Brandon Patton

Online Social Networks

Assignment 05 Graphs: Report

I. Difficulties, General Thoughts

For this assignment I chose to approach the task by splitting up the requirements into as many functions as I could in order to simplify them. Although doing so did help in that regard I also found myself editing each functions parameters to allow for more arguments to perform the desired operations, such as in the function trust_eval. In the future, it may be more optimal to have made one overarching function to perform all the tasks but for my understanding and learning splitting up the tasks between functions did help add simplicity. The epinions.csv file was also too large to process and could not finish after around 30 minutes. The other two were able to finish and I have provided the required information in the screenshots included.

II. Program

I imported networkx as suggested in the specifications, as well as pandas and combinations (from itertools). I used network for any graph specific tasks and for identifying and readying the triads for output. Pandas was useful for reading in each csv file as a DataFrame rather than treating each line as its own string and running through each file line by line. Also, using pandas to read in the csv files allowed me to make custom headers for each column in the files, which I set to “reviewer”, “reviewee”, and “trust” corresponding to each’s respective column in the files.

To ascertain the amount of self-loops, I initially used networkx.graph.nodes_with_selfloops() which would have returned a list containing the nodes with self loops, but the current version of networkx does not support this attribute of the Graph object so I did this instead by checking the first two elements of each graph in the csv file to see if they were the same and increasing the count if they were.

III. Actual vs. Expected Distribution

Since the expected distribution of triad types relies on probabilities, it is natural that it would end up being different than the actual distribution which is based on actual data. In this way, a negative or positive edge could drastically impact the probability of seeing another of either.

IV. Screenshots:

In the included screenshots, I was able to include the triads of the epinions96.csv file but not the epinions_small.csv or epinions.csv files. The epinions_small.csv file’s triad list was far too long to include in a screenshot and the epinions.csv file did not finish listing the triads after around 30 minutes so I could not acquire the actual distribution for the latter, but I was able to include the both the actual and expected distributions for the first two files.