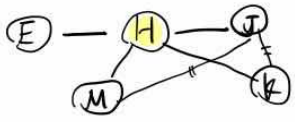
Intelligent Data Analysis CS5152/6052
Homework #3
Due Date: April 22nd, 9PM

Consider the graph shown in the attached file and answer the following questions in its context.

1. What is the clustering coefficient of node H? Show all your work done for deriving yiour answer.

$CC(V)$: V is a node, k_V is its degree, N_V = number of links between neighbors of V

$$CC(V) = \frac{2N_v}{k_v(k_v-1)}, \quad cc(H) = \frac{2\times 2}{4\times 3} = \frac{1}{3}$$

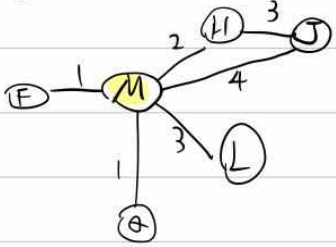


$k_V = 4$

$N_v = 2$

2. What is the eccentricity of node M? Justify your answer by giving reason(s).

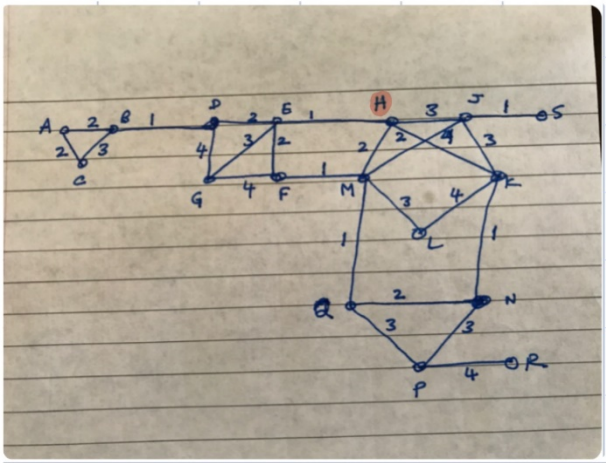
Eccentricity is the maximum distance between a node to all neighbours.



$M,H = 2$
$MJ = 5$
$M,L = 3$
$M,F = 1$
$M,Q = 1$

$\Rightarrow$ Eccentricity of M = 5

3. What is the closeness centrality of node H?



$$\text{closeness} = \frac{1}{\text{Average shoreste path length}}$$

$HA = 4$, $HB = 3$, $HC = 4$, $HD = 2$, $HE = 1$ 14
$HF = 2$, $HG = 2$, $HJ = 1$, $HK = 1$, $HM = 1$ 7
$HL = 2$, $HN = 2$, $HP = 3$, $HQ = 2$, $HR = 4$ 11
$HS = 2$ 2

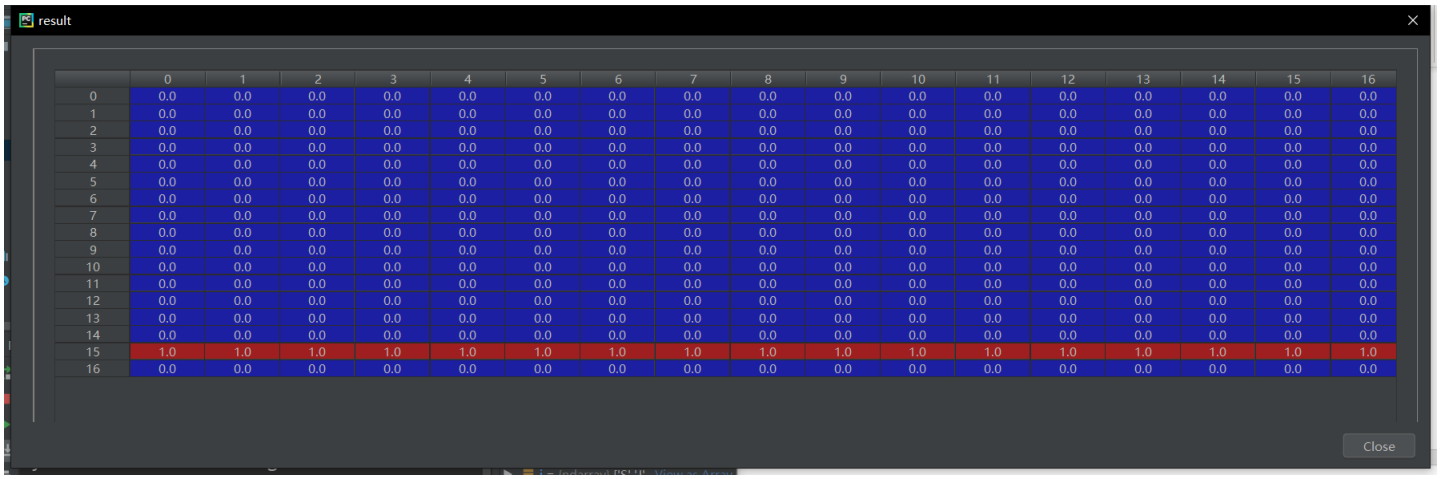
Average shoreste path length $= \frac{1}{(14+7+11+2)/16} = \frac{1}{2.125}$

