



Fundamentals of Database Management

Social Network Analysis

Submitted to:

Prof. Ashok Harnal

Submitted by:

Group 6- BDA 02

Abhishek Das - 025003

Arunita Banerji - 025009

Bagya Shree C M - 025011

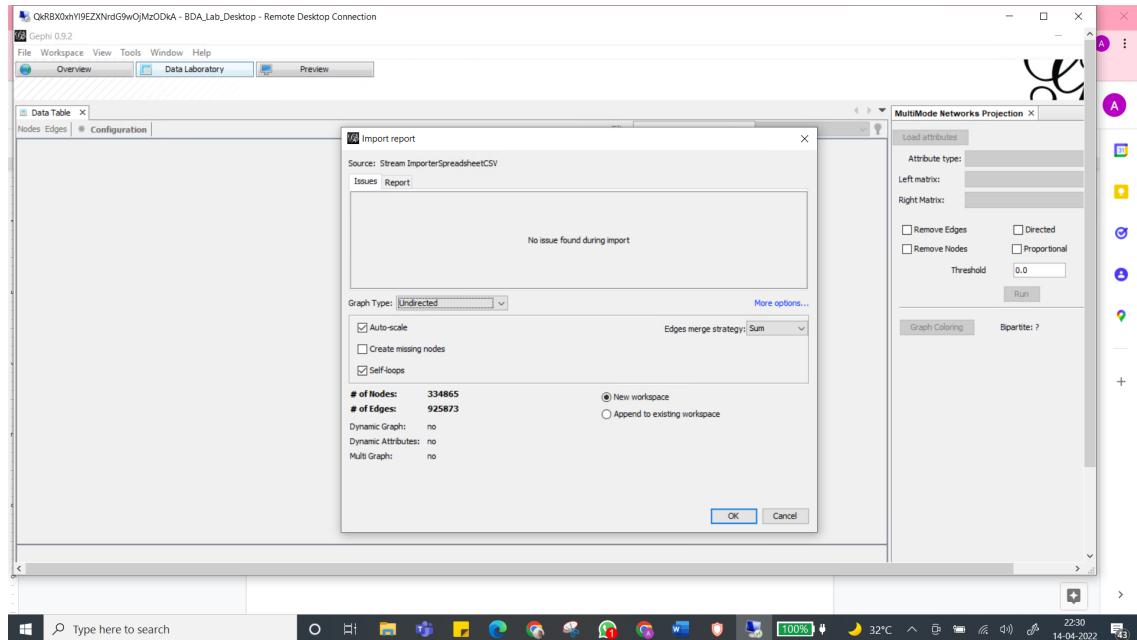
Amazon product co-purchasing network and ground-truth communities

Introduction:

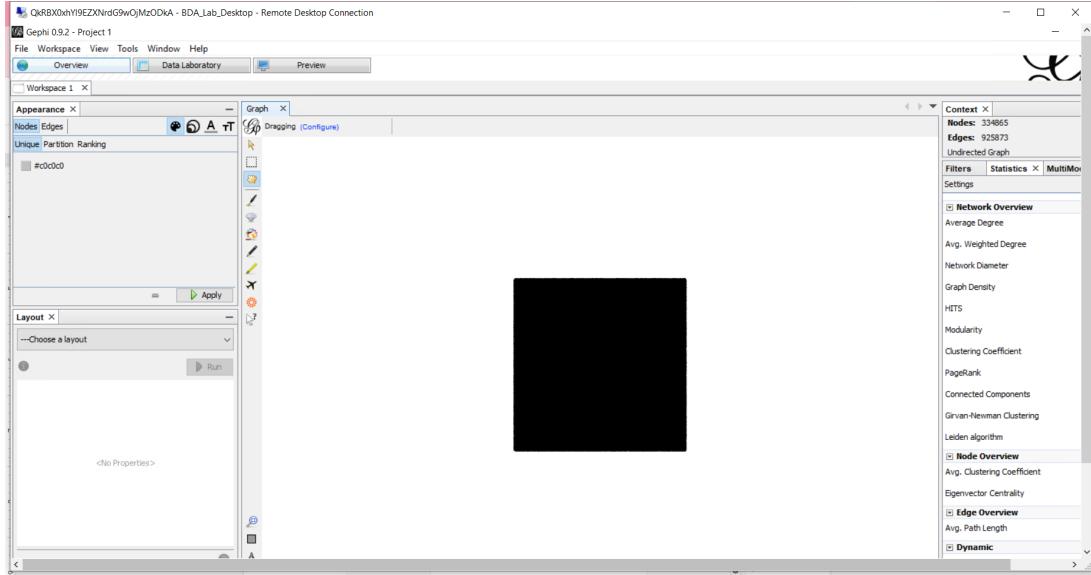
Network was collected by crawling the Amazon website. It is based on Customers Who Bought This Item Also Bought feature of the Amazon website. If a product i is frequently co-purchased with product j, the graph contains an undirected edge from i to j. Each product category provided by Amazon defines each ground-truth community.

