Obligatory Assignment 05

Due 26 April, 23.59

Task 1: In this assignment you will work with NetworkX library. The details are given below:

You have to load a graph representing Zachary's Karate club using the following code:
G = nx.karate_club_graph()
This graph contains nodes representing member of the club and includes attributes named 'club' which represents which group the members belong after the seperation of the club. You will find either Mr. Hi, or Officer... it indicates two different groups.

 # Task 1.1: Print out the list of nodes and edges with attributes.

 # Task 1.2: You have to provide a visualization of the Karate club where the color of nodes will represent the two groups after the seperation.

 # Task 1.3: Using nx.dijkstra_path() find out the shortest path from node 24 to 16. Print the list of nodes.

 # Task 1.4: You have to provide a visualization of the Karate club and modify the color of the nodes by highlighting the selected nodes from the above list. Use red color to indicate the selection of the nodes.

**[Tips: A jupyter notebook is available (sna.ipynb) that you can use as a starting point]**

Task 1: The output of your program should be something similar to the following:

Task 1.1: Expected output is similar to the following...

[(0, {'club': 'Mr. Hi'}), (1, {'club': 'Mr. Hi'}), (2, {'club': 'Mr. Hi'}), (3, {'club': 'Mr. Hi'}), (4, {'club': 'Mr. Hi'}), (5, {'club': 'Mr. Hi'}), (6, {'club': 'Mr. Hi'}), (7, {'club': 'Mr. Hi'}), (8, {'club': 'Mr. Hi'}), ...)]
[(0, 1, {}), (0, 2, {}), (0, 3, {}), (0, 4, {}), (0, 5, {}), (0, 6, {}), (0, 7, {}), (0, 8, {}), (0, 10, {}), (0, 11, {}), (0, 12, {}), (0, 13, {}), (0, 17, {}), (0, 19, {}), (0, 21, {}), (0, 31, {}), (1, 2, {}), (1, 3, {}), (1, 7, {}), (1, 13, {}), (1, 17, {}), (1, 19, {}), (1, 21, {}), (1, 30, {}), (2, 3, {}), (2, 7, {}), (2, 8, {}), (2, 9, {}), (2, 13, {}), (2, 27, {}), (2, 28, {}), ...]

