BLG 223E - Assignment 3 Report

Melike Beşparmak - 150220061

May 2024

1 Introduction

In this assignment, we were given two datasets to create a graph. This graph's nodes are MIDs indicating names and the edges are URLs indicating the paths. Within this set, a map is created and the tasks are completed.

2 Methods

2.1 Part 1 - Hello Neighbor!

In this part, every neighbor of the node given its MID should be printed. Firstly, the wanted node can be accessed using the MID and the graph-map already created in the provided skeleton. Since the node structure has a vector for adjacent nodes, iterating through this "adj" vector is enough to reach every adjacent/neighbor node. Also, the number of neighbors is the size of the vector. The adjacency vector provides only MIDs, therefore I created a second global map to keep the names corresponding to MIDs. Using this map, the names are printed as wanted in the assignment.

2.2 Part 2 - Degree Centrality

Degree centrality considers the degree of the nodes. In other words, if the number of neighbors is x, the degree of the node is also x. So the method is to list every node according to the size of their adjacency vectors. I created a map (counter) that maps integers to strings and while traversing the graph-map, the number of neighbors (multiplied by -1) was mapped to MIDs. With this approach, every node was listed concerning their number of neighbors in descending order. I multiplied the number with -1 because the map structure keeps the entries in ascending order. However, since we want the most central 10 nodes, it is better to keep the number inverted to reach the nodes with maximum centrality starting from the beginning of the newly created map. Finally, by iterating the map 10 times, the most central 10 (can be modified) nodes can be printed. One problem of this approach can be the nodes with the same number of neighbors. They will be overwritten in the map and the

information will be lost, but since this probability is very low, it is not a crucial problem.

2.3 Part 3 - Shortest Distance

As we learned by heart, the shortest distance means a Breadth-First Search (BFS). In this BFS algorithm, the steps are as follows:

1) The starting node will be pushed to the queue

2) Pop top, mark it visited, make it current node

3) If not visited, push the adjacent nodes

4) If the adjacent node is the target node, mark found flag as 1

5) To keep track of the path, create a map (parents) to match the relation with the pushed node-the node used to reach that node

6) Print the map

One important thing to note is that marking the nodes visited is crucial since the nodes can be mapped as parents of each other and cause a loop when printing the path.

3 Results

```
test@oop_docker:~/hostVolume/data/data_hw3$ ./skeleton part1 /m/04mx8h4
29 neighbors
/m/0146mv Nickelodeon (TV channel)
/m/09c7w0 United States
/m/0cc816d Daytime Emmy Award for Outstanding Childrens Animated Program
/m/04mlh8 Jeff Bennett
/m/04mlh8 Jeff Bennett
/m/0dszr0 Nicole Sullivan
/m/022s1m John DiMaggio
/m/0hcr Animation
/m/0cc816d Daytime Emmy Award for Outstanding Childrens Animated Program
/m/04mlh8 Jeff Bennett
/m/0hcr Animation
/m/0ckd1 Executive producer
/m/01htzx Action (fiction)
/m/0pr6f Children's television series
/m/0146mv Nickelodeon (TV channel)
/m/0gkxgfq 38th Daytime Emmy Awards
/m/0347db Neil Patrick Harris
/m/0gkxgfq 38th Daytime Emmy Awards
/m/03k48_ Andy Richter
/m/06n90 Science fiction
/m/04mlh8 Jeff Bennett
/m/0347db Neil Patrick Harris
/m/03k48_ Andy Richter
/m/0725ny Kevin Michael Richardson
/m/01htzx Action (fiction)
/m/0cc816d Daytime Emmy Award for Outstanding Childrens Animated Program
/m/0725ny Kevin Michael Richardson
/m/0ckd1 Executive producer
/m/05p553 Anarchic comedy film
```

Figure 1: Part 1 Example Run

```
test@oop_docker:~/hostVolume/data/data_hw3$ ./skeleton part1 /m/0fk1z
19 neighbors
/m/03ttfc Spanish people
/m/04dbw3 Cuban American
/m/043zg Jennifer Lopez
/m/09v5bdn Puerto Rican people
/m/0146bp Diabetes mellitus type 2
/m/0146bp Diabetes mellitus type 2
/m/03ttfc Spanish people
/m/05zjd Portuguese language
/m/025rpb0 Hispanic and Latino Americans
/m/0c58k Diabetes mellitus
/m/09v5bdn Puerto Rican people
/m/051q39 Rafael Nadal
/m/01g7zj Mexican American
/m/01g7zj Mexican American
/m/0t_2 American English
/m/0c58k Diabetes mellitus
/m/04dbw3 Cuban American
/m/025rpb0 Hispanic and Latino Americans
/m/06nm1 Spanish language
```

Figure 2: Part 1 Example Run

```
test@oop_docker:~/hostVolume/data/data_hw3$ ./skeleton part2
9606 /m/09c7w0 United States
6366 /m/09nqf United States dollar
5526 /m/04ztj Marriage
4512 /m/02hrh1q Actor
3927 /m/0jbk9 United States Department of Housing and Urban Development
3796 /m/02sdk9v Forward (association football)
3743 /m/02nzb8 Midfielder
3566 /m/02_j1w Defender (association football)
3102 /m/0dgrmp Goalkeeper (association football)
2999 /m/05zppz Male
```

Figure 3: Part 2

```
test@oop_docker:~/hostVolume/data/data_hw3$ ./skeleton part3 /m/0fk1z /m/0fhp9
The shortest distance from Hispanic to Vienna is 3
The path:
/m/0fk1z Hispanic
/people/person/ethnicity
/m/043zg Jennifer Lopez
/people/marriage_union_type/unions_of_this_type./people/marriage/spouse
/m/04ztj Marriage
/people/marriage_union_type/unions_of_this_type./people/marriage/location_of_ceremony
/m/0fhp9 Vienna
```

Figure 4: Part 3 Example Run

```
test@oop_docker:~/hostVolume/data/data_hw3$ ./skeleton part3 /m/0fqz6 /m/0fs44
The shortest distance from Cajun to Rabat is 3
The path:
/m/0fqz6 Cajun
/people/ethnicity/languages_spoken
/m/064_8sq French language
/location/country/languages_spoken
/m/04wgh Morocco
/location/country/capital
/m/0fs44 Rabat
```

Figure 5: Part 3 Example Run

4 Discussion

The given skeleton code was very helpful since it was already reading the files, creating the graph and, giving a clue on how to complete the given tasks. Even though, there was a problem in the file reading part. The modified part is below.

```
string ent1 = line.substr(0, line.find_first_of("\t\r"));
string remain = line.substr(line.find_first_of("\t\r")+1,line.length()-ent1.length());
string relationship = remain.substr(0, remain.find_first_of("\t\r"));
remain = remain.substr(remain.find_first_of("\t\r")+1, remain.length()-relationship.length());
string ent2 = remain.substr(0, remain.find_first_of("\t\r"));
```

Figure 6: Added /r and changed the last ”line” as ”remain”

The code does not work very fast and it is due to the dataset, even reading the file takes some amount of time, and traversing through the created maps has the same problem.

5 Conclusion

In this assignment, I learned more about STL structures, especially maps and vectors. I also did some research on the Freebase project and grasped an idea of real-world applications.