$= 0.47058$

4. Consider the MCL algorithm covered in class. Apply the same algorithm on the graph attached with this assignment. Include self-loops with a label of ‘1” at every node. You may have to write a small script in Matlab of Python to execute the MCL algorithm. Submit the script written by you along with your answers. Run the MCL algorithm on the given graph for the following inflation values: 1.1, 1.3, 1.5, 1.7, and 2.1. and stop each run after 10 iterations of inflation and expansion. For each run show the following:

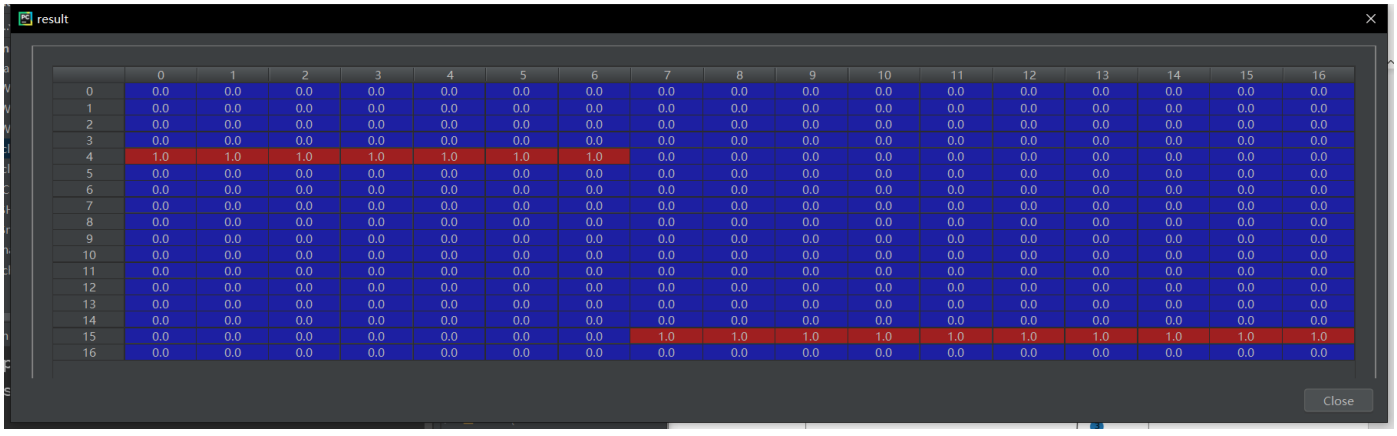
a. The final adjacency matrix.
   The following nodes number (0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16)are according to the string names:A,B,C,D,E,F,G,H,J,K,L,M,N,P,Q,R,S

