



FUNDAMENTALS OF DATABASE MANAGEMENT SYSTEM

Social Network Analysis

SUBMITTED TO

Prof. Ashok Harnal

SUBMITTED BY - Group 10 (Section G)

025024 - Nehal Mehta

025043 - Nandini Gandhi

025045 - Harsh Chauhan

15th April 2022

Facebook Large Page-Page Network

Introduction

This webgraph is a page-page graph of verified Facebook sites. Nodes represent official Facebook pages while the links are mutual likes between sites. Node features are extracted from the site descriptions that the page owners created to summarize the purpose of the site. This graph was collected through the Facebook Graph API in November 2017 and restricted to pages from 4 categories which are defined by Facebook. These categories are: politicians, governmental organizations, television shows and companies. The task related to this dataset is multi-class node classification for the 4 site categories.

Dataset statistics	
Directed	No.
Node features	Yes.
Edge features	No.
Node labels	Yes. Binary-labeled.
Temporal	No.
Nodes	22,470
Edges	171,002
Density	0.001
Transitivity	0.232

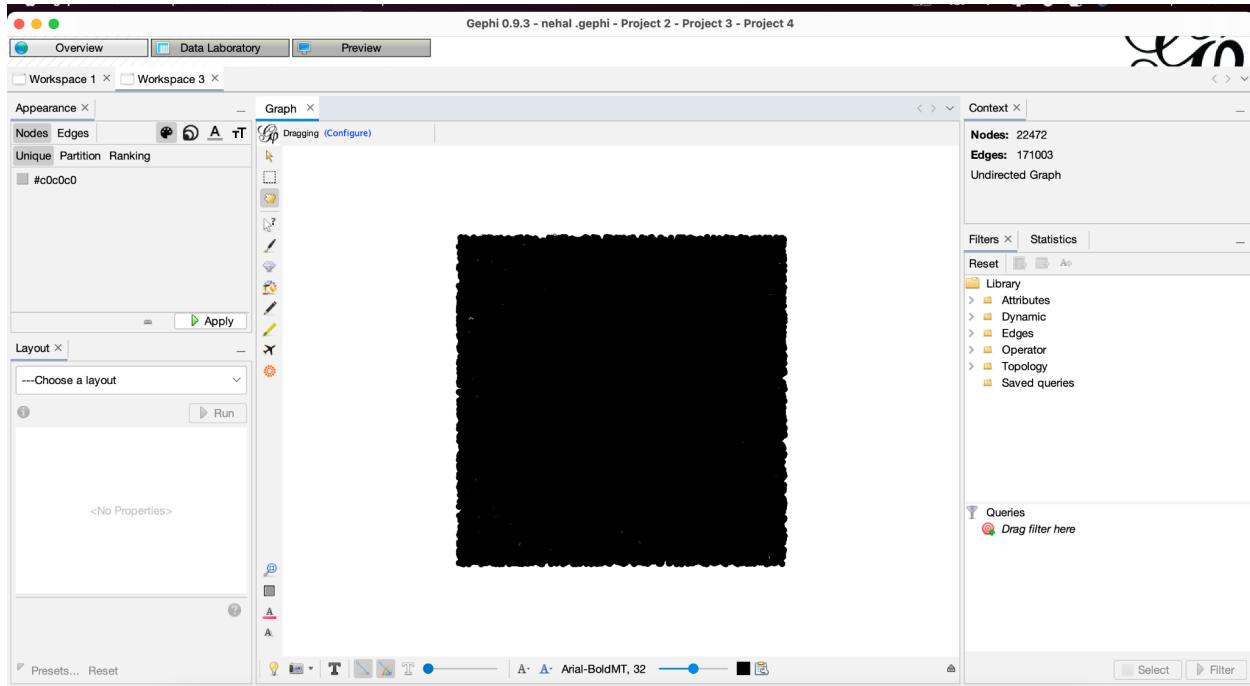
Key Details of Database

- Directed: No.
- Node features: Yes.
- Edge features: No.
- Node labels: Yes. Binary-labeled.
- Temporal: No.

| Nodes - 22,470
| Edges - 171,002
| Density - 0.001
| Transitivity - 0.232

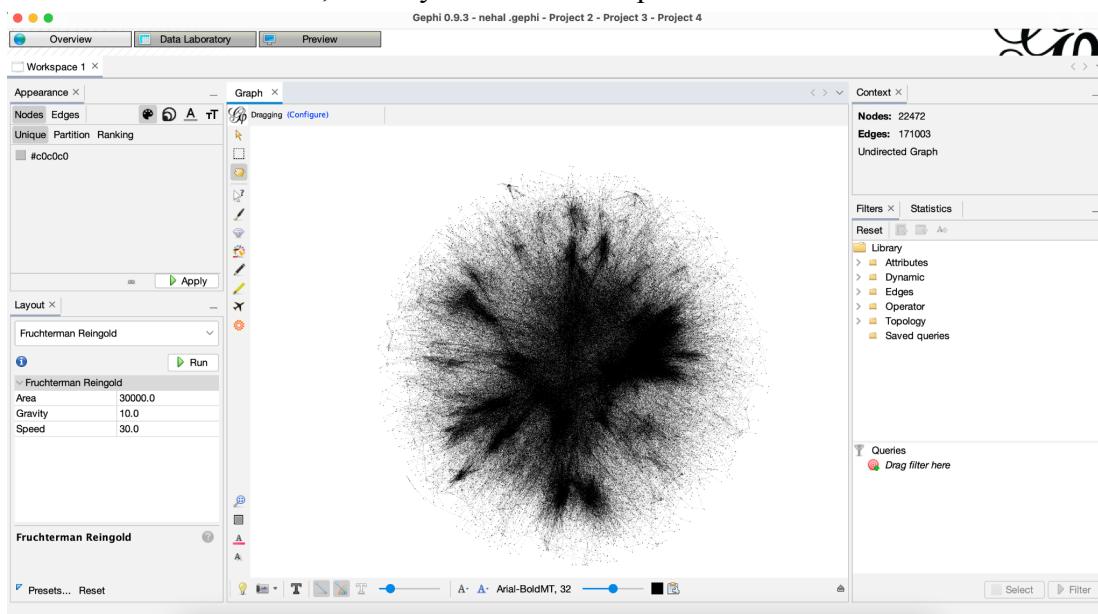
Both the files `musae_facebook_target.csv` and `musae_facebook_edges.csv` were imported into Gephi.

