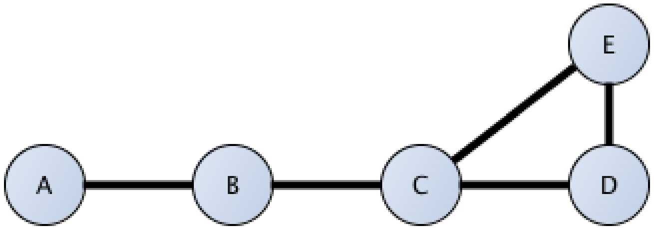
Big Data Management and Analytics
Assignment 12

Assignment 12-1.1

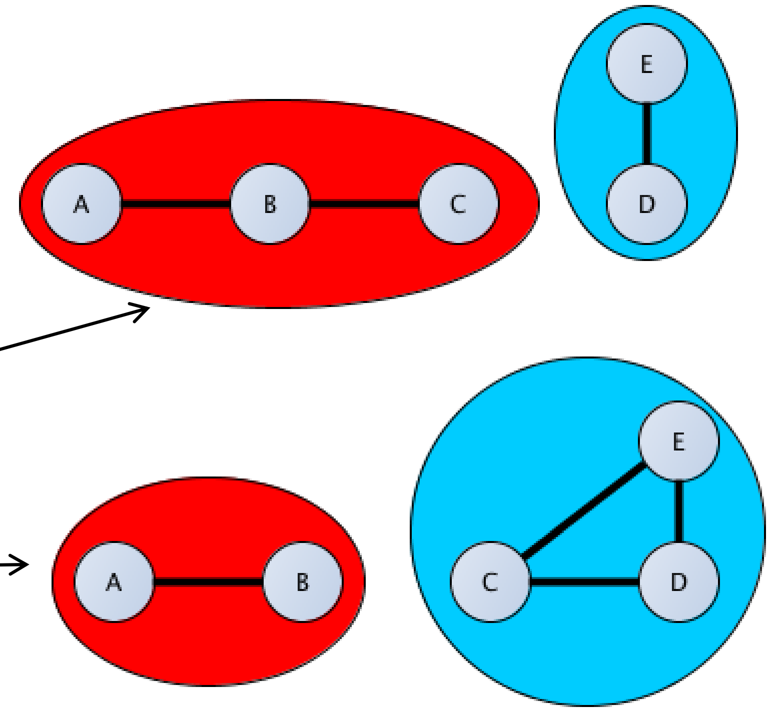
A good partitioning fullfills two conditions:

- Maximizes the number of edges **within a group**
- Minimizes the number of edges **between groups**

Assignment 12-1.1



- Partitions shall be balanced as far as possible → the graph shall be split either in
  - $\{A, B, C\}$ and $\{D, E\}$ or
  - in $\{A, B\}$ and $\{C, D, E\}$



- The second choice has more edges within the groups, and only one edge instead of two have to be removed
  → Best partitioning is $\{\boldsymbol{A}, \boldsymbol{B}\}$ **and** $\{\boldsymbol{C}, \boldsymbol{D}, \boldsymbol{E}\}$

Assignment 12-1.2

RECAP:

The **modularity Q** of a **partitioning S** of a **graph G** is defined as follows:

$$Q \propto \sum_{s \in S} [(\#edges\ within\ group\ s) - (expected\ \#edges\ within\ group\ s)]$$

$$Q(G,S) = \frac{1}{2m} \sum_{s \in S} \sum_{i \in S} \sum_{j \in S} (a_{i,j} - \frac{k_i, k_j}{2m})$$

$Normalizing: -1 < Q < 1$

Assignment 12-1.2

What we already have:

- Number of edges $|n| = 5$
- Number of edges $|m| = 5$
- Degree of nodes:

Node	Degree
k_A	1
k_B	2
k_C	3
k_D	2
k_E	2

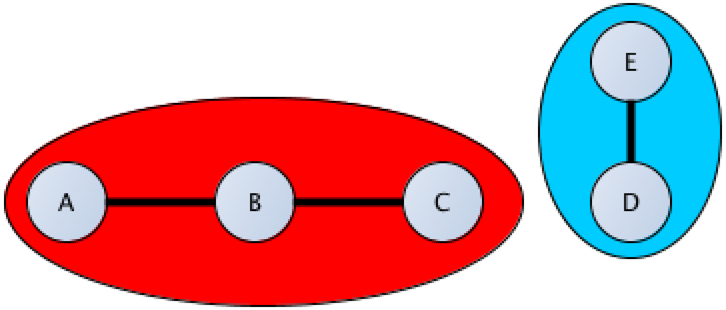
Assignment 12-1.2

Steps for computing Q:

1. Compute the adjacency matrix
2. Compute the modularity matrix: $(B_{ij} = A_{ij} - \frac{k_i \cdot k_j}{2m})$
3. Sum up the entries of the single clusters
4. Sum up the sums of all clusters
5. Normalize the result

Assignment 12-1.2

Steps for computing Q:



1. Compute the adjacency matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Assignment 12-1.2

Steps for computing Q:

2. Compute the modularity matrix: $(B_{ij} = A_{ij} - \frac{k_i \cdot k_j}{2m})$

$$B_{ij} = \frac{1}{10}\begin{pmatrix} -1 & 8 & -3 & - & - \\ 8 & -4 & 4 & - & - \\ -3 & 4 & -9 & - & - \\ - & - & - & -4 & 6 \\ - & - & - & 6 & -4 \end{pmatrix}$$

Assignment 12-1.2

Steps for computing Q:

3. Sum up the entries of the single clusters

$$s_1 = \{A, B, C\} \rightarrow \frac{1}{10}\big((-1) + 8 - 3 + 8 - 4 + 4 - 3 + 4 - 9\big) = \frac{4}{10}$$

$$s_2 = \{D, E\} \rightarrow \frac{1}{10}\big((-4) + 6 + 6 - 4\big) = \frac{4}{10}$$

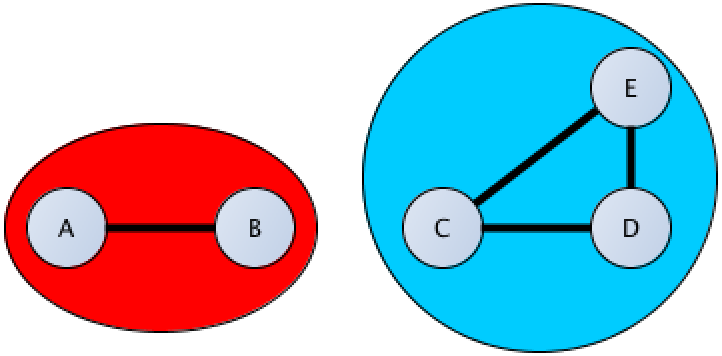
$$\sum_{s \in S} = \frac{4}{10} + \frac{4}{10} = \frac{8}{10} = 0.8$$

$$Q(G, S) = \frac{1}{10} \cdot (0.8) = 0.08$$

Assignment 12-1.3

Steps for computing Q:

1. Compute the adjacency matrix



$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Assignment 12-1.3

Steps for computing Q:

2. Compute the modularity matrix: $(B_{ij} = A_{ij} - \frac{k_i \cdot k_j}{2m})$

$$B_{ij} = \frac{1}{10}\begin{pmatrix} -1 & 8 & - & - & - \\ 8 & -4 & - & - & - \\ - & - & -9 & 4 & 4 \\ - & - & 4 & -4 & 6 \\ - & - & 4 & 6 & -4 \end{pmatrix}$$

Assignment 12-1.3

Steps for computing Q:

3. Sum up the entries of the single clusters

$$s_1 = \{A, B\} \rightarrow \frac{1}{10}\big((-1) + 8 + 8 - 4\big) = \frac{11}{10}$$

$$s_2 = \{C, D, E\} \rightarrow \frac{1}{10}\big((-9) + 4 + 4 + 4 - 4 + 6 + 4 + 6 - 4\big) = \frac{11}{10}$$

$$\sum_{s \in S} = \frac{11}{10} + \frac{11}{10} = \frac{22}{10}$$

$$Q(G, S) = \frac{1}{10} \cdot \frac{22}{10} = 0.22$$

Assignment 12-1.4

- The higher the modularity Q, the better the partitioning
- Removing the edge $\{B, C\}$ yields a higher Q value than the removal of the edges $\{C, E\}$ and $\{C, D\}$
- → The hypothesis from 1 which relies on maximizing the number of edges within the groups and minimizing the number of edges between the groups was correct

