Big Data Analytics – CS7070

Programming Project #2

Phase 2

Submitted By: Ankit Pandey

M13435273

PySpark Code:

```
%spark.pyspark

# read input text file to RDD

rdd = sc.textFile("/tmp/data/tinyDataset.txt")

#Define the List with the list constructor

list_rdd=list()

#Store the rdd in List with the help of collect

list_rdd=rdd.collect()

#Iterating through the Loop to display the graph taken as Input

print("TinyDataSet Graph: (List of edges) as Input:")

for x in range(len(list_rdd)):

   print(list_rdd[x])

#Splitting the RDD at spaces

rdd2=rdd.map(lambda x: x.split())

#FlatMap will flatten multiple list into single list and storing the verices of the graph in two way form assuming it to be a undirected graph

rdd3=rdd2.flatMap(lambda y:[[y[0],y[1]],[y[1],y[0]]])

# reduceByKey Merges the values for each key. It will perform the merging locally on each mapper before sending results to a reducer, similarly to a “combiner” in MapReduce.

rdd4=rdd3.reduceByKey(lambda k,v:k+","+v)

#Mapped to again get the nodes in List format for 2- hop computation

rdd5=rdd4.map(lambda z :[z[0], z[1].split(",")])

#Starting 2 Hop Projection of Graph
```

```
rdd6=rdd5.flatMap(lambda x : [[(x[0],x[1][l]),x[1]] for l in range(len(x[1]))])

rdd7=rdd5.flatMap(lambda x : [[(x[1][m],x[0]),x[1]] for m in range(len(x[1]))])

rdd8 = rdd6.join(rdd7)

rdd9 = rdd8.map(lambda x: [x[0][0] , [x[0][1],x[1][1]]])

rdd10=rdd9.reduceByKey(lambda k,v : k + v)

list_two_hop=rdd10.collect()

#print(list_two_hop)


#print("-------------------------------")

#print(list_two_hop[0])

print("\nOutput for 2 hop project..........\n")

#print(len(list_two_hop[0][0]))

list_temp=list()

final_list=list()

for i in range(len(list_two_hop)):

   tup2=list_two_hop[i]

   print("For Edge:"+tup2[0])

   print()

   list_temp=tup2[1]

   print("2-hop projection is as followed\n")

   for k in range(1,len(list_temp),2):

      print(str(list_temp[k-1])+"->"+str(list_temp[k]))

      print()


   print("------------------------------------------------------------------------------------------")


   #for j in range(len(list_two_hop)):

   #   print("Node:"+tup2[j][0])
```

```
TinyDataSet Graph: (List of edges) as Input:

7  10

7  8

7  4

8  9

8  5

9  5

9 10

10  6

4  5

5  6

4  6

1  4

1  3

2  3

2  6

3  4

3  6
```

Output for 2 hop project..........

```
For Edge:4

2-hop projection is as followed

1->['4', '3']

6->['5', '4', '2', '3', '10']

3->['1', '2', '4', '6']
```

```
7->['10', '8', '4']

5->['6', '8', '9', '4']

------------------------------------------------------------------------------------------
For Edge:10

2-hop projection is as followed

9->['8', '5', '10']

7->['10', '8', '4']

6->['5', '4', '2', '3', '10']

------------------------------------------------------------------------------------------
For Edge:7

2-hop projection is as followed

8->['7', '9', '5']

10->['7', '9', '6']

4->['7', '5', '6', '1', '3']

------------------------------------------------------------------------------------------
For Edge:3
```

```
2-hop projection is as followed

1->['4', '3']

4->['7', '5', '6', '1', '3']

6->['5', '4', '2', '3', '10']

2->['3', '6']

-------------------------------------------------------------------------------------------
For Edge:6

2-hop projection is as followed

10->['7', '9', '6']

4->['7', '5', '6', '1', '3']

3->['1', '2', '4', '6']

5->['6', '8', '9', '4']

2->['3', '6']

-------------------------------------------------------------------------------------------
For Edge:1

2-hop projection is as followed
```

```
4->['7', '5', '6', '1', '3']

3->['1', '2', '4', '6']

-----------------------------------------------------------------------------------------
For Edge:9

2-hop projection is as followed

10->['7', '9', '6']

8->['7', '9', '5']

5->['6', '8', '9', '4']

-----------------------------------------------------------------------------------------
For Edge:8

2-hop projection is as followed

9->['8', '5', '10']

5->['6', '8', '9', '4']

7->['10', '8', '4']

-----------------------------------------------------------------------------------------
For Edge:5
```

```
2-hop projection is as followed

4->['7', '5', '6', '1', '3']

8->['7', '9', '5']

9->['8', '5', '10']

6->['5', '4', '2', '3', '10']

------------------------------------------------------------------------------------------
For Edge:2

2-hop projection is as followed

3->['1', '2', '4', '6']

6->['5', '4', '2', '3', '10']
```

