.

- **Main Function:**

Reads data from "data.csv" into dataVector. Applies Quick Sort, measures the time taken, and displays, saves the sorted data. Applies Merge Sort, measures the time taken, and displays, saves the sorted data. Applies Heap Sort, measures the time taken, and display, saves the sorted data. Each step involves printing the execution time and displaying the sorted data.

- **Time Complexities of Sorting Algorithms Analysis:**

### **Merge Sort:**

Best Case Time Complexity:  $O(n \log n)$

Worst Case Time Complexity:  $O(n \log n)$

Average Case Time Complexity:  $O(n \log n)$

Merge sort has a consistent time complexity of  $O(n \log n)$  for all cases because it always divides the array into two halves and recursively sorts them before merging.

### **Heap Sort:**

Best Case Time Complexity:  $O(n \log n)$

Worst Case Time Complexity:  $O(n \log n)$

Average Case Time Complexity:  $O(n \log n)$

Heap sort also has a consistent time complexity of  $O(n \log n)$  for all cases. The build-heap operation has a time complexity of  $O(n)$ , and the heapify operation in the sorting phase has a time complexity of  $O(\log n)$ .

### **Quick Sort:**

Best Case Time Complexity:  $O(n \log n)$

Worst Case Time Complexity:  $O(n^2)$

Average Case Time Complexity:  $O(n \log n)$

Quick sort has a best-case time complexity of  $O(n \log n)$  when the pivot is always chosen as the median element. However, in the worst case, if the pivot is consistently chosen poorly (e.g., always the smallest or largest element), the time complexity degrades to  $O(n^2)$ . On average, with a good pivot selection strategy, the time complexity is  $O(n \log n)$ .

```
~/Desktop/algo_project/Sorting.cpp - Mousepad
File Edit Search View Document Help
[Icons]
65 bool ComparingTimestamp(const Data& n, const Data& m)
66 {
67     return m.timestamp < n.timestamp;
68 }
69
70 void QuickSorting(vector<Data>& datavector)
71 {
72     sort(datavector.begin(), datavector.end(), ComparingTimestamp);
73 }
74
75 void MergeSorting(vector<Data>& datavector)
76 {
77     stable_sort(datavector.begin(), datavector.end(), ComparingTimestamp);
78 }
79
80 void heapify(vector<Data>& datavector, int n, int i)
81 {
82     int largest = i;
83     int left = 2 * i + 1;
84     int right = 2 * i + 2;
85
86     if (left < n && ComparingTimestamp(datavector[left], datavector[largest]))
87         largest = left;
88
89     if (right < n && ComparingTimestamp(datavector[right], datavector[largest]))
90         largest = right;
91
92     if (largest != i)
93     {
94         swap(datavector[i], datavector[largest]);
95         heapify(datavector, n, largest);
96     }
97 }
98
99 void heapSorting(vector<Data>& datavector)
100 {
101     int n = datavector.size();
102
103     for (int i = n / 2 - 1; i >= 0; --i)
104         heapify(datavector, n, i);
105
106     for (int i = n - 1; i >= 0; --i)
107     {
108         swap(datavector[i], datavector[0]);
109         heapify(datavector, i, 0);
110     }
111 }
112
113 void FileWriting(const vector<Data>& datavector, const string& filename)
114 {
115     ofstream file(filename);
116     if (!file.is_open()) {
117         cout << "Error opening the file: " << filename << endl;
118         return;
119     }
120
121     for (const auto& data : datavector)
122     {
123         file << data.source << " | " << data.destination << " | " << data.weight << " | " << data.timestamp << endl;
124     }
125
126     file.close();
127 }
128
129 void displayGraph(const vector<Data>& datavector)
130 {
131 }
```

## Output:

- Quick Sort Time = 0.000661491

```
kali@kali: ~/Desktop/algo_project
File Actions Edit View Help
(kali@kali)~/Desktop/algo_project
$ g++ Sorting.cpp -o s
(kali@kali)~/Desktop/algo_project
$ ./s
Quick Sort Time: 0.000661491 seconds

Sorted Data using Quick Sort:
Source | Destination | Weight | Timestamp
62 | 79 | 140 | 1335312000
32 | 41 | 8 | 1335312000
3 | 24 | 8 | 1335312000
3 | 12 | 4 | 1335312000
76 | 23 | 105 | 1335312000
31 | 48 | 140 | 1335312000
23 | 17 | 98 | 1335312000
38 | 48 | 14 | 1335312000
37 | 45 | 14 | 1335312000
32 | 41 | 18 | 1335312000
76 | 23 | 105 | 1335312000
78 | 24 | 105 | 1335312000
1 | 2 | 8 | 1337572800
1 | 2 | 8 | 1337572800
1 | 2 | 8 | 1337572800
7 | 2 | 8 | 1339440000
3 | 1 | 2 | 1342584000
3 | 1 | 2 | 1342584000
3 | 1 | 2 | 1342584000
8 | 17 | 7 | 1345219200
0 | 9 | 5 | 1346126400
0 | 9 | 5 | 1346126400
1 | 0 | 5 | 1347854400
1 | 0 | 5 | 1347854400
1 | 0 | 5 | 1347854400
0 | 4 | 2 | 1350014400
0 | 4 | 2 | 1350014400
6 | 1 | 6 | 1350100800
2 | 3 | 5 | 1351742400
2 | 3 | 5 | 1351742400
2 | 3 | 5 | 1351742400
3 | 7 | 2 | 1352592000
2 | 11 | 9 | 1354320000
88 | 99 | 101 | 1356979200
5 | 14 | 3 | 1356979200
```

- Merge Sort Time = 0.000444094



Merge Sort Time: 0.000444094 seconds

Sorted Data using Merge Sort:

Source	Destination	Weight	Timestamp
3	12	4	1335312000
3	24	8	1335312000
32	41	8	1335312000
32	41	18	1335312000
37	45	14	1335312000
38	48	14	1335312000
31	48	140	1335312000
62	79	140	1335312000
23	17	98	1335312000
76	23	105	1335312000
76	23	105	1335312000
78	24	105	1335312000
1	2	8	1337572800
1	2	8	1337572800
1	2	8	1337572800
7	2	8	1339440000
3	1	2	1342584000
3	1	2	1342584000
3	1	2	1342584000
8	17	7	1345219200
0	9	5	1346126400
0	9	5	1346126400
1	0	5	1347854400
1	0	5	1347854400
1	0	5	1347854400
0	4	2	1350014400
0	4	2	1350014400
0	4	2	1350014400
6	1	6	1350100800
2	3	5	1351742400
2	3	5	1351742400
2	3	5	1351742400
3	7	2	1352592000
2	11	9	1354320000
5	14	3	1356979200
5	22	3	1356979200

- Heap Sort Time = 0.000709976

Heap Sort Time: 0.000709976 seconds

Sorted Data using Heap Sort:

Source	Destination	Weight	Timestamp
94	51	127	2070124800
88	50	160	2066124800
62	49	111	2050124800
76	48	130	2040124800
70	47	150	2030124800
64	46	127	2020124800
43	32	120	2010124800
19	49	18	2010124800
91	26	133	2010124800
91	26	133	2010124800
91	26	133	2010124800
19	49	8	2010124800
46	13	11	2010124800
91	26	133	2010124800
58	45	160	2010124800
19	49	10	2000124800
67	73	144	2000124800
47	14	13	2000124800
17	49	4	2000124800
52	44	111	2000124800
17	49	14	2000124800
67	73	144	2000124800
39	33	175	2000124800
67	73	144	2000124800
67	73	144	2000124800
13	14	95	1990124800
48	15	14	1990124800
35	34	150	1990124800
13	14	95	1990124800
15	49	1	1990124800
46	43	130	1990124800
13	14	95	1990124800
13	14	95	1990124800
17	49	11	1990124800
15	49	11	1990124800
38	1	112	1980124800

#### 4. Dijkstra:

- **Introduction:**

The code is written to implement Dijkstra's algorithm, a method for finding the shortest paths between nodes in a directed graph. The graph is defined by a set of nodes and edges, each with an associated weight.

