FUNDAMENTALS OF DATABASE MANAGEMENT

SOCIAL NETWORK ANALYSIS

SUBMITTED TO: PROF. ASHOK HARNAL

SUBMITTED BY:

ACHYUT SINHA (025004)

AKSHAY VASHISHTHA (025005)

TUSHAR SHARMA (025036)

JAZZ MUSICIANS' DATASET

INTRODUCTION

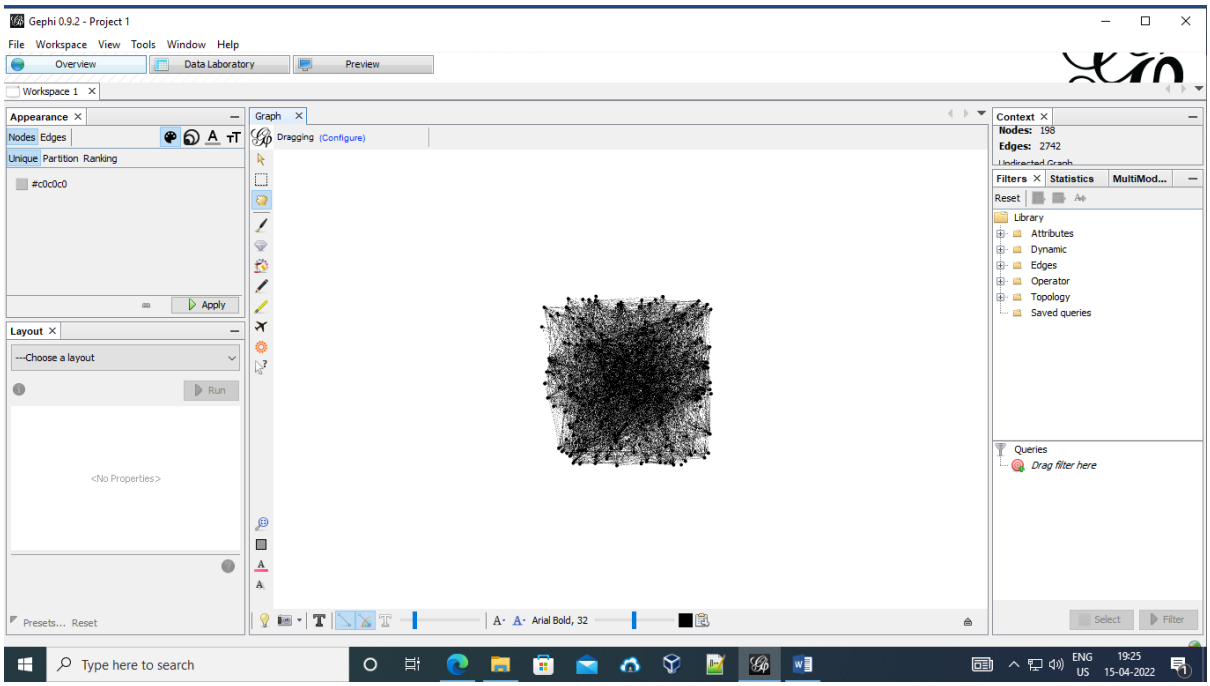
A social network of Jazz Musicians which was collected from the public API. Nodes are Jazz Musicians and edges are mutual follower relationships between them. The vertex features are extracted based on the artists liked by the users. The task related to the graph is binary node classification - one has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