Dataset statistics	
Nodes	334863
Edges	925872
Nodes in largest WCC	334863 (1.000)
Edges in largest WCC	925872 (1.000)
Nodes in largest SCC	334863 (1.000)
Edges in largest SCC	925872 (1.000)
Average clustering coefficient	0.3967
Number of triangles	667129
Fraction of closed triangles	0.07925
Diameter (longest shortest path)	44
90-percentile effective diameter	15

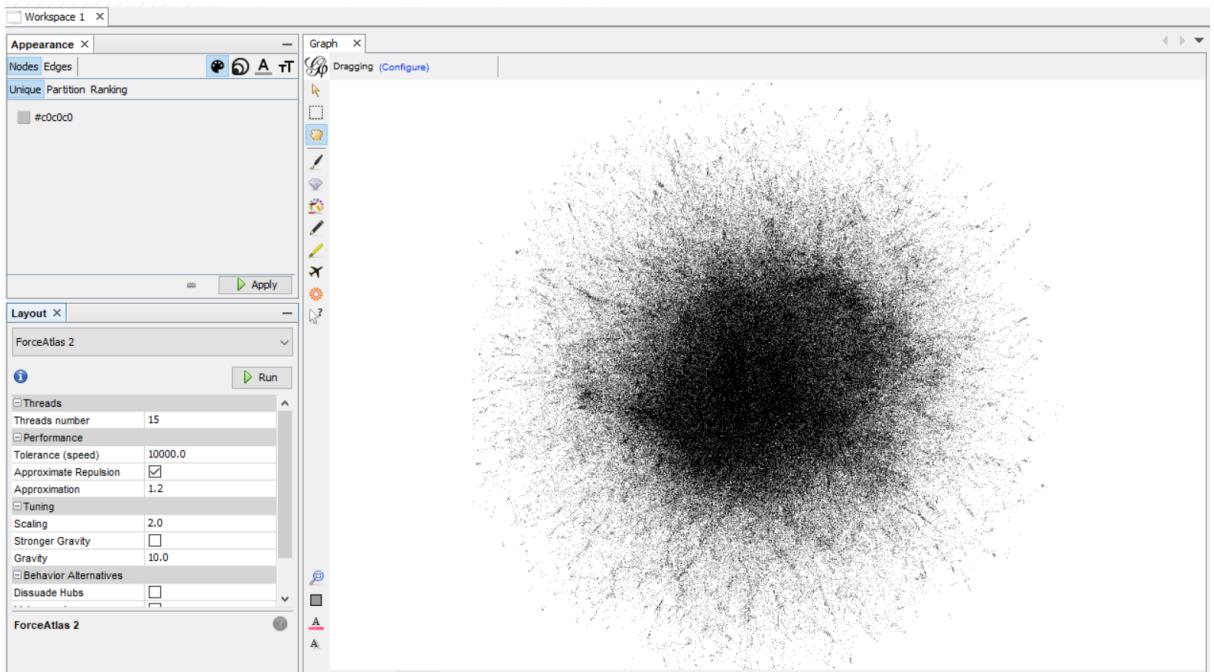
Importing the dataset



Initial graph



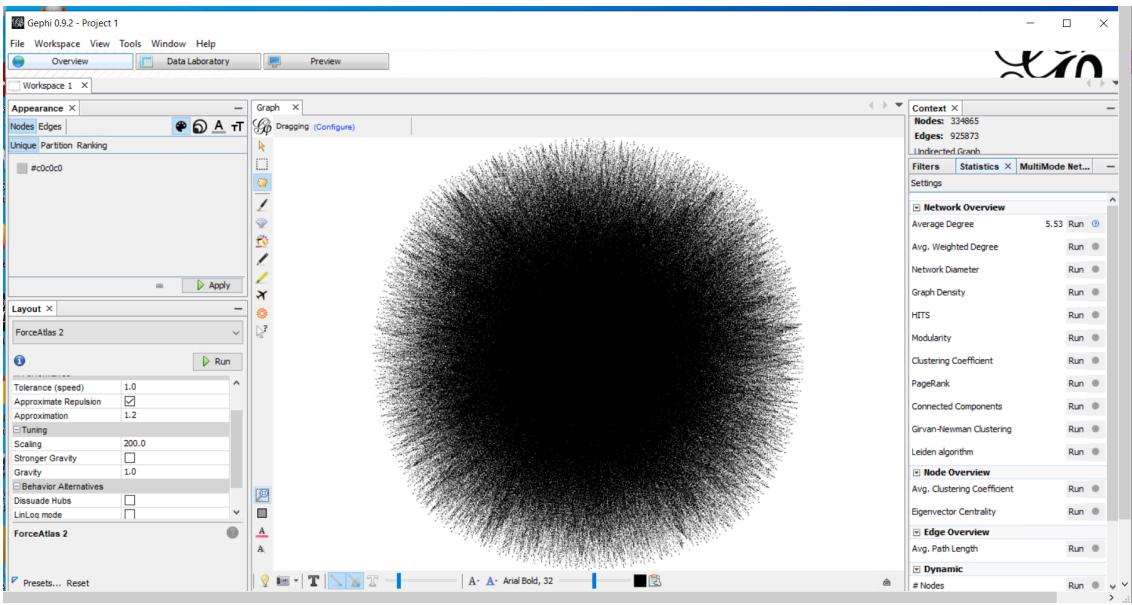
Force Atlas 2



Multigravity Force Atlas 2

Scaling. Control scale of the expansion of the graph.

Prevent overlap. Prevent nodes from overlapping.

