Social Network Analysis for Computer Scientists

Assignment 1

Sevak Mardirosian
s2077086
s.mardirosian@umail.leidenuniv.nl

October 1, 2017

1 INTRODUCTION

This paper is relevant to SNACS course at Leiden University and specifically contains the answers for the first assignment.

1.1 Organization

In section 2 I will try to answer the theoretical questions regarding Graphs. In section 3 I will partially reference the reader to the zip file which contains working scripts. I will, in few words, try to explain the ways I obtained my answers using which tools and techniques.

1.2 Tools and Techniques

For the programing parts I will be using Python 3.5 scripting language explicitly with libraries such as **networkx**, **matplotlib**, **graph tool** and **pickle** which is already built-in in Python 3.*.

2 Theoretical Questions

2.1 Question 1.1: Give a formal definition of the in-degree and out-degree of a node using the notion of a (reversed) neighborhood.

The in-degree of a node (v) basically means the number of edges that are in the reverse neighborhood. edges of the set N'(v). However, The out-degree of node v can be expressed as the number of edges that are in the neighborhood (edges of the set N(v)).

2.2 Question 1.2: Define the reciprocity of a directed graph using the notion of (reversed) neighborhoods.

A vertex pair (A, B) is said to be reciprocal if there are edges between them in both directions. The reciprocity of a directed graph is the proportion of all possible (A, B) pairs which are reciprocal, provided there is at least one edge between A and B.

2.3 Question 1.3: The radius of a connected undirected graph is the minimal eccentricity (which is the maximal distance to another node from that particular node) over all nodes in the graph. How can we use the neighborhood function to define the radius?

I was not able to answer this question.

2.4 Question 1.4: What type of graph could we be dealing with if for all pairs of nodes u, v $\in$ V we have: |N k (u)| = |N k (v)| for any k $\geq$ 0?

In this case we might be dealing with a complete graph, mostly because u and v are directly connected to each other.

2.5 Question 1.5: Give an algorithm that uses the neighborhood function to count the number of cliques of size 4 in a connected undirected graph. What is the time complexity of your algorithm?

I was not able to answer this question.

2.6 Question 1.6: Write an algorithm that computes the density of the graph

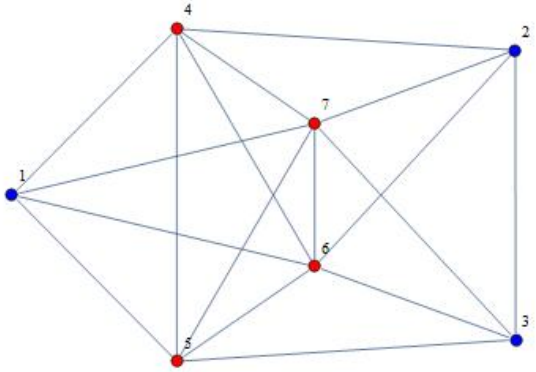
For undirected graphs, the graph density is defined as

$$D = \frac{2|E|}{|V|(|V|-1)}$$

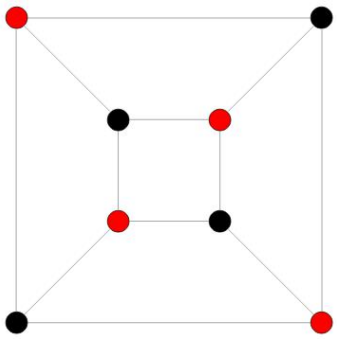
Where |E| is the number of edges and |V| is the number of vertices in the graph.

2.7 Question 1.7: Give an algorithm to check whether the graph is bipartite.

A bipartite graph is possible if the graph coloring is possible using two colors such that vertices in a set are colored with the same color. After, grouping the vertices in two sets, if those vertices in each set are not interconnected, then the graph is bipartite. Below an example.We can verify that it is not bipartite, because we can group 1,2,3 in one set and 4.5.6.7 into another but the vertices in this last set are interconnected.