NODES: 198

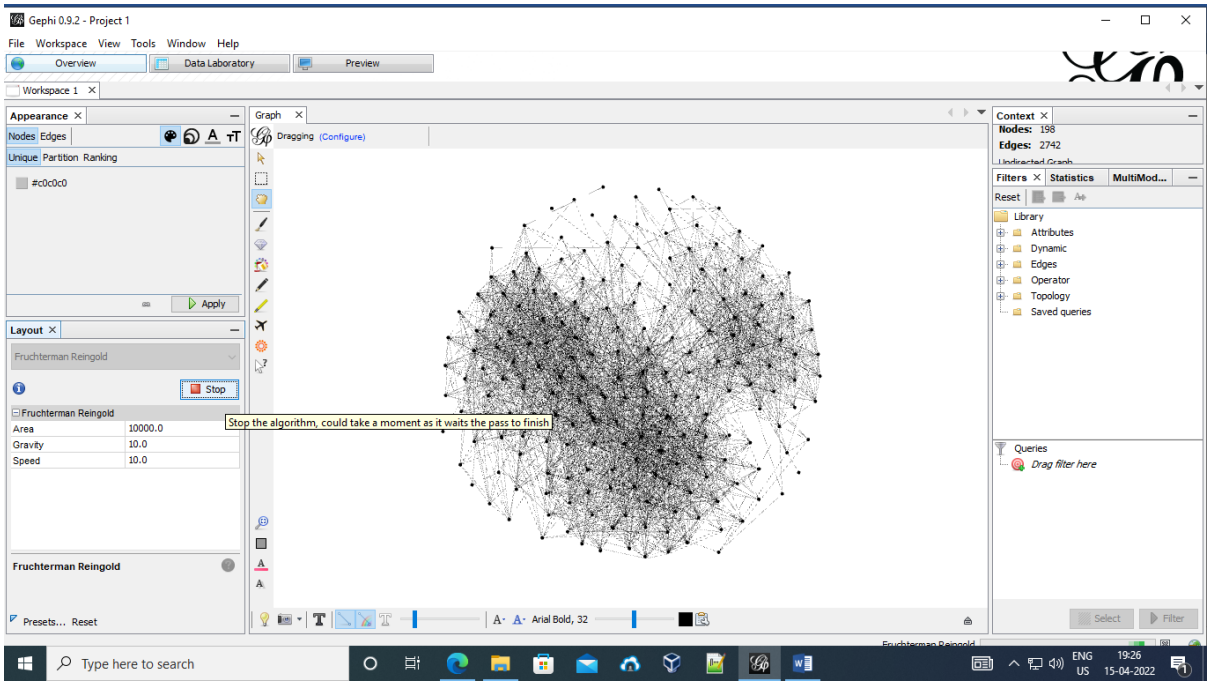
EDGES: 2742

The procedure of social media analysis is as follows:

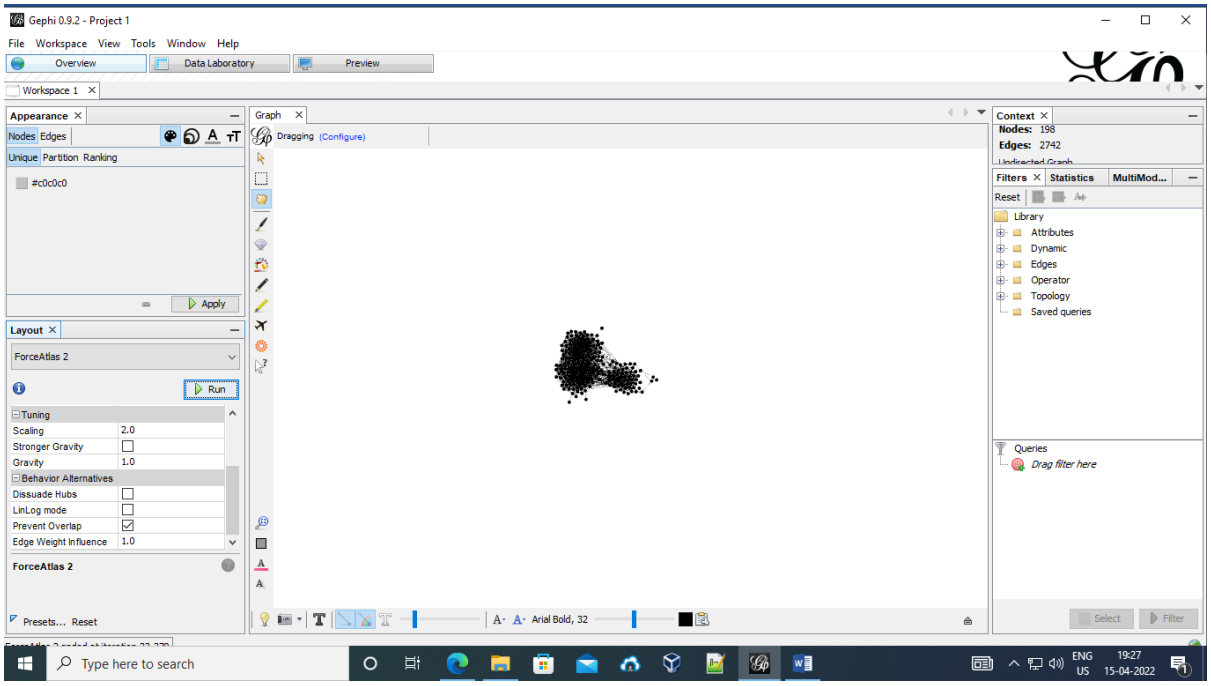
Jazz Musicians Target file and edges files were imported as follows:



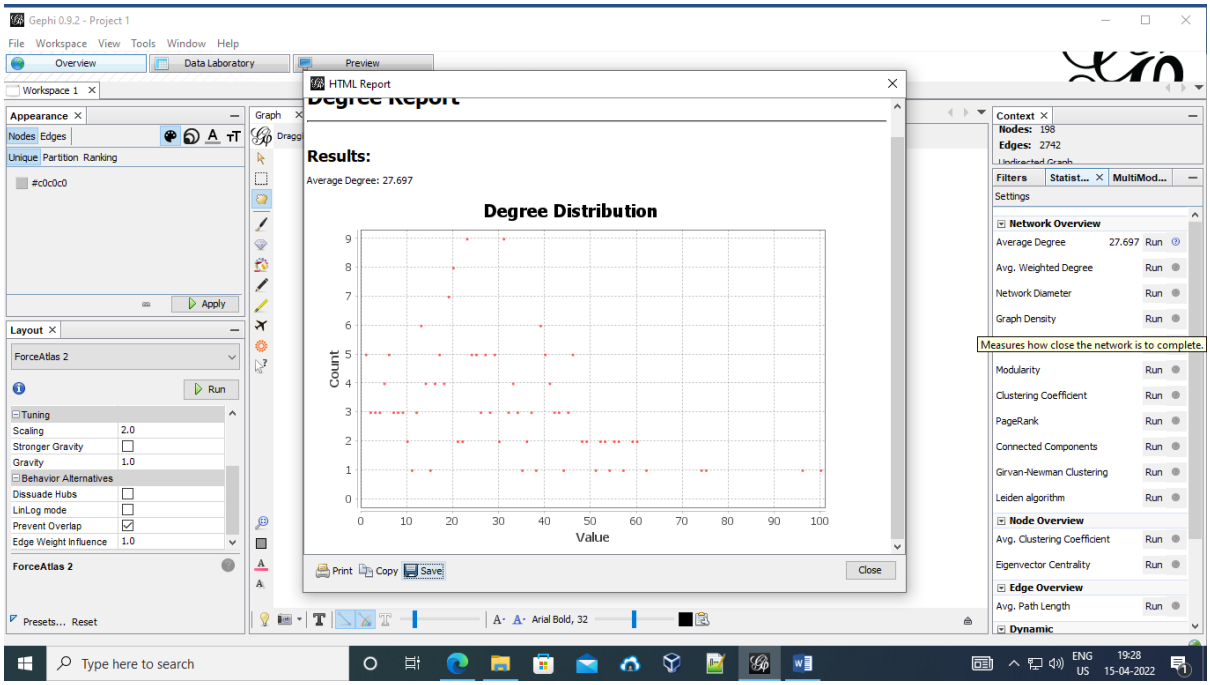
Fruchterman Reingold: For many large social networks, the Fruchterman-Reingold Layout works well, though it may require some tweaking. It's an example of a force-directed algorithm, which uses physical springs as edges to pull connected vertices toward each other and a competing repulsive force to push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



