Project 5 – Centrality of nodes in networks

Due date Nov. 28th 11.59pm

This programming assignment is about computing the centrality properties of the nodes of a Protein-Protein Interaction (PPI) network. Recall that a PPI network is represented by a graph G=(V,E) where nodes of V represent proteins and an edge of E connecting two nodes represents interacting proteins (either physically or functionally).

You will compute two properties: **Closeness centrality** and **Eccentricity**

A detailed description of closeness centrality and eccentricity follows.

The **Closeness centrality** of a node is a measure of centrality that uses information about the length of the shortest paths within a network; it uses the sum of the shortest distances of a node to all other nodes. Precisely, the closeness-centrality of node u is defined as the reciprocal of this sum:

where is the length of a shortest path between u and v.

The **Eccentricity** of a node is defined as:

Define the **center** of a graph as the node with minimum eccentricity.

Your program computes for each node its closeness centrality and eccentricity value.

**Input:** the Protein-Protein Interaction (PPI) graph of the herpes Kaposi virus. The file kshv is available on canvas.

**Output:**

1. Print the top 3 nodes (sorted) with the highest closeness centrality value and for each of them print its closeness centrality value.
2. Print a shortest path between the two nodes with highest and lowest closeness centrality. For that, you need to have *reconstruct*ed the shortest path between the two nodes.
3. Using cytoscape, display:
   1. the top 3 nodes (sorted) with the highest closeness centrality
   2. the center of the graph, i.e. the node with highest eccentricity value. Color the center with a different color.

Electronically submit 1). ipython notebook. 2) screen shot of cytoscape display