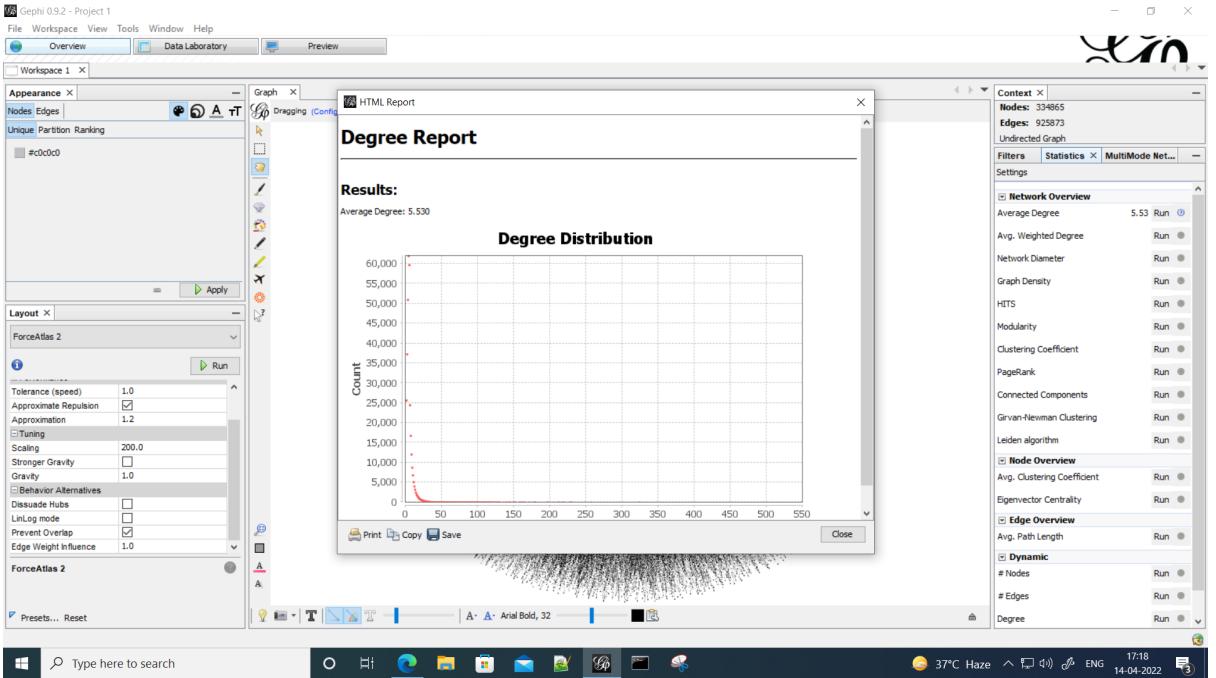


Degree Centrality

Degree: Connectivity of a node

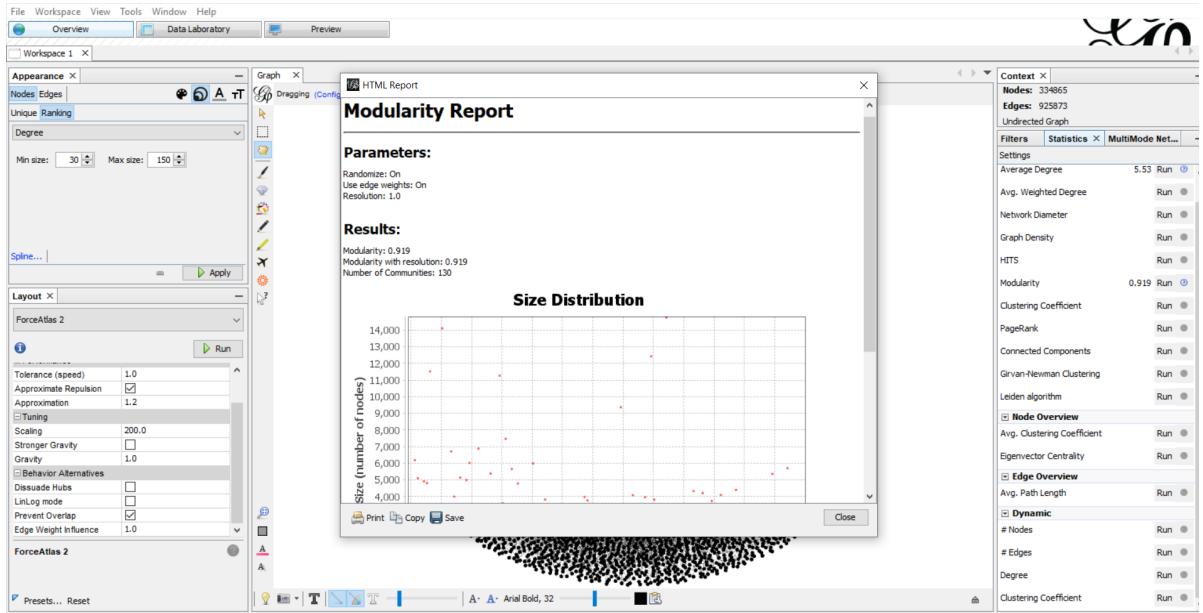
Degree centrality is a simple count of the total number of connections linked to a vertex. It can be thought of as a kind of popularity measure, but a crude one that does not recognize a difference between quantity and quality. Degree is the measure of the total number of edges connected to a particular vertex.

The average degree of our nodes is 5.530, i.e. all the nodes have approximately five connections with other nodes.



Modularity

Modularity is one measure of the structure of networks or graphs. It was designed to measure the strength of division of a network into modules (also called groups, clusters or communities). Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules.



Results:

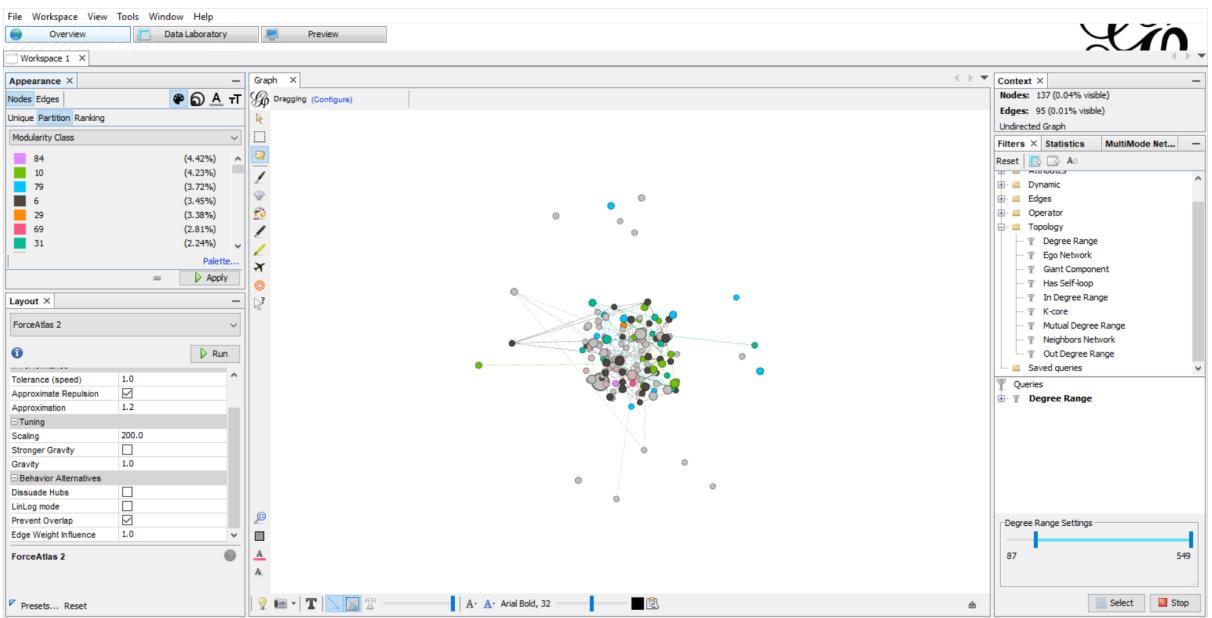
Modularity: 0.919

Modularity with resolution: 0.919

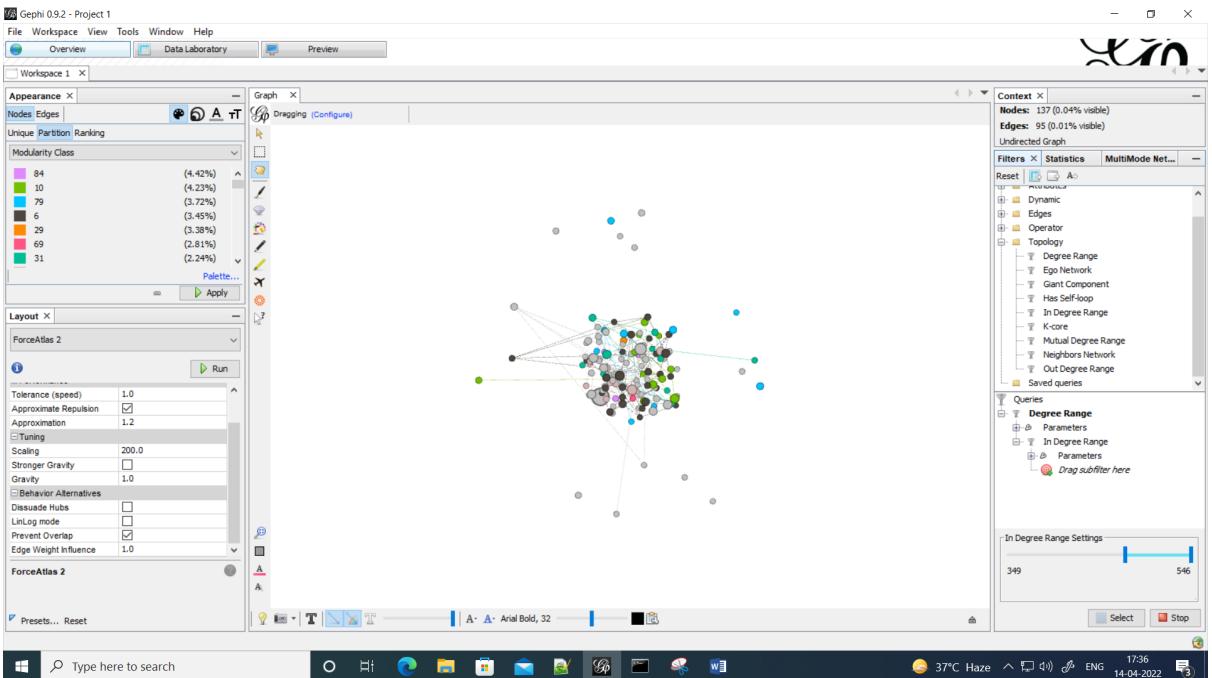
Number of Communities: 130

Topology → Degree Range

Here we selected the nodes whose degree was more than 87