Assignment 12-2

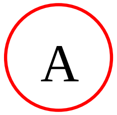
RECAP:

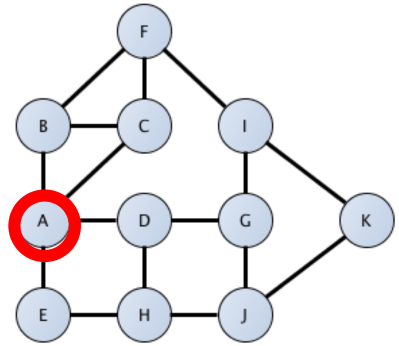
Girven-Newman algorithm:

1. Begin with node A and perform a BFS and construct a DAG (directed acyclic graph)
2. Count the number of shortest paths from A to all other nodes
3. Compute the **betweenness**, by traversing the tree in a bottom-up fashion. If there exist multiple paths, these are counted partially:
   1. $node\ flow = 1 + \sum childEdges$
   2. Split the flow based on the values of the parents (shortest path)

Assignment 12-2

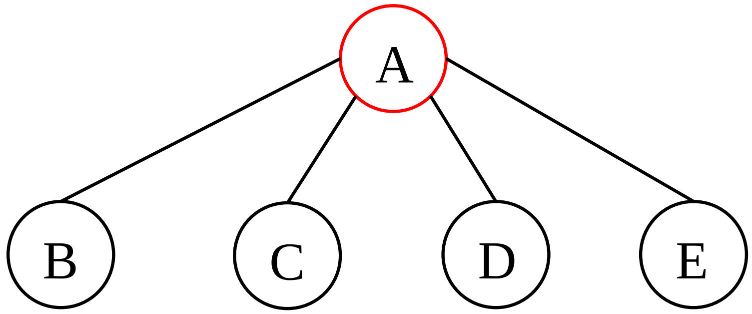
Step 1:

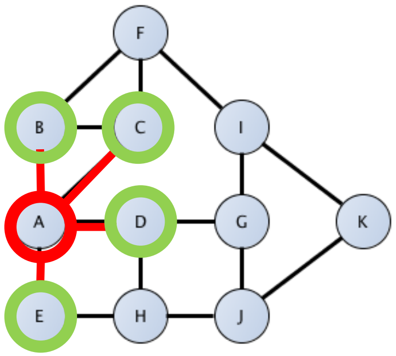




Assignment 12-2

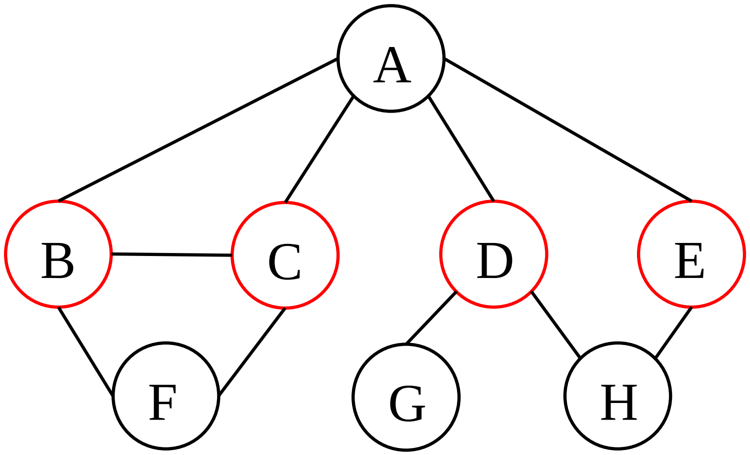
Step 1:

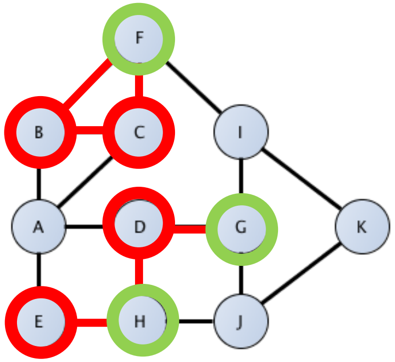




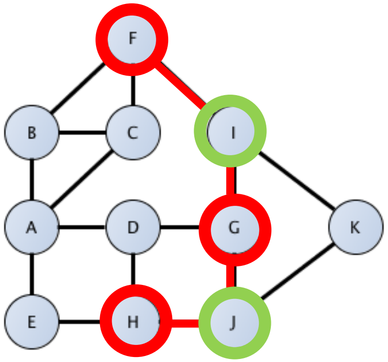
Assignment 12-2

Step 1:

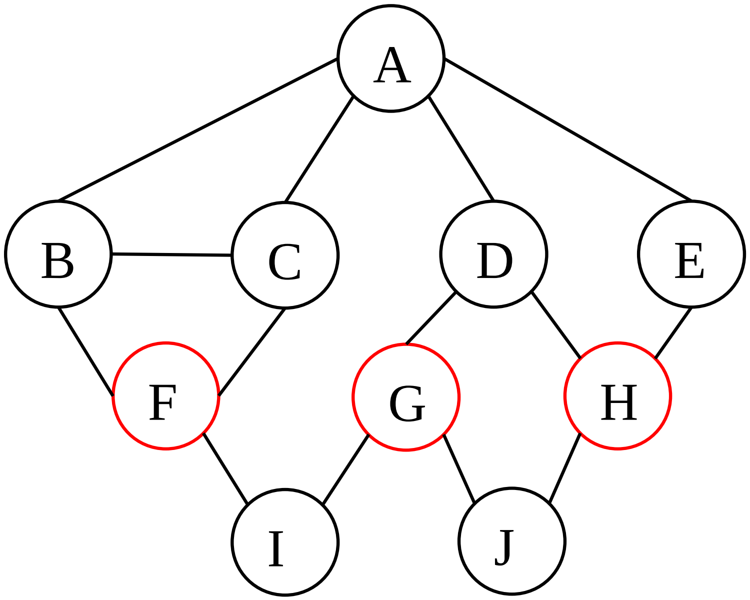




Assignment 12-2

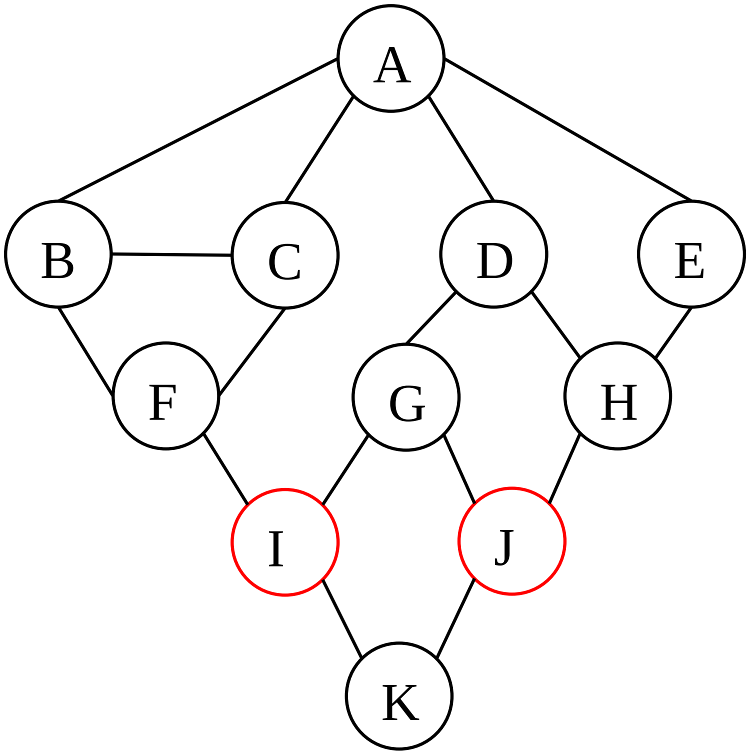


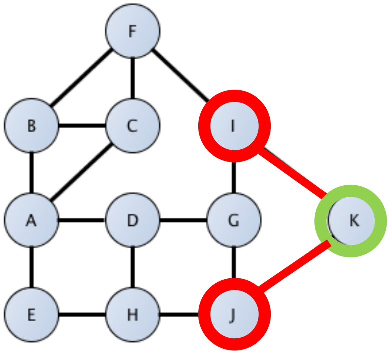
Step 1:



