Department of Computer Science & Engineering, MNIT, Jaipur
Natural Language Processing 21CST822 MTE, February 2023

Max. Marks: 30 Attempt all questions **Time: 90 Minutes**

1. Write regular expressions for the following languages:
   a. To find URLs within a string [1]
   b. For matching a valid date [1]

2. We are given the following corpus:
   a. <s> You are Rio </s> b. <s> Rio You are </s> c. <s> You are Rio </s> d. <s> You like yellow orange Rio </s>
   Using a bigram model with add-one smoothing, what is P(Rio|are) and P(<s> You like yellow Rio </s>) ? Include <s> and </s> in your counts just like any other token. [4]

3. Compute minimum edit distance in between "**kwiliamcohen**" and "**twoiamcohon**". (consider all operations costs as same i.e. 1 for each operation) [show complete table][5]

4. Computer the class of test document using Naïve Bays Classifier (consider add one smoothing). [3]

Training	Doc	Words	Class
	1	Rio sport great India Proud Dish	P
	2	India sport bad player	P
	3	Sam injustice poverty India	N
	4	Rio games feel good sport	P
Test	5	India injustice Sam Rio	?

5. **Computer the cross entropy loss for both classes for the following observation:** [3]
   $X=[2,\ 1,\ 0,\ 2,\ 3.12,\ 1.2]$ $W=[3.5,\ -6,\ -1.2,\ 0.4,\ 3.5,\ 0.3]$ $b=0.20$

6. Use gradient descent learning used in logistic regression to updates the weights and bias after one iteration for the sample $x_1=5$ (count of positive lexicon) $x_2=4$ (count of negative lexicon), initial values of $w_1=1.2$ $w_2=-3.2$, $b=1.5$ and learning rate is 0.1 . [3]

7. Criticize old methods of sentiment classification and propose a novel algorithm for the same with a suitable example. [5]

Term(rows)/Context(Columns)		Automobile	Computer	Money	Household	Politics
1	Car	**10**	**8**	**1**	**0**	**3**
2	Auto	**5**	**1**	**0**	**0**	**1**
3	Insurance	**1**	**0**	**4**	**3**	**0**
4	Window	**1**	**2**	**1**	**2**	**1**
5	Computer	**3**	**2**	**1**	**0**	**0**
6	Technology	**1**	**3**	**1**	**1**	**1**

8. Consider the data (given above) of terms with their context counts given below. Compute final PPMI matrix for these rows. [5]

MALAVIYA NATIONAL INSTITUTE OF TECHNOLOGY JAIPUR

Department of Computer Science & Engineering

Mid Term Exam Spring Semester 2023

Marks: 30 M. Tech (II Semester) Course: Blockchain Technologies **(CST849)** **Time: 1.5 Hrs**

I: Fill-in the blanks [4]

i. 1 Bitcoin(BTC) = _____ Satoshi and _____ BTC ever exists in the bitcoin network.

ii. _____, _____ and _____ fields of a block header are primarily used in the mining process.

iii. Bitcoin uses UTXO based transactions model and Ethereum uses _____

iv. ScriptSig key appears in transaction _____ and ScriptPubKey appears in transaction _____

v. _____ is the process by which the block size limit on a blockchain is increased by removing _____ data from transactions.

vi. Each miner creates a candidate block from Transactions available in its _____.

vii. The signature is in the _____ field of each input of the transaction.

viii. A block with 120 bytes of transactions in bitcoin network requires storage of _____ per year.

II : Write False/True [4.5]

a. Nodes in bitcoin network not keeping track of how much money you have.

b. Only one participant can get privileges to read ledger and verify its consistency, while more than one participant can write into a ledger.

c. Blockchain based system doesn't heavily rely only on the public key cryptography.

d. Transactions are not stored in Merkle tree, rather their data is hashed and resulting hash is stored in each leaf.

e. Blockchain is based on the fact that ledger is distributed and exists only in one node on network.

f. Hash difficulty level decreases in order to control the supply of bitcoin .

g. An orphan block is only created when 51% attack is successful.

h. Locktime makes sure that a transaction is unlocked until a specific block height is reached.

i. A transaction generally references previous transaction outputs as new transaction inputs and dedicates all input values to new outputs.

III. Answer all the following questions. 4.5+4.5+4.5+4.5+2.5

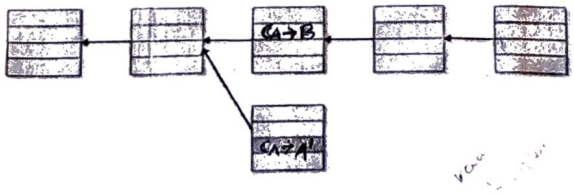
1. **a)** What are the most important problems of the traditional money transfer system which blockchain attempts to solve? How it address the money transfer problem? (only Name the key concepts used).
   b) What are the important basic ingredients which blockchain needs as a support for the working of the blockchain technology?