PySpark Code:

```
%spark.pyspark

# read input text file to RDD

rdd = sc.textFile("/tmp/data/SmallDataSet.txt")

#Define the List with the list constructor

list_rdd=list()

#Store the rdd in List with the help of collect

list_rdd=rdd.collect()

#Iterating through the Loop to display the graph taken as Input

print("SmallDataSet Graph: (List of edges) as Input:")

for x in range(len(list_rdd)):

    print(list_rdd[x])

#Splitting the RDD at spaces

rdd2=rdd.map(lambda x: x.split())

#FlatMap will flatten multiple list into single list and storing the verices of the graph in two way form assuming it to be a undirected graph

rdd3=rdd2.flatMap(lambda y:[[y[0],y[1]],[y[1],y[0]]])

# reduceByKey Merges the values for each key. It will perform the merging locally on each mapper before sending results to a reducer, similarly to a “combiner” in MapReduce.

rdd4=rdd3.reduceByKey(lambda k,v:k+","+v)

#Mapped to again get the nodes in List format for 2- hop computation

rdd5=rdd4.map(lambda z :[z[0], z[1].split(",")])

#Starting 2 Hop Projection of Graph

rdd6=rdd5.flatMap(lambda x : [[(x[0],x[1][l]),x[1]] for l in range(len(x[1]))])

rdd7=rdd5.flatMap(lambda x : [[(x[1][m],x[0]),x[1]] for m in range(len(x[1]))])

rdd8 = rdd6.join(rdd7)

rdd9 = rdd8.map(lambda x: [x[0][0] , [x[0][1],x[1][1]]])

rdd10=rdd9.reduceByKey(lambda k,v : k + v)

list_two_hop=rdd10.collect()
```

```
#print(list_two_hop)


#print("------------------------------")
#print(list_two_hop[0])
print("\nOutput for 2 hop project..........\n")
#print(len(list_two_hop[0][0]))
list_temp=list()
final_list=list()
for i in range(len(list_two_hop)):
    tup2=list_two_hop[i]
    print("For Edge:"+tup2[0])
    print()
    list_temp=tup2[1]
    print("2-hop projection is as followed\n")
    for k in range(1,len(list_temp),2):
        print(str(list_temp[k-1])+"->"+str(list_temp[k]))
        print()


    print("----------------------------------------------------------------------------------------------------")


    #for j in range(len(list_two_hop)):
    #   print("Node:"+tup2[j][0])
```

SmallDataSet Graph: (List of edges) as Input:

1	2
2	3
3	4
4	5

1	3
3	5
1	10
1	6
2	6
2	7
2	11
3	7
3	12
3	8
4	8
5	4
4	9
5	9
6	7
8	7
8	9
9	14
9	13
4	13
8	13
8	12
7	12
7	11
11	6
20	19
5	33
14	34
6	10

13	14
13	12
12	11
11	10
28	29
30	28
31	28
32	28
32	31
30	31
29	30
29	31
32	30
24	28
29	24
20	29
25	29
30	25
21	30
26	30
26	31
27	31
32	27
32	23
23	27
27	22
22	31
22	26
26	27

26	21
21	25
25	20
20	24
24	19
19	28
10	19
19	39
10	15
28	40
15	19
40	39
11	15
16	15
16	11
16	12
16	20
16	21
21	17
17	22
17	16
17	12
17	13
18	13
14	18
14	23
18	23
18	22
23	22