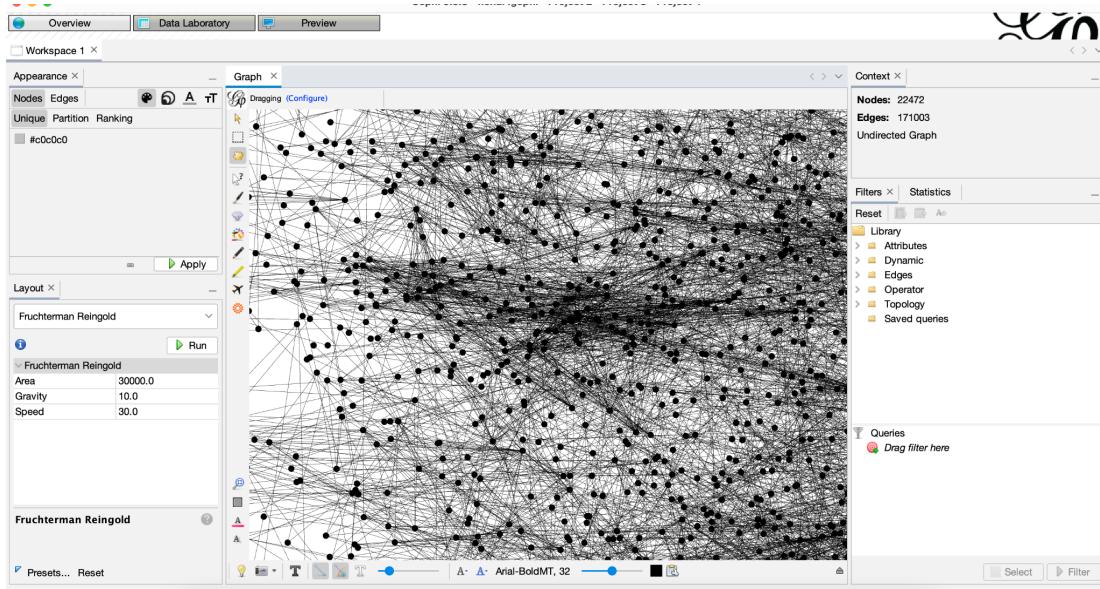
And the initial graph looked like this :



Layout >Fuchterman Reingold

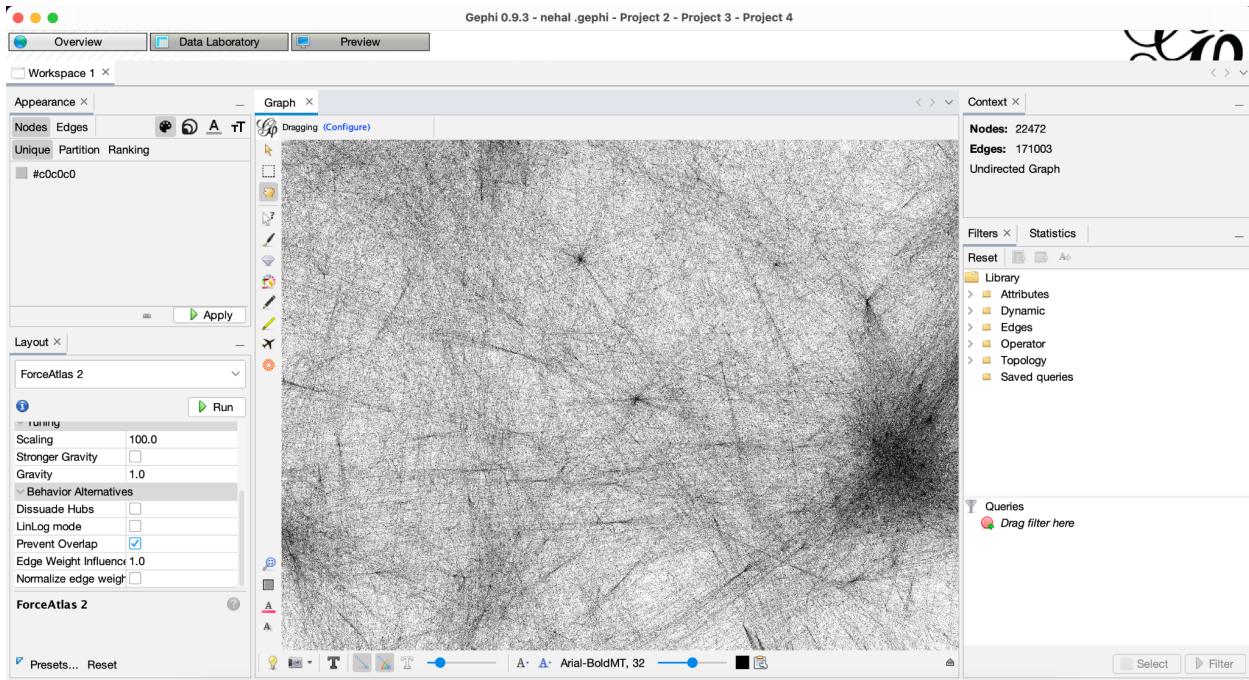
Area selected was 30000, Gravity was 10.0 and Speed was selected as 30.