Task 1: The output of your program should be something similar to the following:
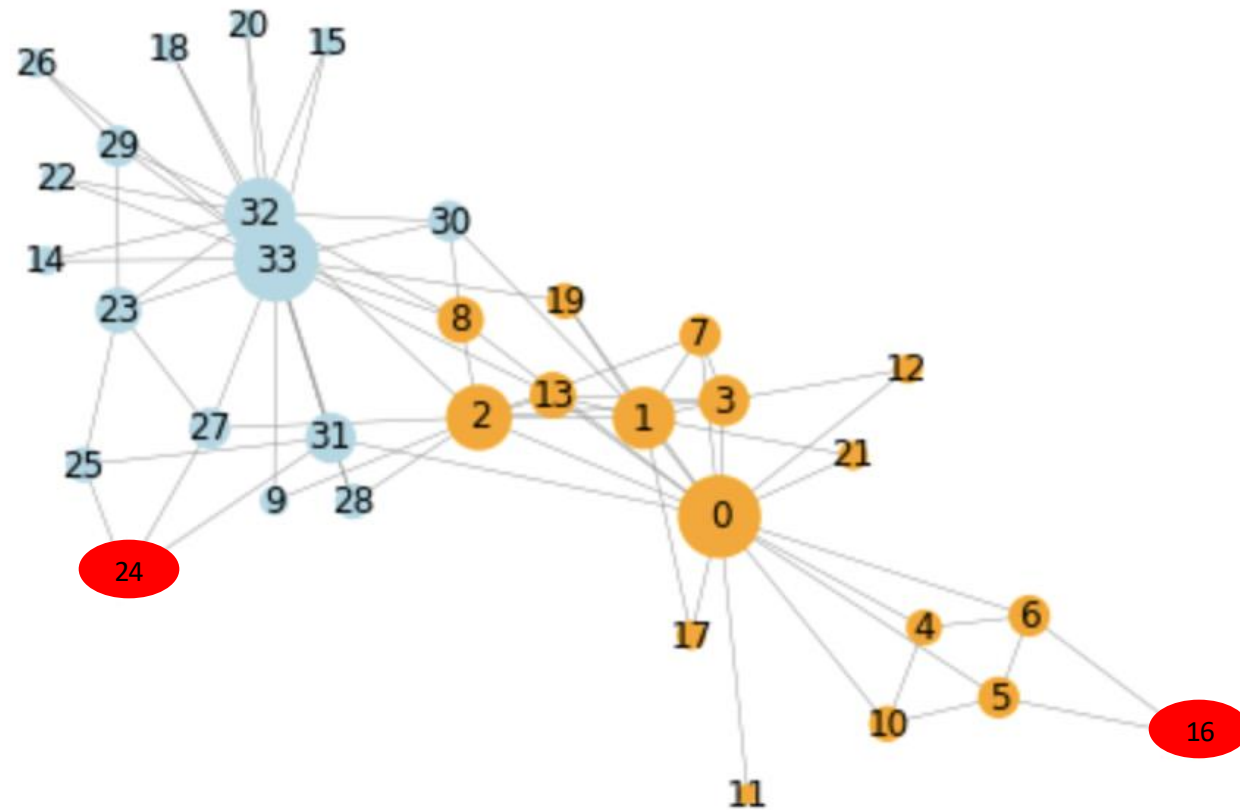
Task 1.2: Expected output:
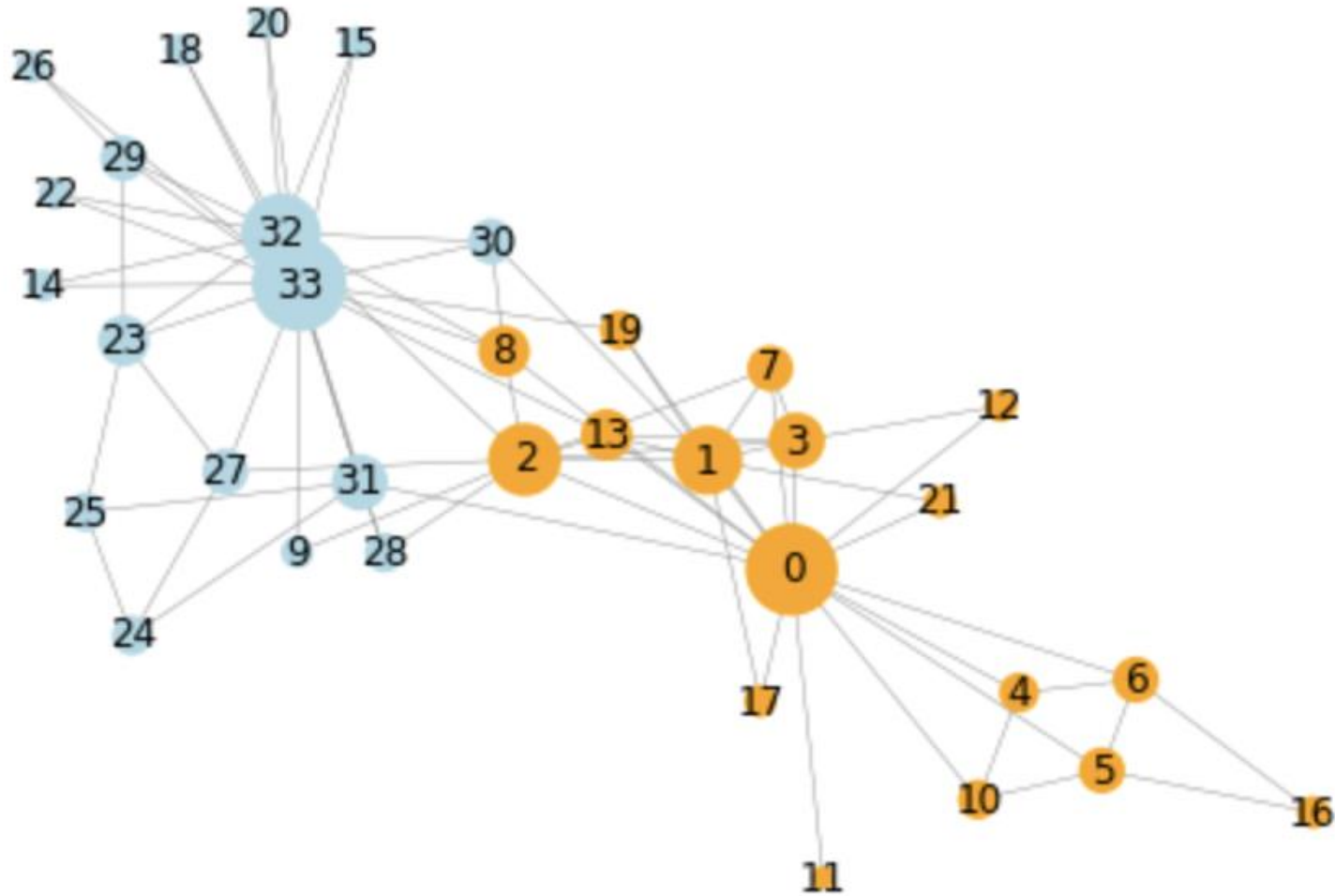
Task 1: The output of your program should be something similar to the following:

Task 1.3: Expected output is a list of integer numbers.

Task 1.4: Expected output is similar to this graph but the **selected nodes from Task 1.3**, need to be highlighted with red color

Task 2: In this assignment you will work with Selenium and Network X library. The details are given below:

Use selenium to follow hashtag links.

Your task is to perform breadth first search in two steps.

- In step1, you will perform crawling starting from a hashtag for example, #Oslo.
- In step 2, you will perform crawling starting from a hashtag for example, #Bergen.

