Roll No. | | | | | | | | | | |

ANNA UNIVERSITY (UNIVERSITY DEPARTMENTS)

B.E. /B.Tech / B. Arch (Full Time) - END SEMESTER EXAMINATIONS, MAY / JUN 2024

Information Science & Technology
Semester VI
IT5020 - Social Network Analysis
(Regulation 2019)

Time:3hrs | Max.Marks: 100

CO1	To gain knowledge about the empirical and theoretical study of social networks, its structure and social network data sources.
CO2	To study about the semantic technologies for social network analysis
CO3	To gain knowledge on visualization of social networks and its applications.
CO4	To gain knowledge about social network analysis software for characterizing the network structure.
CO5	To engage in critical thinking regarding the applicability of social network theory to various sociological phenomena.

BL – Bloom's Taxonomy Levels
(L1-Remembering, L2-Understanding, L3-Applying, L4-Analysing, L5-Evaluating, L6-Creating)

PART- A (10x2=20Marks)
(Answer all Questions)

Q. No.	Questions	Marks	CO	BL
1	Calculate the degree centrality measure for the given mail transactions between given 5 people	2	CO1	L2
2	List any two electronic discussion network used for sharing information.	2	CO1	L1
3	Use 3-means clustering to cluster the given dataset where A, D, G are initial centroids.	2	CO3	L3
4	How will you secure your social network profile? Write any two methods for performing the above security measure.	2	CO4	L2
5	Validate the given RDF file	2	CO5	L4

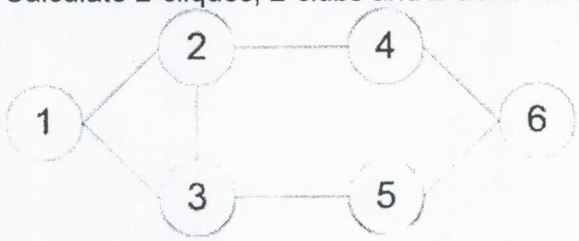
	Cathy	Rosy	Neethu	Tom	Jeni
Cathy	0	1	0	1	0
Rosy	0	0	1	1	1
Neethu	1	0	0	1	1
Tom	1	1	1	0	0
Jeni	1	1	1	1	0

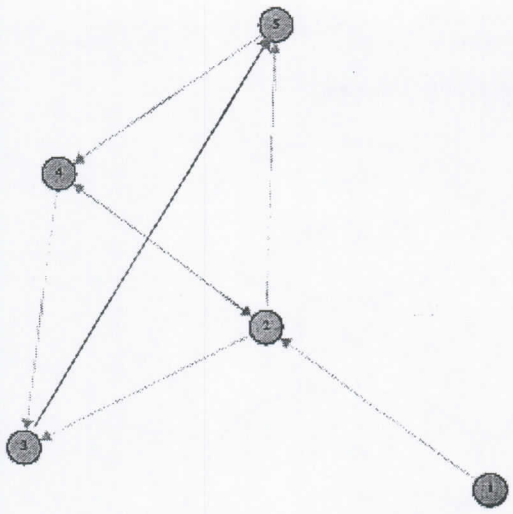
Data Points		
A	2	10
B	2	5
C	8	4
D	5	8
E	7	5
F	6	4
G	1	2
H	4	9

	<?xml version="1.0"?> <River id="Yangtze" xmlns="http://www.geodesy.org/river" xmlns:uom="http://www.measurements.org/units-of-measure#"> **<length uom:units="kilometers">6300</length>** <startingLocation>western China's Qinghai-Tibet Plateau</startingLocation> <endingLocation>East China Sea</endingLocation> </River>			
6	Define RDF Transitive property with an example	**2**	CO1	L3
7	Write down the pros and cons of decentralized Online social network.	**2**	CO2	L1
8	Evaluate precision and recall for the given parameters True positive= 30, True Negative = 930, False Negative =10, False Positive =30.	**2**	CO5	L4
9	How will you use force directed layout to visualize the co-authorship network?	**2**	CO3	L5
10	Represent the given data using node edge diagram	**2**	CO4	L6

	0	**1**	**2**	**3**	**4**	**5**
0	0	1	1	1	1	0
1	1	0	0	1	0	0
2	1	0	0	1	1	1
3	1	1	1	0	0	1
4	1	0	1	0	0	0
5	0	0	1	1	0	0

PART- B (5x 13=65Marks)

