Preet Dabhi 10459151

Assignment 4

PURPOSE: The purpose of this assignment is to extract data from the csv files and do analysis along with creating graphs, calculate all possible triangles, distribution of triads, and the probability and compare those values with actual and expected values.

INPUT: The input for the program will be CVS files, where users can choose from 3 files epinisons0.csv, epinisons1.csv, epinisons2.csv

OUTPUT: The output of the program will consist of a bunch of details starting with the program start time, total number of triangles, total number of edges created using network, number of trusted edges, number of distrust edges, all possible TTT (trust – trust - trust) triads, TTD (trust – distrust - distrust) triads, TDD (trust – distrust - distrust) triads, DDD (distrust – distrust - distrust) triads. The program also calculates the for the possibility of trusted edges and distrust edges, it also calculates the expected and actual distribution of triads, and in the end displays the total time the program took to run

WHAT THE PROGRAM DOES: On running the program it asks the user for a csv file, after that the program iterates over all the data, we initialize a network graph and store all the nodes in the graph and calculate all the possible triads, once all the triads are calculated we assign all the weights/edge values, the ones with -1 it is distrust and ones with 1 is trust. With all this data it stores all the data into an array, and we try to calculate the probability of the trust and distrust edges. Thereafter we calculate the expected and the actual triads values and we see that the expected and actual data differs. In the end we print the calculated data in console.

RESULTS:

The results for the program give all the necessary information that is required such as the number of triangles, probability and expected and actual distribution of triads. We can see that the numbers for the actual and expected do not match. This can be because of when calculating the expected distribution, we assume that there is a uniform distribution of all the values in the csv file.

ANY ADDITIONAL INFORMATION:

Tried running the epinions2.csv it took a lot of time to execute so stopped the program heated up my laptop as well