- **Data Reading:**

The program starts by reading information about nodes and edges from a CSV file called "data.csv". Each line in the file represents a connection between two nodes, with a numerical weight. The code ensures that invalid or out-of-range values are handled properly.

- **Graph Representation:**

The graph is represented using an adjacency list, a data structure that stores each node's neighbors and their corresponding weights. This representation allows for efficient traversal of the graph.

- **Dijkstra's Algorithm:**

The main part of the code is the implementation of Dijkstra's algorithm. This algorithm finds the shortest paths from a specified source node to all other nodes in the graph.

- **Time Measurement:**

The code includes a function to measure the execution time of Dijkstra's algorithm. This helps evaluate the efficiency of the algorithm.

- **Results Display and Storage:**

After running Dijkstra's algorithm, the code displays the shortest paths from a specified source node to all other nodes. It also writes these results, including the trace of the algorithm, to a CSV file named "dijkstra\_shortest\_distances\_with\_trace.csv".

- **Conclusion:**

Finally, the program outputs the execution time of Dijkstra's algorithm, providing insights into the efficiency of the implemented solution.

- **Time Complexity of Dijkstra's Algorithm:**

- ✚ Best Case:  $O((V + E) * \log(V))$

- ✚ Average Case:  $O((V + E) * \log(V))$

- ✚ Worst Case:  $O(V + E)$

- **Analysis:**

- ✚ **Initialization (distances and traces):**  $O(V)$ , Initializing the distance array and trace vector involves iterating over all vertices in the graph. Since the graph has  $V$  vertices, the time complexity is linear in the number of vertices.

- ✚ **Reading Input:**  $O(E)$ , Reading edges from the file involves iterating over each edge once. In the worst case, we have  $E$  edges, leading to a linear time complexity in the number of edges.

- ✚ **Graph Initialization (Adding Edges):**  $O(E)$ , Adding edges to the adjacency list involves iterating over each edge once. In the worst case, we have  $E$  edges, leading to a linear time complexity in the number of edges.

- ✚ **Dijkstra's Algorithm Main Loop:**  $O((V + E) * \log(V))$ , Dijkstra's algorithm with a binary heap priority queue has a worst-case time complexity of  $O((V + E) * \log(V))$ . This dominates the overall time complexity.

- ✚ **Writing Output:**  $O(V)$ , Writing the results involves iterating over all vertices to write distances and traces. Since there are  $V$  vertices, the time complexity is linear in the number of vertices.

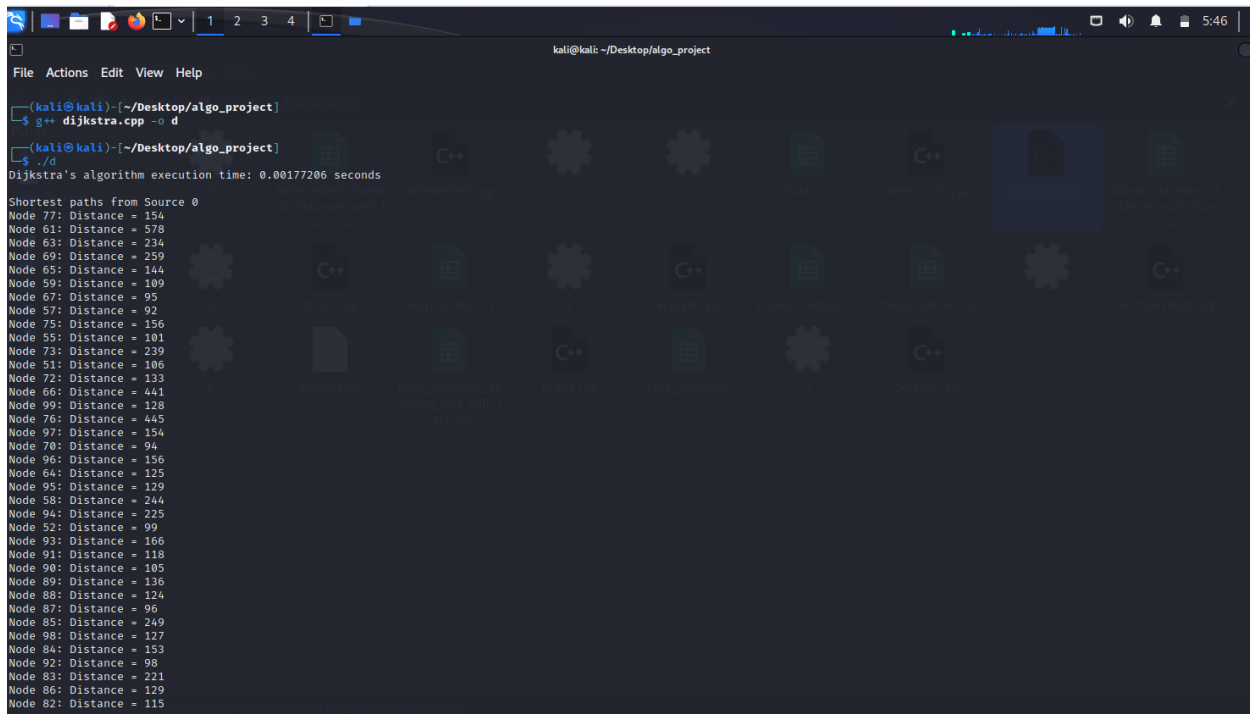
- ✚ **Overall Time Complexity:** The dominant Factor is Dijkstra's Algorithm resulting in a worst case time complexity of  $O((V + E) * \log(V))$ .

```

23 vector<vector<pair<int, int>>> adjlist;
24 int V; // No. of vertices
25
26
27 void readSelectedNodes(const string& filename, vector<Edge>& edges, unordered_set<int>& nodes)
28 {
29     ifstream file(filename);
30     if (!file.is_open()) {
31         cout << "Error opening the file: " << filename << endl;
32         return;
33     }
34     string line;
35     while (std::getline(file, line))
36     {
37         istringstream ss(line);
38         Edge edge;
39         try
40         {
41             string token;
42             getline(ss, token, ',');
43             edge.source = stoi(token);
44             getline(ss, token, ',');
45             edge.destination = stoi(token);
46             getline(ss, token, ',');
47             edge.weight = stoi(token);
48             edges.push_back(edge);
49             nodes.insert(edge.source);
50             nodes.insert(edge.destination);
51         }
52         catch (const Invalid_argument& e)
53         {
54             cout << "Error converting to integer: " << e.what() << std::endl;
55             // skip this line and proceed to the next one
56             continue;
57         }
58         catch (const out_of_range& e)
59         {
60             cout << "Out of range: " << e.what() << endl;
61             continue;
62         }
63     }
64     file.close();
65 }
66
67 void AddingEdge(Graph& graph, const Edge& edge)
68 {
69     graph.adjList[edge.source].emplace_back(edge.destination, edge.weight);
70 }
71
72 void Dijkstra(const Graph& graph, int source, vector<int>& distances, vector<string>& trace, const unordered_set<int>& nodes)
73 {
74     int V = graph.V;
75     distances.assign(V, numeric_limits<int>::max());
76 }
  
