Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent unit of MAHE, Manipal)

VII SEMESTER B.TECH. (COMPUTER SCIENCE AND ENGINEERING)

END SEMESTER MAKEUP EXAMINATIONS, DEC 2019

SUBJECT: GRAPH ANALYTICS FOR BIG DATA [CRA 4008]

REVISED CREDIT SYSTEM
(02/01/2020)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

- Answer **ALL** the questions.
- Missing data may be suitable assumed.

1A. What does the following networks represent? What is the use of graphs in these networks? **5M**
i) Human Information Networks
ii) Biological Networks

1B. Write the Adjacency matrix for the graph in Figure 1B. What is the other way of representing this graph? Represent it. **3M**

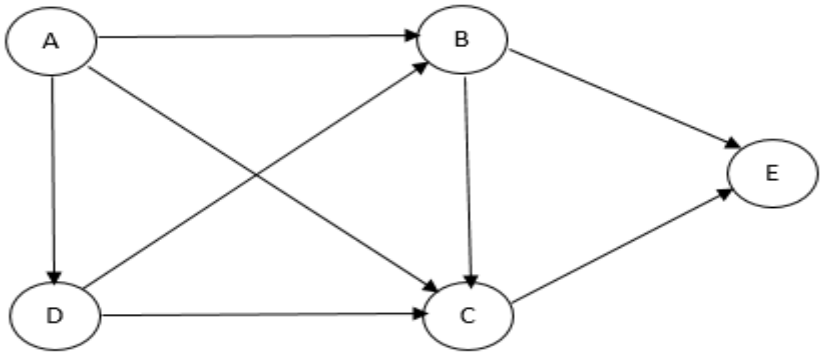

Figure 1B Directed Graph

1C. Explain the different kinds of possible nodes that can be created in a Tweet. What are the different concepts associated with a node? **2M**

2A. What do mean by Group degree centrality, Closeness centrality and Betweenness centrality? Find out the node that is having the highest closeness centrality score in the graph given in figure 2A. **4M**

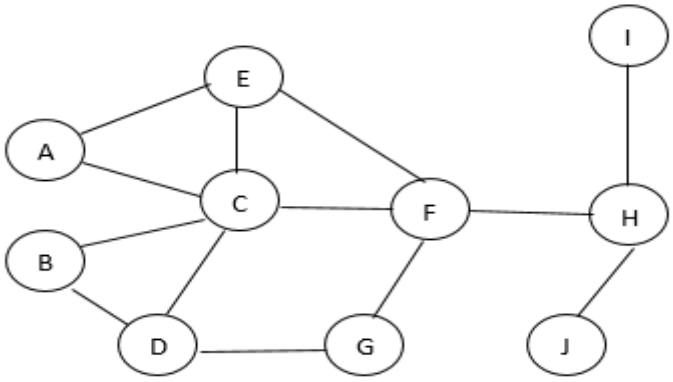

Figure 2A Undirected Graph

2B. Briefly explain how the cluster quality is measured? **2M**

2C. With an example graph define the following terminologies: **4M**
i)Trail ii) Path iii) Strongly connected graph iv) Weakly connected graph

3A. What do you mean by community in a graph? What are the three types of analytic questions that can be asked related to a community? Explain each type with example. **5M**

3B. Write a Cypher query to create a graph for the data given in table 3B. **5M**

Table 3B.

Sl.no	Source	Destination	Distance
1	P1	P2	5
2	P2	P5	50
3	P5	P7	80
4	P1	P3	10
5	P1	P4	30
6	P3	P7	20
7	P4	P3	20
8	P4	P6	10
9	P6	P7	25
10	P2	P3	15

4A. Write a Cypher code to find degree histogram and construct an Adjacency matrix of a graph constructed in question 3B. **5M**

4B. Differentiate between Graphlab and Pregel. **2M**

4C. How does Giraph paradigm efficiently handle the computation of aggregate values with an example function? **3M**

5A. Write a Cypher query which returns a graph not containing a neighborhood of PUNE JN in Table 5B(a). **3M**

5B. Write a GraphX query to create a graph trainGraph, from a list of vertices i.e., the railway station names given in table 5B(a) & a list of edges that will make up the train service from table 5B(b) and print the departing and arrival station and the train number for each row from the trainGraph graph. **5M**

Table 5B(a)

Station ID	Station Name
1	PUNE JN
2	NEW DELHI
3	PURI
4	BANGALORE CITY JN
5	DANAPUR
6	HOWRAH JN
7	JAMMU TAWI
8	DARBHANGA JN

Table 5B(b)

Train number	Source station	Destination station
11177	PUNE JN	JAMMU TAWI
12888	NEW DELHI	PURI
12800	PURI	NEW DELHI
12200	BANGALORE CITY JN	DANAPUR
12134	PUNE JN	HOWRAH JN
12211	DANAPUR	BANGALORE CITY JN
12136	HOWRAH JN	PUNE JN
12347	BANGALORE CITY JN	NEW DELHI
12576	NEW DELHI	DARBHANGA JN

5C. How do you concatenate two sets of nodes created from 5B(a) and 5B(b) into a single RDD? Write a query to filter all the edges from BANGALORE CITY JN (5B(a)) and create a map of destination vertex IDs. **2M**