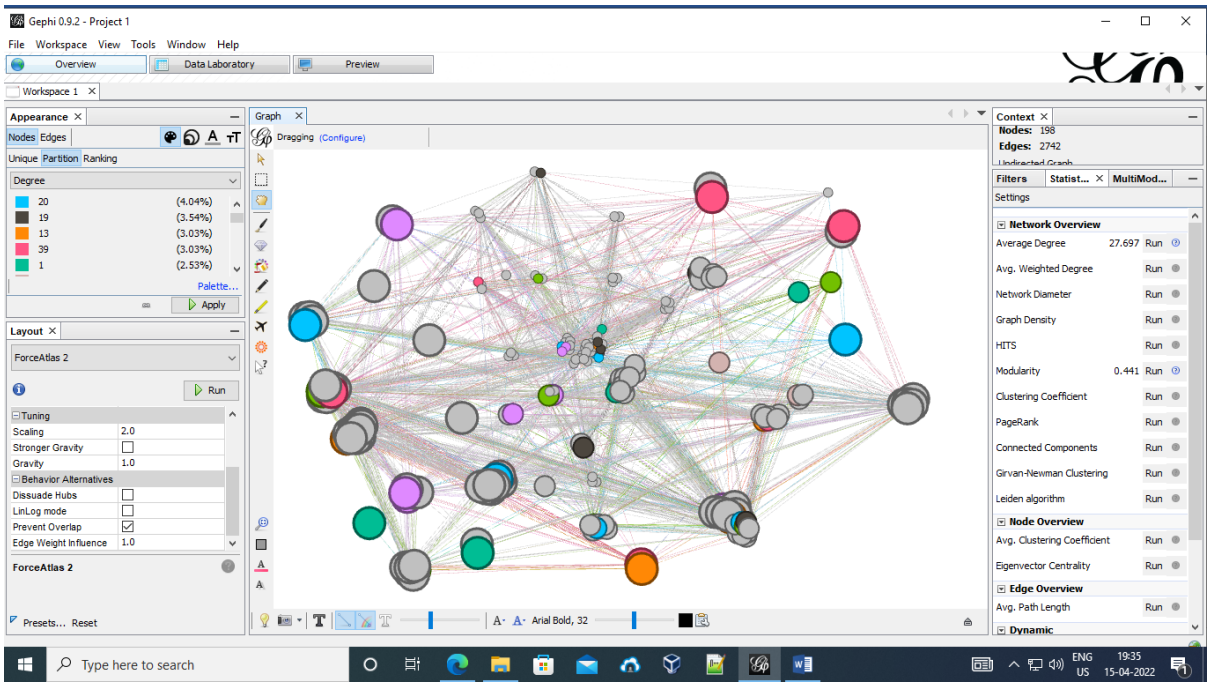
Force Atlas 2: ForceAtlas2 is a rapid force-directed graph layout algorithm. It's used to spatialize a weighted undirected graph in 2D (the edge weight defines the strength of the connection). Scaling helps to control the graph's scale of expansion. We must avoid overlapping. We must also ensure that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

