Assignment -3 :

Communities Detection

Harsh Gupta (Sr.No : 20961) Sept 19, 2022

Here we have two Datasets:

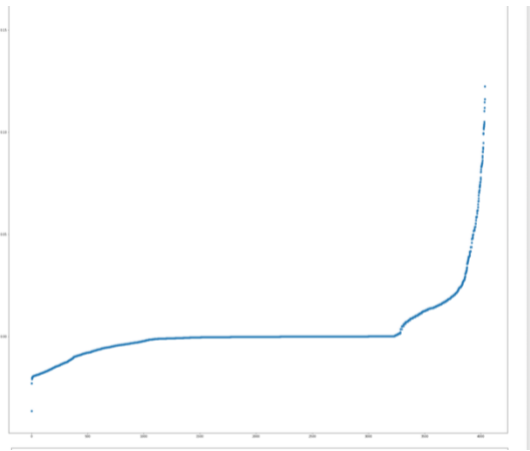
a) Facebook Data (Undirected Unweighted Graph)
b) Bitcoin Data Rating (Directed Weighted Graph and Disconnected Too)

Here , In this Assignment , we have implemented two techniques:
a) Spectral Clustering
b) Louvian Algorithm

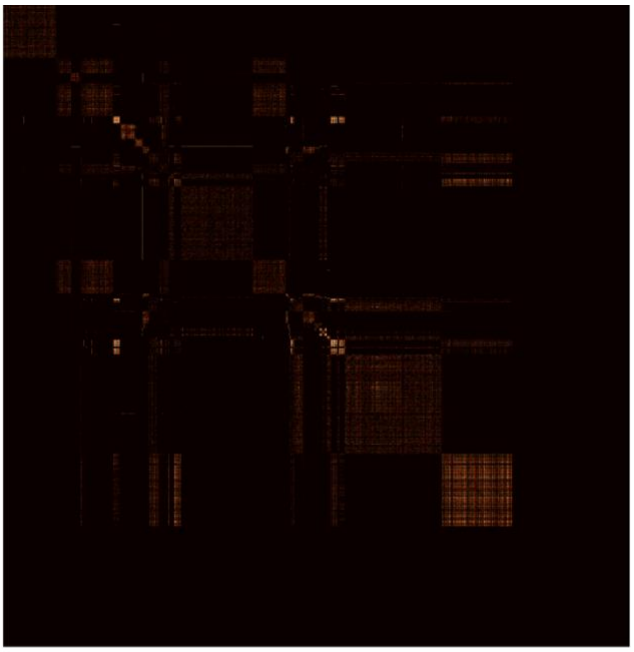
Let's see the different Questions result with respect to different Datasets:
Problem 1:

a) Facebook Dataset ,
   i) Plot of Graph Partition

ii) Plot of Sorted Fiedler Vector:

