FUNDAMENTALS OF DATABASE MANAGEMENT

SOCIAL NETWORK ANALYSIS

SUBMITTED TO: PROF. ASHOK HARNAL

SUBMITTED BY:

ABHIRAM MOLUGU (025002)

MEGHALI DAS (025021)

CHIRAYU JAIN (025049)

DEEZER EUROPE

INTRODUCTION

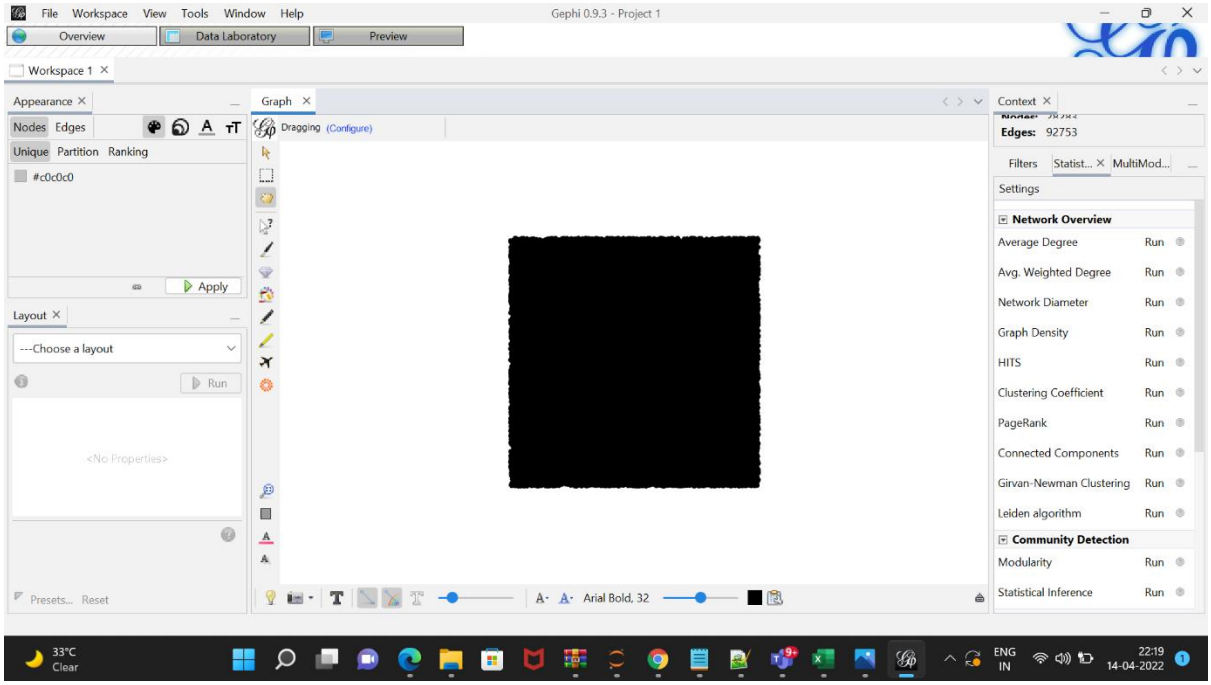
A social network of Deezer users which was collected from the public API in March 2020. Nodes are Deezer users from European countries and edges are mutual follower relationships between them. The vertex features are extracted based on the artists liked by the users. The task related to the graph is binary node classification - one has to predict the gender of users. This target feature was derived from the name field for each user.

NODES: 28,281

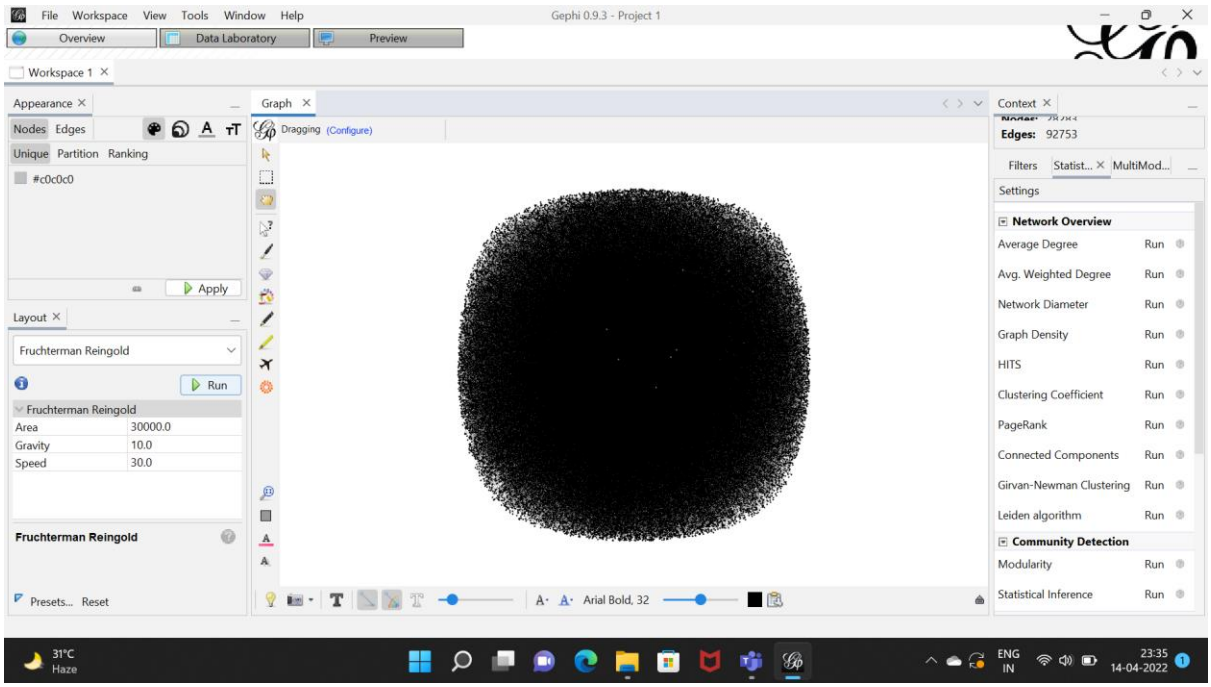
EDGES: 92,752

The procedure of social media analysis is as follows:

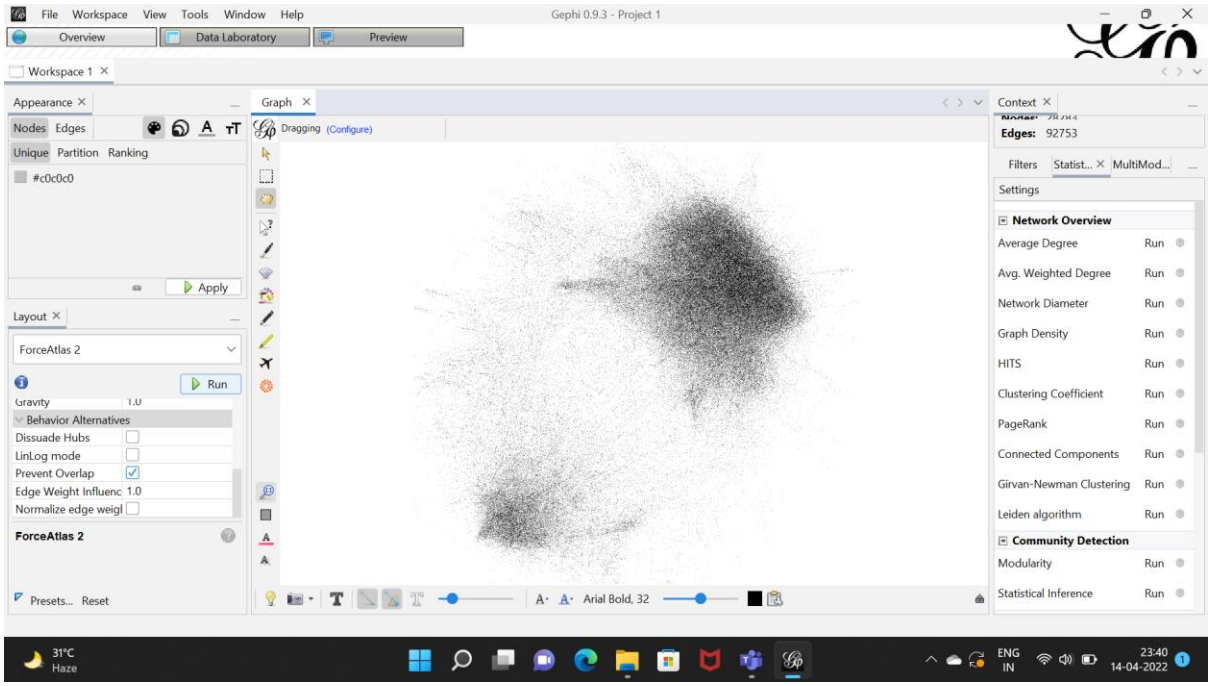
Deezer Europe Target file and Deezer Europe edges files were imported as follows:



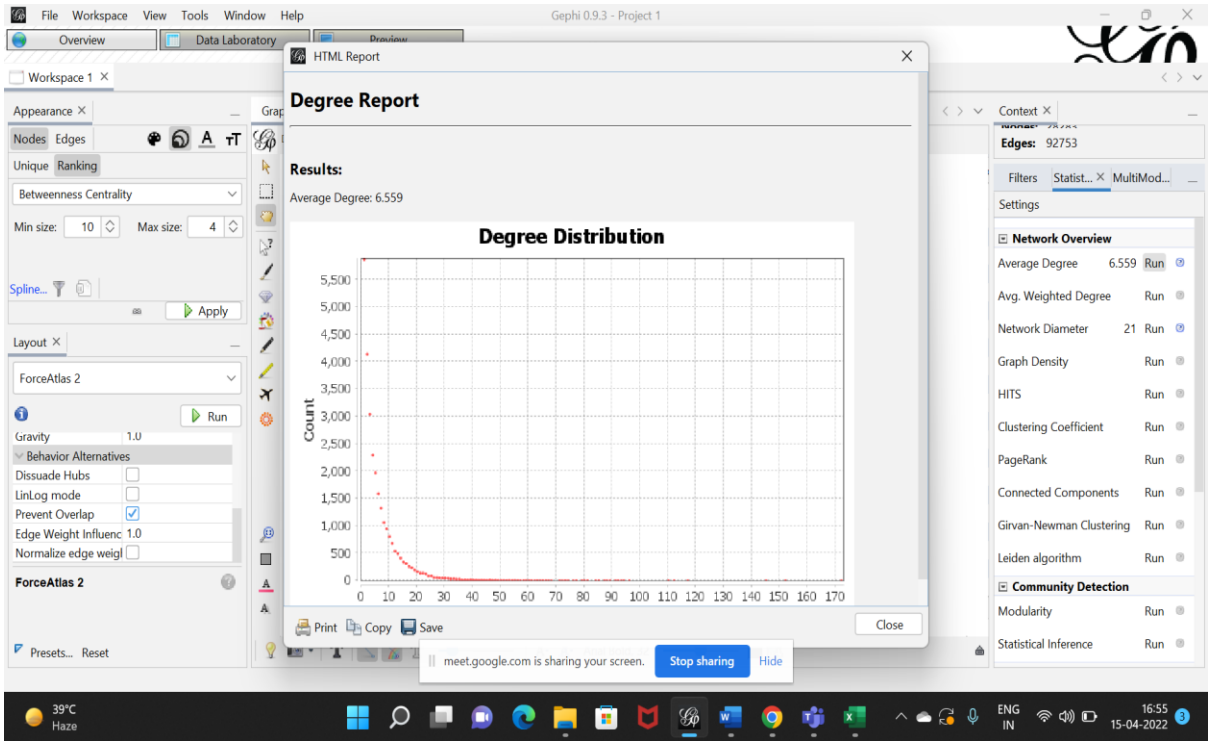
Fruchterman Reingold: For many large social networks, the Fruchterman-Reingold Layout works well, though it may require some tweaking. It's an example of a force-directed algorithm, which uses physical springs as edges to pull connected vertices toward each other and a competing repulsive force to push all vertices away from each other, whether connected or not. Area 30,000, Gravity 10, and Speed: 30