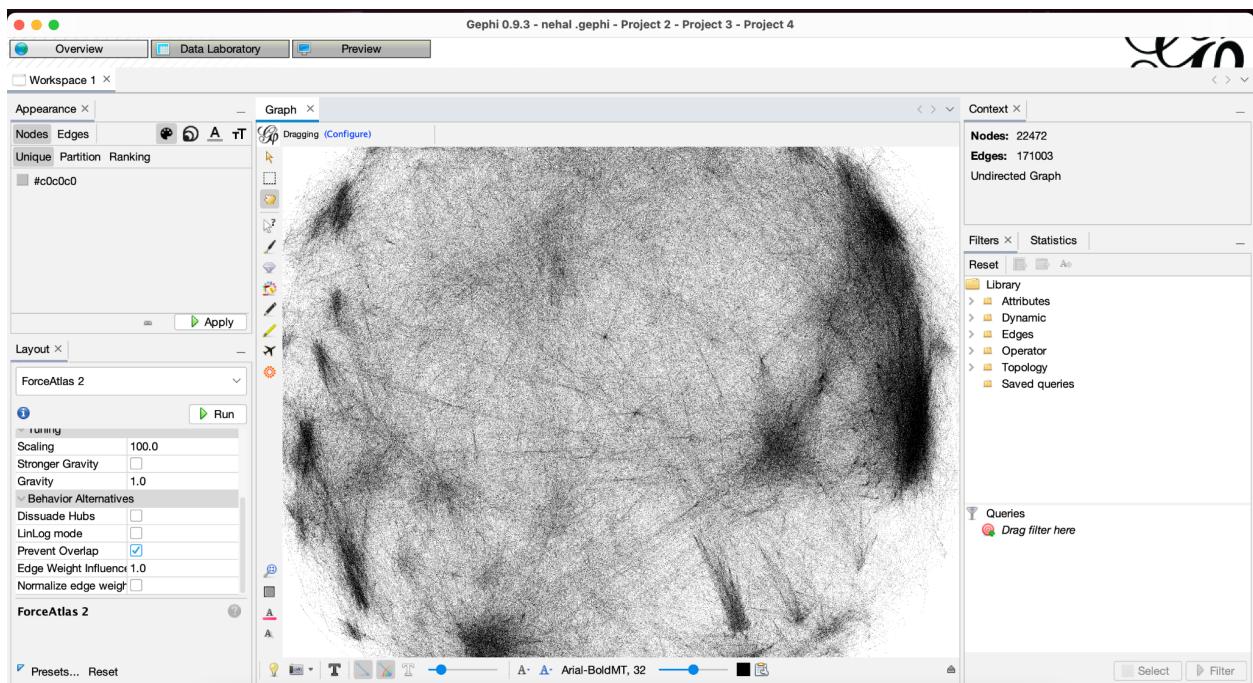


Layout> Force Atlas 2 Q

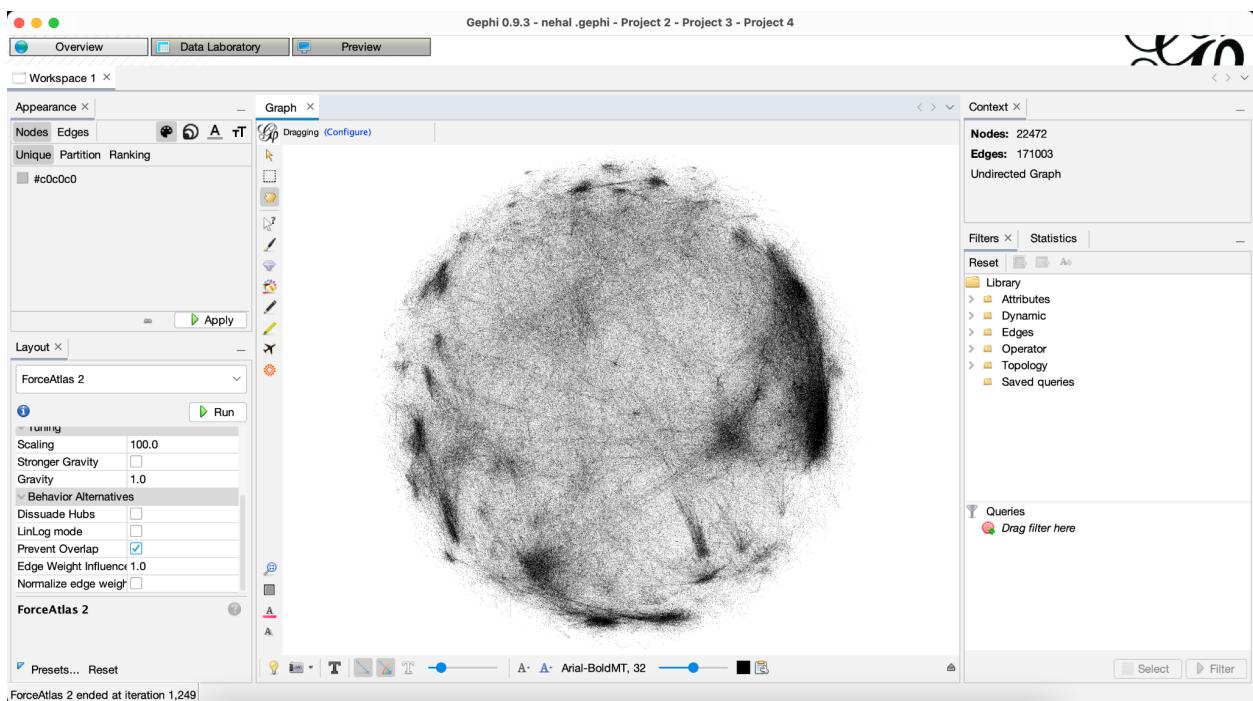
Scaling was selected as 100 and speed as 1. It was executed for 4 minutes by selecting **Run** and the results are as under:



Zoom In View



Zoom Out View