i. Inflation=1.1



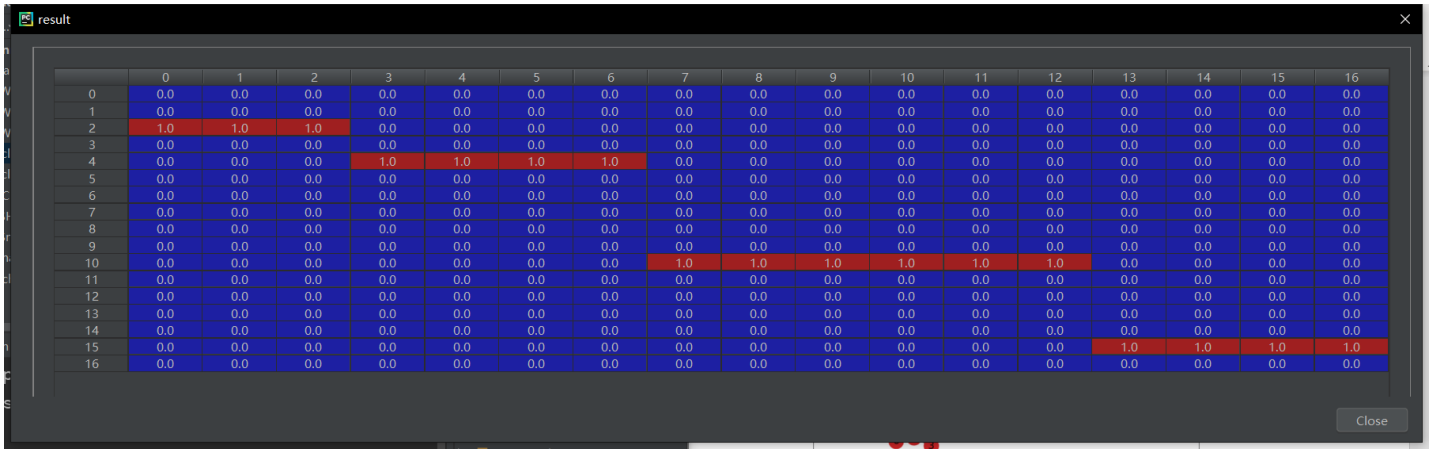
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

