P4 - Graph Data Mining

Vivek Reddy Karri (vkarri, ID: 200315262)

Anomaly Detection in Time Evolving Networks

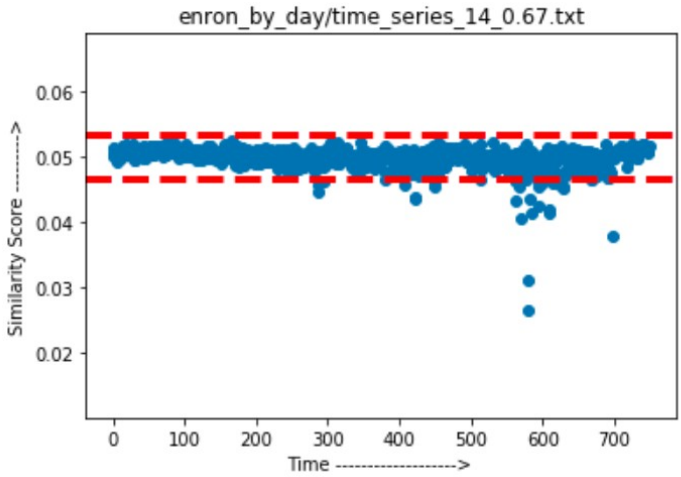
November 5, 2019

A) Choosen Research Paper:
Paper-2 - DELTACON: A Principled Massive-Graph Similarity Function

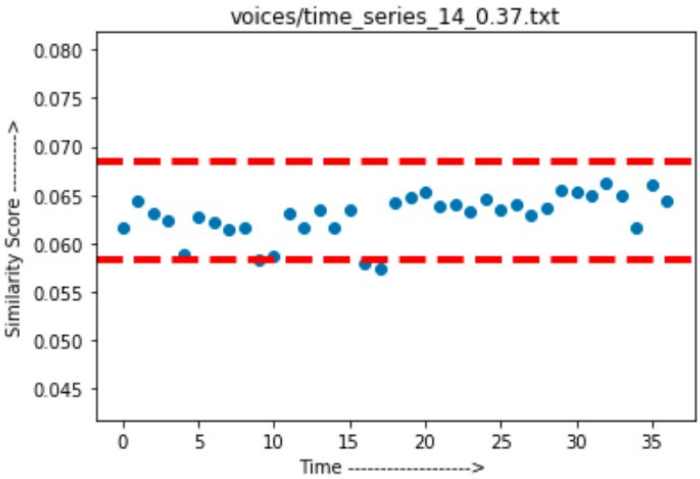
B) Hyper-Param Search:
The Value of g and epsilon are the two hyper-parameters for this similarity function.
According to my observations:
1) G is usually for computational benefits and doesnt affect similarity scores much if it is changed by lower margins.
2) Similarity scores are very sensitive to epsilon.
3) After observing the results over various hyper-params, only one of the combinations used to generate the final output files.

C) Visualization of Similarity Scores along with thresholds

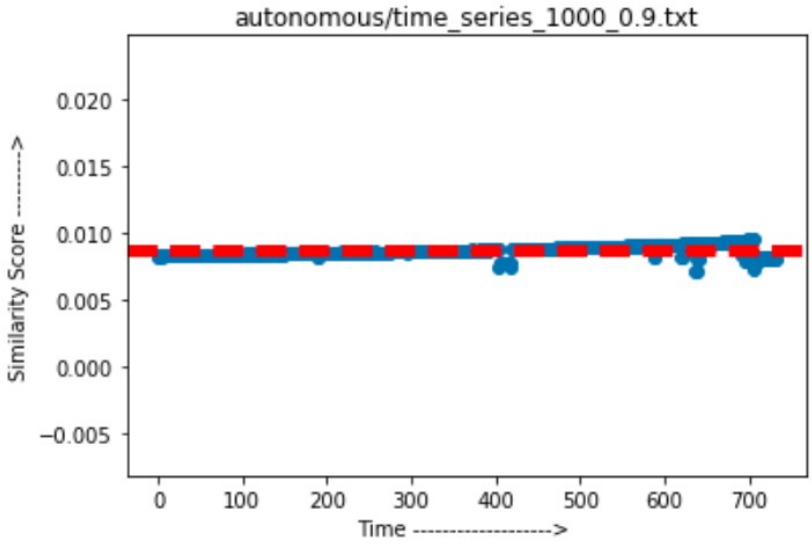
1) Enron_by_day: {g : 14, epsilon : 0.67}



2) Voices: {g : 14 epsilon: 0.37}



3) Autonomous: {g : 1000, epsilon : 0.9}



4) **p2p-Gnutella**: {g: 300, epsilon: 0.19}