We got **Average Degree** of Graph as 15.219

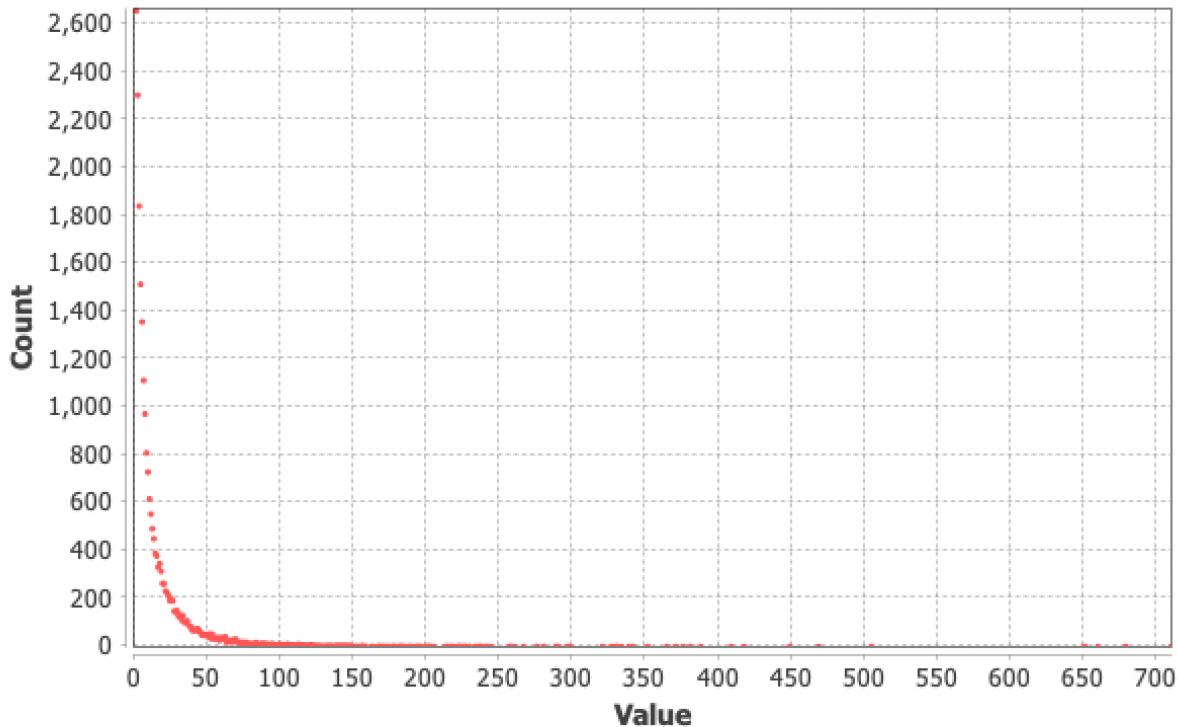
Since the average degree of an undirected graph is used to measure the number of edges compared to the number of nodes, which means that all the nodes have approximately 15 connections with all other nodes.