Q. No.	Questions	Marks	CO	BL
11 (a)(i)	Calculate 2-cliques, 2-clubs and 2-clans for the given network 	**7**	CO1	L4
(ii)	Discuss on the development of social network analysis in detail.	**6**	CO2	L2
	OR			
11 (b) (i)	Calculate the geodesic and diameter for the given network	**7**	CO1	L4



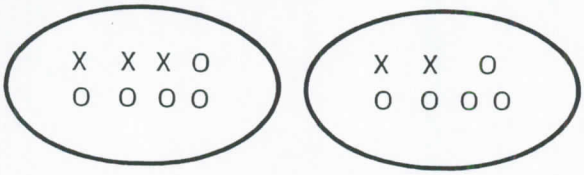
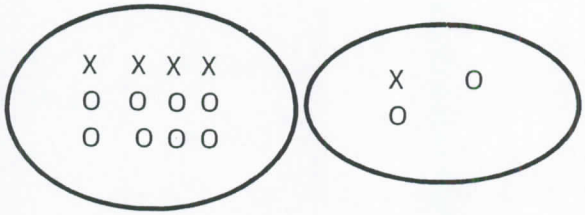
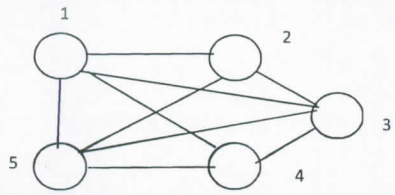
(ii)	Discuss on any four applications of social network analysis in detail.	**6**	CO2	L2
12 (a)	Enumerate and discuss on the quantitative and qualitative analysis of Commercial social network profiles.	**13**	CO4	L3
	OR			
12 (b)	Explain in detail the process of analyzing a log file recorded by the Moodle e-learning system.	**13**	CO4	L3
13 (a)(i)	In web search three documents were found which contain information about the birth place of Tim Corbett. How will you identify which is the correct birth place using semantic web? Explain.	**7**	CO2	L5
(ii)	How will you represent an individual in OWL?	**6**	CO5	L2
	OR			

13 (b)(i)	What do you infer from the below OWL code snippet? Explain.
	`<?xml version="1.0"?>`
	`<Sea rdf:ID="EastChinaSea"`
	`xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"`
	`xmlns="http://www.geodesy.org/water/naturally-occurring#">`
	`<containedIn>`
	`<Sea rdf:about="http://www.china.gov#ChinaSea"/>`
	`</containedIn>`
	`</Sea>`
		7	CO2	L5

(ii)	What is Smushing? Explain how it is performed in social network using an example.	**6**	CO5	L2
14 (a)(i)	Given four documents d1, d2, d3, d4	**7**	CO3	L4

i	Ground Truth		Ranking Function$_1$		Ranking Function$_2$	
	Document Order	r_i	Document Order	r_i	Document Order	r_i
1	d4	2	d3	2	d3	2
2	d3	2	d4	2	d2	1
3	d2	1	d2	1	d4	2
4	d1	0	d1	0	d1	0

Calculate the Normalized Discounted cumulative gain and infer which ranking is more relevant

(ii)	How will you perform community detection using divisive hierarchical clustering method? Explain	**6**	CO4	L3
	OR			
14 (b)(i)	Calculate the normalized mutual information for the below given partitions and infer which is better. Partition A:  Partition B: 	**7**	CO3	L4
(ii)	Calculate the node similarity for the given graph using cosine similarity. 	**6**	CO4	L3
15 (a)	Covid is a disease which has spread in India very quickly. The government would like to analyze the areas where Covid is affected and would also like to predict which are the areas it may spread in near future. Which technique would they use to visualize the data and which methodology do they follow for prediction? Explain.	**13**	CO5	L6
	OR			
15 (b)	Fei Fei Li is a researcher working in the area of Machine Learning. She has published more than 5000 articles with various coauthors and her articles are cited more than 10000 times in other articles. Explain the technique used to visualize the co-citation collaboration network and also explain the kind of analysis that can be performed over the above created social network and the patterns that emerge from the analysis	**13**	CO5	L6

PART- C (1x 15=15Marks)

(Q.No.16 is compulsory)

Q. No.	Questions	Marks	CO	BL
16.	You are appointed as a data analyst to develop and analyze the mail transaction in the company			
	(i) Design the RDF document for mail object using triplet property	**4**	CO1	L3
	(ii) How will you analyse the mail transactions happening between the employees in the company	**4**	CO4	L4
	(iii) Define metrics to measure the similarity between employees based on the mail transactions	**4**	CO4	L3
	(iv) Which visualization technique will you use to visualize the mail transactions. Why?	**3**	CO3	L2