```

**Output:**

- Dijkstra Execution Time = 0.00177206



```
kali@kali: ~/Desktop/algo_project
File Actions Edit View Help

(kali@kali)~/Desktop/algo_project
$ g++ dijkstra.cpp -o d

(kali@kali)~/Desktop/algo_project
$ ./d
Dijkstra's algorithm execution time: 0.00177206 seconds

Shortest paths from Source 0
Node 77: Distance = 154
Node 61: Distance = 573
Node 63: Distance = 234
Node 69: Distance = 259
Node 65: Distance = 144
Node 59: Distance = 109
Node 67: Distance = 95
Node 57: Distance = 92
Node 75: Distance = 156
Node 55: Distance = 101
Node 73: Distance = 239
Node 51: Distance = 106
Node 72: Distance = 133
Node 66: Distance = 441
Node 99: Distance = 128
Node 76: Distance = 445
Node 97: Distance = 154
Node 70: Distance = 94
Node 96: Distance = 156
Node 64: Distance = 125
Node 95: Distance = 129
Node 58: Distance = 244
Node 94: Distance = 225
Node 52: Distance = 99
Node 93: Distance = 166
Node 91: Distance = 118
Node 98: Distance = 105
Node 89: Distance = 136
Node 88: Distance = 124
Node 87: Distance = 96
Node 85: Distance = 249
Node 98: Distance = 127
Node 84: Distance = 153
Node 92: Distance = 98
Node 83: Distance = 221
Node 86: Distance = 129
Node 82: Distance = 115
```

## 5. BellmenFord:

- File Reading:

The code starts by reading a CSV file containing information about edges in a graph (source node, destination node, and edge weight).

- Bellman-Ford Algorithm:

The Bellman-Ford algorithm is then applied to find the shortest paths from a given source node to all other nodes in the graph. The algorithm iterates over all edges multiple times, relaxing the edges by updating the distances whenever a shorter path is found.

- Results Writing:

The results, including the shortest distances and the trace of the algorithm, are written to an output file.

- Displaying Shortest Paths:

The code also prints the shortest paths from the source node to all other nodes on the terminal.

- **Timing Measurement:**

The execution time of the Bellman-Ford algorithm is measured using `<chrono>` library. This helps in understanding how long the algorithm takes to run.

- **Main Function:**

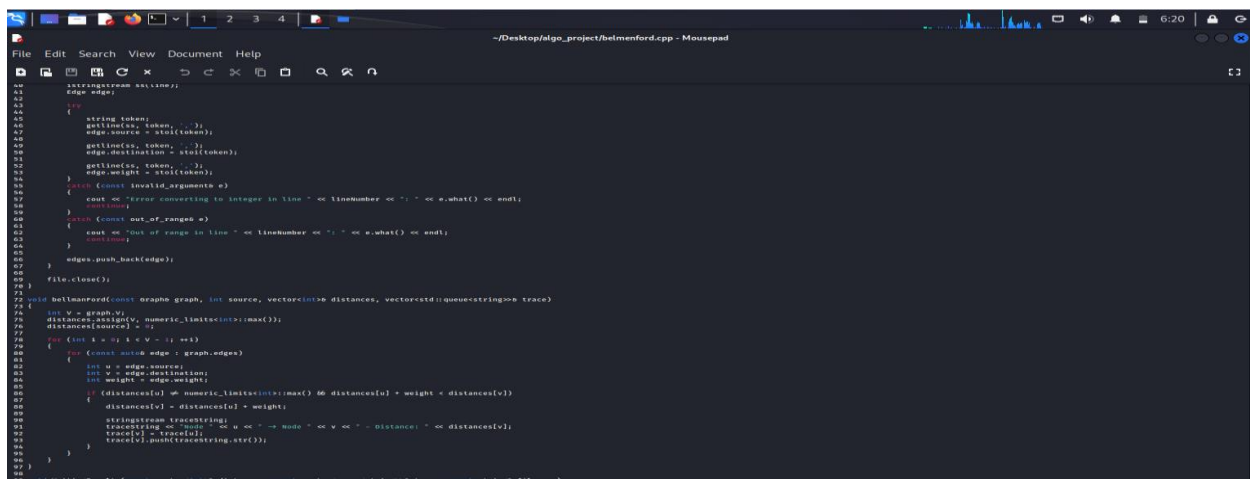
The main function runs the entire process, from reading the file to applying the algorithm, measuring time, and writing the results.

- **Time Complexity of Bellman-Ford Algorithm:**

- ✚ Best Case:  $O(V + E) - O(V)$
- ✚ Average Case:  $O(V * E)$
- ✚ Worst Case:  $O(V * E)$

- **Analysis:**

- ✚ **Initialization (distances and traces):**  $O(V)$  Initializing the distance array and trace vector has a time complexity of  $O(V)$  because it involves iterating over all vertices.
- ✚ **Bellman-Ford Algorithm Main Loop:**  $O(V * E)$ , the outer loop runs  $V-1$  times, where  $V$  is the number of vertices. The inner loop iterates over all edges, which is  $E$  in the worst case (complete graph).
- ✚ **WritingResults Function:**  $O(V)$  Writing the results involves iterating over all vertices to write the distances and traces.
- ✚ **Overall Time Complexity:** The dominant factor is the Bellman-Ford main loop, resulting in a worst-case time complexity of  $O(V * E)$ .