Assignment 12-2

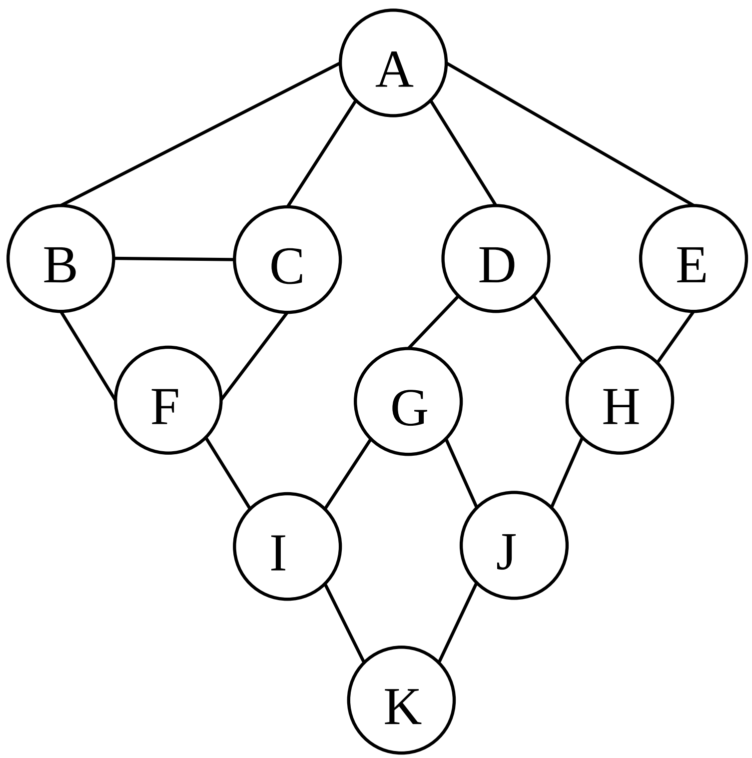
Step 1:

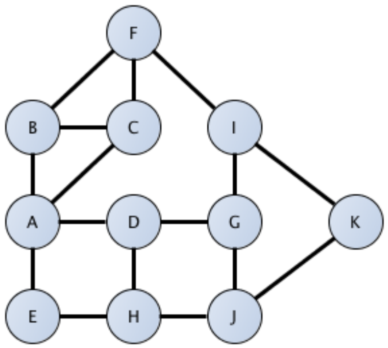




Assignment 12-2

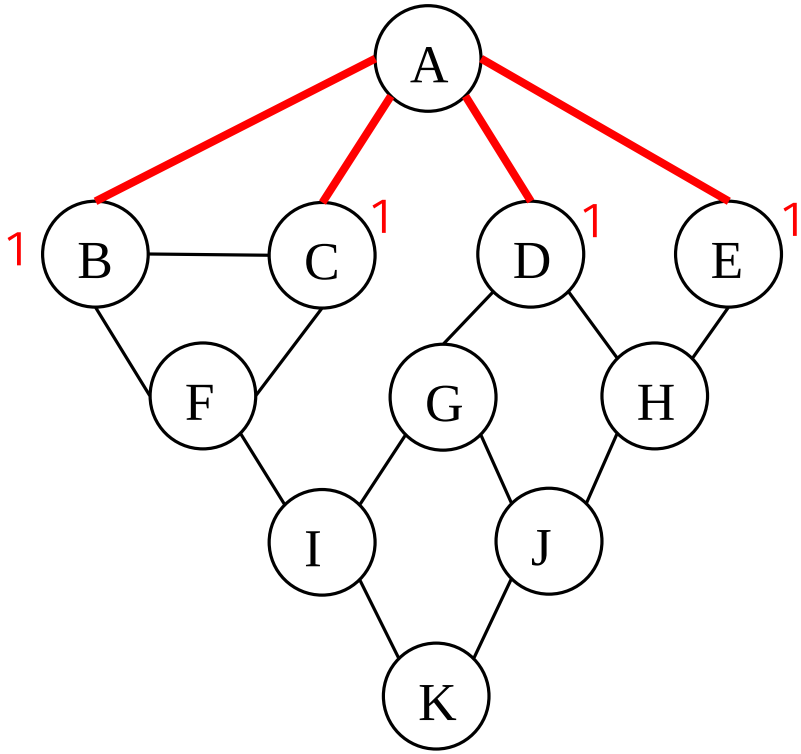
Step 1:





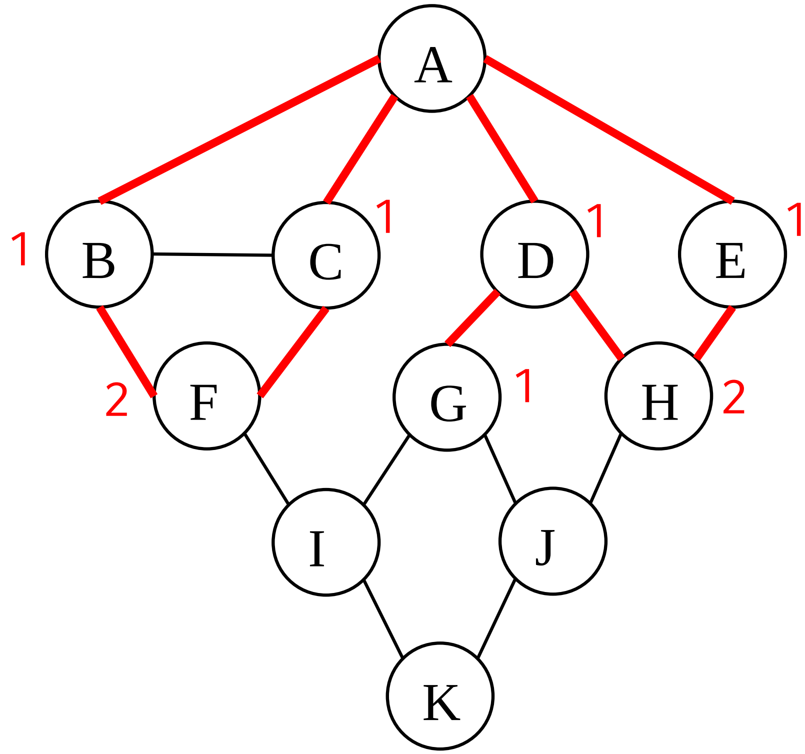
Assignment 12-2

Step 2:



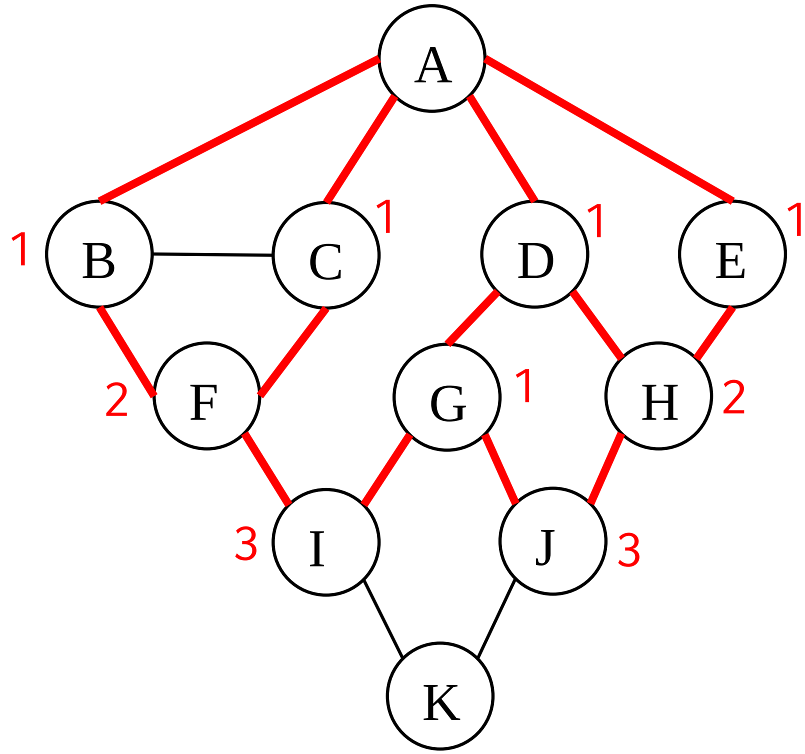
Assignment 12-2

Step 2:



