

Birla Institute of Technology & Science, Pilani
First Semester 2021-2022
Graph Mining (CS F426)
Mid-Semester Exam

Date : Oct 30, 2021
 Nature of Exam : Partially Open Book
 Total Marks : 45

Duration: 90 minutes
 Weightage: 30%

Instructions:

1. This exam is closed book i.e. no laptops, notes, textbooks, etc. during the exam.
2. There are a total of 6 questions. All questions are compulsory.
3. Write important intermediate steps in numerical. Directly writing the final correct answer is not sufficient to obtain full marks.
4. All questions must be attempted
5. Questions pertaining to implementation must be answered in programs only.
6. 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. 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?

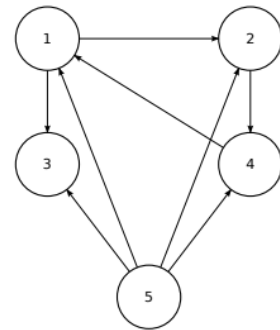


Figure 1.1

Q3. The Figure 1.2 depicts pages A, B and C with no inlinks, and with initial PageRank = 1. Initially, none of the pages link to any other pages and none link to them. Answer the following questions, and calculate the PageRank for each question. [8 marks]

- (a) Link page A to page B.
- (b) Link all pages to each other.
- (c) Link page A to both B and C, and link pages B and C to A.
- (d) Use the previous links from © and add a link from page C to page B.

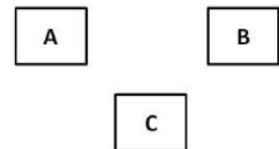


Figure 1.2

Q4. In the below question, **part (c)-(g) should be answered through python program implementation.**

(a) Describe how the shared nearest neighbor algorithm groups vertices. **[2 marks]**

(b) Based on your answer to **part a**, how many total nodes are required at minimum to find a pair of nodes with n shared nearest neighbors? Explain your answer. **[2 marks]**

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

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)

(d) Using the adjacency matrix that you have created, calculate and print the shared nearest neighbor (SNN) graph in the form of a matrix. **[3 marks]**

(e) List all the pairs of nodes that have **at least two shared neighbors** in common. **[1.5 marks]**

(f) List all the pairs of nodes that have **at least three shared neighbors** in common. **[1.5 marks]**

(g) Determine how many clusters will be resulting from the SNN graph if threshold=3. **[3 marks]**

Q5. (a) Describe how the HITS algorithm relates to the Neumann Kernel technique. **[3 marks]**

(b) How does the decay factor γ affect the output of the Neumann Kernel algorithm? **[2 marks]**

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 the purity and entropy of each cluster (using class information available) and determine which is/are the best cluster(s) in terms of entropy and purity both. **(1+4+1 = 6 marks)**

*****END*****