Alexandria University
Faculty of Engineering
Computer and Communications Program

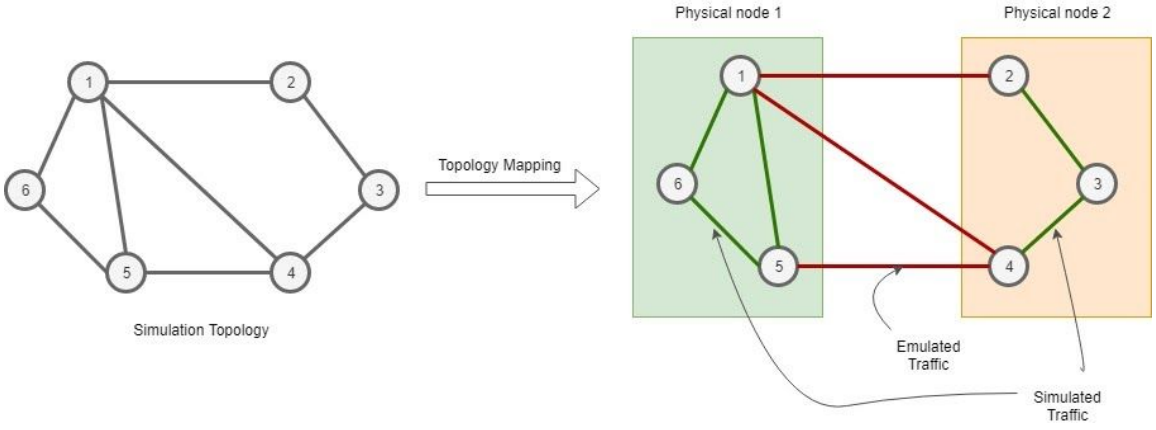
Due : 3/26/2019
CCE :Pattern Recognition

Assignment#2 Topology Mapping (Total 60Points)

Problem Statement

Simulating large-scale network experiments requires powerful physical resources. However, partitioning could be used to reduce the required power of the resources and to reduce the simulation time. Topology mapping is a partitioning technique that maps the simulated nodes to different physical nodes. In this assignment, we will use spectral clustering to partition a given network topology on the available physical nodes.

The network topology is a graph of N nodes communicating with each other by sending data traffic through a set of edges. An edge in the topology is weighted by the traffic (Mbps) passing through it. Our clustering technique should find the cut that minimizes the traffic between different partitions (Emulated traffic in the figure below).



1. Dataset

a. The dataset contains 30 network topology with 3 topology sizes (10 simulated nodes, 50 simulated nodes, and 100 simulated nodes)
b. Each topology is described in a .txt file.
c. Each line in the file is an edge in the topology: <from, to, traffic (Mbps)>
d. Ground truth clustering is available for topologies with sizes 10 and 50. The <ground_truth.txt> file contains a line for each topology ordered by the topology name: t_10_0, t_10_1, …, t_50_0, t_50_1, …, t_50_9
e. The ground truth line consists of N values (number of nodes), where nodes having the same value are in the same cluster.

2. Visualize the Topologies (10 points)

Write a visualization function that takes the description of a topology and visualizes it as a graph with edges labeled by their traffic amount. Furthermore, the function should visualize the topology before and after partitioning, where a group of nodes in the same cluster should have the same color.

3. Topology Mapping (20 Points)(your implementation)

a. Given the topology, build the similarity matrix A with edges' weights, and the degree matrix $\triangle$.
b. Implement and run the spectral clustering algorithm with K values {2,4,6,8,10} for topologies with size 10, and {2,10,12,15,17,20,25} for 50 and 100.

4. Evaluation (20 Points)(your implementation)

a. For topology sizes 10 and 50, use the external evaluation techniques: Conditional Entropy, and F-Measure.
b. For topologies with size 100, use the Normalized-Cut evaluation technique, where the distance between two nodes is $e^{-0.01w}$ where w is the weight of the edge between them, and 1 if there is no edge between them.

5. Big Picture (10 Points)

a. For each given topology, run the clustering with the different values of K.
b. Apply the evaluation for each clustering found.
c. Report and visualize the best clustering based on the mentioned evaluation techniques.
d. Report the amount of traffic running internally in each individual partition, and the total amount of emulated traffic between different partitions (sum of the cut weights between different clusters).

GOOD LUCK