Big Data Analytics – CS7070
Programming Project #2
Phase 4
Submitted by **Ankit Pandey**

Work out on a piece of paper how clustering-coefficient for a node of a graph may be computed from the output obtained after Phase-3. Compute the clustering coefficient for at least 5 nodes of the Tiny Dataset and show all your work.

Out of Phase 1(Adjacency List):

Node	Adjacency List
1	{4-3}
2	{3-6}
3	{1-2-4-6}
4	{7-5-6-1-3}
5	{8-9-4-6}
6	{10-5-4-2-3}
7	{10-8-4}
8	{7-9-5}
9	{8-5-10}
10	{7-9-6}

Out of Phase3(Triangles Formed):

Node	Triangle List
1	[(1,3,4)]
2	[(2,3,6)]
3	[(1,3,4), (3,4,6), (2,3,6)]
4	[(4,5,6), (1,3,4), (3,4,6)]
5	[(5,8,9), (4,5,6)]
6	[(2,3,6), (3,4,6), (4,5,6)]
8	[(5,8,9)]
9	[(5,8,9)]

Algorithm/methodology used to compute the Cluster Coefficient:

1. We have obtained the adjacency list or have the nodes adjacent or connected to each node from phase 1 of the output. We will use this to find the count of the number of nodes say **n**.
2. In phase 3 we have the list of the triangles formed at each node. We will use this to get the count of triangle formed at a given node. Let say this as **nt**.
3. We will use the formulae that is no of triangle formed at the node divided by combination of n to 2 i.e. C(n,2).
   Formula is :- $\frac{\boldsymbol{no.of\ triangles\ (nt)}}{\boldsymbol{no.of\ adj\ nodes(n)\ C2(\ Combination)}}$
   That is,

$$\frac{\boldsymbol{nt}}{\boldsymbol{C(n,2)}}$$

Work showing computation of clustering coefficient for five nodes

a. **Node 8**

nt=1, n=3
C(n,2) =C (3,2) i.e.

$\frac{3!}{2!(3-2)!}$= 3

Cluster Coefficient (CC) for Node 8=$\frac{1}{C(3,2)}$ i.e. $\frac{1}{3}$

Thus, CC for Node8 is 0.33

b. **Node 4**

nt=3, n=5
C(n,2) =C (5,2) i.e.

$\frac{5!}{2!(5-2)!}$= 10

Cluster Coefficient (CC) for Node 4=$\frac{3}{C(5,2)}$ i.e. $\frac{3}{10}$

Thus, CC for Node4 is 0.3

c. **Node 6**

nt=3, n=5
C(n,2) =C (5,2) i.e.

$\frac{5!}{2!(5-2)!}$= 10

Cluster Coefficient (CC) for Node 6=$\frac{3}{C(5,2)}$ i.e. $\frac{3}{10}$

Thus, CC for Node6 is 0.3

d. **Node 9**

nt=1, n=3
C(n,2) =C (3,2) i.e.

$\frac{3!}{2!(3-2)!}$= 3

Cluster Coefficient (CC) for Node 9=$\frac{1}{C(3,2)}$ i.e. $\frac{1}{3}$

Thus, CC for Node9 is 0.33

e. **Node 5**

nt=2, n=4
C(n,2) =C (4,2) i.e.

$\frac{4!}{2!(4-2)!}$= 6

Cluster Coefficient (CC) for Node 5=$\frac{2}{C(4,2)}$ i.e. $\frac{2}{6}$

Thus, CC for Node5 is 0.33