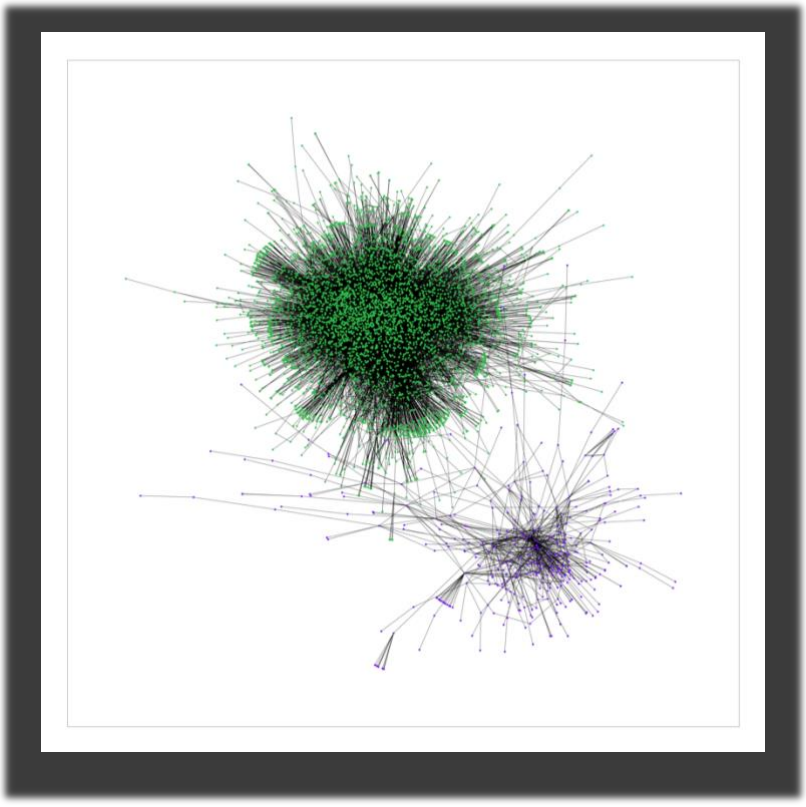


iii) Plot of sorted Adj_Mat:

