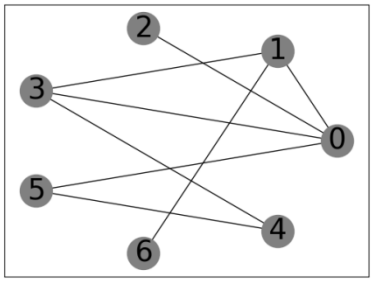
Mid-term Exam (Graph Mining – Spring 2023)

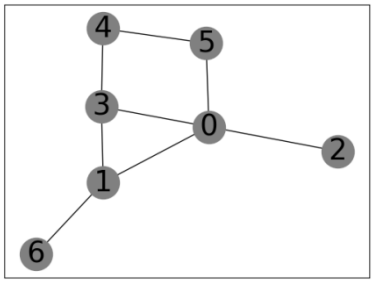
Full Name:
Student ID:

- The formula and solution process should be presented with the answer.
- All the codes must include detail comments in English.

1. Consider an undirected graph G of 7 nodes given in the following figure. Find all neighbors of nodes 0, 1, 3, and 4 in the graph G.

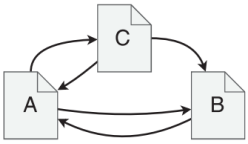


2. Consider an undirected graph of 7 nodes in the following figure.



   a. Calculate the Degree centrality of each node in G.
   b. Calculate the Degree centralization score of graph G (variance and Freeman score).
   c. Calculate the Normalized Betweenness centrality of each node in G.
   d. Calculate the Normalized Closeness centrality of each node in G.

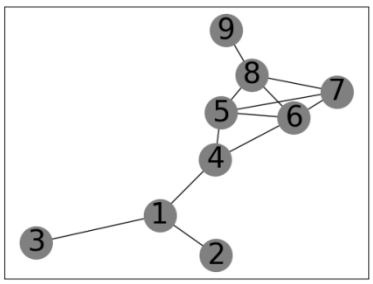
3. Consider a directed graph of 3 nodes (A, B, and C) in the following figure. Initialize the PageRank of each node with the value $R_0 = 1/3$.



$$R_t(i) = 1 - d + d \sum_{(j,i) \in E} \frac{R_{t-1}(j)}{\text{Out\_degree}(j)} \qquad (1)$$

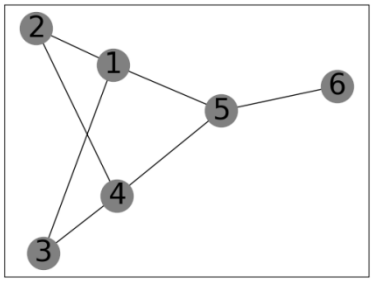
   a. Apply Equation (1) with damping factor (d = 1) to calculate the values of PageRank of each node in the next iteration (t = 1). Out_degree(j) is the number of connections that originate at a node j and point outward to other nodes.
   b. Calculate the values of PageRank of each node when t = 5.

4. Consider an undirected graph G of 9 nodes in the following figure.



   a. Find out all cliques of size 3 and Communities in graph G based on Clique Percolation Method (CPM).
   b. Find 4-clique communities in G using the CPM method.

5. Consider an undirected graph of 6 nodes in the following figure.

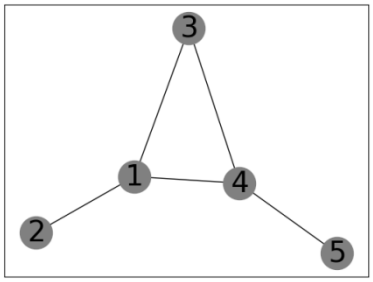


$$s_0(a,b) = \begin{cases} 1, & \text{if } a = b, \\ 0, & \text{if } a \neq b. \end{cases} \quad (2)$$

$$s_t(i,j) = \frac{1}{|N(i)||N(j)|} \sum_{a \in N(i)} \sum_{b \in N(j)} s_{t-1}(a,b) \quad (3)$$

   a. Calculate the Jaccard's coefficient of the following node pairs: (2,3), (3,6), (1,3), (3,5)
   b. Calculate the Preferential attachment (PA) of the following node pairs: (2,4), (4,6), (1,3), (3,5)
   c. When t = 5 by using Equations (2) and (3), calculate the SimRank between the following node pairs: (1,6), (4,1), (2,3), (2,5). $s_t(i,j)$ presents the Simrank similarity between node *i* and *j* at iteration *t*.

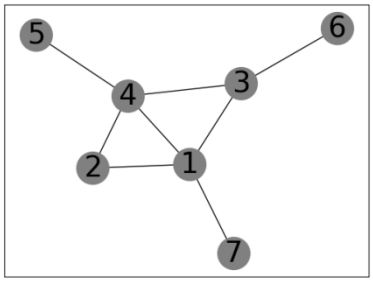
6. Consider an undirected graph of 5 nodes in the following figure. Predict two possible new edges in the graph by using the Jaccard Coefficient metric.



7. Write a pseudo code for a function that accepts a NetworkX graph and a node name and returns the average degree of the target node's neighbors.

```
#Students can only use inbuilt functions: degree() and neighbors() in NetworkX
#Input: Graph G and a node name v in G
#Output: return the average degree of the neighbors of node v
def AverageDegree(G, v):
    #YOUR CODE HERE
    return x
```

8. Consider an undirected graph of 7 nodes in the following figure.



   a. Find out the shortest path length of node pairs: (1,4), (2,7), (5,7), (6,7)
   b. Write a pseudo code for a function that calculate the average shortest path length for all nodes in G.

```
#Input: A NetworkX graph G
#Output: x is the the average shortest path for all nodes in G
#Note: Students can only use Python code (without using inbuilt functions, such as min, max, sum, etc.). In the NetworkX library, students can only use a function 'shortest_path_length()'.
def AverageSP(G):
    #YOUR CODE HERE.
    return x
```

9. If an undirected graph has five nodes of degree 4 and four nodes of degree 3. How many edges does it have?