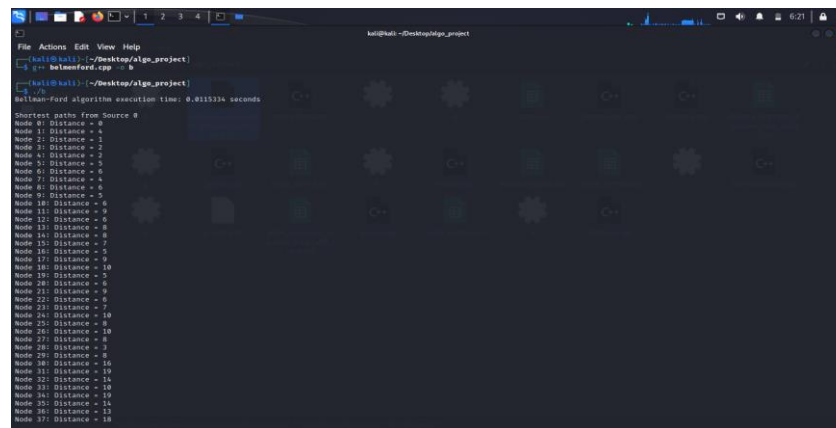


```
File Edit Search View Document Help
~/Desktop/algo_project/belmenford.cpp - Mousepad

44 stringstream s;
45 Edge edge;
46 {
47     string token;
48     getline(s, token, '\n');
49     edge.source = stoi(token);
50     getline(s, token, '\n');
51     edge.destination = stoi(token);
52     getline(s, token, '\n');
53     edge.weight = stoi(token);
54 }
55 // Check for invalid arguments
56 {
57     // Check if source is integer in line
58     if (token.find_first_not_of("0-9") != string::npos)
59         continue;
60     // Check if destination is integer in line
61     if (token.find_first_not_of("0-9") != string::npos)
62         continue;
63     // Check if weight is integer in line
64     if (token.find_first_not_of("0-9") != string::npos)
65         continue;
66 }
67 edges.push_back(edge);
68 }
69 // File close
70 s.close();
71 }
72 // Bellman-Ford
73 void bellmanford(const Graph &graph, int source, vector<int> &distances, vector<string> &traces)
74 {
75     // Initialize distances
76     distances.resize(graph.vertices(), numeric_limits<int>::max());
77     distances[source] = 0;
78     // Relax edges
79     for (int i = 1; i < graph.vertices(); i++)
80     {
81         for (const auto &edge : graph.edges)
82         {
83             int u = edge.source;
84             int v = edge.destination;
85             int weight = edge.weight;
86             // Relax edge
87             if (distances[u] != numeric_limits<int>::max() && distances[u] + weight < distances[v])
88             {
89                 distances[v] = distances[u] + weight;
90                 string trace;
91                 trace += to_string(u) + " -> " + to_string(v) + " -> " + to_string(distances[v]);
92                 trace[v] = trace[u];
93                 trace[v].push_back(to_string(v));
94             }
95         }
96     }
97 }
```

## Output:

- BellmanFord Execution Time = 0.0115334



```
File Actions Edit View Help
kati@kali: ~/Desktop/algo_project
$ g++ bellmanford.cpp -o b
$ ./b
Bellman-Ford algorithm execution time: 0.0115334 seconds

Shortest paths from Source 0
Node 0: Distance = 0
Node 1: Distance = 4
Node 2: Distance = 5
Node 3: Distance = 7
Node 4: Distance = 6
Node 5: Distance = 6
Node 6: Distance = 6
Node 7: Distance = 6
Node 8: Distance = 6
Node 9: Distance = 5
Node 10: Distance = 6
Node 11: Distance = 9
Node 12: Distance = 8
Node 13: Distance = 8
Node 14: Distance = 8
Node 15: Distance = 7
Node 16: Distance = 9
Node 17: Distance = 7
Node 18: Distance = 10
Node 19: Distance = 5
Node 20: Distance = 6
Node 21: Distance = 9
Node 22: Distance = 9
Node 23: Distance = 7
Node 24: Distance = 10
Node 25: Distance = 8
Node 26: Distance = 10
Node 27: Distance = 8
Node 28: Distance = 3
Node 29: Distance = 16
Node 30: Distance = 19
Node 31: Distance = 15
Node 32: Distance = 15
Node 33: Distance = 10
Node 34: Distance = 19
Node 35: Distance = 14
Node 36: Distance = 13
Node 37: Distance = 18
```

## 6. Prim's:

- **File Reading:**

The program starts by reading data from a file named "data.csv" that represents a graph with nodes and weighted edges.

- **Graph Representation:**

The program then creates a graph data structure based on the read data, where nodes are connected by edges, and each edge has a weight.

- **Prim's Algorithm:**

Prim's algorithm is applied to find the Minimum Spanning Tree. It starts from an arbitrary node and systematically adds the smallest edge that connects a new node to the existing tree. The process continues until all nodes are included in the MST.

- **File Writing:**

The results, including the MST and a trace of the algorithm's execution, are written to an output file name "prim\_minimum\_spanning\_tree\_with\_trace.csv"

- **Display Function:**

The program displays the MST on the terminal.

- **Timing Measurement:**

The execution time of Prim's algorithm is measured and printed to the terminal.

- **Main Function:**

The main function runs the entire process, from reading the file to applying Prim's algorithm, measuring time, displaying results, and writing them to an output file.

- **Time Complexity of Bellman-Ford Algorithm:**

- + Best Case:  $O(E \cdot \log V)$
- + Average Case:  $O(E \cdot \log V)$
- + Worst Case:  $O((V+E) \cdot \log V)$

- **Analysis:**

- + **Filereading Function:**  $O(E)$ , where  $E$  is the number of edges. The function reads each line from the file, and in the worst case, it reads all the edges.
- + **AddingEdge Function:**  $O(1)$  (constant time). The function appends the edge to the adjacency list.
- + **primMST Function:**  $O(E + V \log V)$ , where  $E$  is the number of edges and  $V$  is the number of vertices. This is due to the priority queue operations inside the while loop.
- + **WritingResults Function:**  $O(V)$ , where  $V$  is the number of vertices. The function iterates over all vertices to write the results.
- + **displayMST Function:**  $O(V)$ , where  $V$  is the number of vertices. Similar to WritingResults, it iterates over all vertices.
- + **Timemeasurement Function:** Depends on the function passed to it. In this context, it's used to measure the execution time of primMST, so its complexity is  $O(E + V \log V)$ .
- + **Main Function:**  $O(E + V \log V)$ , dominated by the call to primMST.
- + **Overall Time complexity:** The overall time complexity is determined by the primMST function, making it  $O(E + V \log V)$ . The initialization and other operations contribute less to the overall complexity.

```

File Edit Search View Document Help
~/Desktop/algos_project/prim.cpp - Microsoft

1 2 3 4
graph::adjlistEdge::source, replace_backEdge_destination, edge_weight);
graph::adjlistEdge::destination, replace_backEdge_source, edge_weight);
}

void printMST(const graph& graph, vector<int>& parent, vector<string>& trace)
{
    int v = graph.v;
    vector<int> ans(v, numeric_limits<int>::max());
    vector<int> ans(v, 0);
    priority_queue<pair<int, int>, vector<pair<int, int>, greater<pair<int, int>>> pq;
    ans[0] = 0;
    parent[0] = -1;
    pq.push(make_pair(0, 0));
    while (!pq.empty())
    {
        int u = pq.top().second;
        pq.pop();
        ans[u] = 0;
        stringstream ss;
        ss << "Node insertion: u = " << u << " (key value: " << ans[u] << ") result: ";
        trace.push_back(ss.str());
        for (const auto& neighbor : graph.adjList[u])
        {
            int v = neighbor.first;
            int weight = neighbor.second;
            if (ans[v] >= weight + ans[u]) {
                parent[v] = u;
                ans[v] = weight;
                pq.push(make_pair(ans[v], v));
            }
            stringstream ss;
            ss << "Edge insertion: u = " << u << " v = " << v << " (weight: " << weight << ", updated key value: " << ans[v] << ") result: ";
            trace.push_back(ss.str());
        }
    }
}

void writeResult(const vector<int>& parent, const vector<string>& trace, const string& filename)
{
    ofstream file(filename);
    if (!file.is_open())
    {
        cerr << "Error opening the file: " << filename << endl;
        return;
    }
    file << "Minimum Spanning Tree result:" << endl;
    for (int i = 1; i < parent.size(); ++i)
    {
        file << "Edge: " << i << " - " << parent[i] << " result: " << ans[i] << endl;
    }
}

