Birla Institute of Technology & Science, Pilani
First Semester 2022-2023
Graph Mining (CS F426)
Mid-Semester Exam

Date	: Nov 02, 2022	Duration: 90 minutes
Nature of Exam	: Closed Book	Weightage: 30%
Total Marks	: 60	

Instructions:

1. **All questions are compulsory.**
2. **Write important intermediate steps in numerical. Directly writing the final correct answer is not sufficient to obtain full marks.**
3. **Questions pertaining to implementation must be answered in programs only.**
4. **Program code must be written in an optimal way.**

Q1. The unnormalized graph Laplacian does not depend on the diagonal elements of the adjacency matrix. True/False? Prove with the help of an example graph. [**2 marks**]

Q2. Write the two important statistical properties of graphs that can be derived from the eigenvalues and eigenvectors of the adjacency and laplacian matrices? Also explain how?

[**3 marks**]

Q3. For the given graph in the Figure 1.1 **[1+2+2+2=7 marks]**

(a) Compute the adjacency matrix A corresponding to the graph.

(b) Compute the similarity between vertices with respect to outgoing edges.

(c) Compute the similarity between vertices with respect to incoming edges.

(d) How similar is each vertex to vertex 5 compared to each other?

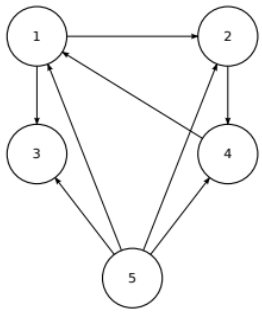


Figure 1.1

Q4. The Figure 1.2 depicts web pages A, B and C and their connectivity with each other. Calculate the pagerank of each page **[4+4=8 marks]**

(a) assuming no taxation.

(b) assuming **β**=0.8.

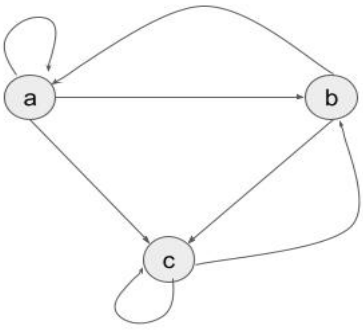
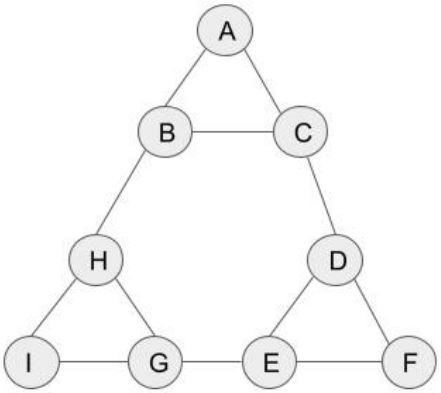


Figure 1.2

Q5. For the below graph with (A-I) nodes, use the Girvan-Newman approach to compute the following **[4+6+2 =12 marks]**



(a) Find the number of shortest paths for the following nodes ***i)*** A and ***ii)*** B.
(b) Compute the betweenness of each edge in the given graph.
(c) Determine the reasonable candidates for the communities by removing all the edges with a betweenness above some threshold.

Q6. In the given Table below, information for the results after applying k-means clustering is provided, where vertices are grouped into 4 clusters.

Vertex ID	Clusters(s) membership	Class membership
1	**1**	+
2	**1**	*
3	**2**	-
4	**3**	*
5	**2**	-
6	**4**	+
7	**4**	*
8	**3**	+

Construct the matching matrix and compute entropy of each cluster (using class information available) and determine which is/are the best cluster(s). **(2+4+2 = 8 marks)**

Q6. This question should be answered through **<u>python program implementation only</u>**.

(a) Given the list of vertices with x and y coordinates shown below, construct and print an adjacency matrix of the vertices, where vertices are considered to be neighbors (adjacent) if they are within a distance of 4, using Euclidean distance.

Node	Position	Node	Position
A	(1, 2)	F	(9, 3)
B	(4, 6)	G	(4, 1)
C	(4, 4)	H	(7, 1)
D	(5, 4)	I	(2, 7)
E	(2, 6)	J	(8, 5)

Note, the vertices should not be considered adjacent to themselves. **[4 marks]**

(d) Using the adjacency matrix that you have created, calculate and print the shared nearest neighbor (SNN) graph in the form of a matrix. **[4 marks]**

(e) List all the pairs of nodes that have **at least two shared neighbors** in common**. [3 marks]**

(f) List all the pairs of nodes that have **at least three shared neighbors** in common. **[3 marks]**

(g) Determine how many clusters will be resulting from the SNN graph if the threshold is equal to ***i)*** 2 ***ii)*** 3. **[3+3=6 marks]**

**********END**********