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

Date	: Dec 11, 2021	Duration: 3 hours
Nature of Exam	: Closed Book	Weightage: 40%
Total Marks	: 70	

Instructions:

1. **This exam is closed book i.e. no laptops, notes, textbooks, etc. during the exam.**
2. **There are a total of 10 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. (Multiple choice questions, 8 marks)

i) Which of the following are true/false?[2 marks]

a) Closeness centrality measures the degree to which a node lies on the geodesics between the other nodes in the graph.
b) Betweenness centrality measures the distance to which a node is from all the other nodes in the graph.
c) If a node is connected to all other nodes in the graph directly, it's CLC score will be 1.
d) A vertex is considered to be a good authority and good hub if it has a high outdegree and high indegree respectively.

ii) The unnormalized Laplacian matrix L of a simple graph holds the following properties? True/False [2 marks]

a) L is symmetric and positive definite.
b) For a vector x, $x^T L\, x \leq 0$ identity holds.
c) The multiplicity of 0 as an eigenvalue of Laplacian is equal to the (number of connected components - 1) of a graph.
d) The eigenvectors of 0 eigenvalue in L_{SYM} are scaled by $D^{1/2}$ in comparison to eigenvectors of 0 eigenvalue in L_{RW}.

iii) In GCN, filter parameters are shared across all nodes: [1 mark]

a) True
b) False

iv) The layer wise propagation rule used in graph convolutional network is: [1 mark]
a) H(l+1) = σ(AH(l)W(l) + b (l))
b) H(l+1) = σ(W(l)H(l)A + b (l))
c) Both
d) None of these

Hint: W(l) and b (l) are weight and bias matrices, σ is a non-linear function. H(l) and H(l+1) are the output from the lth and (l + 1)th layers respectively.

v) In order to learn the structure of the 3rd-order neighborhood of every node (all nodes up to 3 "hops" away), how many layers do we need to create in GNN? [1 mark]
a) one input, two hidden, one output
b) one input, three hidden, one output
c) one input, one hidden, one output
d) None of these

vi) Suppose you have a graph with 500 nodes and 220 edges and each node has a feature vector of size 20. In GCN, if the first hidden layer has 20 neurons. How many parameters (including bias) does this hidden layer will have? [1 mark]
a) 80
b) 400
c) 420
d) none of these

Q2. [2 marks] You are asked to design a deep learning system to detect the traces of a disease in the human body. It is crucial that your model detects disease if it exists, to start early medication. Which of the following is the most appropriate evaluation metric: Accuracy, Precision, Recall. Explain your choice.

Q3. [5 marks] Assume you have defined a decoder based on $L2$ distance to learn the node embeddings in a given graph G as follows

$$DEC\left(z_u, z_v\right) = \left|\left|z_u - z_v\right|\right|_2^2$$

n is the number of nodes in G.
The loss function weights the pairs of nodes based on the edges present in adjacency matrix $A[u, v]$ using the below equation:

$$L = \sum_{(u,v)\in D} DEC\left(z_u, z_v\right) A[u,v]$$

$A[u,v] = 1$ if nodes are connected, 0 otherwise.

Write the expression for gradient $\frac{\partial L}{\partial z_k}$?

Q4. [6 marks] Write down any two multi-relational decoder equations and compare their representational abilities.

Q5. **[4 marks]** Give two benefits of using graph neural network models over shallow embedding approaches to generate node embedding for solving downstream tasks such as classification, link prediction etc.

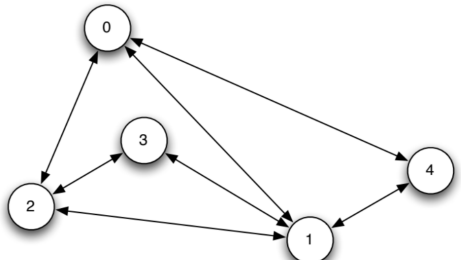
Q6. [5 marks] You are asked to design a Graph Convolution Network architecture with an input layer, two hidden layers H_1 and H_2 and one output layer. Write down the graph level equation for node representation using UPDATE and AGGREGATION function at H_1 and H_2. X and A represent the input node feature and adjacency matrix respectively. Also explain the strategy to merge UPDATE and AGGREGATION steps together and limitations if any?

Q7. [5 marks] Max-margin loss (given below) not only enforces the decoder to produce a higher score for true pairs, but at the same time ensures the smaller score value for negative pairs. "**Do you agree with this statement**"? Why/Why not? Justify your answer with an example. For example with Dec(negative)=1, what is the greatest value of decoder (true ?) that will minimize the loss?

$$L = \sum_{(u,\tau,v)\in E} \sum_{v_n \in P_{n,u}} (0, -DEC\left(z_u, \tau, z_v\right) + DEC(z_u, \tau, z_{v_n}))$$

Q8 [10 marks]. Suppose you have a multi-relational knowledge graph with 1000 nodes and 200 relation types. You come up with a RGCN model to learn the embedding of nodes with three hidden layers (32, 64 and 128) neurons. For each layer, calculate the number of parameters to be learned, and the size of the associated feature maps assuming the effect of the self node and its neighbors on the final embeddings differently.

Q9. [2+2+3+3+2+3=15 marks] In the below question (**answered through python program implementation**), a directed graph is shown in the Figure



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

(b) Calculate the transpose of the adjacency matrix, A^T corresponding to the graph.

c) Calculate the document correlation matrix and term correlation matrix.

d) Calculate the maximum indegree and outdegree of the graph.

e) Calculate the valid range of the decay factor γ in the neumann kernel?

f) What does it mean for a matrix to be symmetric? Are the document correlation and term correlation matrices symmetric? Justify your answer.

Q10. a) [**3 marks]** List the formula for accuracy, precision and recall used to evaluate the performance of classification algorithms.

b) [**7 marks]** Using the information provided in table (Predictions by models M1 and M2), compute the confusion matrices, and compare the performance of models M1 and M2 using accuracy, precision and recall. Which model has a better performance and why?

Datapoint	Actual Class	Model M_1	Model M_2
1	+	+	+
2	−	−	+
3	−	+	−
4	+	+	+
5	+	−	−
6	+	+	−
7	−	+	+
8	+	+	+

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