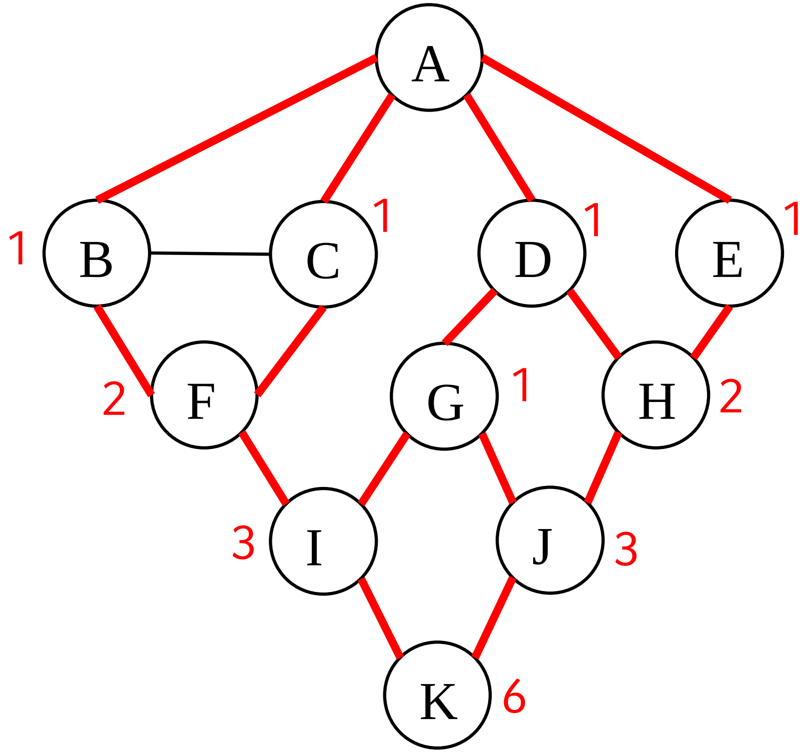
Assignment 12-2

Step 2:



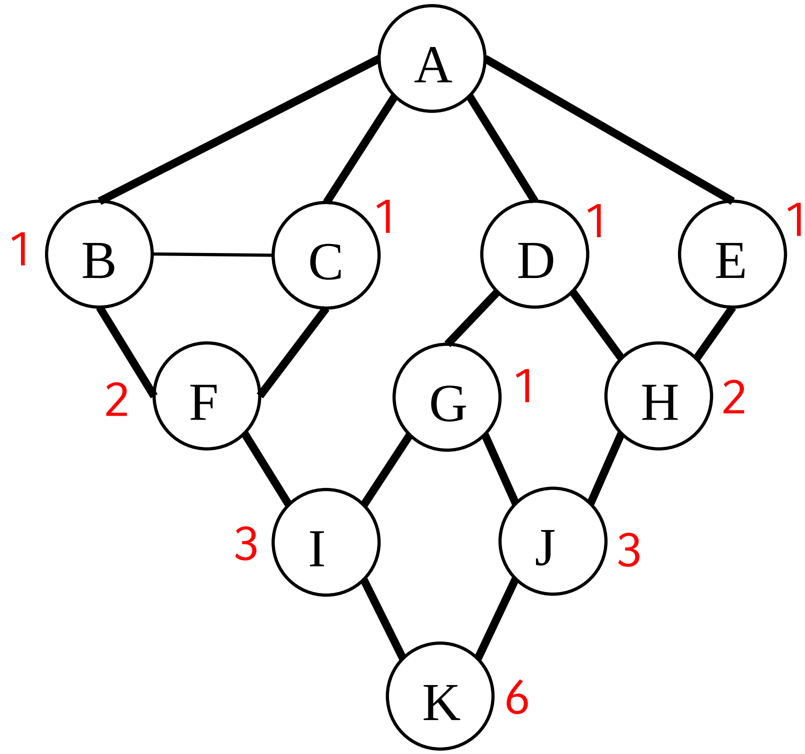
Assignment 12-2

Step 2:

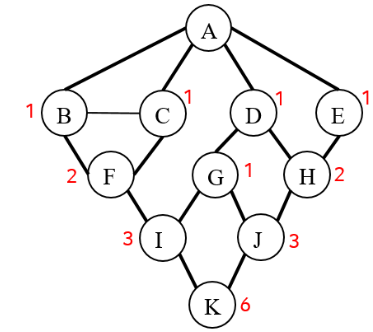


Assignment 12-2

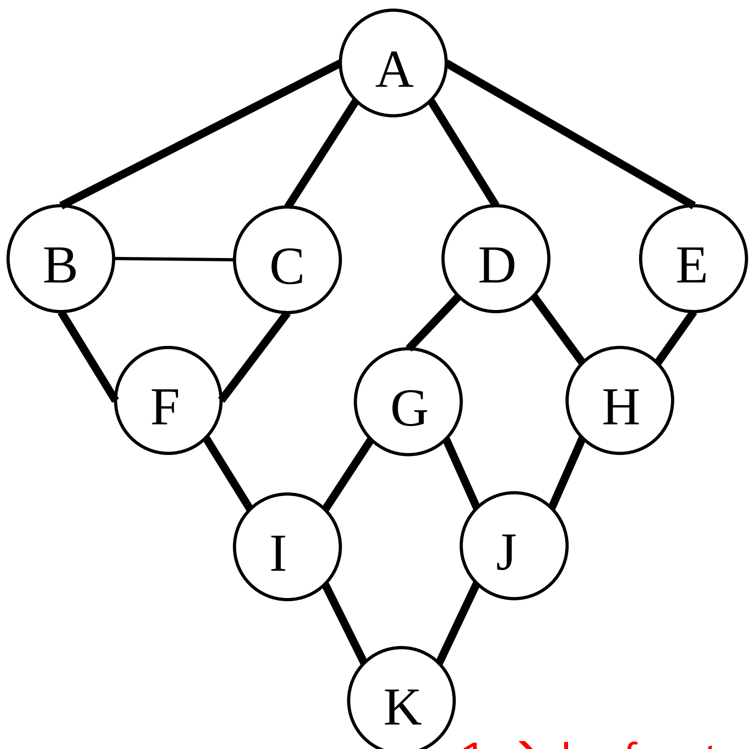
Step 2:



Assignment 12-2



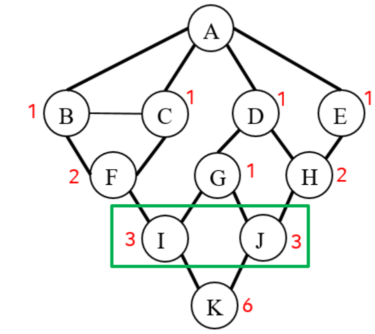
Step 3:

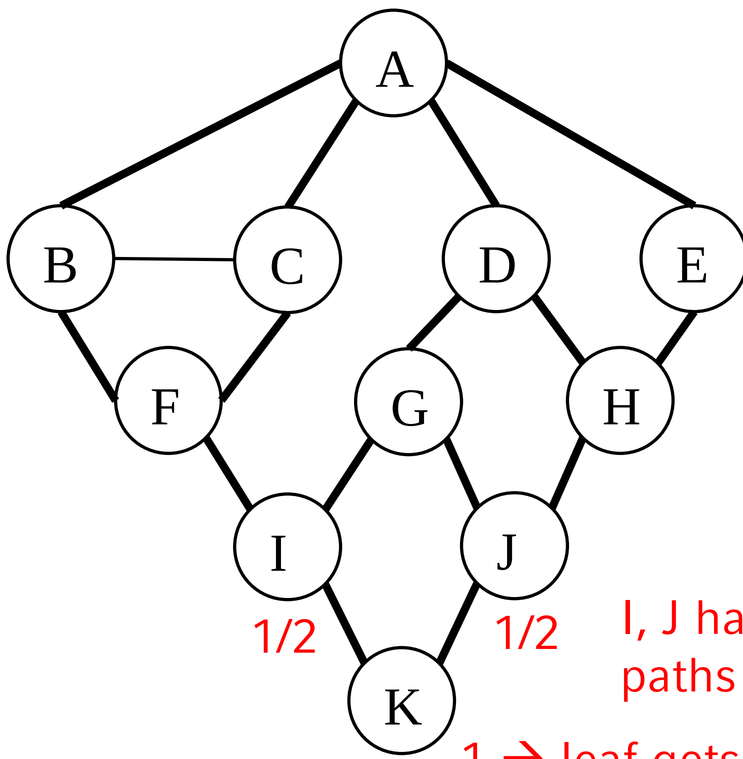


1 → leaf gets a credit of 1.