While crawling hashtags, you will keep track of already visited hashtags. Your program should not visit a hashtag twice.

Follow minimum 20 hashtags in both steps.

While visiting hashtag pages, your program will add new connections in an undirected graph.

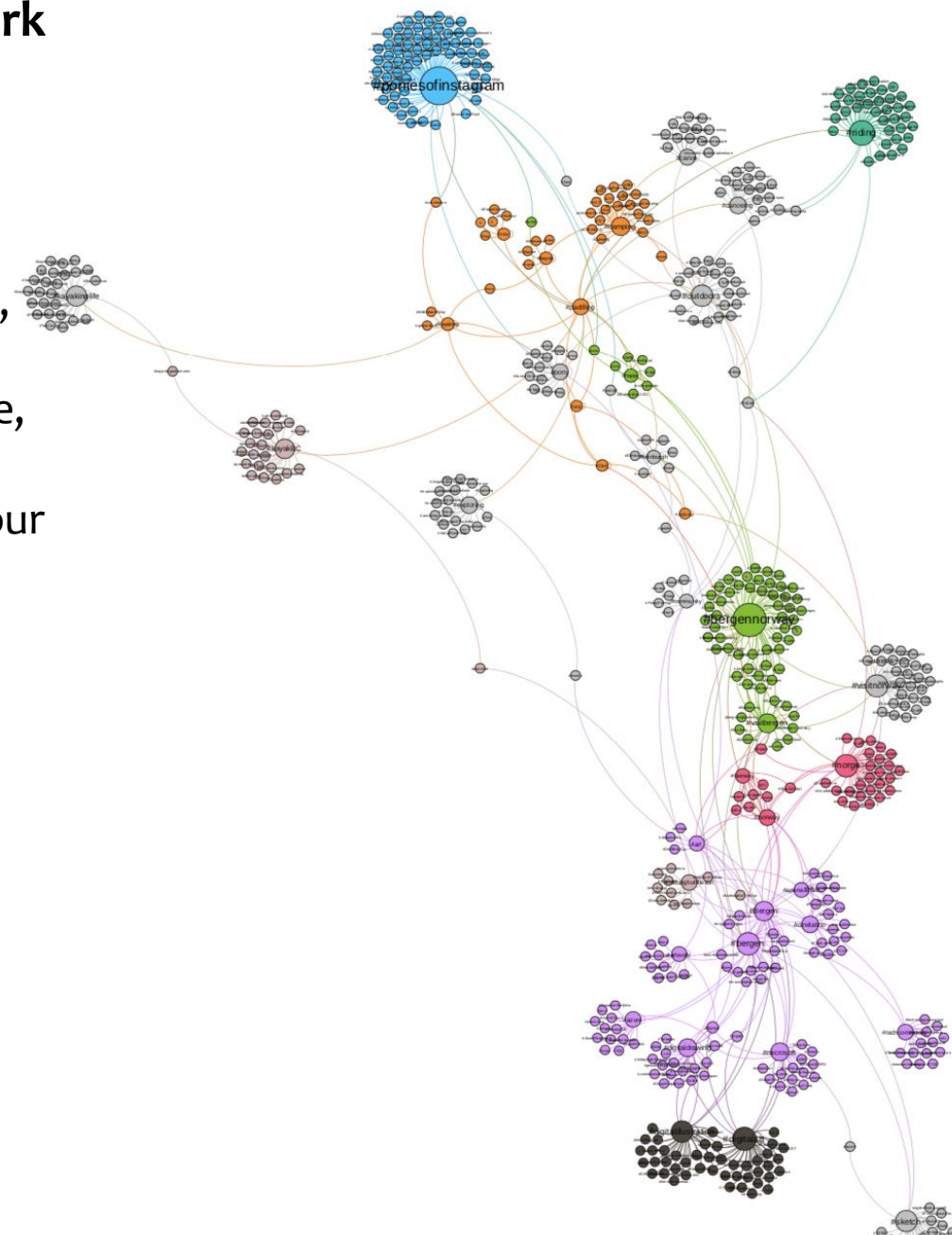Your program should include error handling.

Output the graph in GraphML format.

Import the graph in Gephi as undirected graph.

Change the size of the graph based on degree.

Change the color of the nodes based on modularity.

Fix the layout to make the graph better presentable.

- You have to submit 3 files that includes solution for tasks 1 and 2.
  - Task1->
    - File1: Containing python program (You can deliver jupyter notebook file with .ipynb extention)
  - Task2 ->
    - File2: Containing selenium program which includes the graph construction.
    - File3: A pdf file exported from Gephi. The pdf contains the output of your network visualization.

  - It is very important that your program should not have any syntax error at the time of submission.

  - Note:: Do not delete the projects from your computer until your assignment is approved. Seminar leader will contact you if they need more information about your projects. For example, you may be asked to fix your program if the program is not behaving as it should – such as the output is different or the selenium program is not following the hashtags properly, etc.