

BLG223E

Homework-3

02.06.2024

Dr. Yusuf Huseyin Sahin
sahinyu@itu.edu.tr

- Please write your own codes, copying code parts from books, websites or any other source including your friends is considered as plagiarism and results in penalty. Also disciplinary actions will be taken.
- Do not upload your codes to any public platform (e.g. Github) until the deadline of homework passes.
- Your code should be able to be compiled with default g++ compiler and run under Ubuntu OS. **Even if you write your code on a different OS, you should check it via ITU SSH.**
- Do not forget to comment your code.
- Submit your source codes and report files on Ninova before the deadline, late submissions and submissions via e-mail will not be accepted.
- If you have any questions, please use the message board.

1 Overview

WALTER: You know someone like that?

SAUL: Let's just say I know a guy who knows a guy, who knows another guy.

Breaking Bad, S2 E11

Freebase, developed by Metaweb, was a large collaborative knowledge base consisting of data composed mainly by its community members. It was an online collection of structured data harvested from many sources. Freebase, was acquired by Google in 2010. Google utilized it to enhance its Knowledge Graph, improving search results with semantic information from diverse sources. In 2015, Google announced the discontinuation of Freebase.

Freebase consists of triplets, where each record shows two entities and the relationship between them. For example, the following record indicates that The Sopranos (*m\0kfv9*) contains the actor Steven Van Zandt (*m\01l1sq*).

```
1 /m/0kfv9 /tv/tv_program/regular_cast./tv/regular_tv_appearance/actor /m/01
  1lsq
```

All entities are represented by MIDs, however multiple strings could be represented by the same MID if they direct the user to the same Wikipedia page. For example, for the searches "Erehwon Prison" and "Face/off", Wikipedia directs the user to the same page. In the following list, all strings used to direct the user to this page is given. In this homework, you can use only the first appearance.

```
1 /m/02_fz3 Erehwon Prison
  /m/02_fz3 Face/off
3 /m/02_fz3 Face off (film)
  /m/02_fz3 Sean Archer
5 /m/02_fz3 Castor Troy
```

With the homework document two different text files are given:

- **"mid2name.tsv"** contains MIDs and their textual equivalents.
- **"freebase.tsv"** contains the records for the relationship graph.

A Node structure could be created to store every node of the graph.

```
1 struct Node {
    string MID;
3    vector<Node*> adj;
    vector<string> relation;
5 };
```

Here *adj* stores the pointers to the its neighbors while *relation* stores the type of the relationship.

To make the search by MID in a easy way, we could define a map from MIDs and Node pointers. (**Hint:** You are free to create other maps to store more specific information.)

```
1 map<string, Node*> graph_map = {};
```

2 Hello neighbor! (30 pts.)

For a given MID, print MIDs and names of all its neighbors. For this part, your code should be run by ".\main part1 MID1" command where MID1 represents an MID. An example run is given below.

```
1 .\part1 /m/04mx8h4
  29 neighbors
3 /m/0146mv Nickelodeon (TV channel)
  /m/09c7w0 United States
5 /m/0cc816d Daytime Emmy Award for Outstanding Childrens Animated Program
  /m/04mlh8 Jeff Bennett
7 /m/0dszr0 Nicole Sullivan
```

```
9 /m/022s1m John DiMaggio
  /m/0hcr Animation
  ...
```

3 Degree centrality (30 pts.)

Find the most central 10 entities according to degree centrality and print them on the screen. For this part, your code should be run by `".\main part2"` command.

4 Shortest distance (40 pts.)

For given two MIDs, find the shortest distance between them. Print both the distance and the full path. Your code should be run by `".\main part3 MID1 MID2"` command.