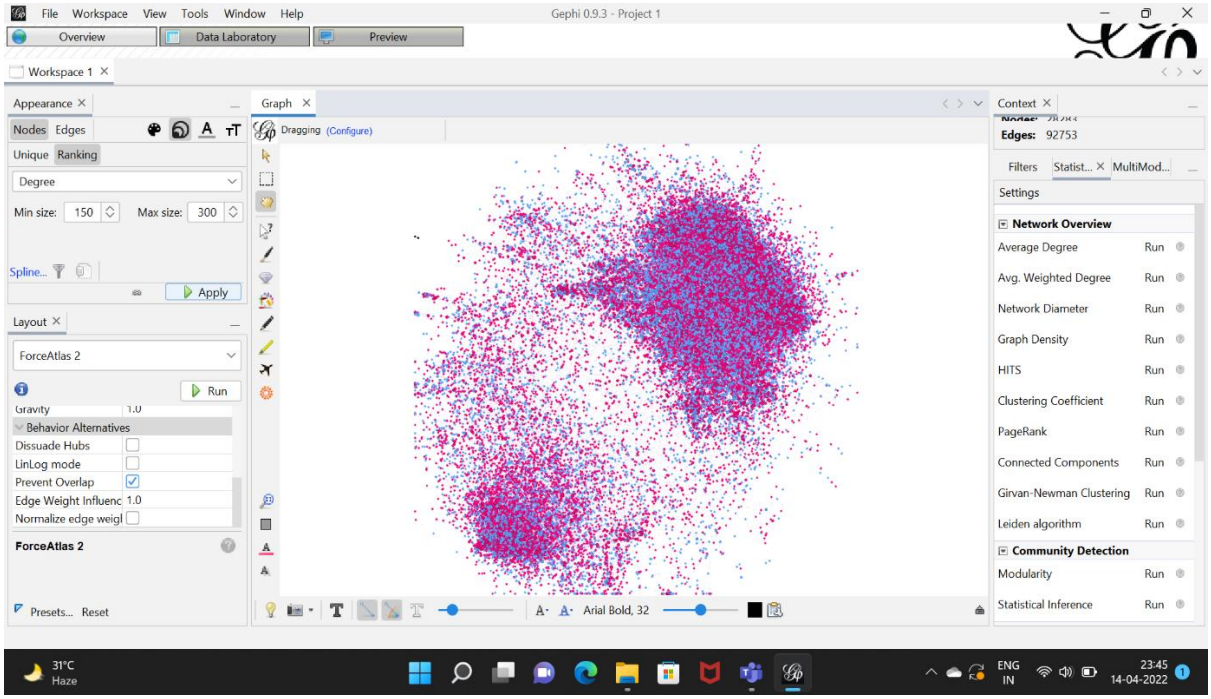
Force Atlas 2: ForceAtlas2 is a rapid force-directed graph layout algorithm. It's used to spatialize a weighted undirected graph in 2D (the edge weight defines the strength of the connection). Scaling helps to control the graph's scale of expansion. We must avoid overlapping. We must also ensure that nodes do not overlap.



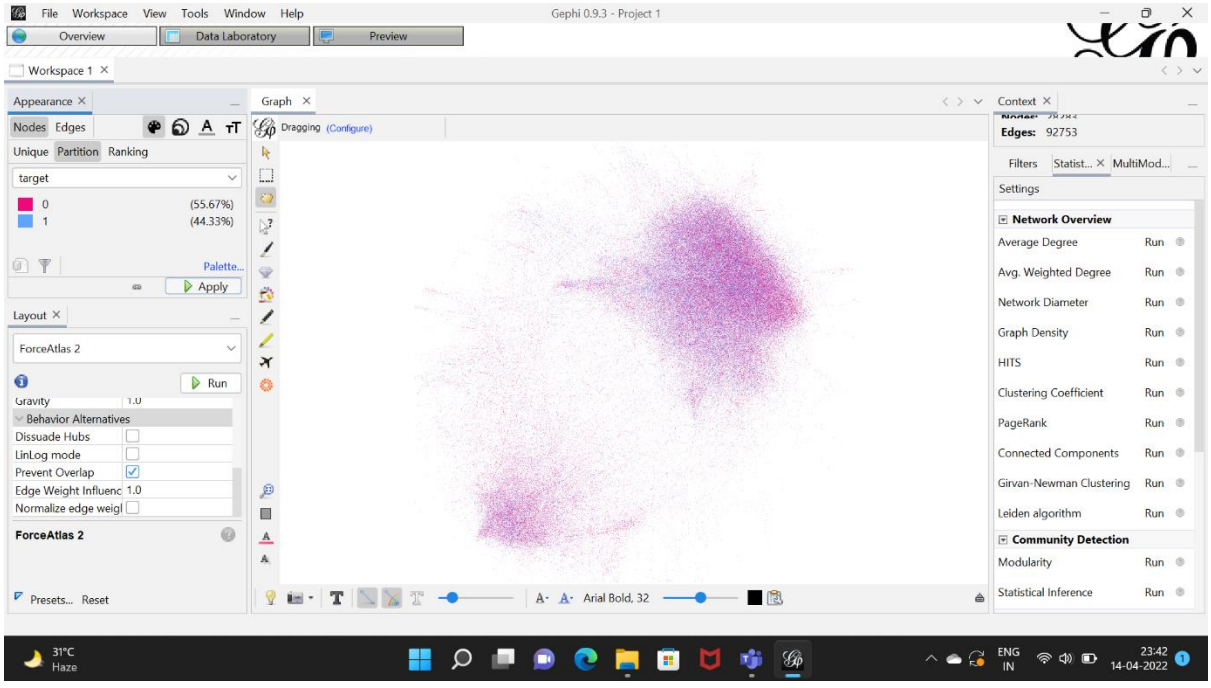
Modularity + Degree Partitions (without filter): The degree of a vertex in graph theory is the number of edges linking it. The easiest centrality metric to calculate is degree centrality. We must remember that the degree of a node is simply the number of social connections (edges) it possesses. A node's degree determines its degree centrality. The degree centrality of a node with ten social connections is 10. The degree centrality of a node with one edge is 1. Our node's average Degree is **6.559**.



