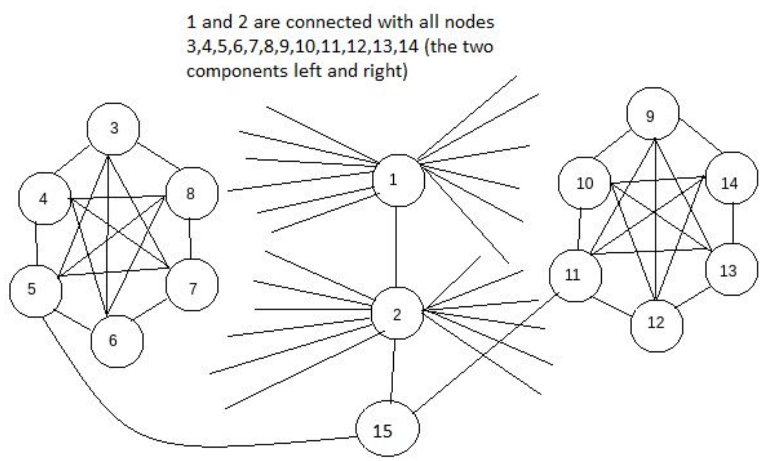
Social Network Analysis

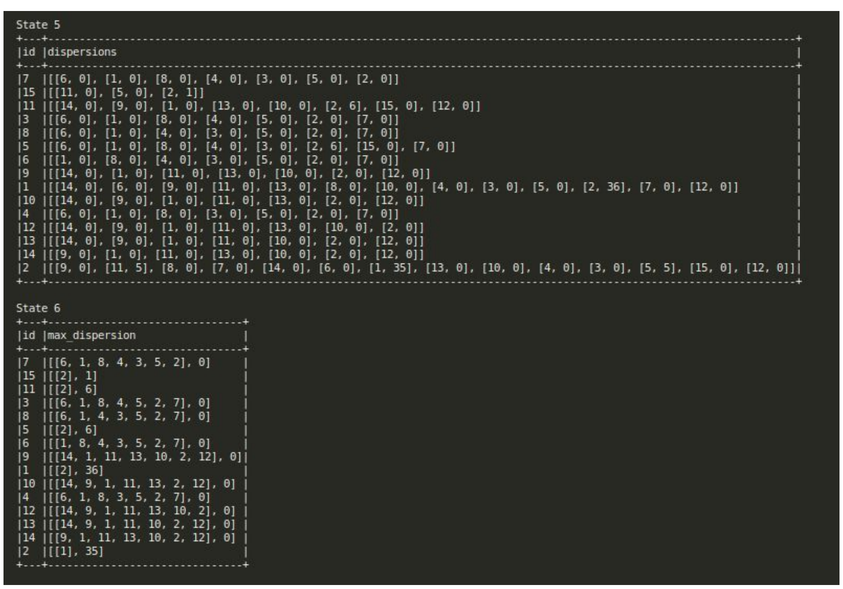
Project 1 Report

Stefou Theodoros – Στέφου Θεόδωρος
cs2190002@di.uoa.gr / theo.stefou@gmail.com
AM: CS2190002

For this project I have constructed the following undirected graph on paper and created the corresponding csv files of the vertices and edges manually.



The following picture shows the final results of my code for the dispersion of every node with its neighbors, as well as the neighbors with maximum dispersion value.



State 5

id	dispersions
7	[[6, 0], [1, 0], [8, 0], [4, 0], [3, 0], [5, 0], [2, 0]]
15	[[11, 0], [5, 0], [2, 1]]
11	[[14, 0], [9, 0], [1, 0], [13, 0], [10, 0], [2, 6], [15, 0], [12, 0]]
3	[[6, 0], [1, 0], [8, 0], [4, 0], [5, 0], [2, 0], [7, 0]]
8	[[6, 0], [1, 0], [4, 0], [3, 0], [5, 0], [2, 0], [7, 0]]
5	[[6, 0], [1, 0], [8, 0], [4, 0], [3, 0], [2, 6], [15, 0], [7, 0]]
6	[[1, 0], [8, 0], [4, 0], [3, 0], [5, 0], [2, 0], [7, 0]]
9	[[14, 0], [1, 0], [11, 0], [13, 0], [10, 0], [2, 0], [12, 0]]
1	[[14, 0], [6, 0], [9, 0], [11, 0], [13, 0], [8, 0], [10, 0], [4, 0], [3, 0], [5, 0], [2, 36], [7, 0], [12, 0]]
10	[[14, 0], [9, 0], [1, 0], [11, 0], [13, 0], [2, 0], [12, 0]]
4	[[6, 0], [1, 0], [8, 0], [3, 0], [5, 0], [2, 0], [7, 0]]
12	[[14, 0], [9, 0], [1, 0], [11, 0], [13, 0], [10, 0], [2, 0]]
13	[[14, 0], [9, 0], [1, 0], [11, 0], [10, 0], [2, 0], [12, 0]]
14	[[9, 0], [1, 0], [11, 0], [13, 0], [10, 0], [2, 0], [12, 0]]
2	[[9, 0], [11, 5], [8, 0], [7, 0], [14, 0], [6, 0], [1, 35], [13, 0], [10, 0], [4, 0], [3, 0], [5, 5], [15, 0], [12, 0]]

