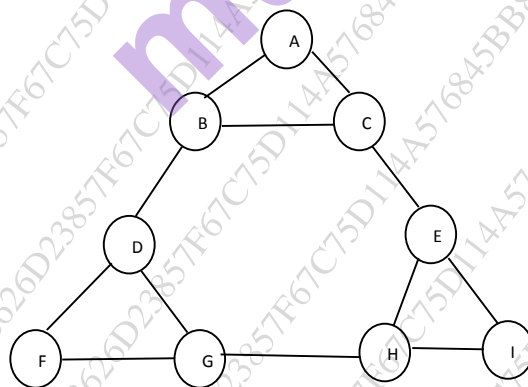


[Time: 3 Hours]

[Mks: 80]

- NB : 1) **Question 1 is compulsory.**  
 2) Attempt any **three** questions from the **remaining** questions.  
 3) **Assume** suitable **data** wherever applicable.

- 1 (a) Explain characteristics of big data. Analyze the tourism data and identify the characteristics which are difficult in storing in RDBMS and the need for big data techniques for storing and analyzing them. 5
  - (b) Explain any five components of Hadoop ecosystem. 5
  - (c) Give problems in Flajolet-Martin (FM) algorithm to count distinct elements in a stream. 5
  - (d) Explain the nearest neighbor problem. What similarity measure can be used in an application to find plagiarism in documents? 5
  - 2 (a) Discuss Matrix-Matrix Multiplication. Perform Matrix Multiplication with 1-step Map Reduce method 10
- $$\begin{array}{ccc} 1 & 2 & 3 & 4 \\ 5 & 6 & * & 1 & 2 \\ 1 & 0 & & & \end{array}$$
- (b) Explain with example Collaborative based filtering in a recommendation system. 10
  - 3 (a) Explain the concept of Parallel Decision Trees with the help of an example. 10
  - (b) Recall all NoSQL design patterns with example. Justify CAP with suitable example. 10
  - 4 (a) For the graph given below use Clique percolation and find all communities 10

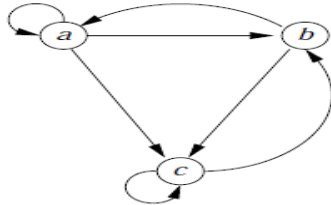


- (b) Employ the DGIM algorithm. Shown below is a data stream with  $N=14$  and the current bucket configuration. New elements enter the window at the right. Thus, the oldest bit of the window is the left-most bit shown 100110101010111 10
- i) Show one way of how the above initial stream will be divided into buckets and count distinct 1's.

ii) The following bits enter the window one at a time: 10101. What is the bucket configuration in the window after this sequence of bits has been processed by DGIM?

- 5 (a) Explain working of all phases of Map Reduce with one common example. 10  
(b) Explain Park-Chen-Yu algorithm. How memory mapping is done in PCY. 10

- 6 (a) Compute the page rank of each page in the following figure, assuming  $\beta = 0.8$ . 10



- (b) Explain how Hadoop goals are covered in Hadoop Distributed File System. 10

\*\*\*\*\*