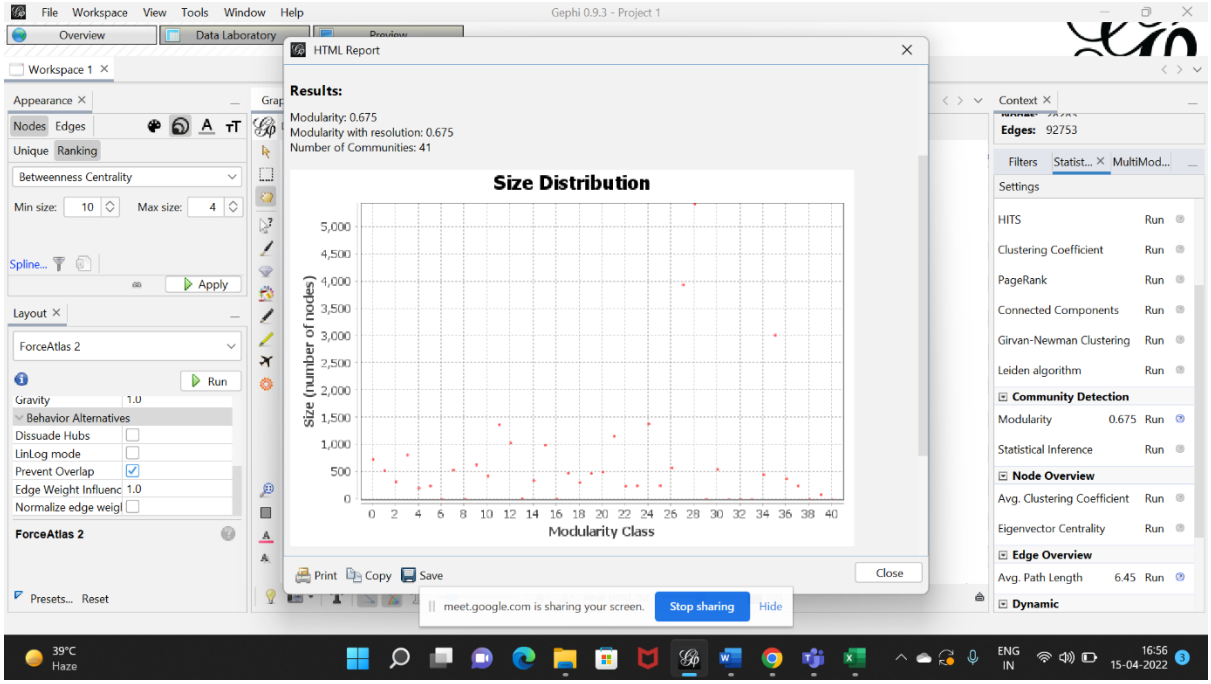
NODES> RANKING> DEGREE

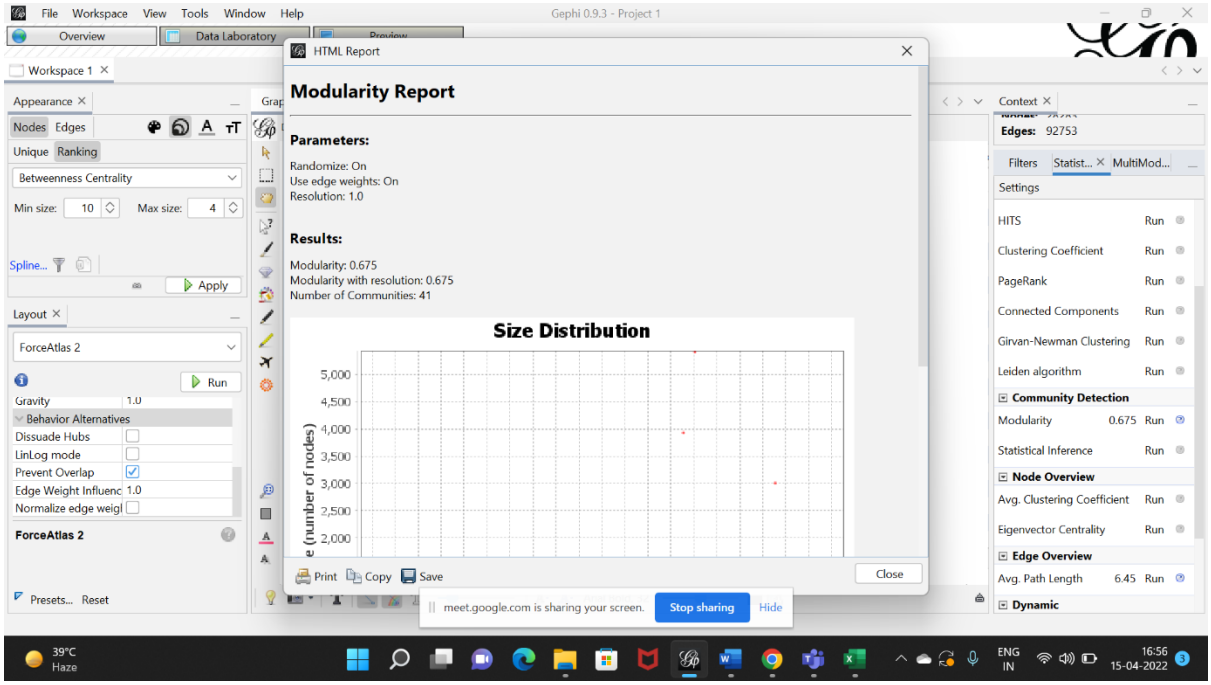


PARTITION

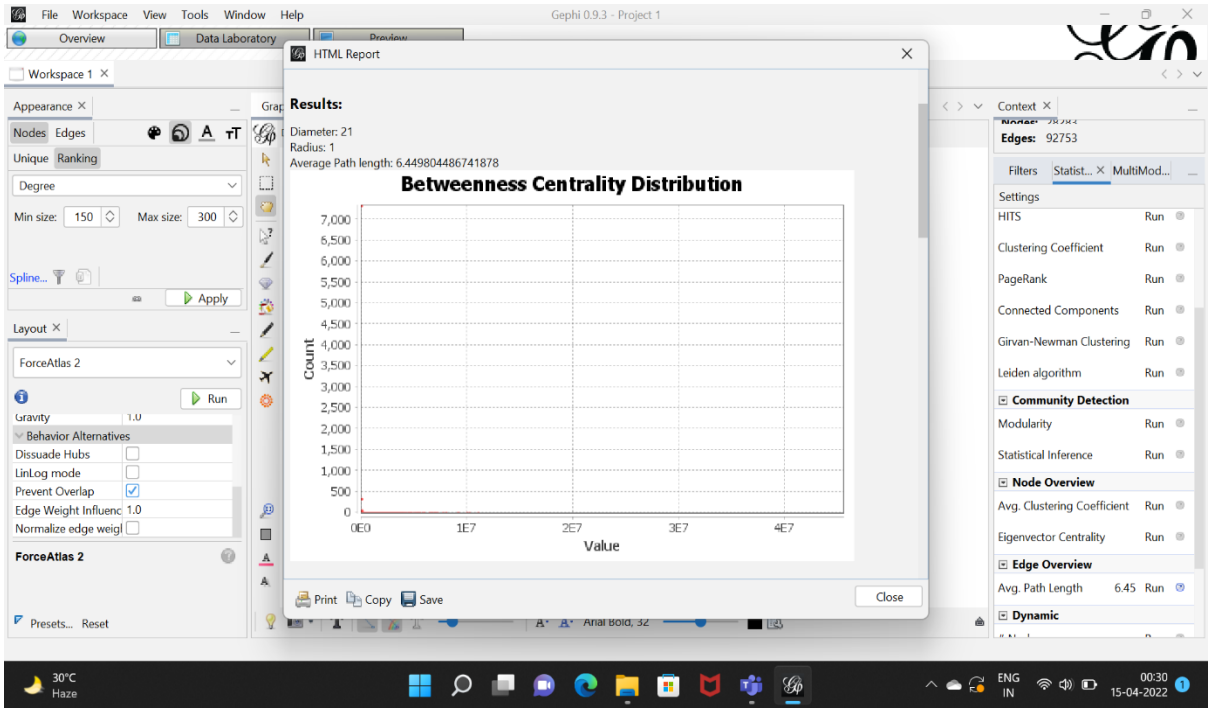


Modularity is a system feature that assesses the degree to which a system's densely connected compartments can be dissociated into discrete communities or clusters that interact more with one another than with other communities. A shock to one compartment in a highly interconnected system with low degrees of modularity may cascade to another compartment, increasing the danger of a system-wide collapse. Modularity = 0.675

