ReadMe

Anomaly Detection in Time-Evolving Networks

Name: Venkata Pranav Singaraju
UnityId: pvsingar
Student Id: 200404472

Research Paper: I have implemented an algorithm from paper 2 (DELTACON: A Principled Massive-Graph Similarity Function)

Datasets:

There are four datasets in the datasets folder: autonomous, enron_by_day, p2p-Gnutella, and voices. This data was primarily taken from the Stanford Large Network Dataset Collection. These are four time-evolving graphs from different domains.

Python Version: Python 3.5.6

Python Libraries: numpy, scipy, random, matplotlib

Instructions to execute:

Please ensure that the data folders (autonomous, voices, enron_by_day, p2p-Gnutella) are in the dataset/dataset folders, (relative path: /dataset/dataset/autonomous)

Please execute the anomaly.py file as

1. To run on autonomous data files: **python anomaly.py autonomous**
2. To run on voices data files: **python anomaly.py voices**
3. To run on enron_by_day data files: **python anomaly.py enron_by_day**
4. To run on p2p-Gnutella data files: **python anomaly.py p2p-Gnutella**

By executing the above commands, a corresponding output text file with similarity value time series will be generated. A png file will be generated as well which contains the plot. For example, if we execute for the autonomous dataset, then autonomous_time_series.txt and autonomous_time_series.png will be generated. All these are stored in the output folder.

Algorithm Running Time:

Algorithm	**Time (sec)**
autonomous	1730
voices	1.325

ReadMe

enron_by_day	36
p2p-Gnutella	3460.18

Depending on the size of the graph and the number of graphs, the runtime may vary. From the above table, we observe that for the voices dataset, the running time is the least, and for p2p-Gnutella it is the highest. These were executed on Macbook Pro (16GB RAM, M1 Chip)

By changing the g (group count) value on the voices dataset, the algorithm is executed and all the corresponding files are generated and stored in the output_g folder.

g (group count) value	Time (sec)
10	1.389
20	1.325
30	1.326
40	1.362
50	1.314