2. **a)** What are the different types of blockchains and what are the benefits and drawbacks of such classifications?
   b) Show the architecture of a blokchain? Why merkle tree is beneficial in blockchain technology?

3. **a)** How nonce is used in mining? What is selfish mining attack and what are the risk with selfish mining?
   b) Why do we need consensus mechanisms? Why does the consensus considered to be a hard problem?

4. **a)** What is the use of bitcoin scripts? Write the pros and cons of different types of bitcoin transaction scripts?
   b) Suppose Z creates a digital coin and transfers it to X which further transfer it to Y. Create and show sequence of coin transfer events that occurs using only private and public key concepts to demonstrate that finally it owns by Y.

5. Consider Bob the merchant deciding whether or not to accept the $C_A \rightarrow B$ transaction. What Bob is really interested in is whether or not the other chain will catch up. Why, then, does he simply check how many confirmations $C_A \rightarrow B$ has received, instead of computing the difference in length between the two chains?



Department of Computer Science & Engineering, MNIT, Jaipur

Natural Language Processing 21CST822 ETE, April 2023

Max. Marks: 50 Attempt all questions. **Time: 150 Minutes**

1. **(a)** Consider the following grammar and generate the CKY parsing table for the sentence "**I eat sushi with chopsticks with you**" : [3]

S → NP VP	**VP → Verb NP**	**NP → you**
NP → NP PP	**VP → VP PP**	**PP → Prep NP**
NP → sushi	**NP → chopsticks**	**Verb → eat**
NP → I	**Prep → with**	

(b) Convert the following grammar in Chomsky Normal Form (CNF): [6]

S→ASB **A→aASA | a |ε** **B→SbS | A | bb**

2. **(a)** Consider the following senses of the words:

Word: **ash**

Sense 1 Tree of the olive family with pinnate leaves, thin furrowed bark and gray branches.
Sense 2 The solid residue left when combustible material is thoroughly burned or oxidized.
Sense 3 To convert into ash

Word: **coal**

Sense 1 A piece of glowing carbon or burnt wood.
Sense 2 charcoal.
Sense 3 A black solid combustible substance formed by the partial decomposition of the vegetable matter without free access to air and under the influence of moisture and often increased pressure and temperature that is widely used as a fuel for burning

Which one is the winner sense using Lesk's algorithm of "ash" for the sentence "On burning coal we get ash". [4]

(b) Suggest additional information that can to be extracted from the question while Question Processing in Question-Answering systems, which can further improve the performance. Explain with suitable example. [4]

3. **(a)** Consider the following table of test results of seven questions which gives ranked answers as output. Compute the accuracy of the system using Mean Reciprocal Rank method. [2]

Question Number	1	2	3	4	5	6	7
Correct Answer Number	4	3	6	2	8	1	2

(b) Compute the ROUGE-2 score for the following text: [2]

Human Summary: Water spinach is a commonly eaten leaf vegetable of Asia.
System Summary: Water spinach is a leaf vegetable commonly eaten in tropical areas of Asia.

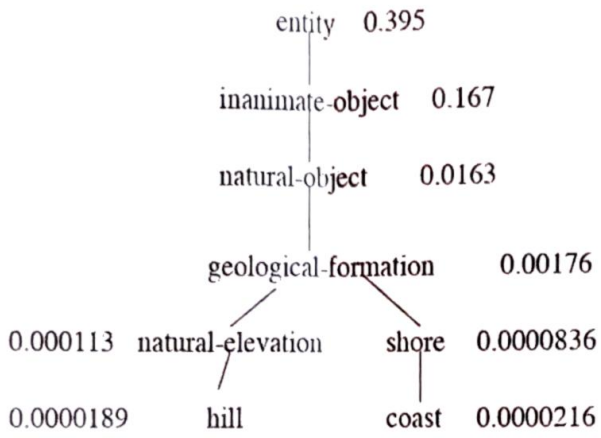
4. Consider the following sentences **(1,2,3,4)** as set of reference sentences from document collection and sentence number **5** as a query sentence. Consider value of lambda (**λ**) **0.3** . Use cosine similarity to compute similarity between sentences. Find the maximal set of sentences to be selected in set of summary sentences (find only 3 sentences in final summary set) using maximal marginal relevance. [5]

Doc 1/sentence 1 **new home sales top forecasts**
Doc 2/sentence 2 **home sales rise in july**
Doc 3/sentence 3 **increase in home sales in july**
Doc 4/sentence 4 **july new home sales rise**
Doc 5/sentence 5 **sales home is very bad**

5. **(a)** Consider the training data given below for machine translation from English to French. Assume uniform initial translation probabilities. Compute translation probabilities and normalized alignment probabilities after one iteration. [5]