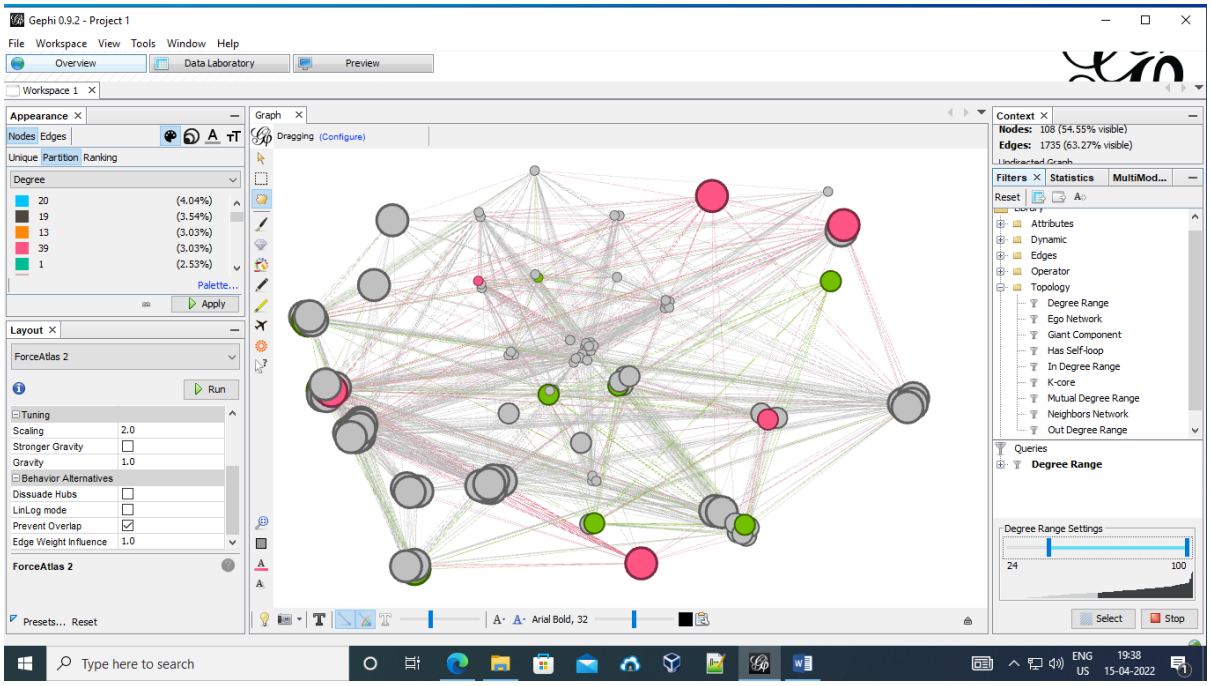


Modularity + Degree Partitions (without filter): The degree of a vertex in graph theory is the number of edges linking it. The easiest centrality metric to calculate is degree centrality. We must remember that the degree of a node is simply the number of social connections (edges) it possesses. A node's 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 **27.697**.