21	22
20	21
33	34
23	35
34	35
32	36
35	36
1	37
37	38
10	38
38	39

Output for 2 hop project..........

```
For Edge:4

2-hop projection is as followed

8->['3', '4', '7', '9', '13', '12']

9->['4', '5', '8', '14', '13']

5->['4', '3', '4', '9', '33']

5->['4', '3', '4', '9', '33']

5->['4', '3', '4', '9', '33']

5->['4', '3', '4', '9', '33']
```

```
13->['9', '4', '8', '14', '12', '17', '18']

3->['2', '4', '1', '5', '7', '12', '8']

------------------------------------------------------------------------------------------
For Edge:10

2-hop projection is as followed

1->['37', '2', '3', '10', '6']

19->['24', '28', '10', '39', '15', '20']

11->['2', '7', '6', '12', '10', '15', '16']

38->['37', '10', '39']

6->['1', '2', '7', '11', '10']

15->['10', '19', '11', '16']

------------------------------------------------------------------------------------------
For Edge:12

2-hop projection is as followed

8->['3', '4', '7', '9', '13', '12']
```

```
17->['21', '22', '16', '12', '13']

13->['9', '4', '8', '14', '12', '17', '18']

11->['2', '7', '6', '12', '10', '15', '16']

3->['2', '4', '1', '5', '7', '12', '8']

7->['2', '3', '6', '8', '12', '11']

16->['15', '11', '12', '20', '21', '17']

-----------------------------------------------------------------------------------------

For Edge:20

2-hop projection is as followed

19->['24', '28', '10', '39', '15', '20']

29->['28', '30', '31', '24', '20', '25']

21->['26', '25', '16', '17', '22', '20', '30']

25->['29', '30', '21', '20']

24->['20', '19', '28', '29']

16->['15', '11', '12', '20', '21', '17']
```

```
------------------------------------------------------------------------------------------

For Edge:24

2-hop projection is as followed

29->['28', '30', '31', '24', '20', '25']

19->['24', '28', '10', '39', '15', '20']

28->['29', '30', '31', '32', '24', '19', '40']

20->['25', '24', '16', '21', '19', '29']

------------------------------------------------------------------------------------------

For Edge:26

2-hop projection is as followed

22->['27', '31', '26', '17', '18', '23', '21']

21->['26', '25', '16', '17', '22', '20', '30']

27->['31', '32', '23', '22', '26']

30->['28', '31', '29', '32', '25', '21', '26']

31->['28', '32', '30', '29', '26', '27', '22']

------------------------------------------------------------------------------------------
```

```
For Edge:16

2-hop projection is as followed

21->['26', '25', '16', '17', '22', '20', '30']

17->['21', '22', '16', '12', '13']

11->['2', '7', '6', '12', '10', '15', '16']

15->['10', '19', '11', '16']

20->['25', '24', '16', '21', '19', '29']

12->['16', '17', '3', '8', '7', '13', '11']

--------------------------------------------------------------------------------------------
For Edge:40

2-hop projection is as followed

28->['29', '30', '31', '32', '24', '19', '40']

39->['19', '40', '38']

--------------------------------------------------------------------------------------------
For Edge:23

2-hop projection is as followed
```

```
14->['18', '23', '9', '34', '13']

22->['27', '31', '26', '17', '18', '23', '21']

27->['31', '32', '23', '22', '26']

35->['23', '34', '36']

32->['28', '31', '30', '27', '23', '36']

18->['13', '14', '23', '22']

-----------------------------------------------------------------------------------------
For Edge:31

2-hop projection is as followed

22->['27', '31', '26', '17', '18', '23', '21']

29->['28', '30', '31', '24', '20', '25']

28->['29', '30', '31', '32', '24', '19', '40']

27->['31', '32', '23', '22', '26']

30->['28', '31', '29', '32', '25', '21', '26']

32->['28', '31', '30', '27', '23', '36']
```

```
26->['22', '27', '21', '30', '31']

----------------------------------------------------------------------------------------
For Edge:25

2-hop projection is as followed

21->['26', '25', '16', '17', '22', '20', '30']

29->['28', '30', '31', '24', '20', '25']

30->['28', '31', '29', '32', '25', '21', '26']

20->['25', '24', '16', '21', '19', '29']

----------------------------------------------------------------------------------------
For Edge:15

2-hop projection is as followed

19->['24', '28', '10', '39', '15', '20']

11->['2', '7', '6', '12', '10', '15', '16']

10->['19', '15', '38', '1', '6', '11']

16->['15', '11', '12', '20', '21', '17']
```

```
-----------------------------------------------------------------------------------------

For Edge:18

2-hop projection is as followed

14->['18', '23', '9', '34', '13']

22->['27', '31', '26', '17', '18', '23', '21']