	1		2		3	
English	play	cricket	the	cricket	cricket	team
French	jouer	croquet	la	grillon	croquet	equipe

(b) Consider following sub tree with probabilities. Compute $\mathbf{sim_{path}}$, $\mathbf{sim_{Resnik}}$ and $\mathbf{sim_{Lin}}$ between "hill" and "shore". [3]



6. Consider the following Tagged sentences:

S	Rahul	can	play	chess	E
	N	M	V	N	
S	Karan	should	clean	table	E
	N	M	V	N	
S	will	Shyam	play	chess	E
	M	N	V	N	
S	Karan	should	beat	Rahul	E
	N	M	V	N	

Use Hidden Markov Model (HMM) to predict the probability of tagging the sentence "Rahul should beat Karan" with tag sequence "N M V N". [5]

7. Suggest a novel methodology that can handle co-references in English Text and evaluate it also. Explain it properly using a suitable example. [5]

8. Design and evaluate a novel recognition system capable of recognizing temporal expressions of the kind appearing in emails or conversations. Sample text is given below: [6]
"I am free next week. Let's meet on Saturday." "Are you free on Tuesday?"

Malaviya National Institute of Technology Jaipur

Social Network Analysis (CST836)

Mid Term Examination

Date: February 27, 2023

Timing: 3:30 to 5 PM | Max marks: 30

Attempt all the questions

1. (a) For the given graph H, compute the followings:
   1. Clustering Coefficient of nodes B and G
   2. Betweenness Centrality of node G
   3. Average Path Length for graph H
   4. Closeness Centrality of node G
   5. Diameter of the graph H
   6. Maximum Matching and size of the maximum matching (Matching Number)
   7. Maximum Clique and size of maximum clique (Clique Number)
   8. Maximum Independent Set and size of the maximum independent set(Vertex Independence Number)
   9. Minimum Vertex Cover and size of the minimum vertex cover (Vertex Covering Number)
   10. Minimal Dominating Set of graph H

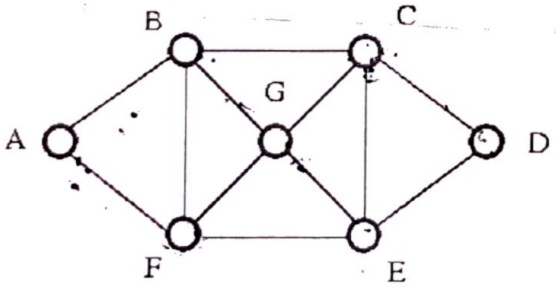


Figure 1: Graph H

2. (a) Calculate PageRank scores for the Graph G (shown on Page 2) using power iteration method (show calculations for the first three iterations only). **(10)**

(4)

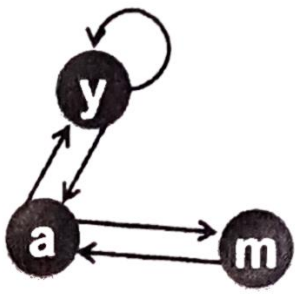


Figure 2: Graph G

(b) For graphs, consider joining adjacency matrix and features (feature matrix) and then feed it into a deep neural net (ANN/CNN). In other words, apply CNN on graphs by considering adjacency matrix and feature matrix (instead of pixel matrix which is the traditional case of images). List down 3 limitations of this approach and resolution mechanism(s). (4)

(c) Give an example of herd behavior. (2)

3. (a) For each of the questions below, write your answer either True or False. No explanations are needed.

1. Dropout is a technique used in deep learning to prevent overfitting by randomly dropping out neurons (units) in the neural network during training.
2. Hyperparameters are the parameters of the model that are learned during training.
3. Learning rate, epochs and weights of the neural network are hyper parameters.
4. Location invariant feature is present in artificial neural network.
5. The small world model offers network generation with low diameter and low clustering coefficient.

(5)

(b) Explain node embedding generation process (with equations) in Graph Convolutional Networks (GCN). (3)

(c) Describe: Sampling and aggregation mechanism in GraphSAGE. (2)

Best wishes

Malaviya National Institute of Technology Jaipur

Social Network Analysis (CST836)

End Term Examination

Date: April 28, 2023

Timing: 8:00 to 10:30 AM | Max marks: 50

Attempt all the questions

1. (a) For the given graph H, compute the followings:
    1. Chromatic number of graph H
    2. Number of edges and vertices in the complement of H
    3. Clustering Coefficient of nodes 5 and 9
    4. Betweenness Centrality of node 10
    5. Average Path Length for graph H
    6. Closeness Centrality of node 8
    7. Diameter of the graph H
    8. Maximum Matching and size of the maximum matching (Matching Number)
    9. Maximum Clique and size of maximum clique (Clique Number)
    10. Maximum Independent Set and size of the maximum independent set(Vertex Independence Number)
    11. Minimum Vertex Cover and size of the minimum vertex cover (Vertex Covering Number)
    12. Minimal Dominating Set of graph H

