Title: Do an analysis of the social graph as per direction provided   : 20 Marks (will be scaled down to 10)

Goal: *Goal of this assignment is to have a hands on exposure to Social Network Analytics using Networkx and Matplotlib in Python.*

Team :
1. **It is a group activity by MAXIMUM <u>THREE</u> members.**
2. **THE TEAM HAS TO BE THE SAME TEAM FOR ASSIGNMENT1, ASSIGNMENT 2 and PROJECT.**
3. ***You need to mail a python notebook clearly indicating your team members' names and SRN(s).***

Required background:
   a. Google Drive -> code_used_in_class
   b. Dataset provided in Google Drive

Instruction:

(a) **Dataset :** You will attempt to do a social network analysis of the famous French novel **Les Miserables (**Victor Hugo) using its dataset. The file lesmis.gml contains the weighted network of coappearances of characters. Nodes represent characters as indicated by the labels and edges connect any pair of characters that appear in the same chapter of the book. The values on the edges are the number of such coappearances. The data on coappearances were taken from D. E. Knuth, The Stanford GraphBase.
(b) **First read up about the novel** i.e. overall plot from the given link or elsewhere on the net (http://www.sparknotes.com/lit/lesmis/summary/). You will need this in the analysis.
(c) Provide the followings in the python notebook:
   a. Do a visualization with networkx layouts
   b. List the nodes and edges with their attributes
   c. Do a centrality analysis (degree, closeness, betweenness, eigenvector) of all the nodes
   d. Calculate Pearson Correlation coefficient between each pair of centrality (vectors)
   e. Draw a degree distribution histogram and a log-log plot
   f. Find degree assortativity for the nodes
   g. Find the shortest paths between the pairs of nodes
   h. Find the diameter of the graph
   i. Find the network density
   j. Find the top edges in terms of edge-betweenness
   k. For the top 3 nodes (in terms of degree centrality), draw the ego network and list their alters
   l. Find the similarity between nodes. List the top five pairs in terms of structural similarity

    m. Find the clustering coefficients of each node and the global clustering coefficient
    n. Detect sub-structures (k clique, K core etc)
    o. Detect community using Louvain and Clique percolation methods

(d) **Provide the analysis in submitted python notebook trying to compare your findings with the novel plot that you are aware of** :
    a. Analyse centrality (contrast degree, closeness, betweenness, Eigenvector) and "centrality correlation" plots. Make comments about characters (any 4 and use your judgement)
    b. Make a comment about the community/components detected in the context of the plot
    c. Do you draw any conclusion or insight from structural similarity that you calculated ?
    d. Do the clustering analysis, density, diameter etc tell you anything ?
    e. Any other interesting observation ?

Marking Guidelines

Marking scheme	Timeliness (5 marks)	Functionality (5 marks)	Analysis (5 marks)	Visualization (5 marks)
Max (5)	on time	working code without issue	Well done	Well done
Mid (3)	on time + 3 days	working with minor issues	Somewhat done	Somewhat done
Min (2)	delay more than a week	working with major issues	Very poorly done	Very poorly done