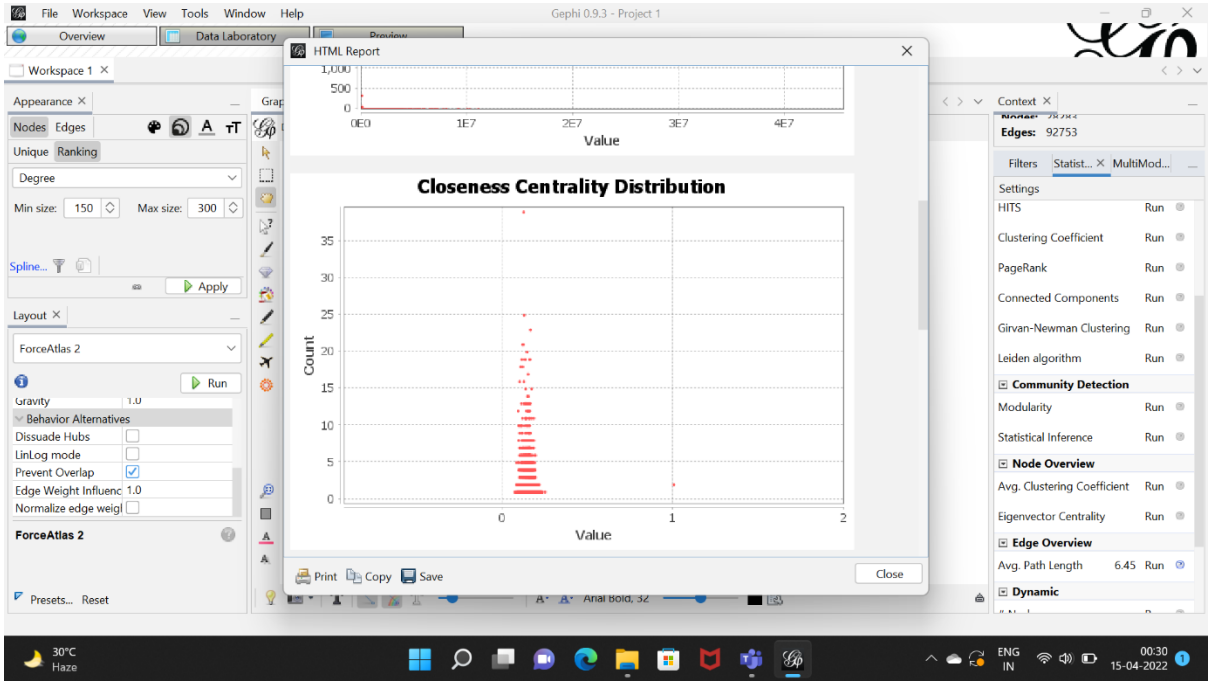




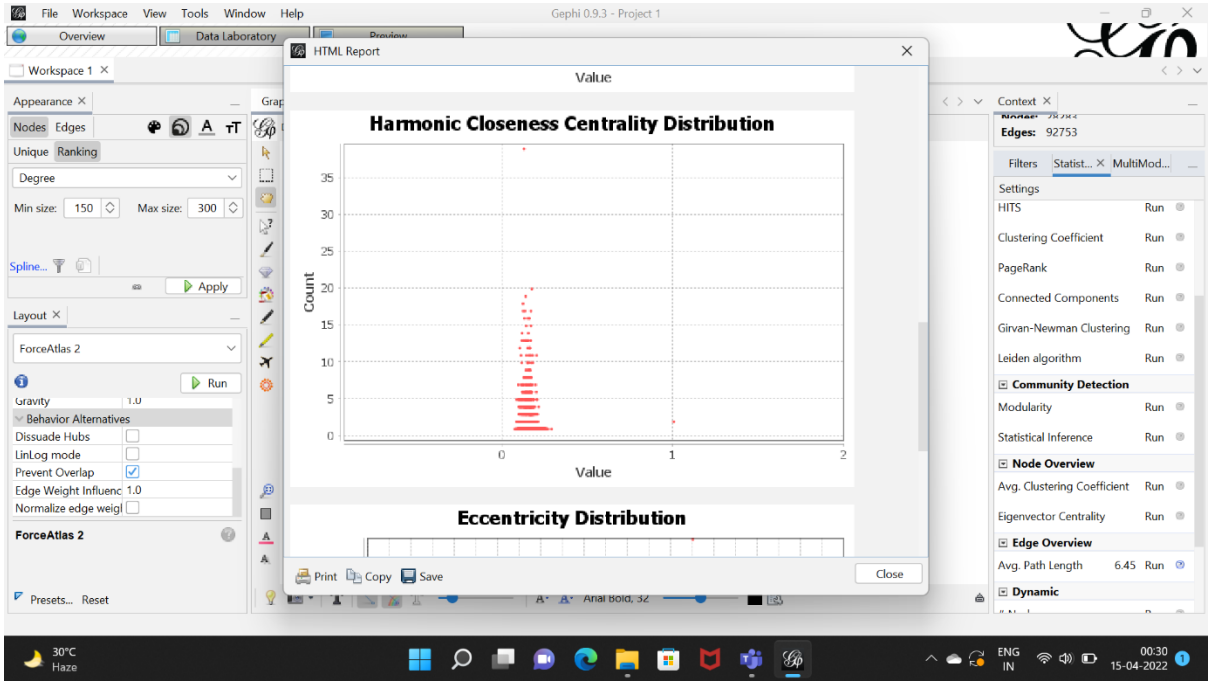
Betweenness Centrality Distribution



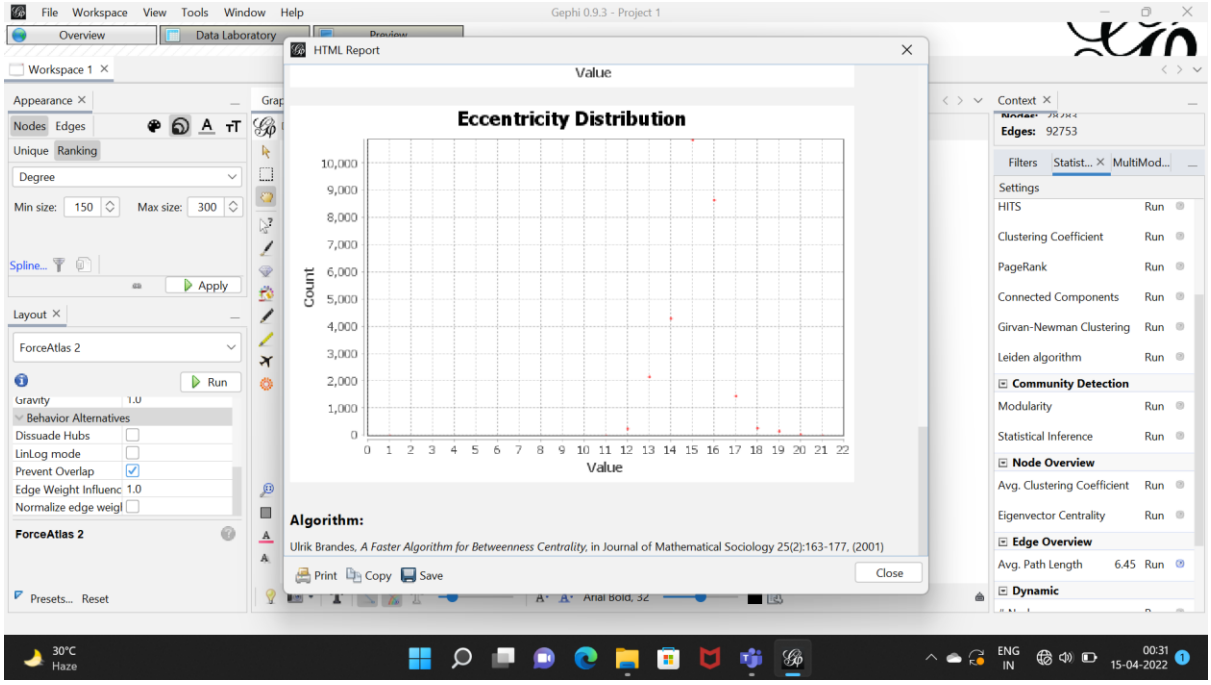
Closeness Centrality Distribution