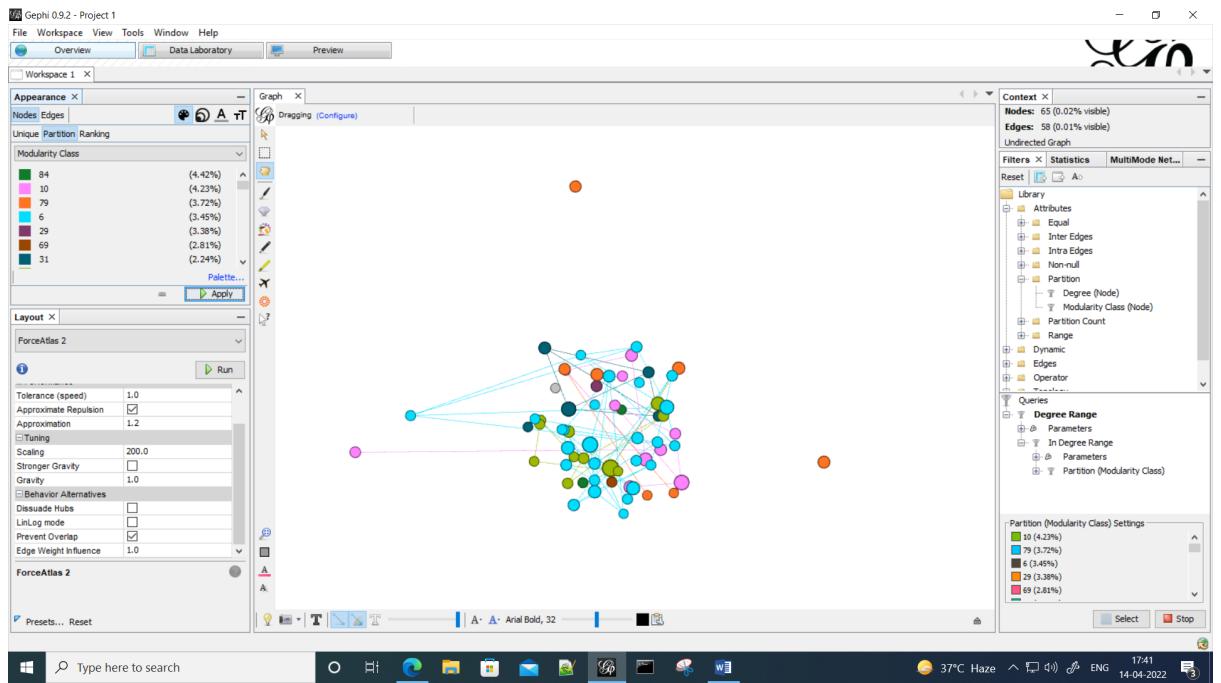


We tried putting in-degree filter for this graph, but as it is an undirected graph it made not much difference, fro degree range selection



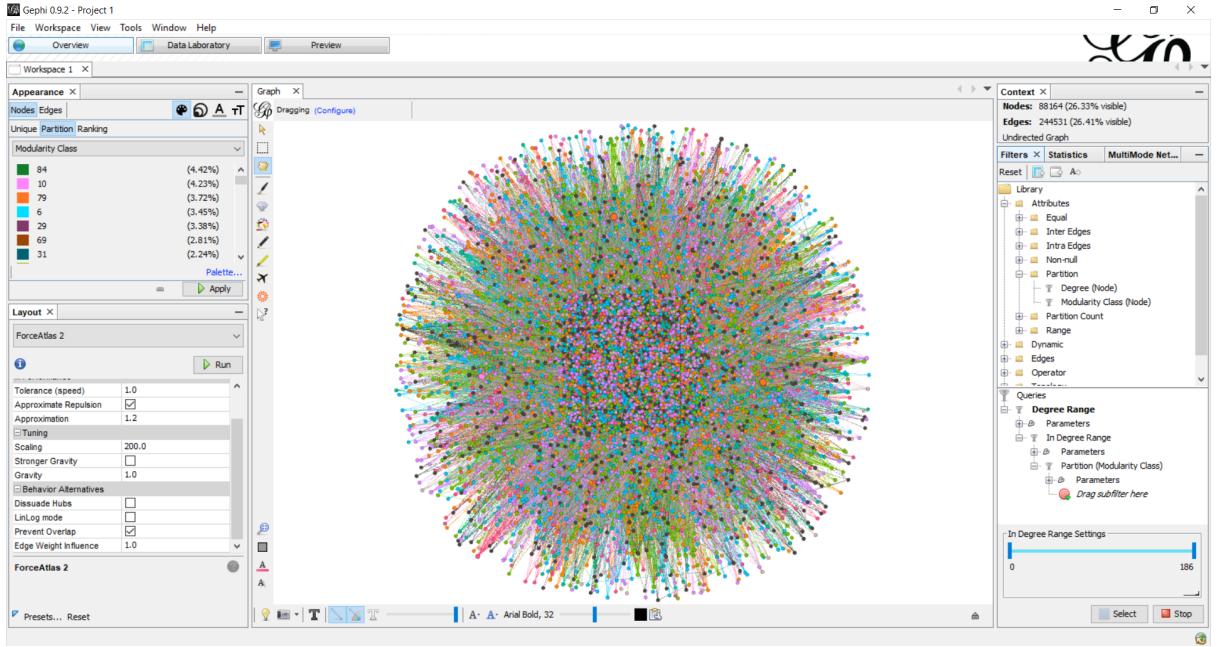
Attributes → Partition → Modularity

Partitioning was done on the basis of Modularity, the top six communities were selected, i.e. the communities with Degree above 87 and forming the major communities were selected