Degree Report

Results:

Average Degree: 15.219

Degree Distribution



Modularity

Modularity is a measure of the structure of networks. It measures the strength of division of a network into modules or groups or communities. Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules. Modularity is often used in detecting community structure in networks

Modularity Report

Parameters:

Randomize: On

Use edge weights: On

Resolution: 1.0

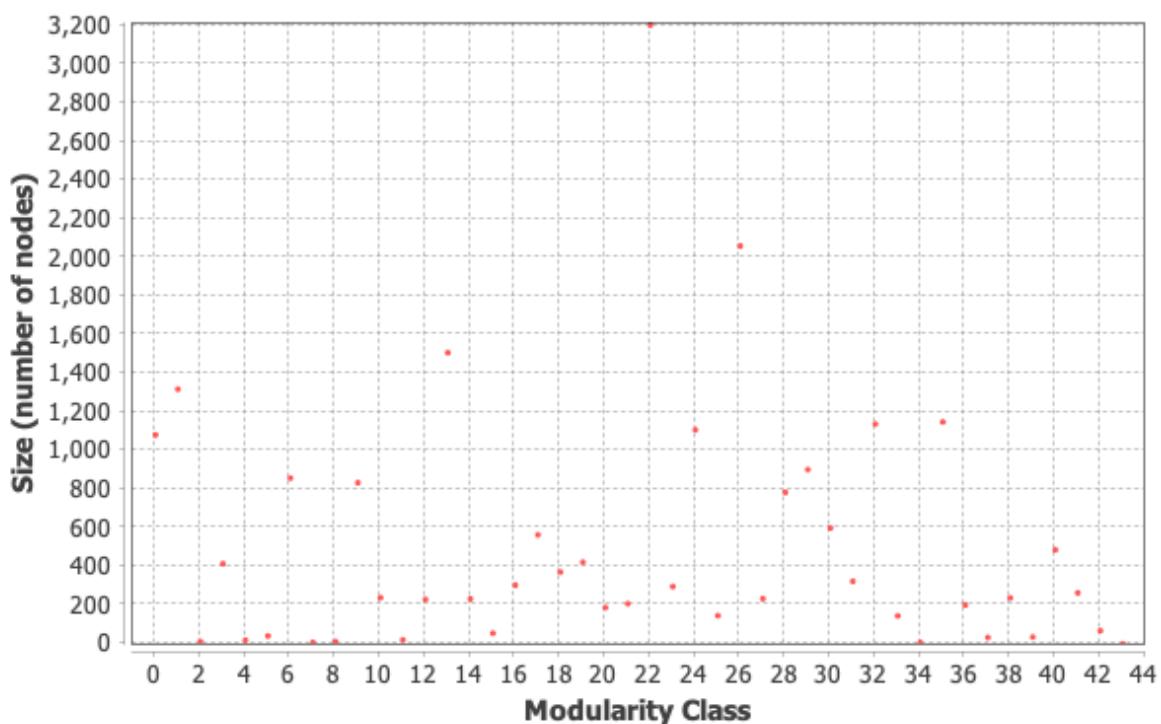
Results:

Modularity: 0.811

Modularity with resolution: 0.811

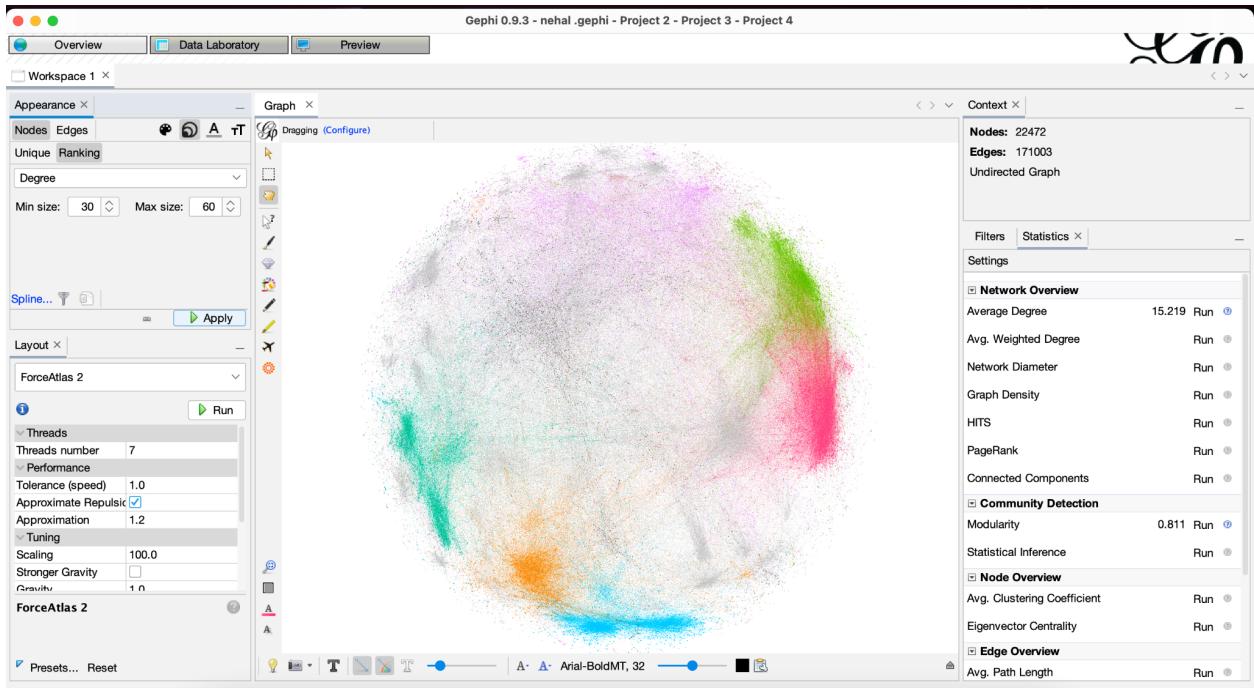
Number of Communities: 44

Size Distribution



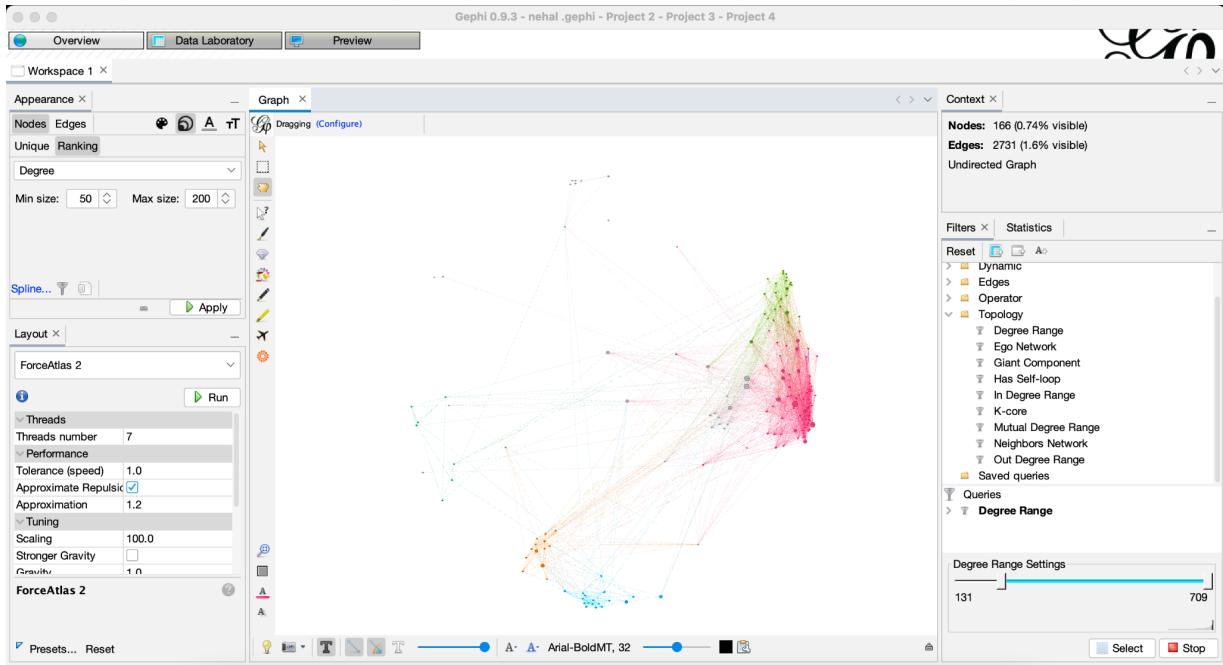
Nodes > Ranking > Degree

Degree was selected with minimum size as 30 and maximum size as 60



Topology > Degree range

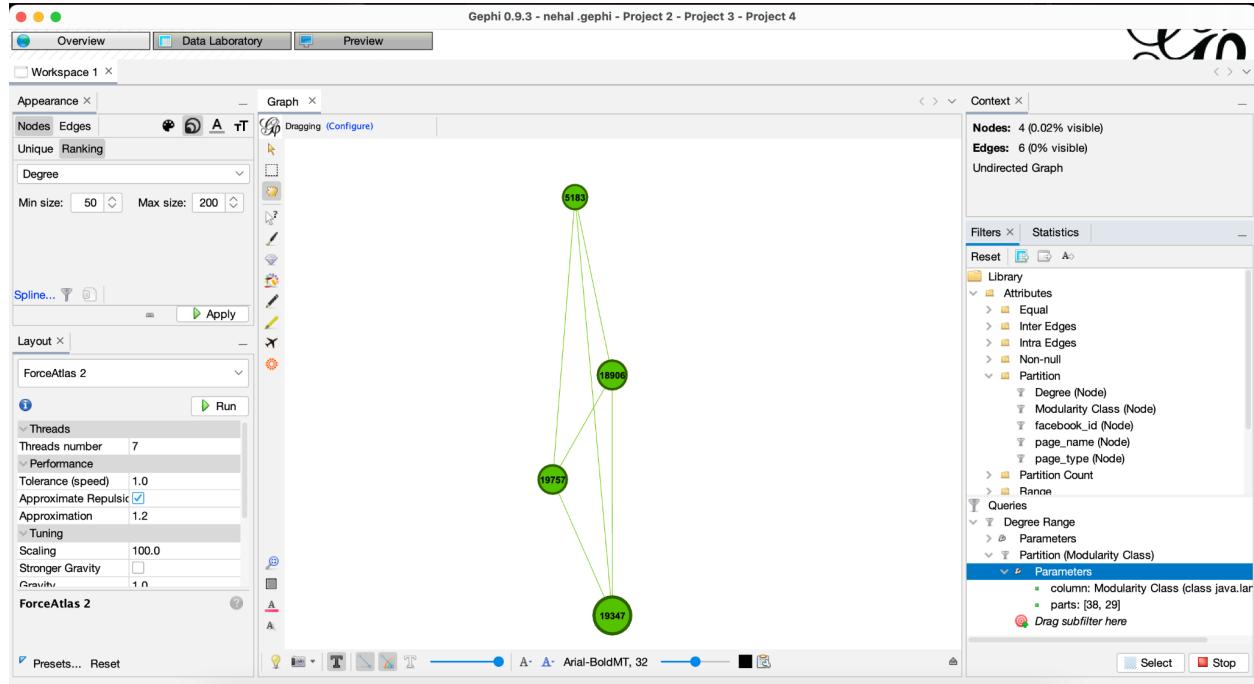
Degree Range was selected and range is set between 131 and 709, which means all the nodes whose degree above 131 have been selected.