State 6

id	max_dispersion
7	[[6, 1, 8, 4, 3, 5, 2], 0]
15	[[2], 1]
11	[[2], 6]
3	[[6, 1, 8, 4, 5, 2, 7], 0]
8	[[6, 1, 4, 3, 5, 2, 7], 0]
5	[[2], 6]
6	[[1, 8, 4, 3, 5, 2, 7], 0]
9	[[14, 1, 11, 13, 10, 2, 12], 0]
1	[[2], 36]
10	[[14, 9, 1, 11, 13, 2, 12], 0]
4	[[6, 1, 8, 3, 5, 2, 7], 0]
12	[[14, 9, 1, 11, 13, 10, 2], 0]
13	[[14, 9, 1, 11, 10, 2, 12], 0]
14	[[9, 1, 11, 13, 10, 2, 12], 0]
2	[[1], 35]

As the project requires, I have manually calculated the dispersion of node 2 (chosen as the most interesting) with all of its neighbors.

Dispersion calculation for the node with id = 2 against all its neighbors

Node 2 neighbors: [1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]

- $disp_2(1)$

Their common neighbors are: [3,4,5,6,7,8,9,10,11,12,13,14] and initially $disp_2(1) = 0$.
For each unordered pair of common neighbors, in order to add +1 to $disp_2(1)$, the two nodes must not be connected and share no other common neighbors except 1 and 2 in the ego-network of 2.
The following pairs are connected and therefore do not contribute to the dispersion:
(3,4), (3,5), (3,6), (3,7), (3,8)
For (3,9): (3,9) are a) not connected and b) share no other neighbors except 1 and 2 in the ego-network of 2. Therefore, $\underline{disp_2(1) \mathrel{+}= 1}$. The same goes for (3,10), (3,11), (3,12), (3,13), (3,14). $\underline{disp_2(1) \mathrel{+}= 5}$
In a symmetric fashion we continue for the rest of the nodes.
(4,5), (4,6), (4,7), (4,8) all connected and do not contribute, but (4,9), (4,10), (4,11), (4,12), (4,13), (4,14) all not connected and share only 1 and 2 => $\underline{disp_2(1) \mathrel{+}= 6}$
The same pattern continues for all pairs starting with 6,7,8 => 6+6+6 = 18 => $disp_2(1) \mathrel{+}= 18$
However, for pairs that start with 5, the pair (5,11) does not contribute to the total dispersion because they have an extra common neighbor, which is 15. $disp_2(1) \mathrel{+}= 5$.
The rest of the pais, which start with 9, 10, 11, 12, 13, 14 are all connected and do not cotribute to the total dispersion.
Concluding, $\underline{disp_2(1) = 35}$

- $disp_2(3)$

Common neighbors: [4, 5, 6, 7, 8, 1]
These nodes are all connected to each other. No pair contributes to the total dispersion.
Therefore,$\underline{disp_2(3) = 0}$

For the same reason as above, disp = 0 for the following pairs of nodes:
(2,4), (2,6), (2,7), (2, 8), (2, 9), (2, 10), (2, 12), (2, 13), (2, 14)

- $disp_2(5)$

Common neighbors: [1, 3, 4, 6, 7, 8, 15]

Pairs that start with 1, 3, 4, 6, 7 are all connected and do not contribute to the total dispersion with the following exceptions:
(1,15) are not connected but share 3 common neighbors in the ego-network of 2, which are 2, 5, 11. Therefore, this pair does not contribute to the total dispersion.
The rest of the pairs that end with 15, which are (3,15), (4, 15), (6, 15), (7, 15), (8, 15) are not connected and only share 2 and 5 as common neighbors in the ego-network of 2. Therefore, each one of these pairs contributes +1 to the total dispersion.
Concluding, $\underline{disp_2(5) = 5}$

- $disp_2(11)$

This is symmetric to the previous calculations.
$\underline{disp_2(11) = 5}$

- $disp_2(15)$

Common neighbors: [5,11]

Only one pair here, (5,11)
5 and 11 are not connected but share 3 neighbors in the ego-network of 2, which are 1, 2, 15, therefore no contribution to the total dispersion.
Concluding, $\underline{disp_2(15) = 0}$

<u>Conclusion</u>

We have presented an undirected graph and using the Spark and Graphframes frameworks in python we have computed in a distributed way the dispersion of each directed pair of connected nodes in the graph, as well as the neighbors with the maximum dispersion value for each node. Then in this project report we manually computed the dispersion of node 2 with all of its neighbors. This node was chosen because it is the most interesting one in the graph. You may find the code for this project in this github repository
https://github.com/TheoStefou/graph-frames-spark-dispersion