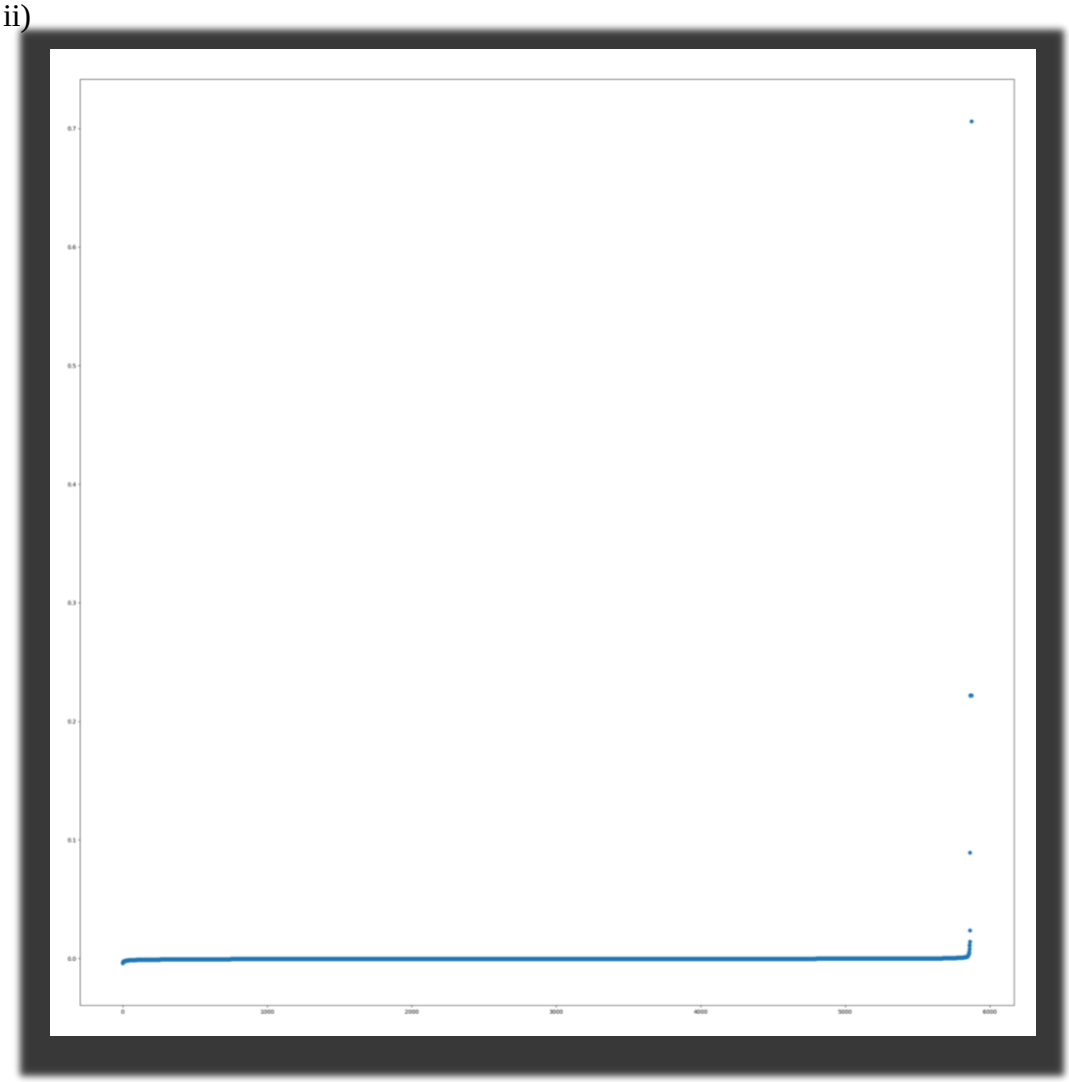


b) Bitcoin Dataset:

i) Plot of Graph Partition



ii)



Problem 2:

The above algorithm is for one Iteration i.e. it breaks the graph into two communities. So we can have algorithm based on our requirement

a) Based on No of Communities required
b) Based on minimum Cut as a stopping Condition

Here, I have implemented first Algorithm, in which I ran the spectral_decompostion on the basis of largest community I have till that Point

Algorithm:

Spectral_Decomposition(edge_list, n ->no of clusters):

communities= []. #list of nodes of a particular community

While(len(communities)<n):

t_Communities = spectral_decomposition_one_itr(Edge_list)

Communites += t_communities)

Largest_community = Communities.pop(largest_community)

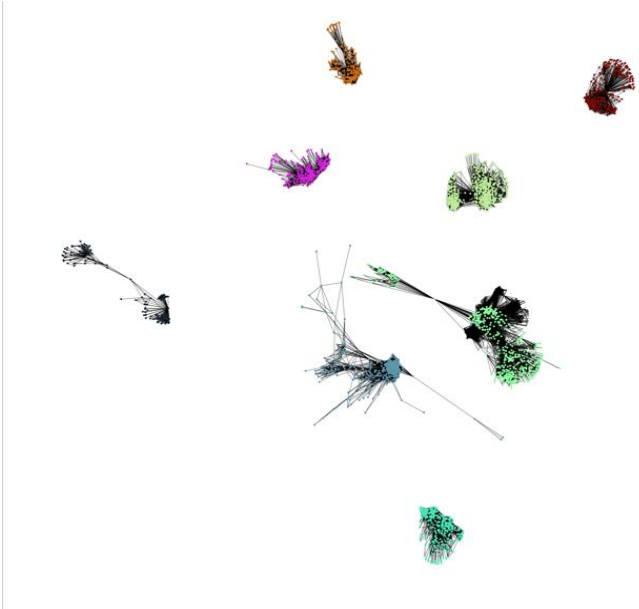
Prepare edge_list for largest_community

Repeat the process until condition is true

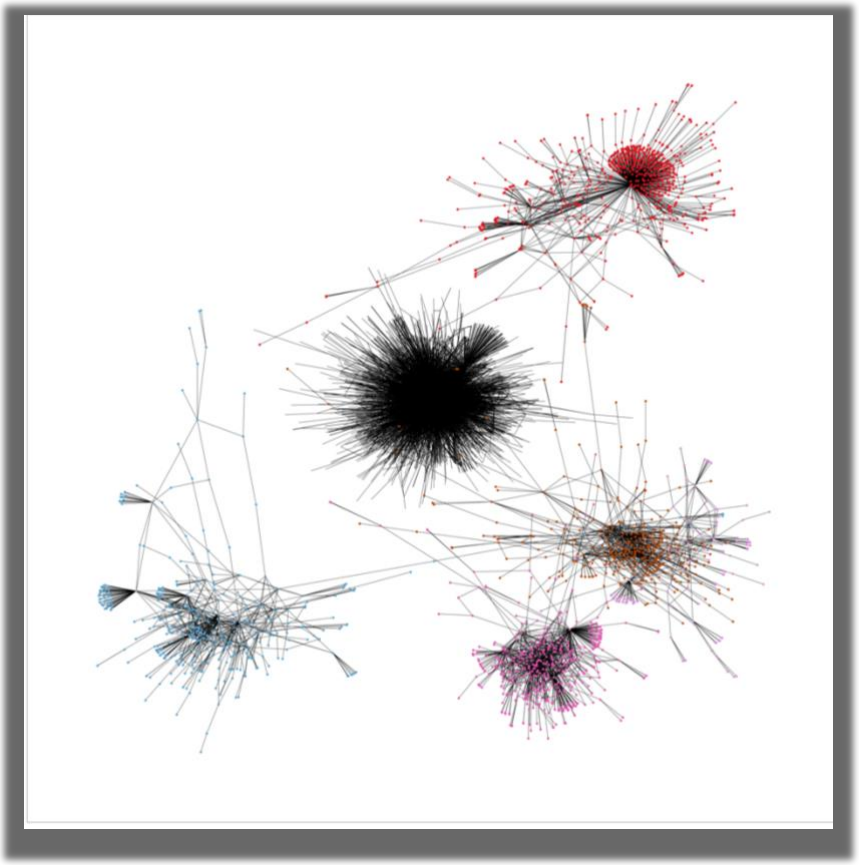
Here Stopping Criteria is to be required no of clusters to be fulfilled