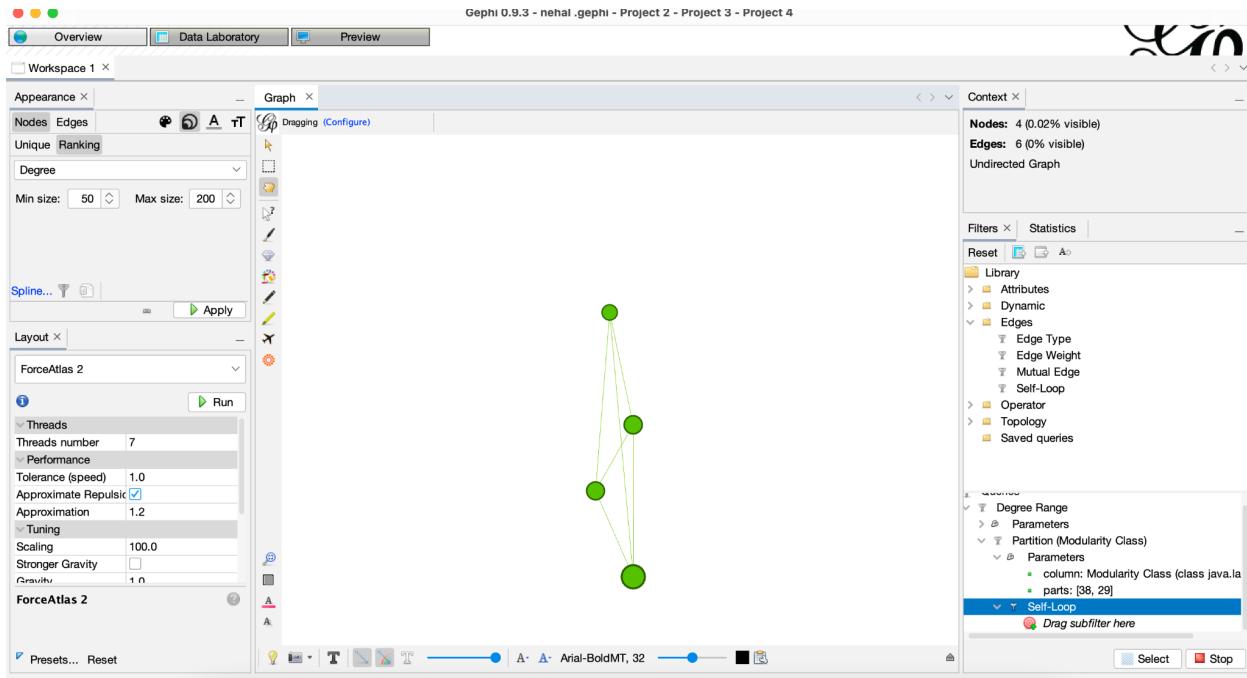
* This graph demonstrates the **Important Entities : Facebook Users**, who are highly **active** on Facebook.

Partition > Modularity Class

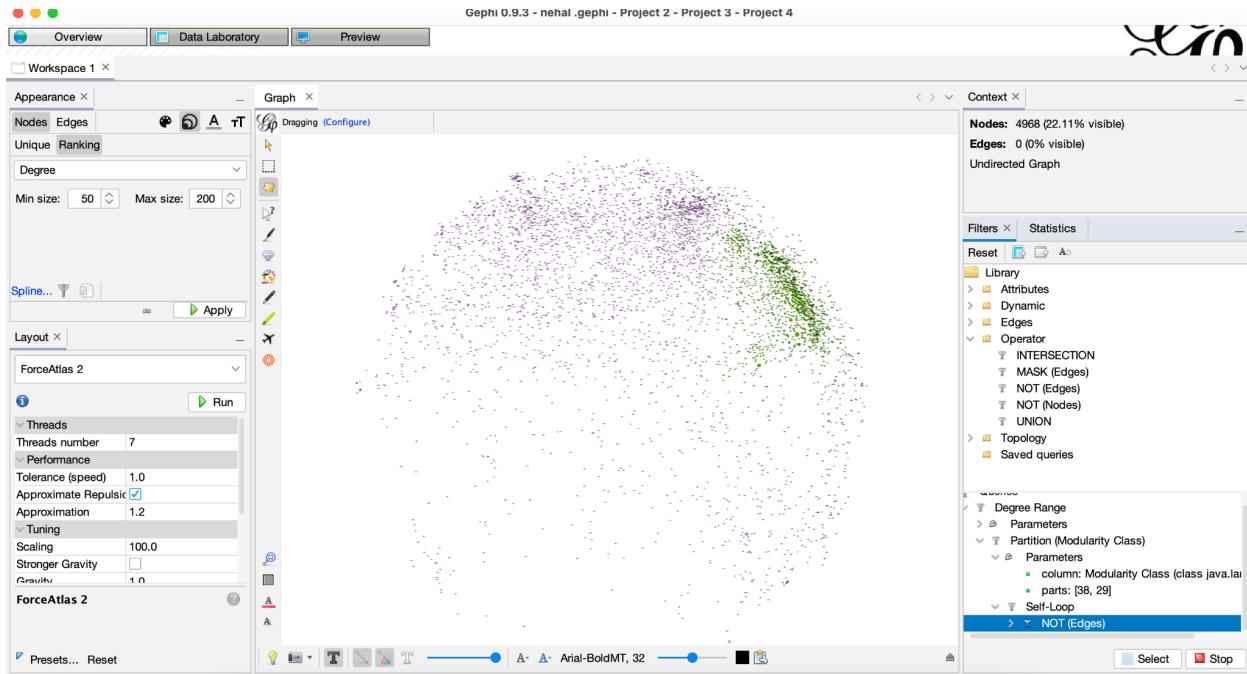
And top 4 communities were selected, those communities with degrees above 131.