13->['9', '4', '8', '14', '12', '17', '18']

23->['32', '27', '14', '18', '22', '35']

-----------------------------------------------------------------------------------------

For Edge:3

2-hop projection is as followed

1->['37', '2', '3', '10', '6']

8->['3', '4', '7', '9', '13', '12']

2->['1', '3', '6', '7', '11']

5->['4', '3', '4', '9', '33']

7->['2', '3', '6', '8', '12', '11']

4->['3', '5', '8', '5', '9', '13']
```

```
12->['16', '17', '3', '8', '7', '13', '11']

------------------------------------------------------------------------------------------
For Edge:6

2-hop projection is as followed

1->['37', '2', '3', '10', '6']

2->['1', '3', '6', '7', '11']

11->['2', '7', '6', '12', '10', '15', '16']

7->['2', '3', '6', '8', '12', '11']

10->['19', '15', '38', '1', '6', '11']

------------------------------------------------------------------------------------------
For Edge:7

2-hop projection is as followed

8->['3', '4', '7', '9', '13', '12']

2->['1', '3', '6', '7', '11']

11->['2', '7', '6', '12', '10', '15', '16']
```

```
3->['2', '4', '1', '5', '7', '12', '8']

6->['1', '2', '7', '11', '10']

12->['16', '17', '3', '8', '7', '13', '11']

------------------------------------------------------------------------------------------
For Edge:30

2-hop projection is as followed

29->['28', '30', '31', '24', '20', '25']

21->['26', '25', '16', '17', '22', '20', '30']

28->['29', '30', '31', '32', '24', '19', '40']

31->['28', '32', '30', '29', '26', '27', '22']

32->['28', '31', '30', '27', '23', '36']

25->['29', '30', '21', '20']

26->['22', '27', '21', '30', '31']

------------------------------------------------------------------------------------------
For Edge:32

2-hop projection is as followed
```

```
27->['31', '32', '23', '22', '26']

28->['29', '30', '31', '32', '24', '19', '40']

30->['28', '31', '29', '32', '25', '21', '26']

31->['28', '32', '30', '29', '26', '27', '22']

23->['32', '27', '14', '18', '22', '35']

36->['32', '35']

---------------------------------------------------------------------------------------
For Edge:36

2-hop projection is as followed

35->['23', '34', '36']

32->['28', '31', '30', '27', '23', '36']

---------------------------------------------------------------------------------------
For Edge:1

2-hop projection is as followed

10->['19', '15', '38', '1', '6', '11']
```

```
3->['2', '4', '1', '5', '7', '12', '8']

6->['1', '2', '7', '11', '10']

2->['1', '3', '6', '7', '11']

37->['1', '38']

------------------------------------------------------------------------------------------
For Edge:8

2-hop projection is as followed

4->['3', '5', '8', '5', '9', '13']

12->['16', '17', '3', '8', '7', '13', '11']

3->['2', '4', '1', '5', '7', '12', '8']

7->['2', '3', '6', '8', '12', '11']

13->['9', '4', '8', '14', '12', '17', '18']

9->['4', '5', '8', '14', '13']

------------------------------------------------------------------------------------------
For Edge:9

2-hop projection is as followed
```

```
4->['3', '5', '8', '5', '9', '13']
5->['4', '3', '4', '9', '33']
13->['9', '4', '8', '14', '12', '17', '18']
14->['18', '23', '9', '34', '13']
8->['3', '4', '7', '9', '13', '12']
------------------------------------------------------------------------------------------
For Edge:19
2-hop projection is as followed
20->['25', '24', '16', '21', '19', '29']
24->['20', '19', '28', '29']
10->['19', '15', '38', '1', '6', '11']
15->['10', '19', '11', '16']
28->['29', '30', '31', '32', '24', '19', '40']
39->['19', '40', '38']
------------------------------------------------------------------------------------------
```

```
For Edge:29

2-hop projection is as followed

24->['20', '19', '28', '29']

20->['25', '24', '16', '21', '19', '29']

30->['28', '31', '29', '32', '25', '21', '26']

31->['28', '32', '30', '29', '26', '27', '22']

25->['29', '30', '21', '20']

28->['29', '30', '31', '32', '24', '19', '40']

------------------------------------------------------------------------------------------
For Edge:21

2-hop projection is as followed

26->['22', '27', '21', '30', '31']

16->['15', '11', '12', '20', '21', '17']

20->['25', '24', '16', '21', '19', '29']

30->['28', '31', '29', '32', '25', '21', '26']
```

```
25->['29', '30', '21', '20']

22->['27', '31', '26', '17', '18', '23', '21']

17->['21', '22', '16', '12', '13']

-----------------------------------------------------------------------------------------
For Edge:22

2-hop projection is as followed

26->['22', '27', '21', '30', '31']