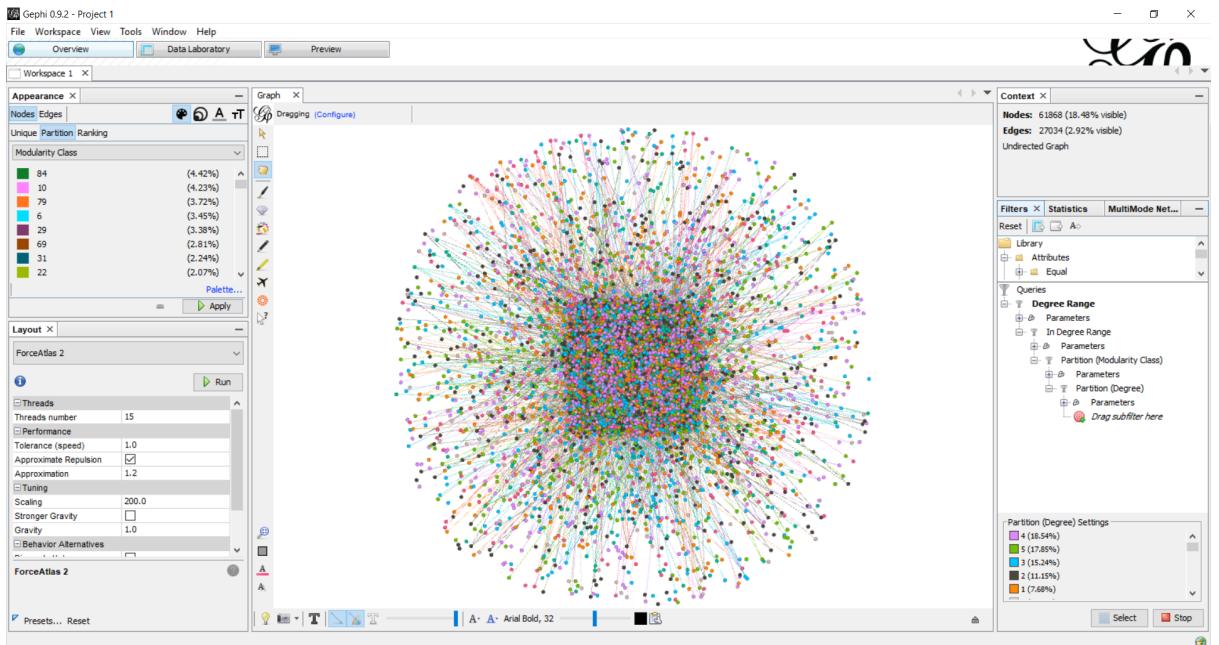
Attributes → Partition → Modularity

Partitioning was done on the basis of Modularity, the top six communities were selected, i.e. the communities with Degree below 187 and forming the major communities were selected