An example of a bipartite graph



Below the steps of an algorithm to check if a graph is a bipartite using Breadth first search (BFS):

1. Assign RED color to the source vertex (putting into set U).
2. Color all the neighbors with BLUE color (putting into set V).
3. Color all neighbor's neighbor with RED color (putting into set U).
4. This way, assign color to all vertices such that it satisfies all the constraints of m way coloring problem where m = 2.
5. While assigning colors, if we find a neighbor which is colored with same color as current vertex, then the graph cannot be colored with 2 vertices (or graph is not Bipartite).

If graph is represented using adjacency list, then the complexity becomes O(V+E).

3 Practical Questions

3.1 Question 2.1 How many directed links does this network have?

Using the Networkx library I will be using number-of-edges function to retrieve number of edges from each dataset.

- medium: 6452
- Large: 1751463

For the implementation part I refer you to the **Question2-1.py** python file which could be found under the Questions directory.

3.2 Question 2.2 How many nodes does this social network have?

For this part I have used number-of-nodes function to retrieve number of nods in each dataset.

- medium: 2239
- Large: 384413

For the implementation part I refer you to the **Question2-2.py** python file which could be found under the Questions directory.

3.3 Question 2.3 Give the in-degree and out-degree distribution of this graph.

For this I have used the in-degree and out-degree function which Networkx has to return number of edges pointing in to the node and the other way around. Using these two functions I fetch the values of each of those sets and feed each one of them into a list and afterwards to a plot.
For further exploration I refer you to the **Question2-3.py** python file which could be found under the Questions directory.
In **Diagrams** folder the following PNG files are relevant to this question.

- Medium In Degree Distribution
- Medium Out Degree Distribution
- Large In Degree Distribution
- Large Out Degree Distribution

3.4 Question 2.4 How many weakly connected components and how many strongly connected components does this network have? How many nodes and edges are in the largest strongly connected component of this graph?

I have used number weakly connected components function to fetch the following answers.

- Number of weakly connected components Medium: 9
- Number of weakly connected components Large: 6751
- Number of strongly connected components Medium: 2218
- Number of strongly connected components Large: 361187

In combination with number of nods/edges I have used strongly connected component subgraphs to get answers to the following questions.
How many nodes are in the largest strongly connected component?

- Medium Network: 8
- Large Network: 16208

How many edges are in the largest strongly connected component?

- Medium Network: 18
- Large Network: 70802

For further exploration I refer you to the **Question2-4.py** python file which could be found under the Questions directory.

3.5 Question 2.5 Give the exact or approximated distance distribution of the largest weakly connected component of this graph as a diagram.

At first I was struggling getting the required answer for this question using the networkx lib. After a bit of research I discovered graph tool which is essentially written in C++ but has a python library available. I have also used pickle python library to save the data as a byte stream, given how big the **large.in** file is so I can run it smoothly without running out of memory which was one of issues I experiencing as well.
In **Diagrams** folder the following are relevant to this question.

- Medium Distance Distribution
- Large Distance Distribution

For further exploration I refer you to the **Question2-5.py** python file which could be found under the Questions directory.

3.6 Question 2.6 Visualize the social network in medium.in

For this part I have made use of gephi software to make the visualization. beforehand I have wrote a small python script to add directed property along with source and target. After I imported the file I applied the force Atlas 2 algorithm to the dataset and afterwards added gravity to avoid escaping few nodes to the void (I had issues such as no matter what parameters I used a few nodes just kept escaping into the void and wasn't able to get the proper image I was after). I chose In-Degree param for the nodes and re-run the algorithm again to take shape. After stabilizing I added 'Prevent overlap' option to create a nice clustered nodes. For the rendering I added thickness to the edges in order to have a better look overall and checked the curves option as well. The PNG file and the gephi files could be found under gephi folder in the zip file.

References

For question 1.7 I made use of the following links to obtain an answer to the question.
socratic.org
stackexchange-math
hrishikeshmishra.com
geeksforgeeks.org

For question 1.6 I made use of the following links to obtain an answer to the question.
What is the definition of the density of a graph?
Wiki - Dense graph

For question 1.2 I made use of the following links to obtain an answer to the question.
Reciprocity of graphs