```

## Output:

- Prims' Execution Time = 0.00123901

```

File Actions Edit View Help
kali@kali:~/Desktop/algos_project

--(kali@kali)~/Desktop/algos_project
$ g++ prim.cpp -o p
--(kali@kali)~/Desktop/algos_project
$ ./p
Prim's algorithm execution time: 0.00123901 seconds

Minimum Spanning Tree:
Edge: 3 - 1
Edge: 0 - 2
Edge: 0 - 3
Edge: 3 - 4
Edge: 15 - 5
Edge: 50 - 6
Edge: 3 - 7
Edge: 3 - 8
Edge: 11 - 9
Edge: 10 - 10
Edge: 40 - 11
Edge: 19 - 12
Edge: 22 - 13
Edge: 7 - 14
Edge: 8 - 15
Edge: 7 - 16
Edge: 21 - 17
Edge: 5 - 18
Edge: 7 - 19
Edge: 5 - 20
Edge: 23 - 21
Edge: 10 - 22
Edge: 12 - 23
Edge: 21 - 24
Edge: 20 - 25
Edge: 2 - 26
Edge: 15 - 27
Edge: 50 - 28
Edge: 8 - 29
Edge: 44 - 30
Edge: 9 - 31
Edge: 30 - 32
Edge: 15 - 33
Edge: 50 - 34
Edge: 3 - 35
Edge: 35 - 36
Edge: 0 - 37

```

## 7. Kruskal:

- File Reading Function:**

This function reads data from a CSV file containing information about edges in a graph. Each line of the file represents an edge with source node, destination node, and weight. The function handles potential errors while reading the file, such as invalid data or out-of-range values.



- **Edge Structure:**

The code defines a structure (Edge) to represent an edge in the graph. It stores information about the source node, destination node, and the weight of the edge.

- **Disjoint Set Data Structure:**

This structure is used to implement the disjoint-set data structure, which helps in efficiently detecting and merging connected components in a graph.

- **Kruskal's Algorithm:**

The main algorithm implemented in the code is Kruskal's algorithm for finding the Minimum Spanning Tree (MST) of a graph. The algorithm iteratively selects edges in ascending order of their weights, adding them to the MST if they don't create a cycle. The progress of the algorithm and its decisions (acceptance or rejection of edges) are recorded in a trace.

- **Time Measurement Function:**

This function measures the execution time of a given function using <chrono> library. In this case, it measures the time taken by Kruskal's algorithm to find the MST.

- **Main Function:**

Reads data from the CSV file using the Filereading function. Determines the number of nodes in the graph. Applies Kruskal's algorithm to find the MST and records the execution time. Outputs the MST and a trace of the algorithm's steps to a CSV file (kruskal\_output.csv). Outputs the MST to the terminal. Outputs the execution time of Kruskal's algorithm.

- **Time Complexity of Kruskal's Algorithm:**

- ✚ Best Case:  $O(E \log E)$
- ✚ Average Case:  $O(E \log E)$
- ✚ Worst Case:  $O(E \log E)$

- **Analysis:**

✚ **File Reading:**  $O(E)$ , where  $E$  is the number of edges. This is because each edge needs to be processed once while reading from the file.

✚ **Kruskal's Algorithm:**  $O(E \log E)$ , where  $E$  is the number of edges. This is the dominant factor and represents the time taken to sort the edges.

- ✚ **Disjoint Set Operations:**  $O(E \log^* V)$ , where  $V$  is the number of vertices. The  $\log^*$  function is the iterated logarithm and grows extremely slowly.
- ✚ **Constructing Minimum Spanning Tree:**  $O(E \log^* V)$
- ✚ **Writing Results to Output File:**  $O(E)$  for writing the MST and trace to the output file.
- ✚ **Time Measurement:**  $O(T)$ , where  $T$  is the time taken by the provided function `kruskalMST`.
- ✚ **Overall Time Complexity:** The dominant factor is sorting edges ( $O(E \log E)$ ), resulting in a time complexity of  $O(E \log E)$ .

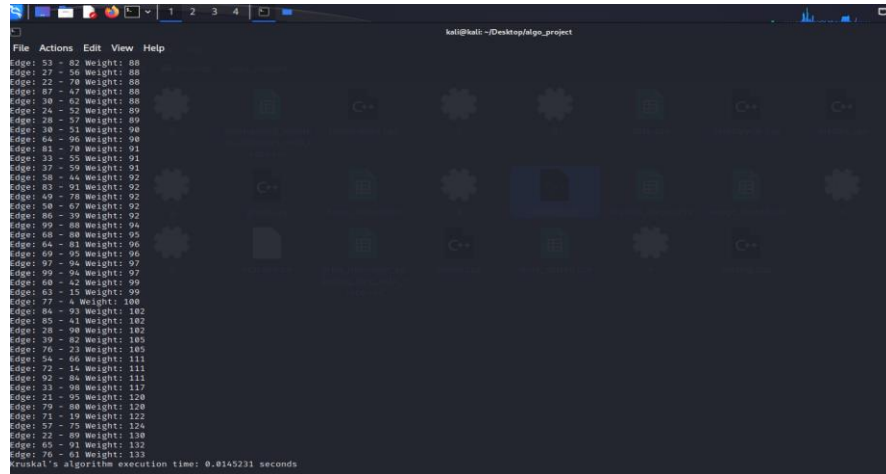
```

65     edges.push_back(edge);
66 }
67
68 file.close();
69 }
70
71 bool ComparingEdges(const Edge& m, const Edge& n)
72 {
73     return m.weight < n.weight;
74 }
75
76 int finding(DisjointSet& ds, int i)
77 {
78     if (ds.parent[i] == i) {
79         ds.parent[i] = finding(ds, ds.parent[i]);
80     }
81     return ds.parent[i];
82 }
83
84 void Union(DisjointSet& ds, int x, int y)
85 {
86     int xRoot = finding(ds, x);
87     int yRoot = finding(ds, y);
88
89     if (ds.rank[xRoot] < ds.rank[yRoot])
90     {
91         ds.parent[xRoot] = yRoot;
92     }
93     else if (ds.rank[xRoot] > ds.rank[yRoot])
94     {
95         ds.parent[yRoot] = xRoot;
96     }
97     else
98     {
99         ds.parent[yRoot] = xRoot;
100         ds.rank[xRoot]++;
101     }
102 }
103
104 void kruskalMST(const vector<Edge>& edges, int V)
105 {
106     vector<Edge> result;
107     DisjointSet ds;
108     ds.parent.resize(V);
109     ds.rank.resize(V, 0);
110
111     for (int i = 0; i < V; ++i)
112     {
113         ds.parent[i] = i;
114     }
115
116     vector<Edge> sortedEdges = edges;
117     sort(sortedEdges.begin(), sortedEdges.end(), ComparingEdges);
118
119     stack<string> traceStack;
120
121     for (const auto& edge : sortedEdges)
122     {
123         int x = finding(ds, edge.source);
124         int y = finding(ds, edge.destination);
125
126         stringstream trace;
127         trace << "Edge: " << edge.source << " - " << edge.destination << " weight: " << edge.weight;
128         trace << "\n";
129     }