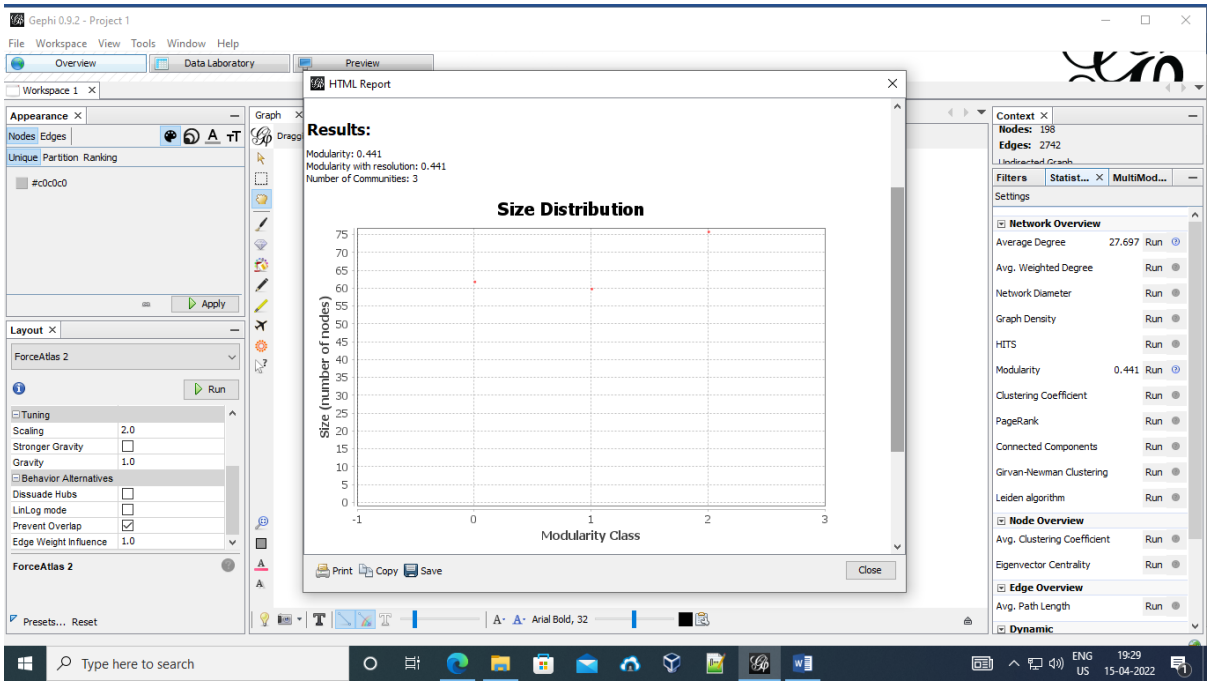

Gephi 0.9.2 - Project 1
File Workspace View Tools Window Help
Overview
Data Laboratory
Preview
Workspace 1
HTML Report
Results:
Average Degree: 27.697
Degree Distribution
Count
Value
Print
Copy
Save
Close
Appearance
Nodes Edges
Unique Partition Ranking
Apply
Layout
ForceAtlas 2
Run
Tuning
Scaling 2.0
Stronger Gravity
Gravity 1.0
Behavior Alternatives
Dissuade Hubs
LinLog mode
Prevent Overlap
Edge Weight Influence 1.0
Presets... Reset
Graph
Context
Nodes: 198
Edges: 2742
Filters
Statist...
MultiMod...
Settings
Network Overview
Average Degree 27.697 Run
Avg. Weighted Degree Run
Network Diameter Run
Graph Density Run
Measures how close the network is to complete.
Modularity Run
Clustering Coefficient Run
PageRank Run
Connected Components Run
Girvan-Newman Clustering Run
Leiden algorithm Run
Node Overview
Avg. Clustering Coefficient Run
Eigenvector Centrality Run
Edge Overview
Avg. Path Length Run
Dynamic
Type here to search
ENG US
19:28
15-04-2022

PARTITION


Gephi 0.9.2 - Project 1
File Workspace View Tools Window Help
Overview
Data Laboratory
Preview
Workspace 1
Appearance
Nodes Edges
Unique Partition Ranking
Degree
20 (4.04%)
19 (3.54%)
13 (3.03%)
39 (3.03%)
1 (2.53%)
Palette...
Apply
Layout
ForceAtlas 2
Run
Tuning
Scaling 2.0
Stronger Gravity
Gravity 1.0
Behavior Alternatives
Dissuade Hubs
LinLog mode
Prevent Overlap
Edge Weight Influence 1.0
Presets... Reset
Graph
Dragging (Configure)
Context
Nodes: 198
Edges: 2742
Filters
Statist...
MultiMod...
Settings
Network Overview
Average Degree 27.697 Run
Avg. Weighted Degree Run
Network Diameter Run
Graph Density Run
HITS Run
Modularity 0.441 Run
Clustering Coefficient Run
PageRank Run
Connected Components Run
Girvan-Newman Clustering Run
Leiden algorithm Run
Node Overview
Avg. Clustering Coefficient Run
Eigenvector Centrality Run
Edge Overview
Avg. Path Length Run
Dynamic
Type here to search
ENG US
19:35
15-04-2022

