Nidhi Parekh

Professor Dugas

Online Social Networks

9 March 2022

I pledge my honor that I have abided by the Stevens Honor System. – Nidhi Parekh

Assignment 4: Graphs

PURPOSE:

Parekh.py is a program that processes .csv files and considers positive(trust) and negative (distrust) relationships in a network with triads. The data given from the epinions csv files creates a graph that represents an undirected, signed network and calculates any sort of specific data from that graph.

INPUT:

When running the program, the user is first asked what file the user would like to read. Here is an example:

```
----------------------------------------------ASSIGNMENT 4: GRAPHS----------------------------------------------

What file should I process? epinions1.csv
```

OUTPUT:

The output in the console prints out the data specified in the requirements. The output reveals the number of triangles (as well as the number of triangles per type) and edges used to identify triads, the number of trust and distrust edges, and the probability of that an edge is positive or negative. Additionally, the console prints out of the expected distribution and actual distribution of triad type. Here is an example from one of the screenshots:

```
*** START ***

triangles: 59680
TTT: 46522
TTD: 6837              trust edges:     26559                probability %:  84.87%
TDD: 5557              distrust edges:  4736                 probability %:  15.13%
DDD: 764              total:           31295                                100.00%

Expected Distribution                Actual Distribution
       Percent     Number           Percent           Number
 TTT:  61.12         36478.68           TTT:  77.95          46522
 TTD:  32.70        19514.63           TTD:  11.46           6837
 TDD:  5.83         3479.85           TDD:  9.31           5557
 DDD:  0.35          206.84           DDD:  1.28            764
 Total:  100.00     59680.00          Total:  100.00         59680

Duration: 5.501293182373047 seconds

*** END ***
```

Nidhi Parekh

Professor Dugas

Online Social Networks

9 March 2022

I pledge my honor that I have abided by the Stevens Honor System. – Nidhi Parekh

WHAT THE PROGRAM DOES:

The program starts off by asking the user to input a file that the user would like to process. First, the user has to write the name of the csv file. Afterwards, the program processes(or reads) the selected file. While traversing through the dataset, the first and second column of the csv file are generated as nodes and given that the third column is weights, which is either -1 or 1, it gets split into two conditions: if the weight of the edge between two nodes is -1, it creates an edge between the two nodes, with weight being -1, and it keeps a count on the number of edges created as well as the positive and negative edges. This also applies for when the weight is -1.

Next, the program traverses through the list of all cliques made in graph G, and finds all of the triads, which is if any list inside the list of cliques has a length of 3 and keeps a count on the number of triads made. Then, it is time to distinguish the type of triads being formed (TTT, TTD,TDD,DDD). Depending on the weights, for example, if they add up to 3, that means the type of triangle is TTT. The program checks for the total weight for each triangle and categorizes by the sum.

Afterwards, the program prints the outcomes for the number of triangles and the number of triad types. Then, the expected and actual distribution of the triad types are calculated, both percentages and the expected/actual number of triad types. The result of the distributions gets printed out into the console. Finally, since the assignment asks for the run time(duration) of the file. This can also be seen in the screenshots.

RESULTS:

When looking at the output printed on the console, it can be seen that the actual distribution of triad type differs from the expected distribution based on random assignment. Usually, in statistics, the actual value is the value obtained by observation/measuring available data. In this case, the expected distribution is calculated based on the probability of the trust and distrust edges. Because we are using that probability, it is does not necessarily mean that it will be that way in the data. However, the actual distribution is calculated directly from the dataset. Another thing noticed is that the gap between expected distribution and actual distribution results in epinions1.csv is closer than for epinions0.csv. It's likely that the more data provided, the closer the gap between the expected and actual distribution.

Nidhi Parekh

Professor Dugas

Online Social Networks

9 March 2022

I pledge my honor that I have abided by the Stevens Honor System. – Nidhi Parekh

ANY ADDITIONAL INFORMATION:

When I first looked at this assignment, I thought that we had to create everything like the graph from scratch, so I had trouble with figuring out how to start writing the code. Then I realized about the library, Networkx, and that very much helped with the assignment. The main issue I have is getting epinions2.csv to work. Due to the size of the file, I keep receiving a 'Memory Error'. Luckily, I was still able to use epinions0 and epinions1.