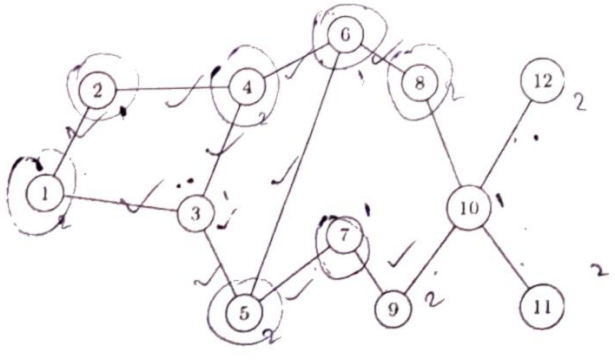


Figure 1: Graph H

(12)

2. (a) Consider the game theoretic model of cascades for Graph H. Assume that nodes 2, 5 and 6 are early adopters of the new technology (lets say iOS). Assume threshold value of 0.5 i.e. If more than $q \geq 50\%$ of my friends are using iOS then I'll also use iOS. Explain how the technology adoption will cascade in the network.

(5)

(b) Perform K-core decomposition on Graph H. (3)

3. (a) For each of the questions below, write your answer either True or False. No explanations are needed.

1. A set of vertices is a Vertex Cover if and only if its complement is an Independent Set.
2. The vertices of any Maximal Matching form a (not necessarily minimal) Vertex Cover.
3. For exam scheduling problems, maximum matching is the most relevant graph parameter.
4. For the room assignment problems, clique is the most relevant graph parameter.
5. Every tree is a bipartite graph. (5)

(b) Answer the following questions briefly (in one/two lines only):

- Why hoax article detection is more effective using network theory as compared to the traditional/manual approaches?
- Why citation networks are acyclic?
- What should be the value of p in $G_{n,p}$ so that exactly 1 giant component appears?
- What is the reference model in network theory?
- List down the main limitation of the small world model. (5)

4. (a) Explain the overall architecture and working of GraphRNN in detail. (5)

(b) Identify and summarize the key learning outcomes that you gained from your CST836 project. (5)

5. (a) Provide a detailed comparison between GCN, GAT, and GraphSAGE (highlighting their differences and unique characteristics). (5)

(b) Explain the concept of "The Strength of Weak Ties" with real-life examples. (3)

(c) Explain the following terms briefly a) Softmax and b) Dropout (2)

Best wishes

MALAVIYA NATIONAL INSTITUTE OF TECHNOLOGY JAIPUR

Mid Term Examination, Spring Semester 2022-23

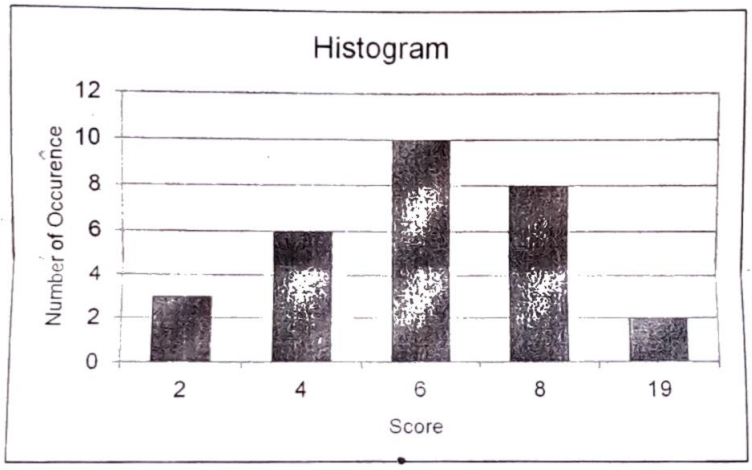
M. Tech Computer Science and Engineering (II Semester)
M. Tech Computer Science and Information Security (II Semester)
Doctor of Philosophy (Ph.D) Course Work

Marks: 30 **21CST507-Research Methodology** Time: 90 mins

21CST901-Research Methodology - II

Answer all questions. All questions carry equal marks.

1. Define measures of central tendency and Quartiles. Compute mean and median of the histogram shown below:



2. Prove that a sequence Z constructed from z-score of sequence A shall have zero mean and unity variance.

3. How is bin-boundary applied for data cleaning? Explain with help of an example.

4. Give the algorithm for Breadth First Search and traverse a sample graph with 10 nodes.

5. Write short notes on the following:
   a. Impact Factor
   b. Quartile Index
   c. Conference Ranking categories
   d. h-index and i-10 index
   e. Author, Year reference style

6. List the statistical scales of measurement and explain them. Discuss the various classes of statistical data.