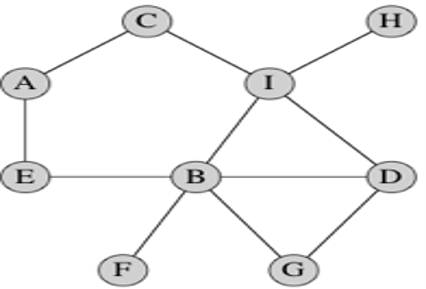
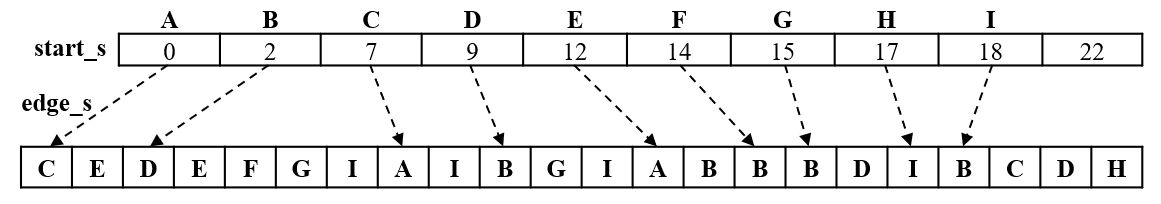
1. (4 pts) Given a Graph G below, Draw its Adjacency Array Representation.



**Answer:**



1. (2 pts) A Map task has? Mappers whereas the Reducer task has? Reducers.

**Answer:** 1) many and 2) one or more

1. (4 pts) There are two principal costs,
   * 1. **Computation cost**of mappers, reducers, and the system. System cost depends on the operations,  
        a. ? key-value pairs by key and b. ? the key-value pairs
     2. **Communication cost**depends on transferring key-value pairs from? to?

**Answer:** 1) sorting, 2) merging, 3) mappers and 4) reducers

1. (2 pts) Communication cost often dominates the computation cost. Why?

**Answer:** Ethernet is being the bottleneck because moving data among tasks takes time. Also, Gigabit Ethernet can often fell to keep up when many compute nodes are generating and consuming data on the same network.

1. (2 pts) Replication rate is ?? of key-value pairs created by each mapper denoted by **r.r** represents ????

**Answer:** 1) the average number and 2) the communication cost per input

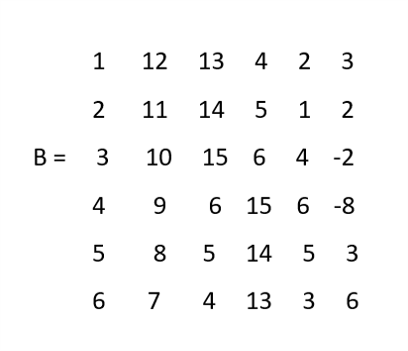
1. (2pts) If each reducer has size **q** if there are **p** reducers, and **I** (capital i) is the number of inputs for the problem, then **r=??/I.**

**Answer:** pq

1. (2 pts) A mapping schema for a problem with a reducer of size **q** is an assignment of inputs to sets of reducers, with two conditions:
   1. Maximum **q** inputs per reducer.
   2. For every output, there is some reducer that receives all of the inputs? with that?

**Answer:** 1) associated and 2) output

1. a. (8 pts) Assuming a reducer size of q = 2, trace the map-reduce algorithm discussed on module 8 to multiply the two matrices below. If you think that this value of q is not feasible choose the smallest q that is feasible and trace the algorithm with this value of q. Show all your work.



**Answer:**

For q = 2, g = 2n2/q = 2 \* 62 / 2= 36. But, dividing rows of the first matrix into g groups of n/g rows each is not impossible if g = 36. So, we chose the minimum possible q which is q = 2n.

q = 2n = 12 and n = 6, so g = 2n2/q = 2 \* 62 / 12= 6

1st Step: Dividing rows of the matrix A into g groups of n/g rows each that is 6 groups of 1 row each.

2nd Step: Dividing columns of the matrix B into g groups of n/g columns each that is 6 groups of 1 column each.

3rd Step: There are g2 = 36 reducers, each with 12 inputs consisting of a group of rows and a group of columns. Each reducer computes dot product of one row from the matrix A and one column from the matrix B.

4th Step: Each reducer provides each element in the matrix C where C = AB.

b. (2 pts) Specify the communication cost of this algorithm.

**Answer:** The total communication cost is 4n4/q = 4 \* 64 / 12 = 432.

1. (22 pts **extra credit**) Provide the details of the proof that a one-pass matrix-multiplication algorithm requires replication rate at least r ≥ 2n2/q, including:
2. The proof that, for a fixed reducer size, the maximum number of outputs are covered by a reducer when that reducer receives an equal number of rows of M and columns of N.
3. The algebraic manipulation needed, starting with ∑ki=1q2i ≥ 4n4.

**Answer:**