ii. Inflation=1.3

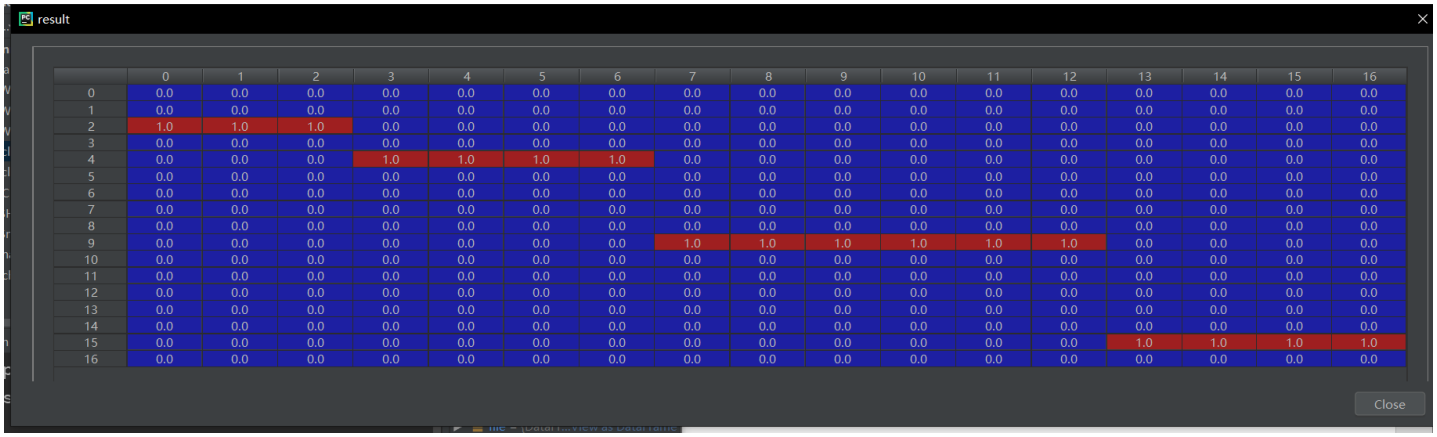


	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

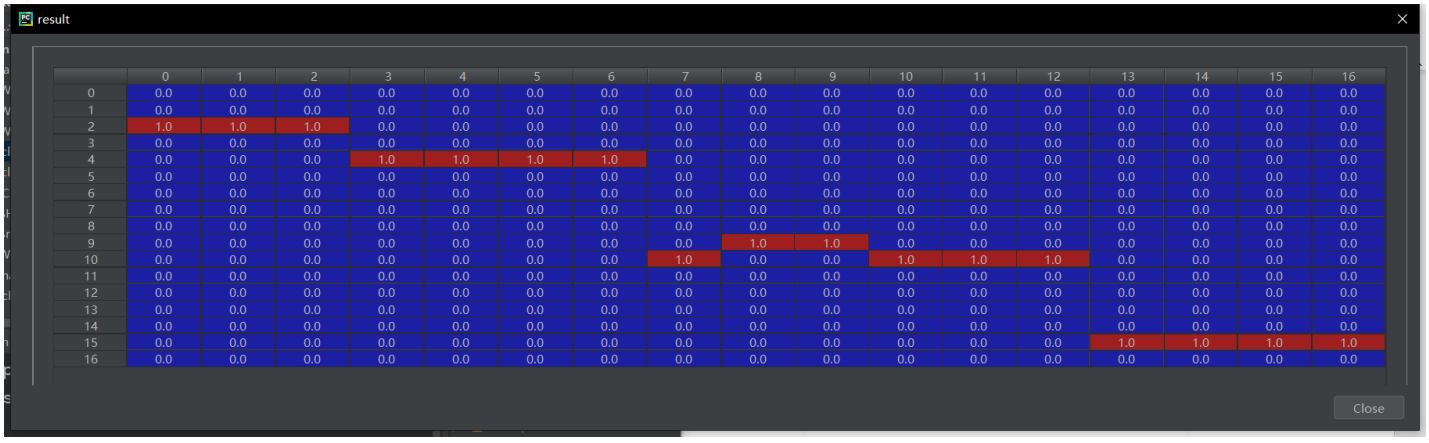
iii. Inflation=1.5



iv. Inflation=1.7

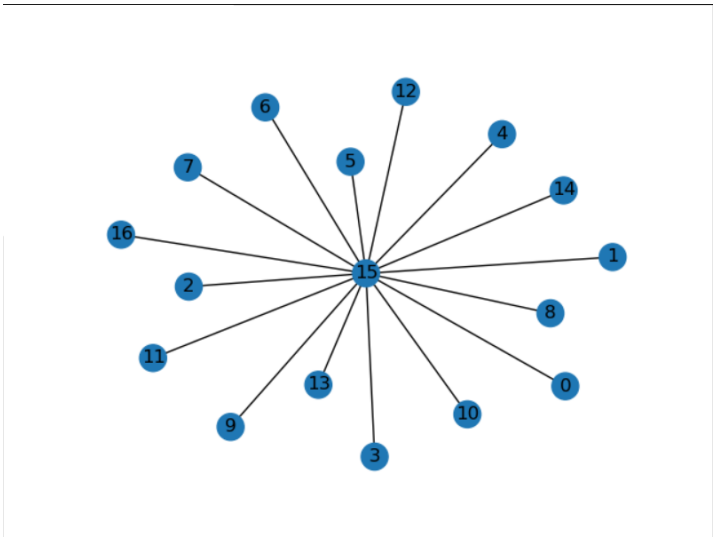


v. Inflation=2.1

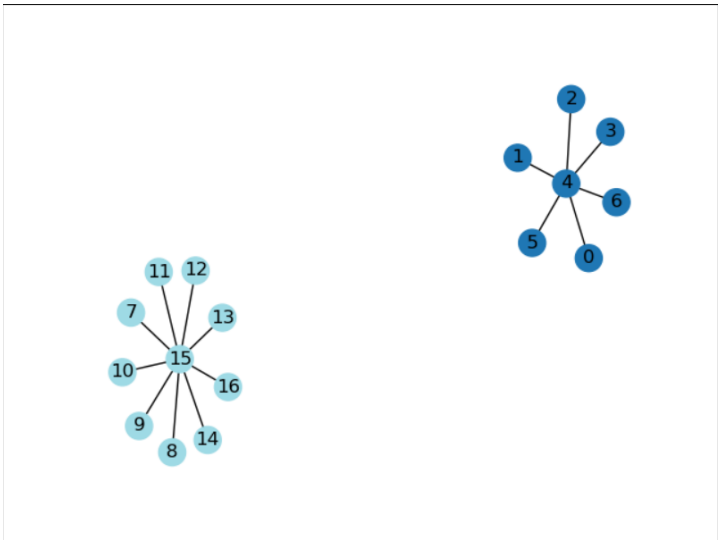


b. The communities of nodes that you see in this final version of the adjacency matrix.