31->['28', '32', '30', '29', '26', '27', '22']

18->['13', '14', '23', '22']

23->['32', '27', '14', '18', '22', '35']

27->['31', '32', '23', '22', '26']

21->['26', '25', '16', '17', '22', '20', '30']

17->['21', '22', '16', '12', '13']

-----------------------------------------------------------------------------------------
For Edge:17

2-hop projection is as followed
```

```
16->['15', '11', '12', '20', '21', '17']

12->['16', '17', '3', '8', '7', '13', '11']

13->['9', '4', '8', '14', '12', '17', '18']

22->['27', '31', '26', '17', '18', '23', '21']

21->['26', '25', '16', '17', '22', '20', '30']

------------------------------------------------------------------------------------------

For Edge:14

2-hop projection is as followed

18->['13', '14', '23', '22']

23->['32', '27', '14', '18', '22', '35']

13->['9', '4', '8', '14', '12', '17', '18']

34->['33', '35', '14']

9->['4', '5', '8', '14', '13']

------------------------------------------------------------------------------------------

For Edge:33
```

```
2-hop projection is as followed

5->['4', '3', '4', '9', '33']

34->['33', '35', '14']

------------------------------------------------------------------------------------------
For Edge:34

2-hop projection is as followed

35->['23', '34', '36']

14->['18', '23', '9', '34', '13']

33->['34', '5']

------------------------------------------------------------------------------------------
For Edge:27

2-hop projection is as followed

26->['22', '27', '21', '30', '31']

32->['28', '31', '30', '27', '23', '36']

23->['32', '27', '14', '18', '22', '35']

31->['28', '32', '30', '29', '26', '27', '22']
```

```
22->['27', '31', '26', '17', '18', '23', '21']

-----------------------------------------------------------------------------------------

For Edge:28

2-hop projection is as followed

40->['28', '39']

24->['20', '19', '28', '29']

30->['28', '31', '29', '32', '25', '21', '26']

31->['28', '32', '30', '29', '26', '27', '22']

32->['28', '31', '30', '27', '23', '36']

19->['24', '28', '10', '39', '15', '20']

29->['28', '30', '31', '24', '20', '25']

-----------------------------------------------------------------------------------------

For Edge:39

2-hop projection is as followed

40->['28', '39']
```

```
38->['37', '10', '39']

19->['24', '28', '10', '39', '15', '20']

-------------------------------------------------------------------------------------------
For Edge:11

2-hop projection is as followed

16->['15', '11', '12', '20', '21', '17']

12->['16', '17', '3', '8', '7', '13', '11']

10->['19', '15', '38', '1', '6', '11']

15->['10', '19', '11', '16']

7->['2', '3', '6', '8', '12', '11']

6->['1', '2', '7', '11', '10']

2->['1', '3', '6', '7', '11']

-------------------------------------------------------------------------------------------
For Edge:13

2-hop projection is as followed

4->['3', '5', '8', '5', '9', '13']
```

```
12->['16', '17', '3', '8', '7', '13', '11']

18->['13', '14', '23', '22']

17->['21', '22', '16', '12', '13']

9->['4', '5', '8', '14', '13']

8->['3', '4', '7', '9', '13', '12']

14->['18', '23', '9', '34', '13']

-----------------------------------------------------------------------------------------

For Edge:38

2-hop projection is as followed

10->['19', '15', '38', '1', '6', '11']

39->['19', '40', '38']

37->['1', '38']

-----------------------------------------------------------------------------------------

For Edge:5

2-hop projection is as followed
```

```
4->['3', '5', '8', '5', '9', '13']

4->['3', '5', '8', '5', '9', '13']

4->['3', '5', '8', '5', '9', '13']

4->['3', '5', '8', '5', '9', '13']

3->['2', '4', '1', '5', '7', '12', '8']

9->['4', '5', '8', '14', '13']

33->['34', '5']

----------------------------------------------------------------------------------------
For Edge:35

2-hop projection is as followed

23->['32', '27', '14', '18', '22', '35']

36->['32', '35']

34->['33', '35', '14']

----------------------------------------------------------------------------------------
For Edge:2

2-hop projection is as followed
```

```
3->['2', '4', '1', '5', '7', '12', '8']

6->['1', '2', '7', '11', '10']

7->['2', '3', '6', '8', '12', '11']

11->['2', '7', '6', '12', '10', '15', '16']

1->['37', '2', '3', '10', '6']

------------------------------------------------------------------------------------------
For Edge:37

2-hop projection is as followed

38->['37', '10', '39']

1->['37', '2', '3', '10', '6']

------------------------------------------------------------------------------------------
```