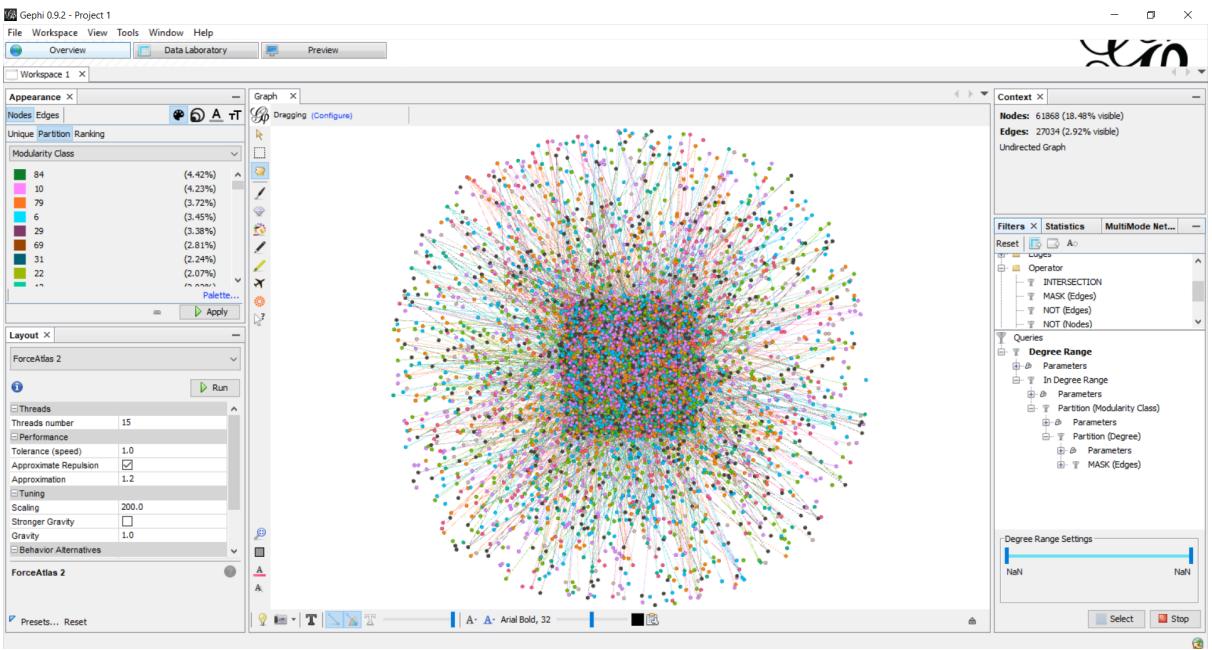
Attributes → Partition → Degree

Then we selected top 5 degrees in the



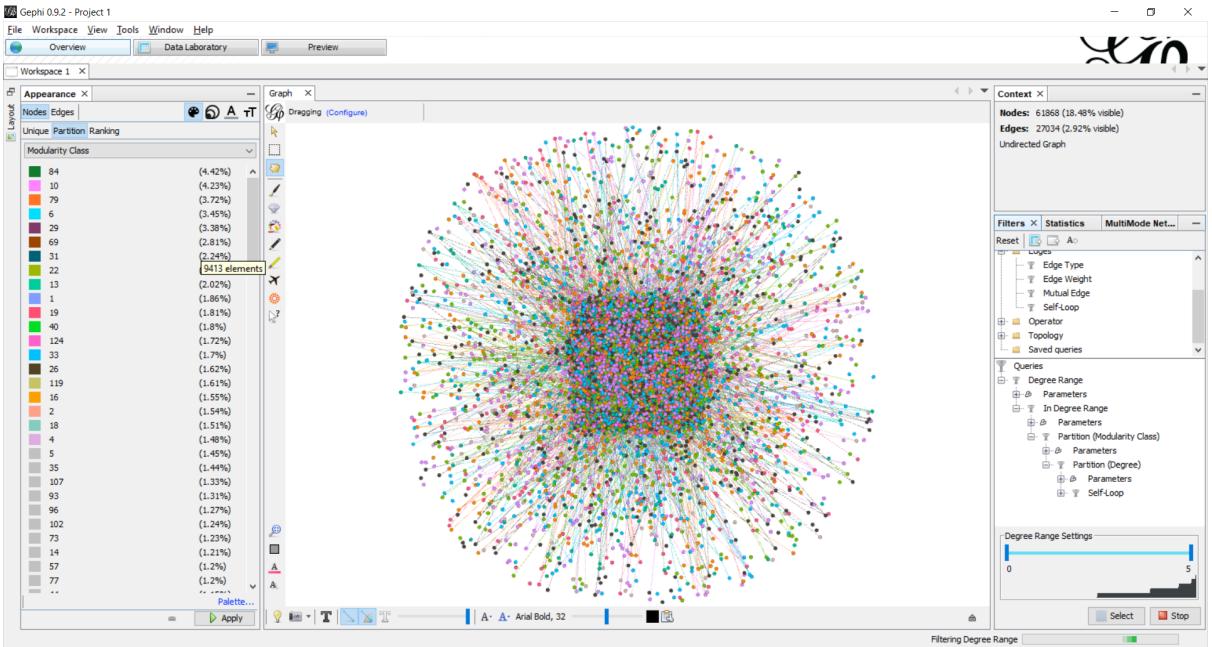
Operator → MASK (edge)

The mask operator is used to show relations between specific nodes, such as SOurce node to target node here, but the relations are high so the graph is not much clear



Edges → self loop

This filter is used to remove self loops of edges to nodes, what we could infer was there are no self nodes



Conclusion

The last image suggests that the dataset has many items that are bought frequently, and the item bought along with those items, are random and are not strongly connected. There are 130 different communities and filtering them suggests the top six communities are also not bonded strongly enough.