```

## Output:

- Kruskal's Execution time = 0.0145231



```
File Actions Edit View Help
Edge: 53 - 82 Weight: 88
Edge: 27 - 56 Weight: 88
Edge: 22 - 78 Weight: 88
Edge: 87 - 47 Weight: 88
Edge: 38 - 82 Weight: 88
Edge: 24 - 52 Weight: 89
Edge: 28 - 57 Weight: 89
Edge: 39 - 51 Weight: 90
Edge: 64 - 96 Weight: 90
Edge: 81 - 78 Weight: 91
Edge: 33 - 55 Weight: 91
Edge: 37 - 59 Weight: 91
Edge: 58 - 44 Weight: 92
Edge: 83 - 91 Weight: 92
Edge: 49 - 78 Weight: 92
Edge: 58 - 67 Weight: 92
Edge: 86 - 39 Weight: 92
Edge: 99 - 88 Weight: 94
Edge: 68 - 88 Weight: 95
Edge: 64 - 81 Weight: 96
Edge: 69 - 95 Weight: 96
Edge: 97 - 94 Weight: 97
Edge: 99 - 84 Weight: 97
Edge: 68 - 42 Weight: 99
Edge: 62 - 15 Weight: 99
Edge: 77 - 4 Weight: 100
Edge: 84 - 93 Weight: 102
Edge: 85 - 41 Weight: 102
Edge: 28 - 98 Weight: 102
Edge: 39 - 82 Weight: 105
Edge: 76 - 23 Weight: 105
Edge: 54 - 66 Weight: 111
Edge: 72 - 14 Weight: 111
Edge: 92 - 84 Weight: 111
Edge: 31 - 88 Weight: 117
Edge: 21 - 95 Weight: 120
Edge: 78 - 88 Weight: 120
Edge: 71 - 19 Weight: 122
Edge: 57 - 75 Weight: 124
Edge: 22 - 89 Weight: 128
Edge: 65 - 91 Weight: 132
Edge: 76 - 61 Weight: 133
Kruskal's algorithm execution time: 0.0145231 seconds
```

## 8. DetectCycle:

- **Edge Structure:**

Represents a connection between two points (vertices). Each connection has a starting point (source), an ending point (destination), and a weight.

- **File Reading:**

Reads information about connections from a file. Each line in the file represents a connection and contains details such as the starting point, ending point, and the cost of the connection.

- **Sorting Edges:**

Sorts the connections based on their costs in ascending order.

- **Union-Find Operations:**

Implements a method to check for loops in the connections without drawing the entire map. Uses a technique called union-find to keep track of connected points. If connecting two points would create a loop, it is detected during this process.

- **Detecting Cycles:**

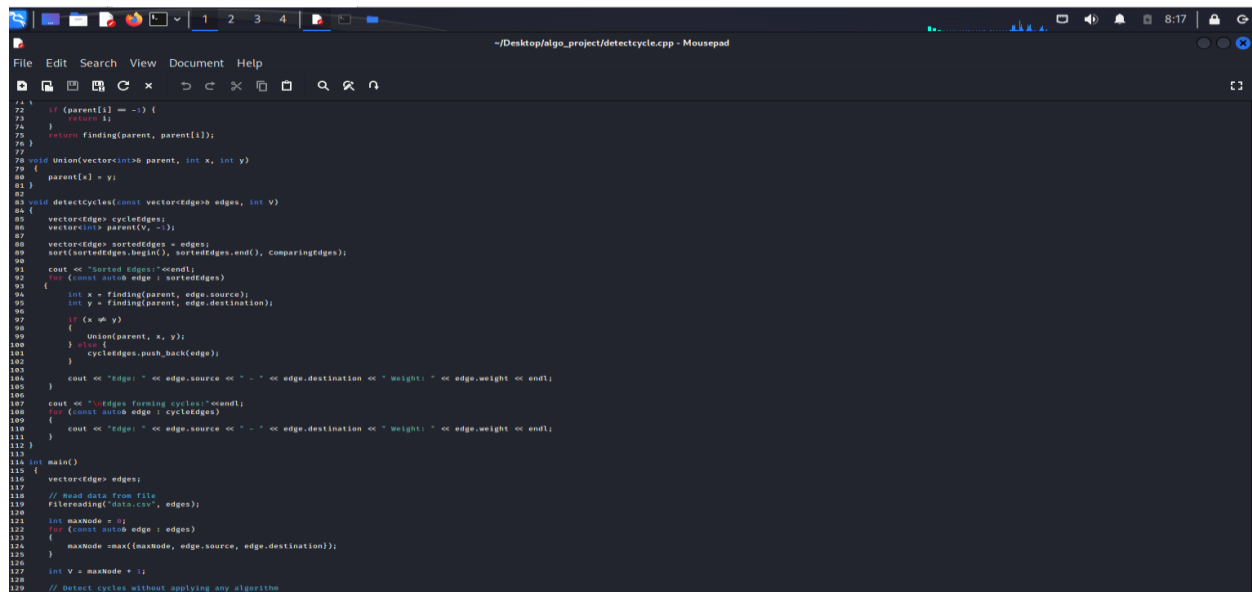
Outputs the connections that would create loops in the network. Highlights specific connections that, if taken, would form a loop.