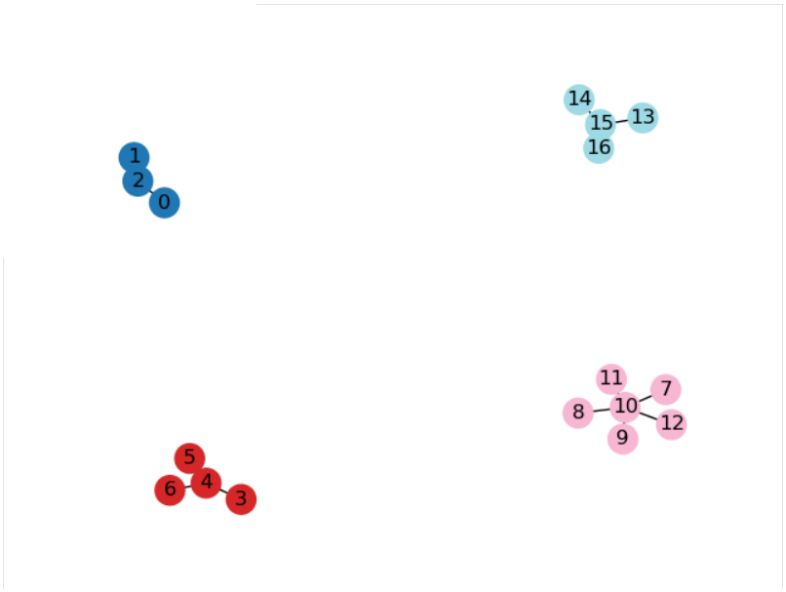
i. Inflation=1.1



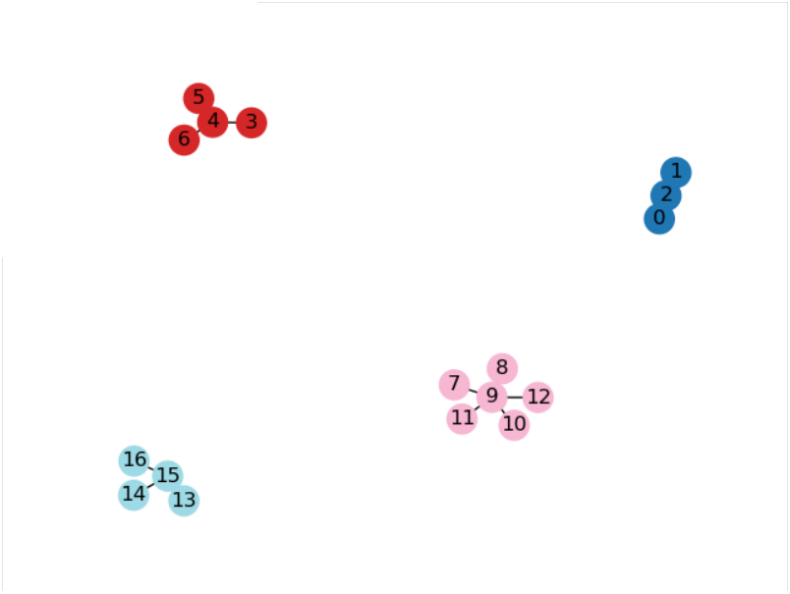
ii. Inflation=1.3



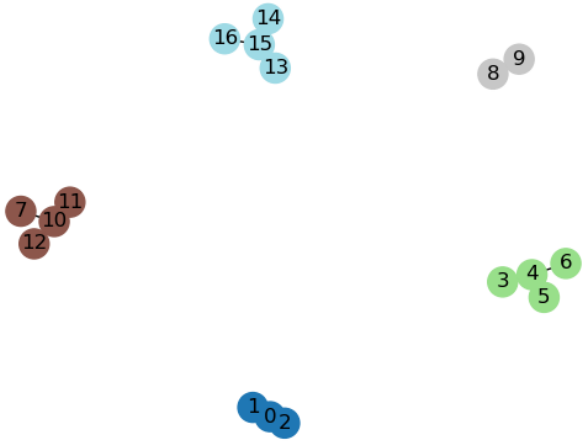
iii. Inflation=1.5



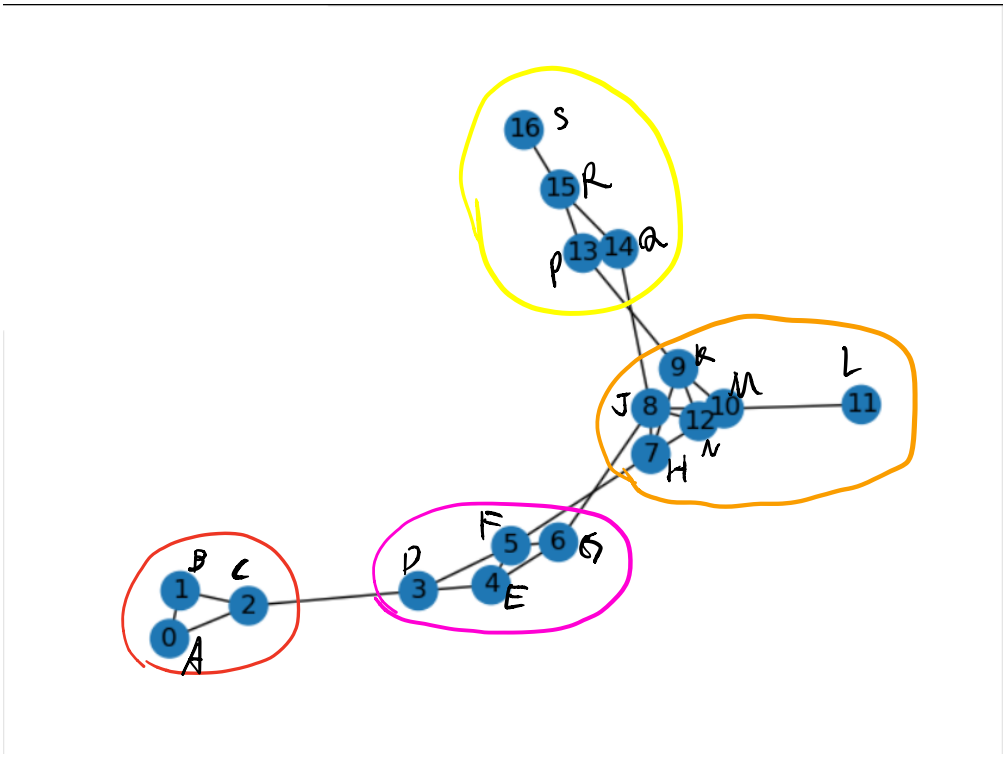
iv. Inflation=1.7



v. Inflation=2.1



c. On a copy of the original graph (enclosed with this assignment) encircle each cluster found by your algorithm.



d. Is this clustering intuitive according to you? Would you have preferred to see a somewhat different community structure? If yes, then what communities would you have liked to find and how would you achieve that? Justify your answer.

This clustering is intuitive. The structure generated by the algorithm is almost the same as I thought. According to the figure in b) part iv and v, when the inflation value = 1.5, 1.7. the four clusters are shown clearly. However, when inflation value keep increasing, the model will be over fitted.