HARMONIC CLOSENESS CENTRALITY DISTRIBUTION

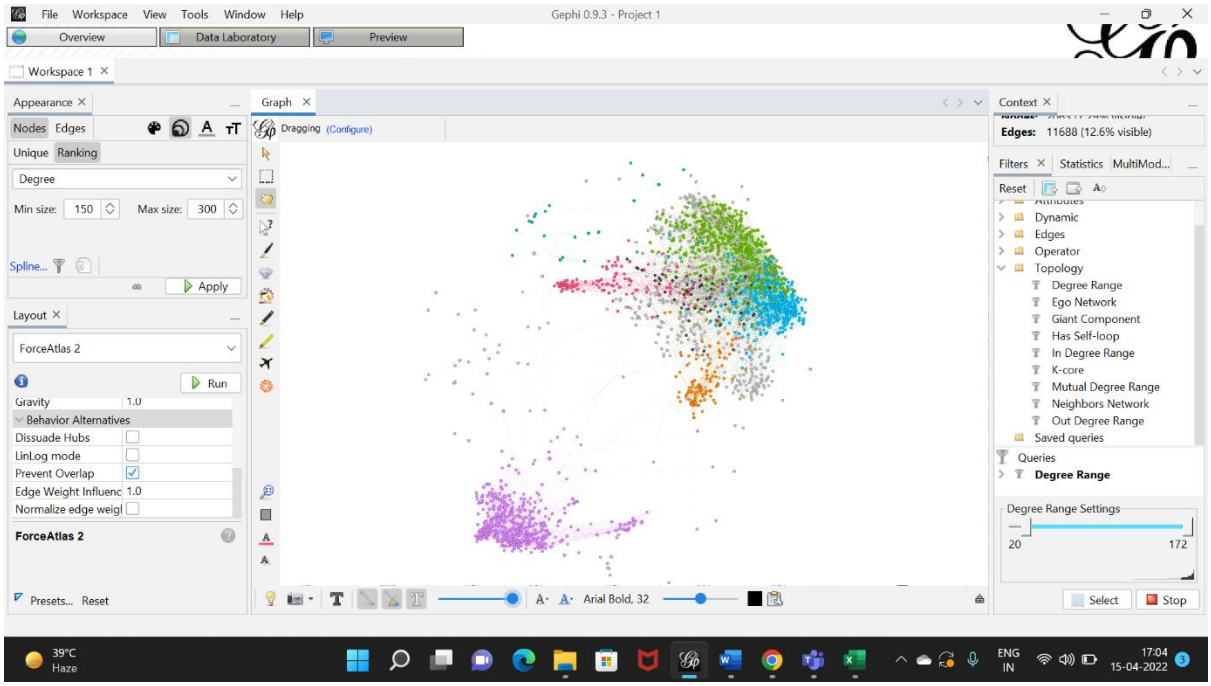


ECCENTRICITY DISTRIBUTION

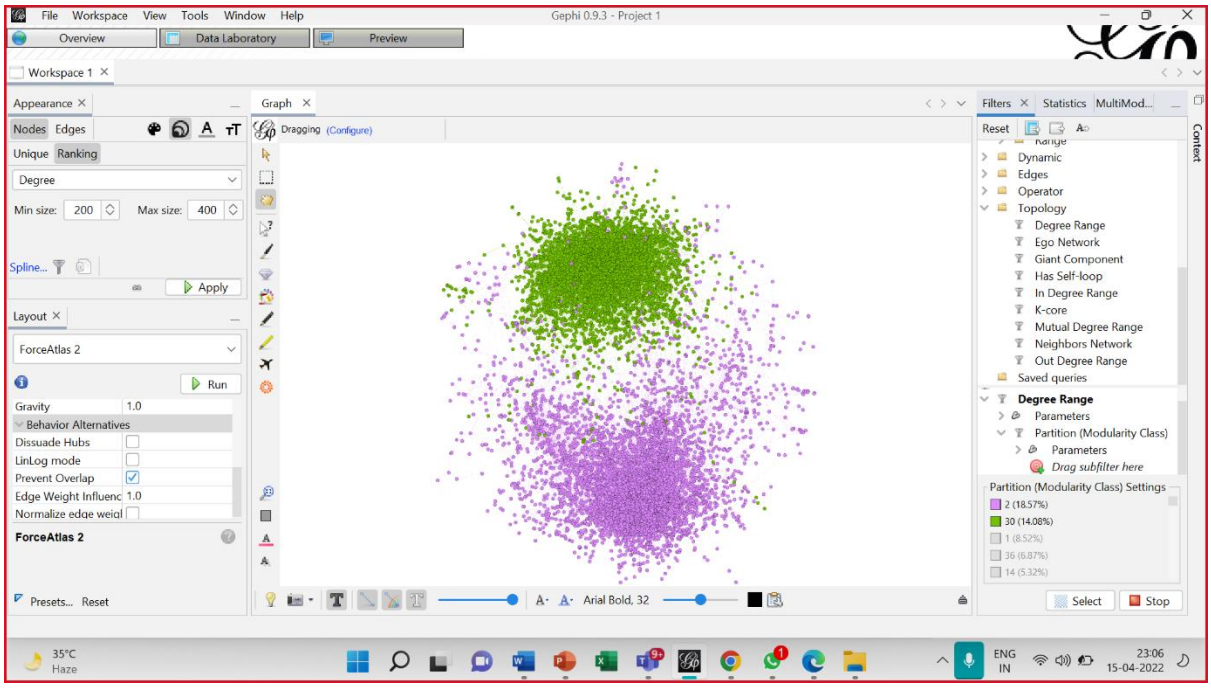


FILTERS:

Topology: Degree Range

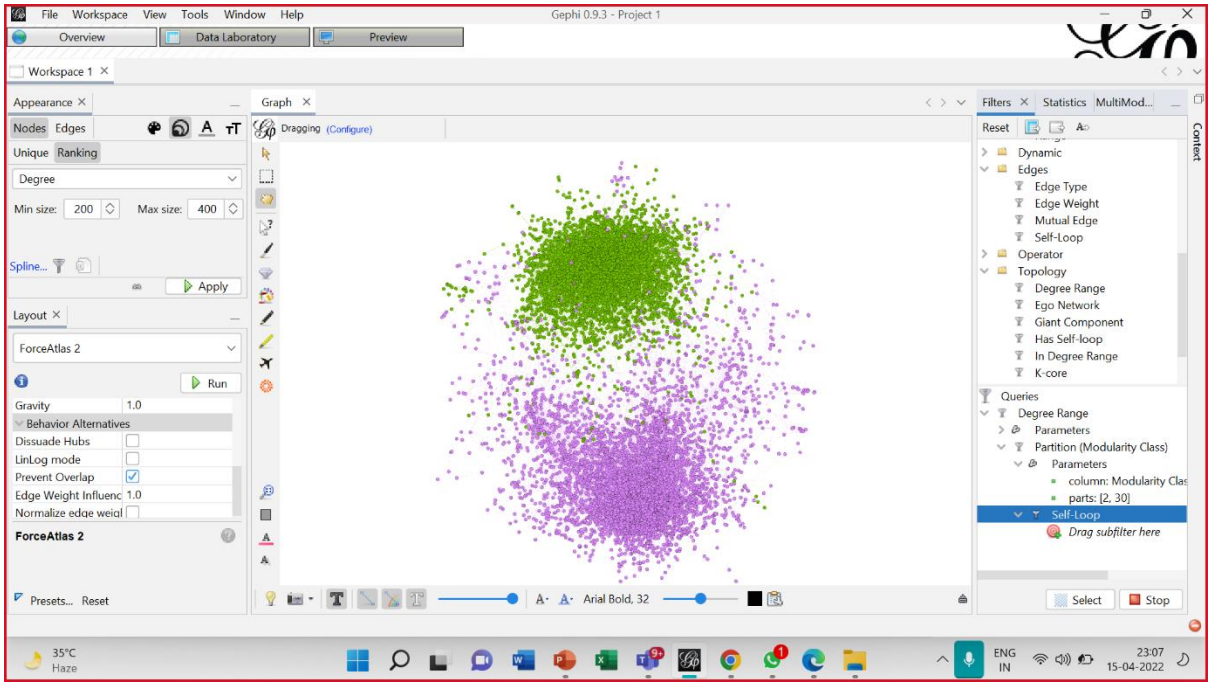


Attributes: Modularity Class



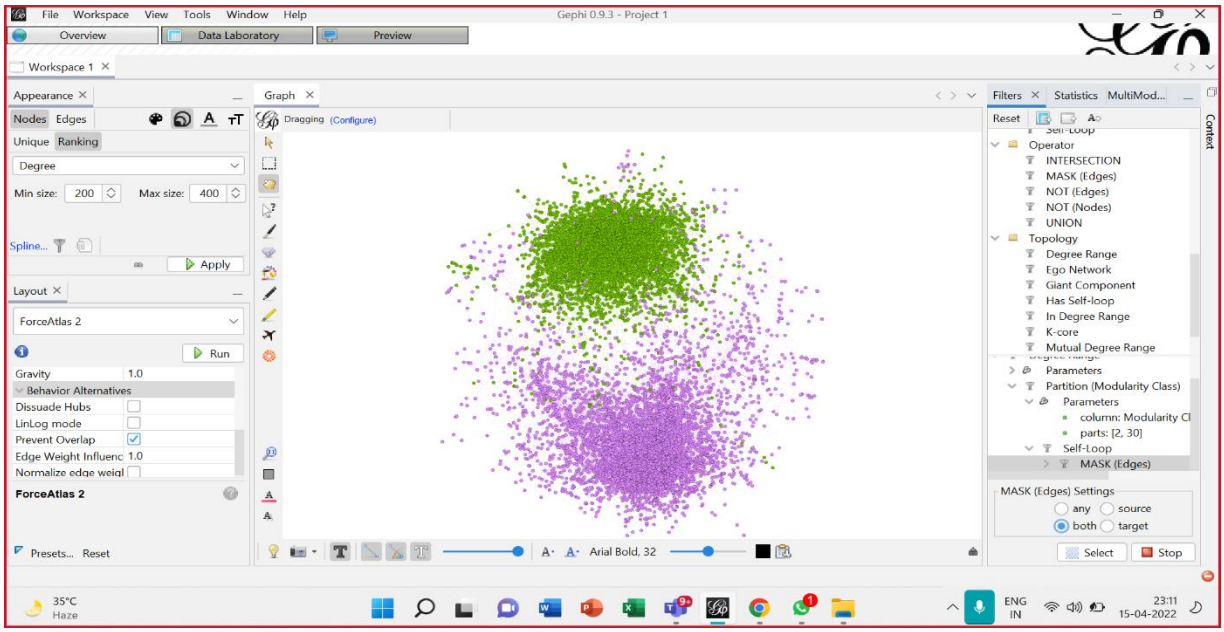
Edges: Self Loop

This filter removes all the self-loops that exist in the network. After proper comparison, we can see the difference between networks.



Operator: Edge Mask

Only edges from the node filter sub-query are returned by this filter, which returns the entire graph. For example, we can maintain edges with a source node degree greater than 5. Both of the networks developed have a little difference.



- The entities are Deezer users from European countries
- The total number of communities are 41

CONCLUSION

This experiment suggests that the dataset has many items which are interrelated to each other and links between European countries and mutual followers that have strong relationship with each other. There are 41 communities and filtering them suggests that the top 2 communities are bonded strong enough.