- Time Complexity of Detect Cycle Algorithm:

- Best Case:  $O(E)$
- Average Case:  $O(E \log E)$
- Worst Case:  $O(E \log E + E)$  or  $O(E \log E + E \log V)$

- Analysis:

- File Reading:**  $O(E)$ , the function reads the edges from a file line by line and processes each edge once. The time complexity is linear with respect to the number of edges.
- Sorting Edges:**  $(E \log E)$ , the edges are sorted based on their weights using the sort function. This is the dominant factor in the overall time complexity. Sorting has a time complexity of  $O(E \log E)$ .
- Disjoint Set Operations:**  $O(E \log^* V)$ , the disjoint set operations involve finding the representative of a set (finding) and union operation. The time complexity is determined by the union-find operations during the Minimum Spanning Tree construction. The  $\log^* V$  term accounts for the efficiency of disjoint set operations.
- Detecting Cycles:**  $O(E \log E)$  the detect Cycles function utilizes sorting and disjoint set operations to find cycles in the graph.
- Writing Results to Output File:**  $O(E)$  The output of sorted edges and edges forming cycles involves printing each edge once.
- Overall Time Complexity:** The overall time complexity is dominated by the sorting of edges. Resulting time complexity of  $O(E \log E)$ .



```

1  2  3  4
~/Desktop/algo_project/detectcycle.cpp - Mousepad

File Edit Search View Document Help

72 if (parent[i] == -1) {
73     parent[i] = i;
74 }
75 return finding(parent, parent[i]);
76 }
77
78 void Union(vector<int>& parent, int x, int y)
79 {
80     parent[x] = y;
81 }
82
83 void detectCycles(const vector<Edge>& edges, int V)
84 {
85     vector<Edge> cycleEdges;
86     vector<int> parent(V, -1);
87     vector<Edge> sortedEdges = edges;
88     sort(sortedEdges.begin(), sortedEdges.end(), ComparingEdges);
89
90     cout << "Sorted Edges:" << endl;
91     for (const auto& edge : sortedEdges)
92     {
93         int x = finding(parent, edge.source);
94         int y = finding(parent, edge.destination);
95         if (x == y)
96         {
97             Union(parent, x, y);
98             cycleEdges.push_back(edge);
99         }
100         cout << "Edge: " << edge.source << " - " << edge.destination << " weight: " << edge.weight << endl;
101     }
102     cout << "Edges forming cycles:" << endl;
103     for (const auto& edge : cycleEdges)
104     {
105         cout << "Edge: " << edge.source << " - " << edge.destination << " weight: " << edge.weight << endl;
106     }
107 }
108
109 int main()
110 {
111     vector<Edge> edges;
112     // Read data from file
113     FileReading("data.csv", edges);
114     int maxNode = 0;
115     for (const auto& edge : edges)
116     {
117         maxNode = max(maxNode, edge.source, edge.destination);
118     }
119     int V = maxNode + 1;
120     // Detect cycle without applying any algorithm
121     detectCycles(edges, V);
122 }

```

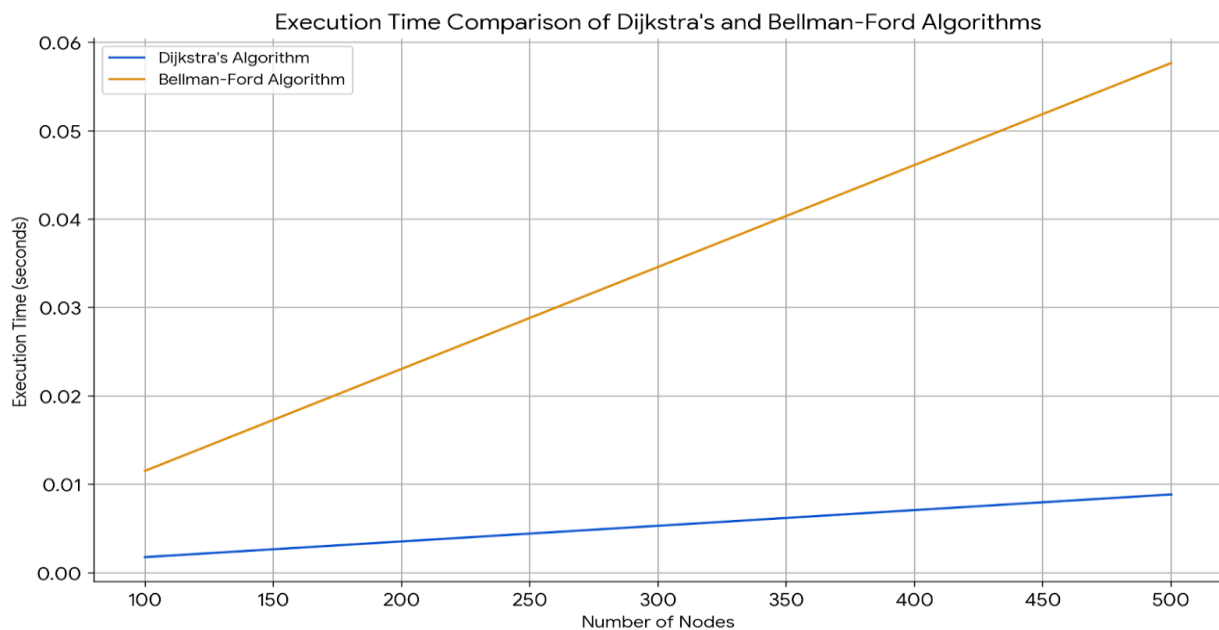
Output:

```
File Actions Edit View Help
Edge: 68 - 30 Weight: 160
Edge: 98 - 35 Weight: 160
Edge: 28 - 40 Weight: 160
Edge: 58 - 45 Weight: 160
Edge: 48 - 10 Weight: 160
Edge: 4 - 42 Weight: 160
Edge: 98 - 85 Weight: 160
Edge: 3 - 42 Weight: 160
Edge: 12 - 4 Weight: 160
Edge: 12 - 2 Weight: 160
Edge: 88 - 50 Weight: 160
Edge: 78 - 15 Weight: 160
Edge: 8 - 20 Weight: 160
Edge: 38 - 21 Weight: 160
Edge: 38 - 25 Weight: 160
Edge: 35 - 9 Weight: 175
Edge: 40 - 33 Weight: 175
Edge: 23 - 12 Weight: 175
Edge: 70 - 97 Weight: 175
Edge: 32 - 94 Weight: 175
Edge: 39 - 33 Weight: 175
Edge: 2 - 30 Weight: 175

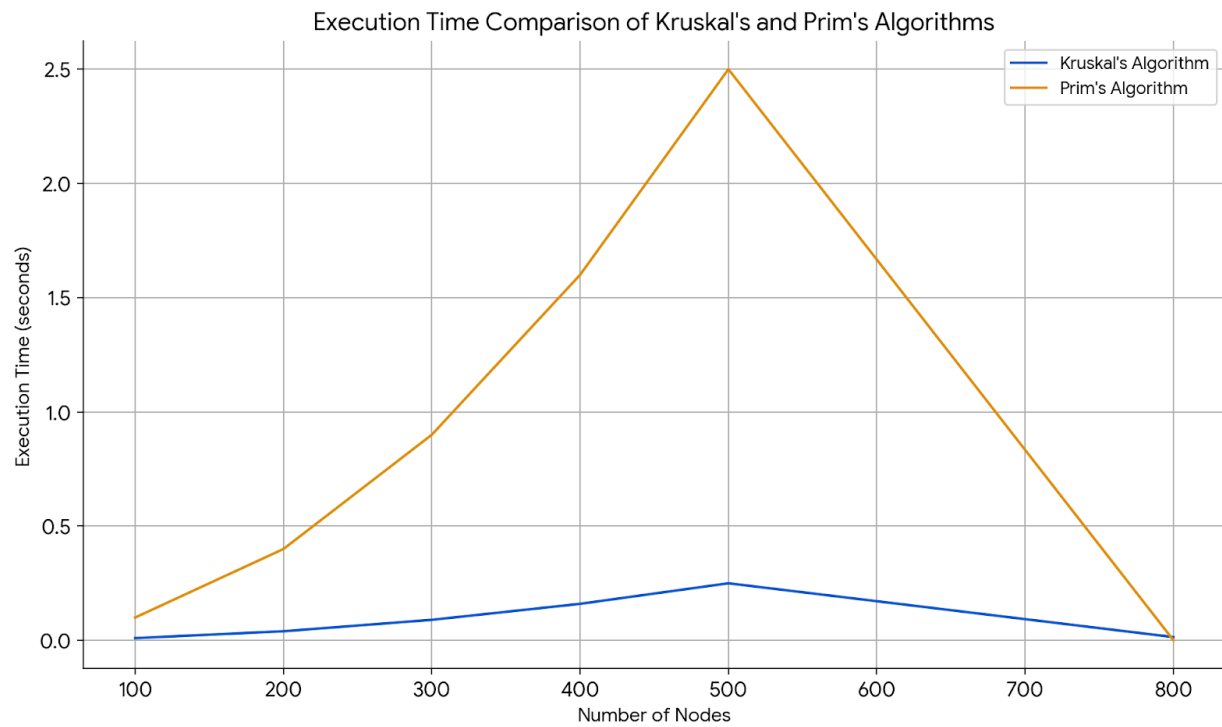
Edges forming cycles:
Edge: 2 - 0 Weight: 1
Edge: 8 - 3 Weight: 1
Edge: 2 - 0 Weight: 1
Edge: 4 - 3 Weight: 1
Edge: 8 - 3 Weight: 1
Edge: 4 - 3 Weight: 1
Edge: 8 - 7 Weight: 2
Edge: 8 - 16 Weight: 2
Edge: 0 - 4 Weight: 2
Edge: 12 - 50 Weight: 2
Edge: 0 - 4 Weight: 2
Edge: 15 - 5 Weight: 2
Edge: 18 - 25 Weight: 2
Edge: 3 - 1 Weight: 2
Edge: 1 - 10 Weight: 2
Edge: 0 - 3 Weight: 2
Edge: 3 - 1 Weight: 2
Edge: 8 - 14 Weight: 2
Edge: 15 - 7 Weight: 2
Edge: 3 - 7 Weight: 2
Edge: 4 - 5 Weight: 3
Edge: 5 - 0 Weight: 3
```