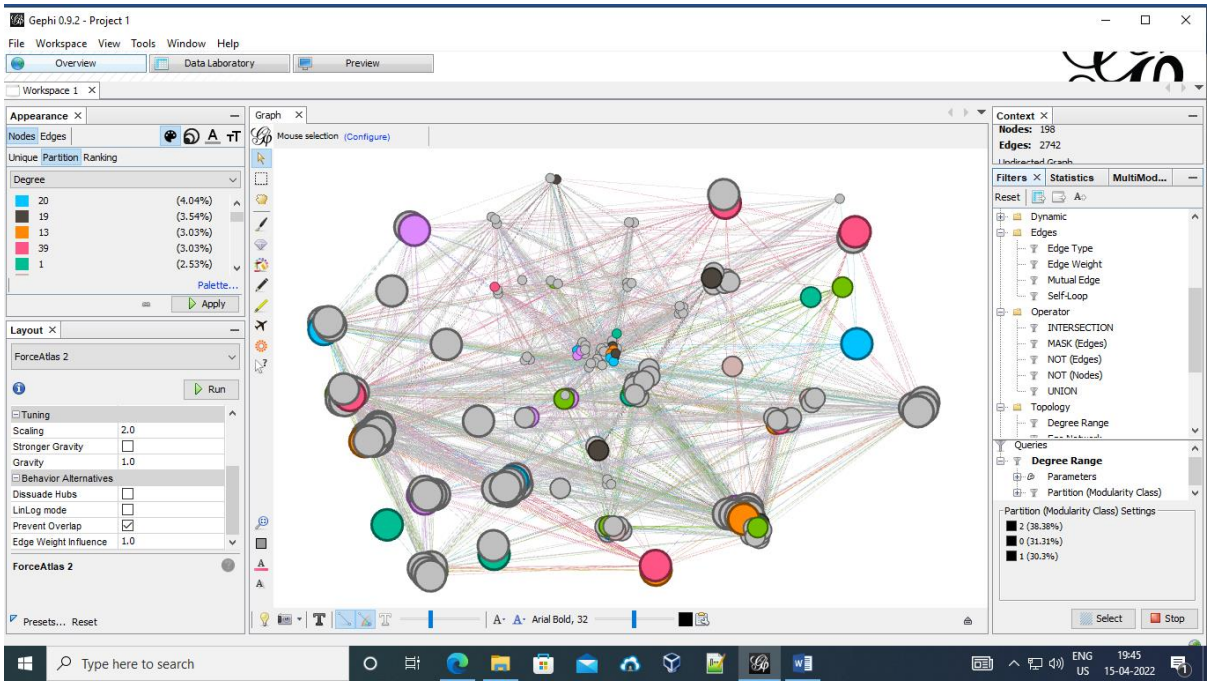


Modularity is a system feature that assesses the degree to which a system's densely connected compartments can be dissociated into discrete communities or clusters that interact more with one another than with other communities. A shock to one compartment in a highly interconnected system with low degrees of modularity may cascade to another compartment, increasing the danger of a system-wide collapse. Modularity = 0.441