Assignment 12-2

Step 3:

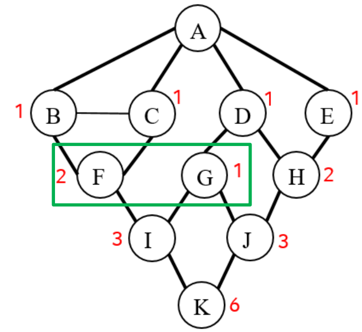




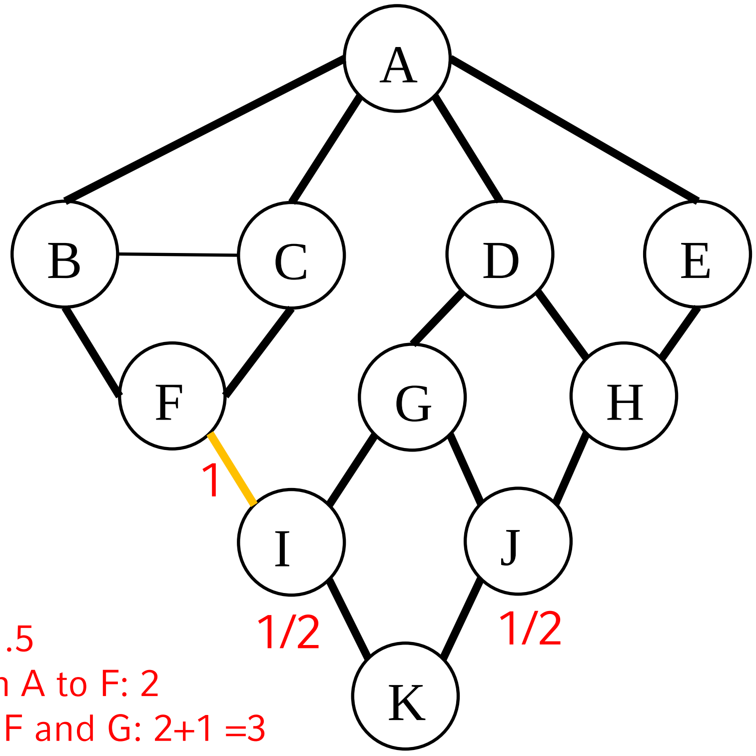
I, J have both equally 3 shortest paths → 3/(3+3) = 1/2

1 → leaf gets a credit of 1.

Assignment 12-2



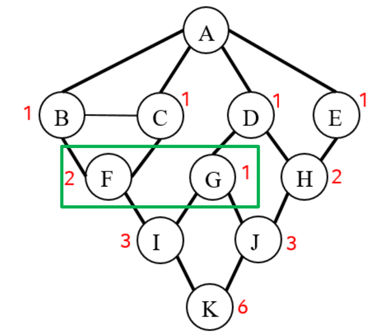
Step 3:



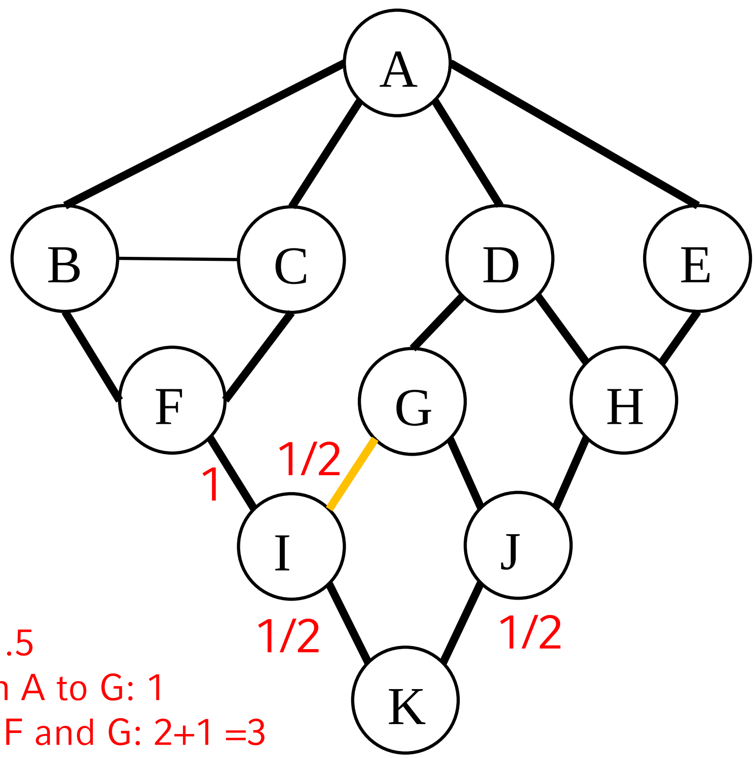
Credit c of I: 1+0.5=1.5
shortest paths from A to F: 2
sum #shortest paths F and G: 2+1 =3

betweenness $B(I,F) = {}^{(1.5*2)}/_{3} = 1$

Assignment 12-2



Step 3:

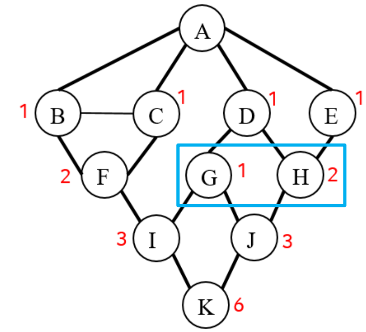


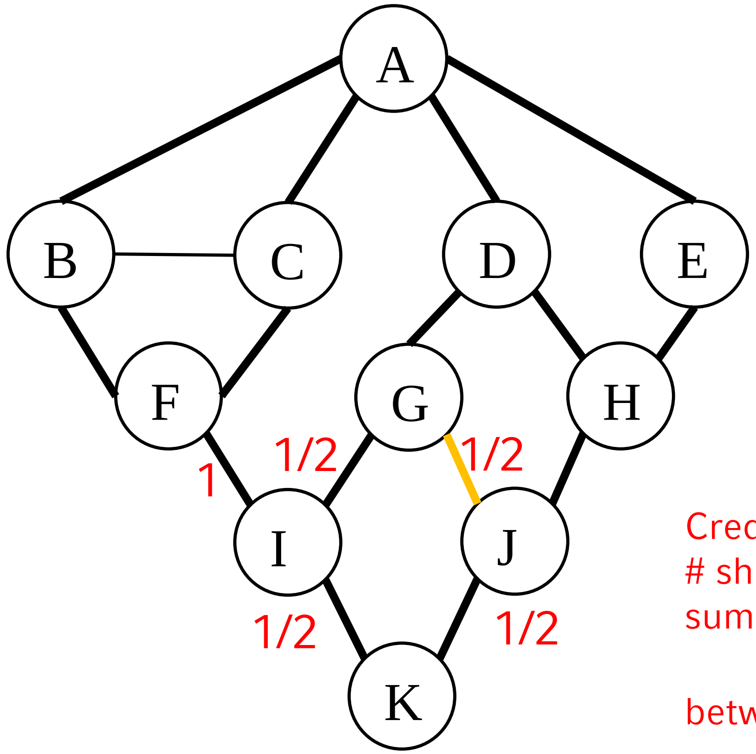
Credit c of I: 1+0.5=1.5
shortest paths from A to G: 1
sum #shortest paths F and G: 2+1 =3

betweenness $B(I,G) = {}^{(1.5 * 1)}/_{3} = {}^{1}/_{2}$

Assignment 12-2

Step 3:



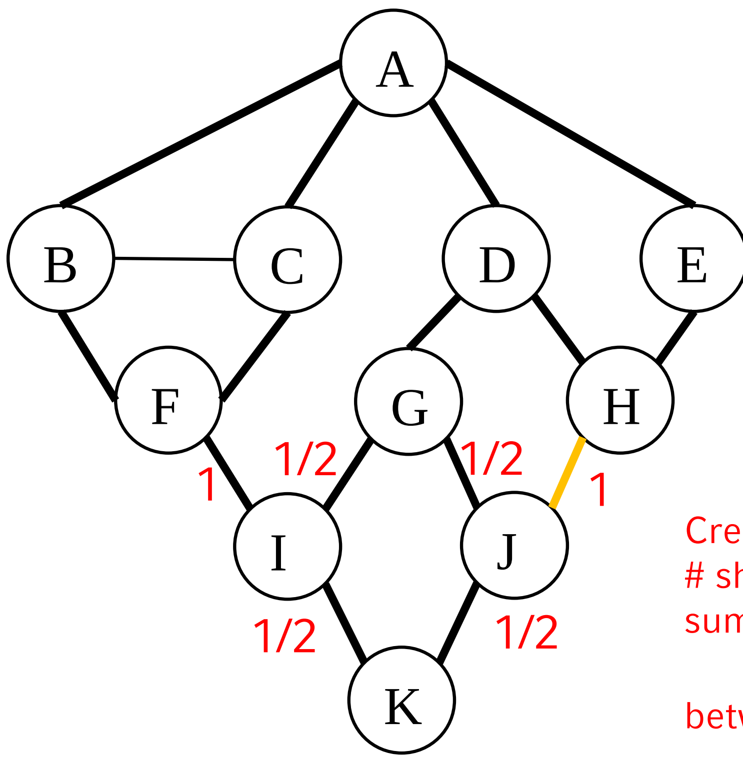


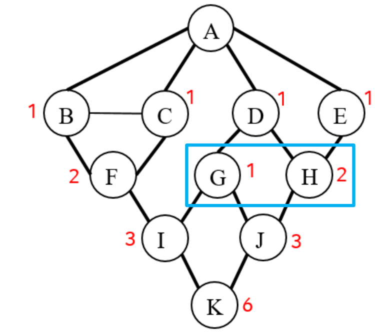
Credit c of J: 1+0.5=1.5
shortest paths from A to G: 1
sum #shortest paths G and H: 1+2 =3

betweenness $B(J,G) = {}^{(1.5\,*1)}/_{3} = {}^{1}/_{2}$

Assignment 12-2

Step 3:

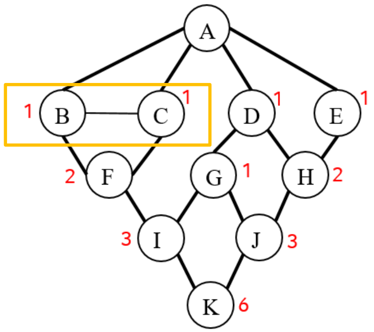




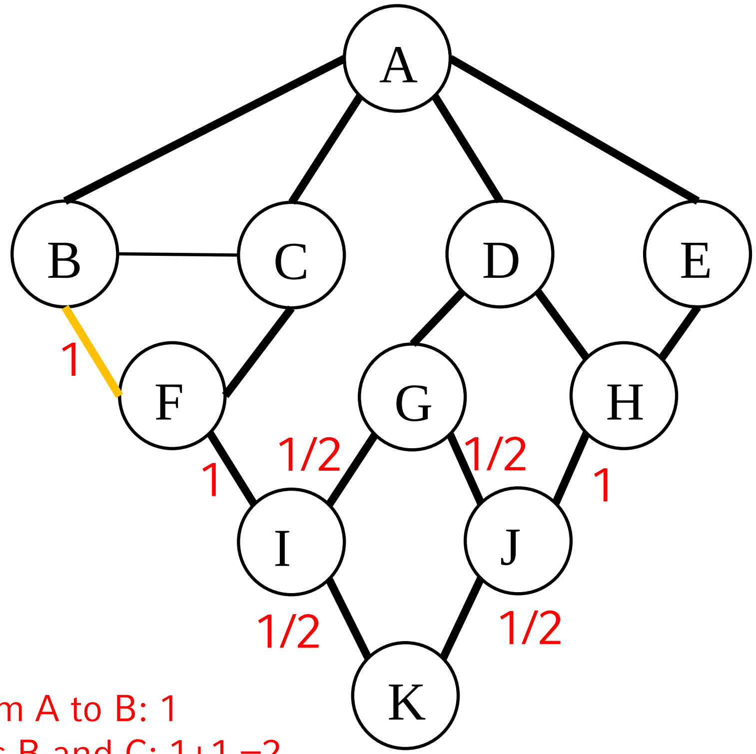
Credit c of J: 1+0.5=1.5
shortest paths from A to H: 2
sum #shortest paths G and H: 1+2 =3

betweenness $B(J,H) = {}^{(1.5 *2)}/_{3} = 1$

Assignment 12-2



Step 3:

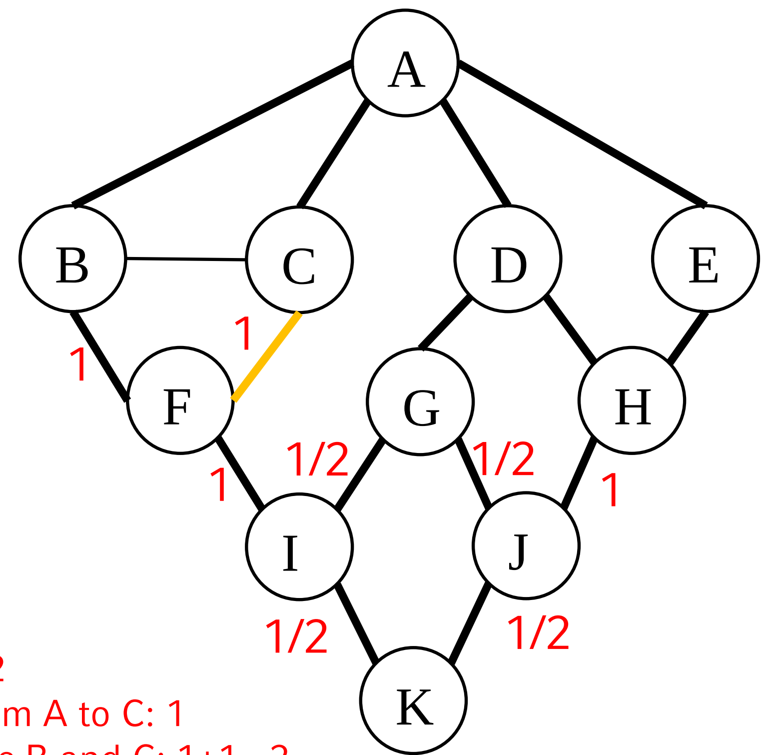


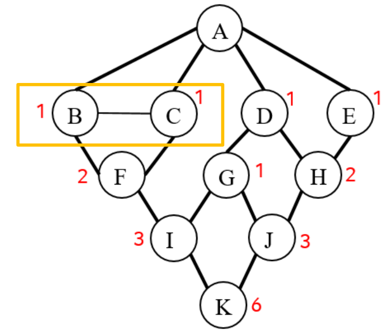
Credit c of F: 1+1=2
shortest paths from A to B: 1
sum #shortest paths B and C: 1+1 =2

betweenness $B(F,B) = {}^{(2*1)}/_{2} = 1$

Assignment 12-2

Step 3:



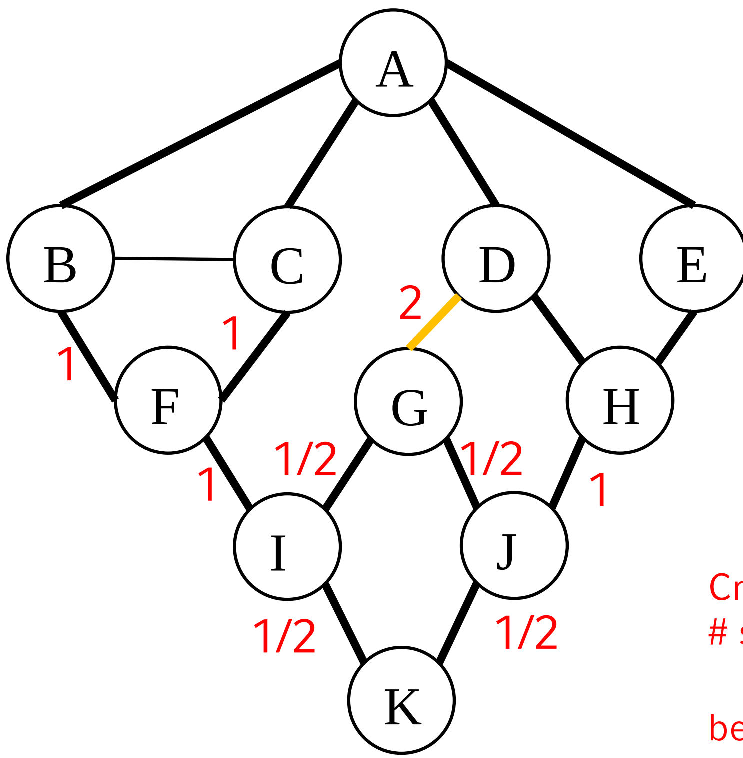


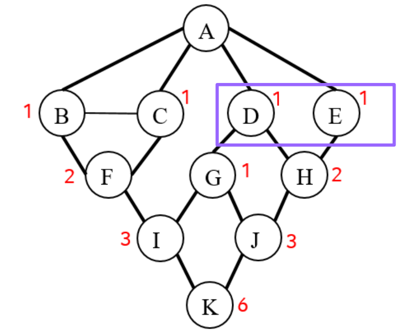
Credit c of F: 1+1=2
shortest paths from A to C: 1
sum #shortest paths B and C: 1+1 =2

betweenness $B(F,C) = {}^{(2*1)}/_{2} = 1$

Assignment 12-2

Step 3:



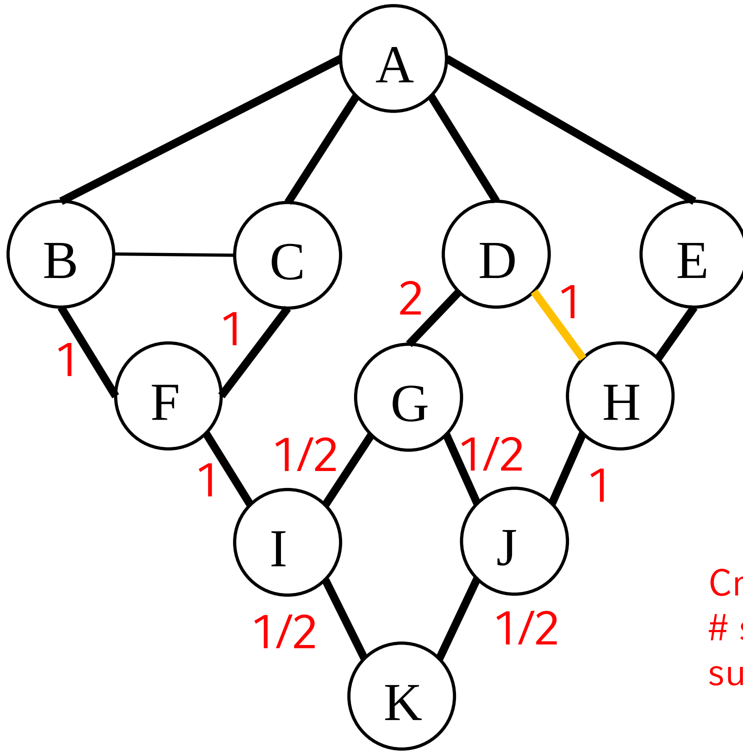


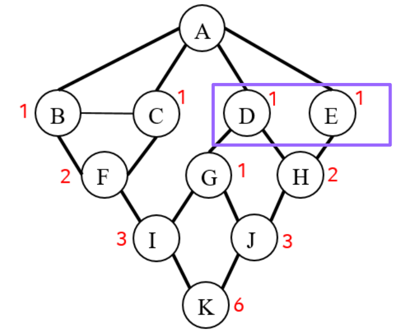
Credit c of G: 1+0.5+0.5=2
shortest paths from A to D: 1

betweenness $B(G,D) = {}^{(2*1)}/_{1} = 2$

Assignment 12-2

Step 3:



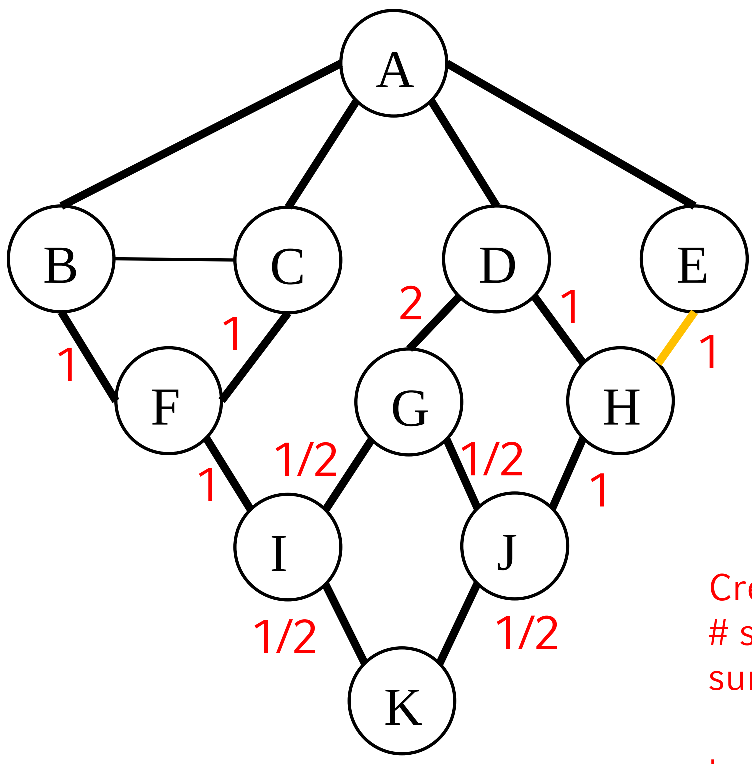


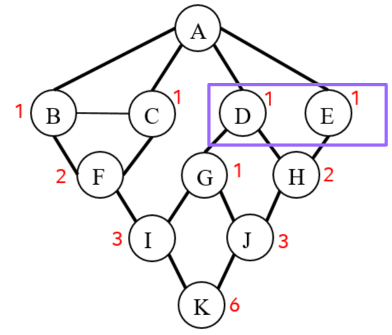
Credit c of H: 1+1=2
shortest paths from A to D: 1
sum #shortest paths D and E: 1+1 =2

betweenness $B(H, D) = {}^{(2*1)}/_{2} = 1$

Assignment 12-2

Step 3:



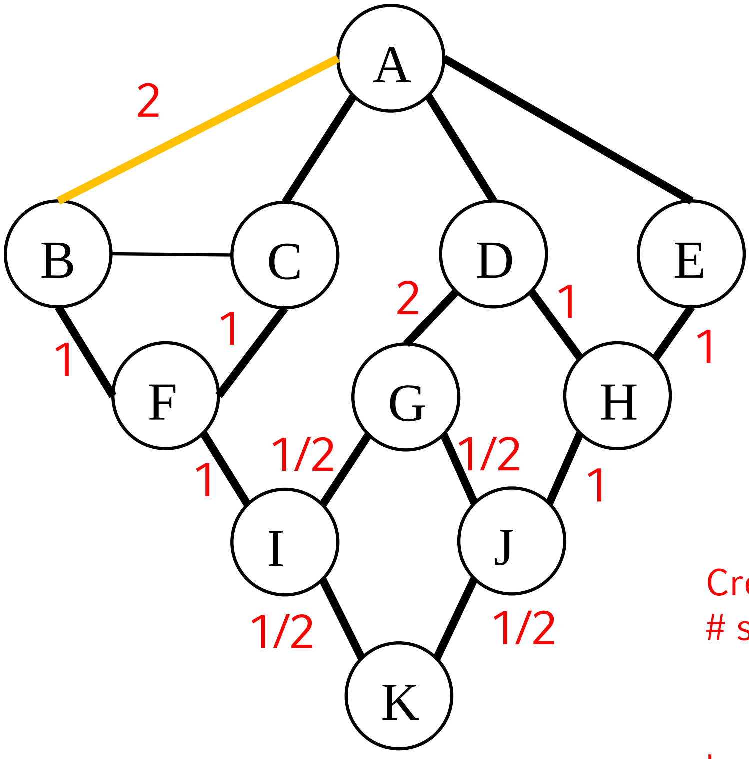


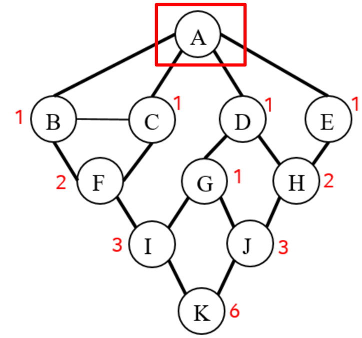
Credit c of H: 1+1=2
shortest paths from A to E: 1
sum #shortest paths D and E: 1+1 =2

betweenness $B(H, E) = {}^{(2*1)}/_{2} = 1$

Assignment 12-2

Step 3:



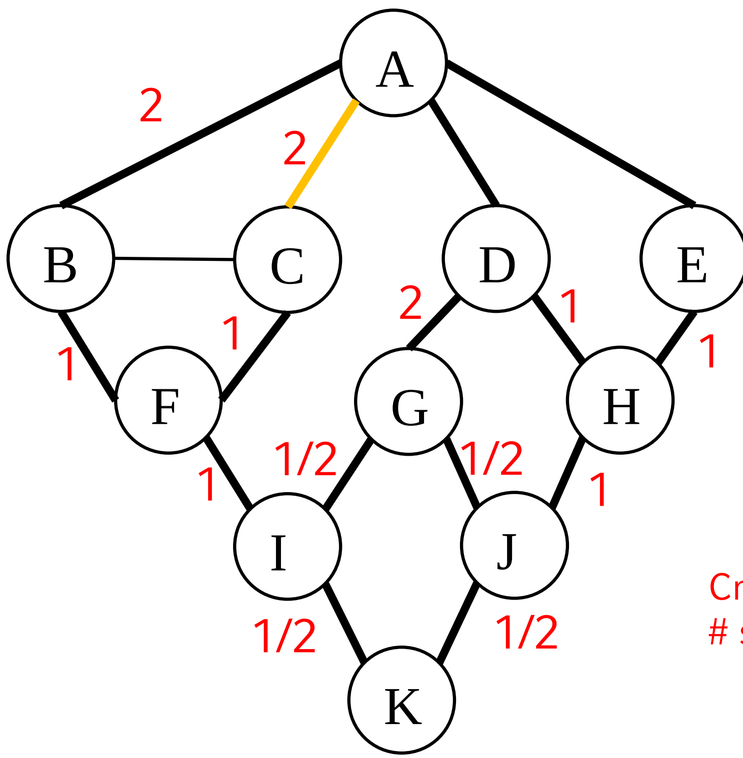


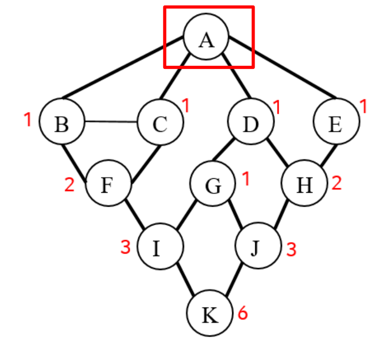
Credit c of B: 1+1=2
shortest paths from A to A: 1

betweenness $B(B, A) = {}^{(2*1)}/_{1} = 2$

Assignment 12-2

Step 3:

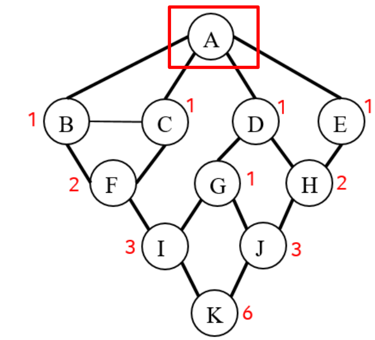




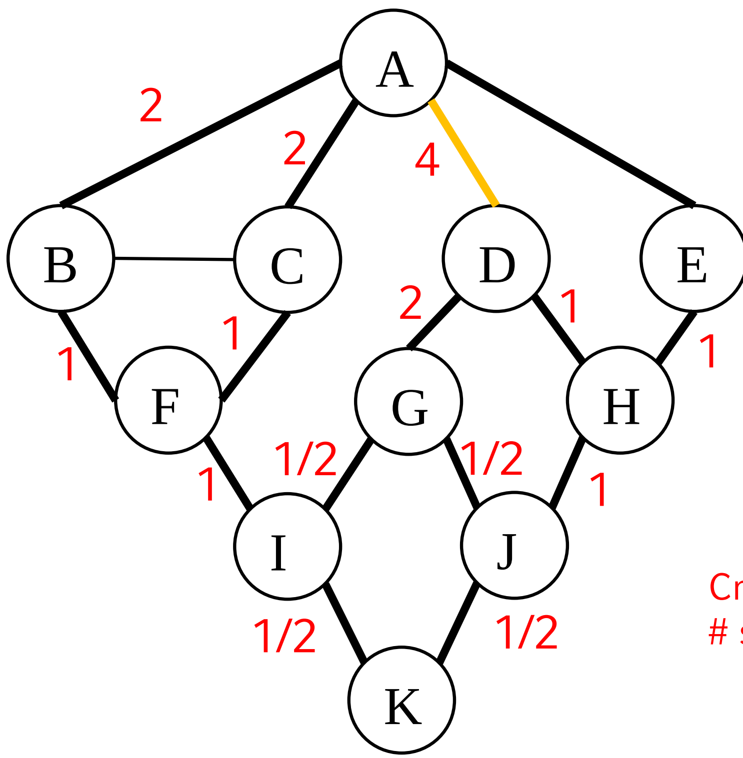
Credit c of C: 1+1=2
shortest paths from A to A: 1

betweenness $B(C,A) = {}^{(2*1)}/_{1} = 2$

Assignment 12-2



Step 3:

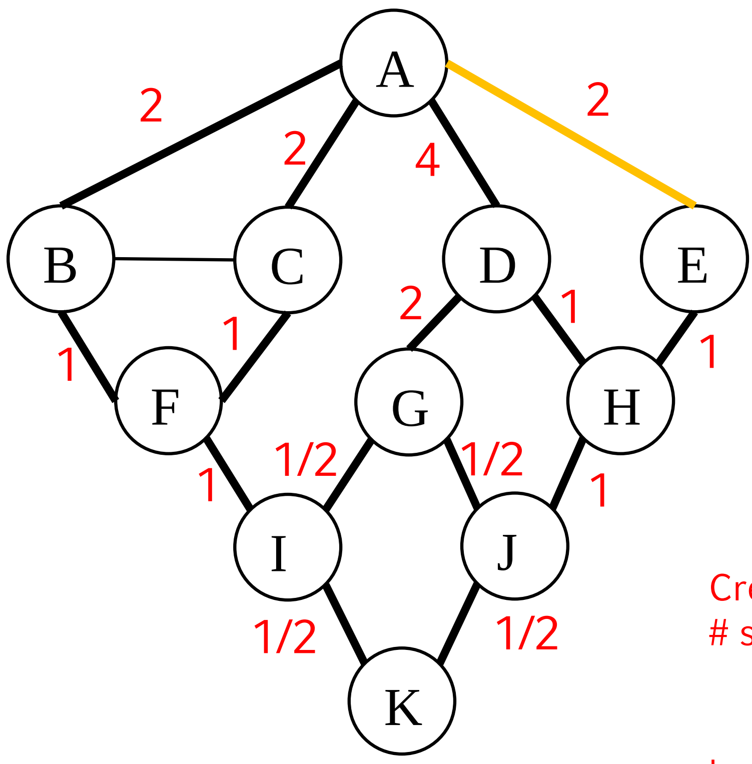


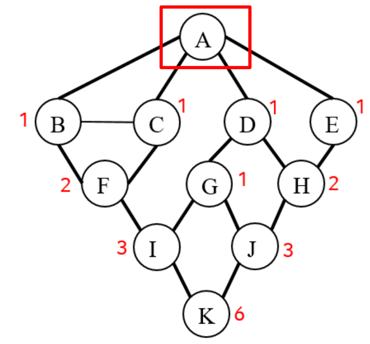
Credit c of D: 1+2+1=4
shortest paths from A to A: 1

betweenness $B(D,A) = {}^{(4*1)}/_{1} = 4$

Assignment 12-2

Step 3:





Credit c of E: 1+1=2

shortest paths from A to A: 1

betweenness $B(E, A) = {}^{(2*1)}/_{1} = 2$