## GRAPHICAL COMPARISON

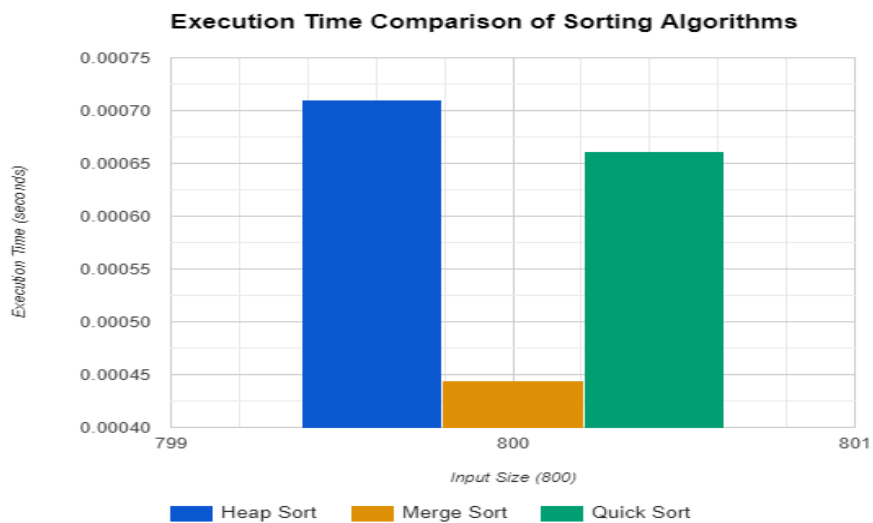
*Dijkstra and BellmanFord:*



### Prims and Kruskal:



### Merge Sort, Heap Sort and Quick Sort:



### *Implementation structure:*

- Dijkstra:

In Dijkstra code, stack was not used but priority queue is used. Using of queues is crucial in Dijkstra as it heavily relies on selecting and relaxing edges based on their weights. Absence of queues to handle vertex selection can degrade the algorithm's efficiency. In addition, we used priority queues that decreases the time complexity of the code. With a simple queue, the time complexity of Dijkstra's algorithm using an adjacency list is  $O(V^2)$  where  $V$  is the number of vertices. Using a priority queue the time complexity reduces to  $O((V + E) * \log V)$ , where  $E$  is the number of edges. This improvement arises from the faster extraction of minimum-distance vertices.

- Bellman-Ford:

In the Bellman-Ford algorithm, queues are used to trace the path of the shortest distance from the source node to each node in the graph. However, a stack is used along with the queue to help optimize the code in certain scenarios, especially when dealing with the tracing of paths.

- Prim's:

Prim's algorithm employs a priority queue to efficiently select the minimum-weight edge in each iteration. The priority queue stores pairs of vertices and their corresponding edge weights, facilitating the selection of the minimum-weight edge efficiently. The optimization in this code primarily lies in the utilization of the priority queue data structure within the context of Prim's algorithm, thereby contributing to the overall time complexity optimization.

- Kruskal's:

In the Kruskal's algorithm implementation there isn't explicit uses of queues or stacks for optimizing the time complexity. Instead, the optimization lies in the algorithmic approach itself. Kruskal's algorithm operates by sorting the edges based on their weights and then greedily selecting the smallest weighted edges that do not form a cycle in the Minimum Spanning Tree (MST). Kruskal's algorithm uses a disjoint set data structure with the "Union" and "Find" operations. These operations have nearly constant time complexity in practical scenarios, amortized to nearly  $O(1)$  per operation, leading to efficient merging of sets and finding connected components. The stack is used to keep track of the accepted or rejected edges along with their details. However, its primary purpose here is for creating the trace output rather than optimizing the algorithm's time complexity.

### *Conclusion:*

In summary, this comprehensive analysis of various algorithms and data structures showcases their diverse applications and optimizations. Dijkstra's algorithm, leveraging priority queues, excels in efficiently finding shortest paths in weighted graphs. Bellman-Ford algorithm, employing queues and stacks, stands resilient in handling scenarios with negative-weight edges. Prim's and Kruskal's algorithms, utilizing priority queues and disjoint sets, respectively, demonstrate their prowess in finding minimum spanning trees with efficiency. The sorting algorithms—Merge Sort, Heap Sort, and Quick Sort—highlight their varying time complexities, with Merge Sort standing out for its stability. The implementation structure emphasizes the significance of queues, priority queues, and disjoint sets in optimizing algorithmic efficiency. Overall, we underscore the importance of choosing the right algorithm and data structure for specific tasks, considering their time complexities and application contexts.

THE END....