Problem 3:

i) After Running Spectral Decomposition on Facebook Dataset and no of clusters =8 , I have got:



ii) After Running Spectral Decomposition on Bitcoin Dataset and no of clusters =5 , I have got:

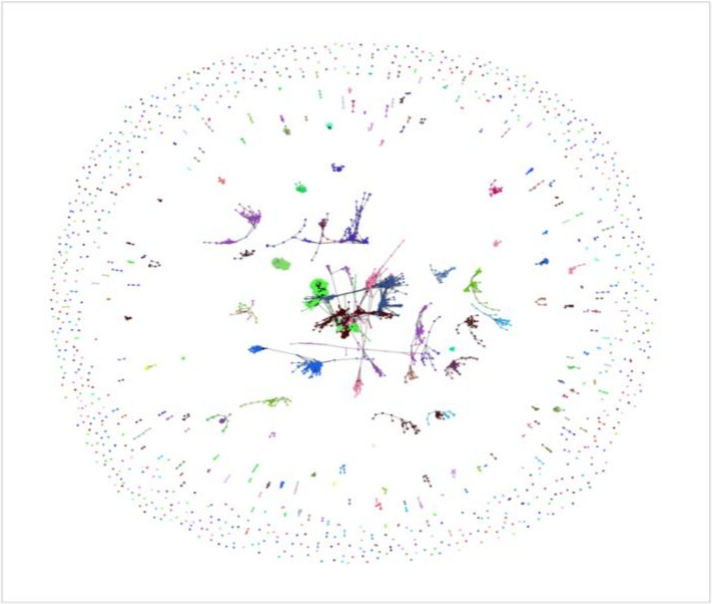


Problem 4:

For Louvain Algorithm , the visualisation of the dataset for 93 clusters that formed during one Iteration is as follows:

Here, we have reduced cluster from around 4000 ->647 -> 174 -> 104 -> 98 ->93->93

After one Iteration, the result for Facebook Dataset looks like this:



Problem 5:

In the Louvain algorithm, the best decomposition of nodes into communities is performed on the basis of the maximum modularity change of nodes with respect to various communities, whereas in spectral clustering, it is performed on the basis of the balanced minimum cut, which we have collectively referred to as a community component if all nodes of communities have a strong hold on one another.

We will divide them into small communities for Best Decomposition, and then combine them to form larger clusters based on the maximum modularity value for the specified communities in the Louvain Algorithm.

Problem 6:

Time Complexity of Spectral Clustering is :

T(n) = O(n^3) // it is due to its calculation of eigen_vectors for large datasets

It is computationally expensive

The Time Complexity of the Louvain Algorithm is :

T(n) =. O(N log N) + O(n^2) // as we are using the divide and conquer approach and O(n^2) is for modularity matrix preparation that can also be counted as data pre-processing time

Problem 7:

Spectral Clustering Gave the best rise to communities but it is expensive for the large dataset as,it works on the basis of best partition, it can find, whereas the Louvain algorithm needs a higher no of iterations to update the modularity of each community.