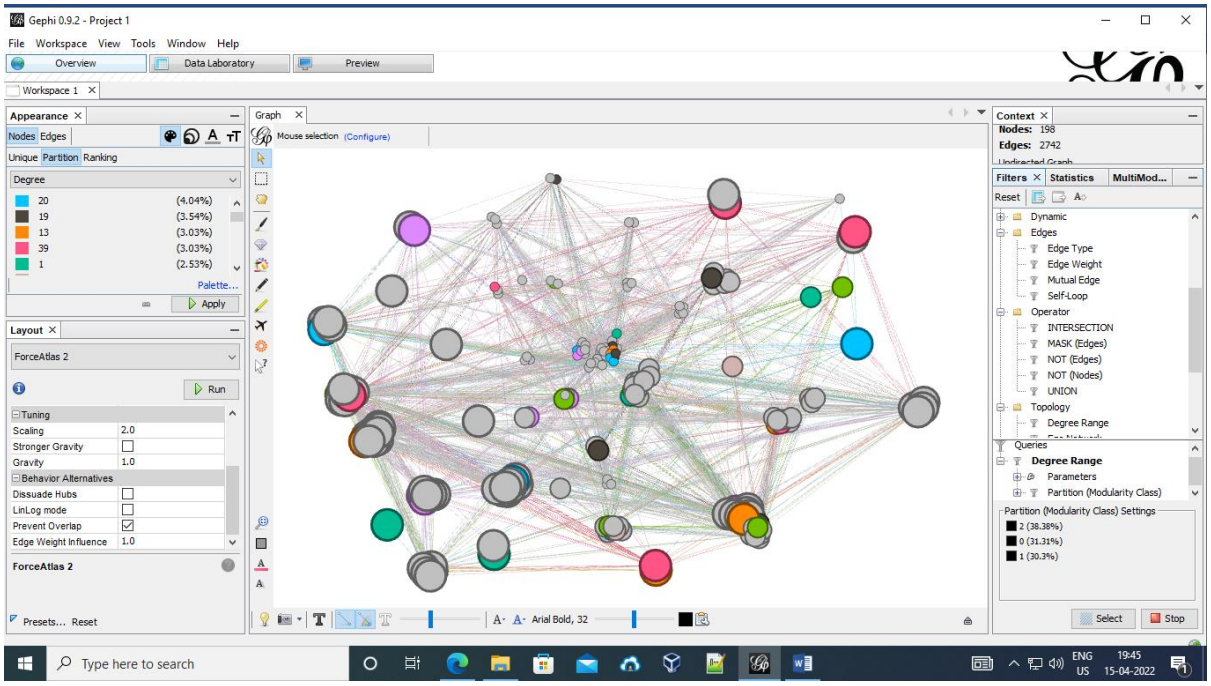


FILTERS:

Topology: Degree Range

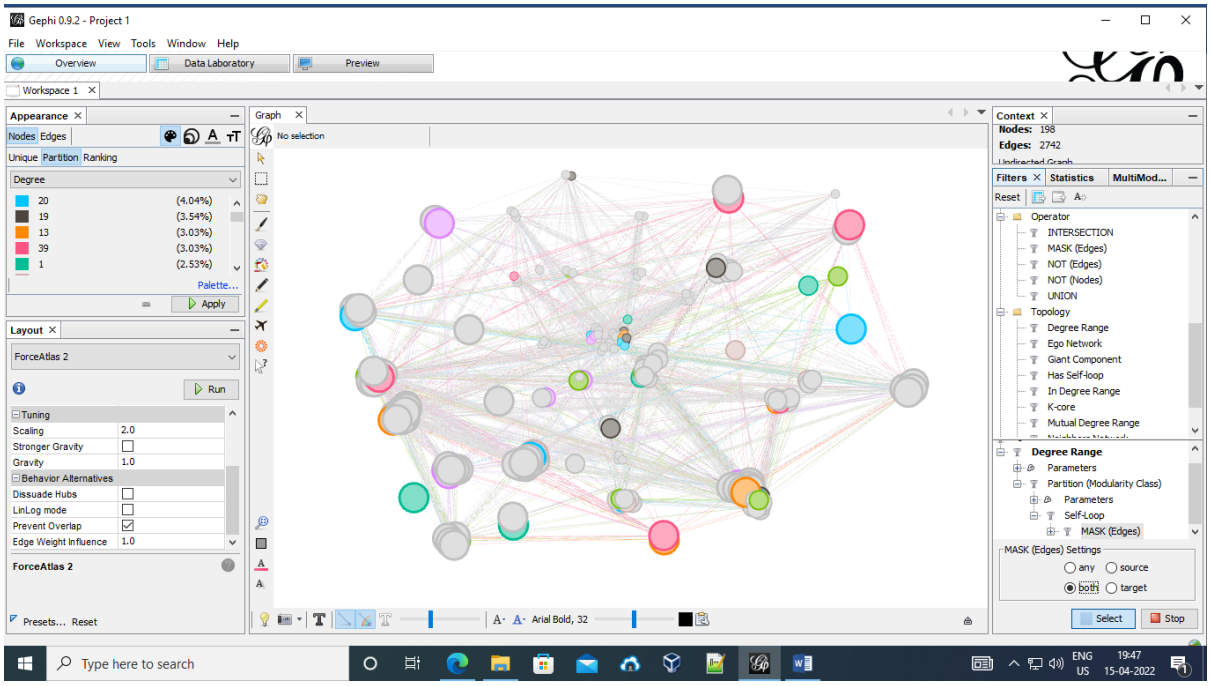


Attributes: Modularity Class



Edges: Self Loop & Operator: Mask Edges

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



CONCLUSION

This experiment suggests that the dataset has many items which are interrelated to each other and links between Jazz Musicians 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.