Self loop filter is applied under the **edges** which gives the same result as above and no self loop self loop was present in the data set.



Under the **operator** we used the NOT operator on an **edge** filter and The NOT operator flips the result of a filter: what was hidden becomes visible and vice versa.



Data

Laboratory

Gephi 0.9.3 - facebook_.large.zip.gephi

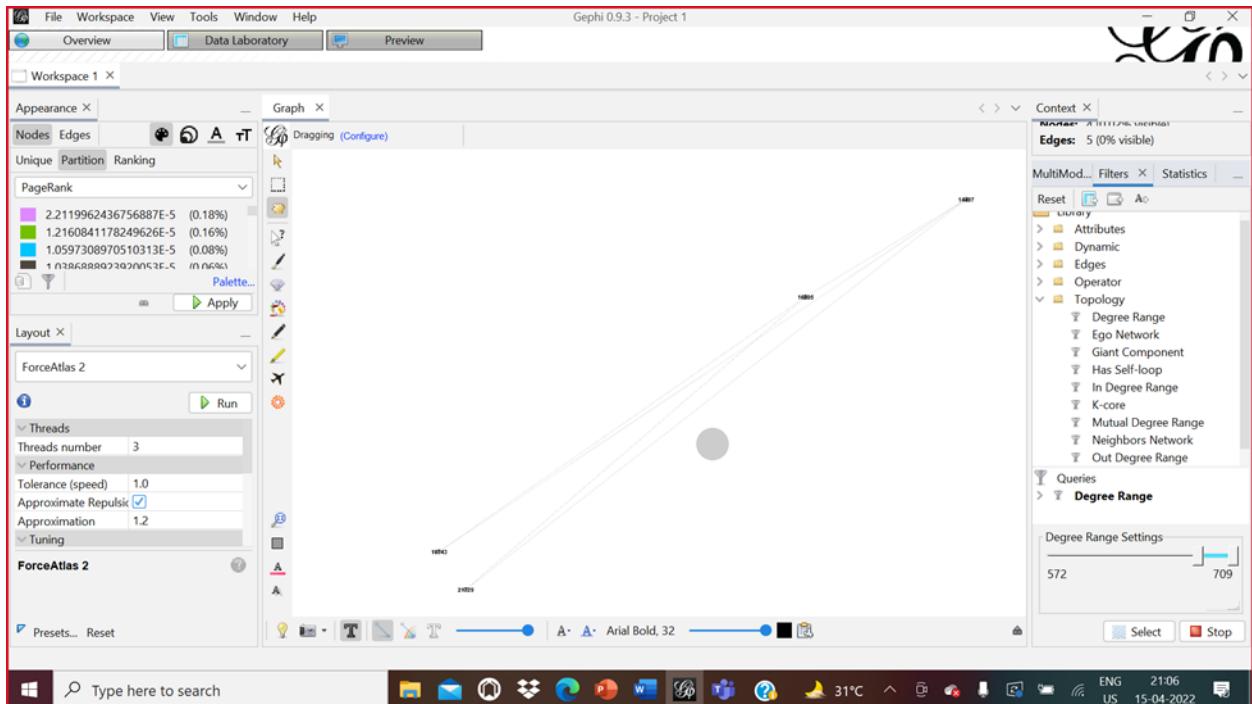
Nodes	Edges	Configuration	Add node	Add edge	Search/Replace	Import Spreadsheet	Export table	More actions	Filter:	Id	
16052	16052										
904	904										
290	290										
15644	15644										
11611	11611										
22403	22403										
1193	1193										
3300	3300										
10439	10439										
12831	12831										
15362	15362										
4000	4000										
9759	9759										
10500	10500										
16291	16291										
1964	1964										
1037	1037										
21803	21803										
7108	7108										
1668	1668										
17247	17247										
15470	15470										
14988	14988										
3533	3533										
17860	17860										
15647	15647										

Q1: Who are important entities from different points of view.

Node ID - 16895

Label ID - 16895

Degree - 709



14497, 19743 and 21729 are the important entities as their range is >572 and <=709.

Q2: How many communities exist within the network? Examine the characteristics of each community. Why a community is different from other communities. For this purpose, examine community attributes in Data Laboratory of gephi.

There are 44 Communities that exist (Already explained above)

Modularity Report

Parameters:

Randomize: On

Use edge weights: On

Resolution: 2.0

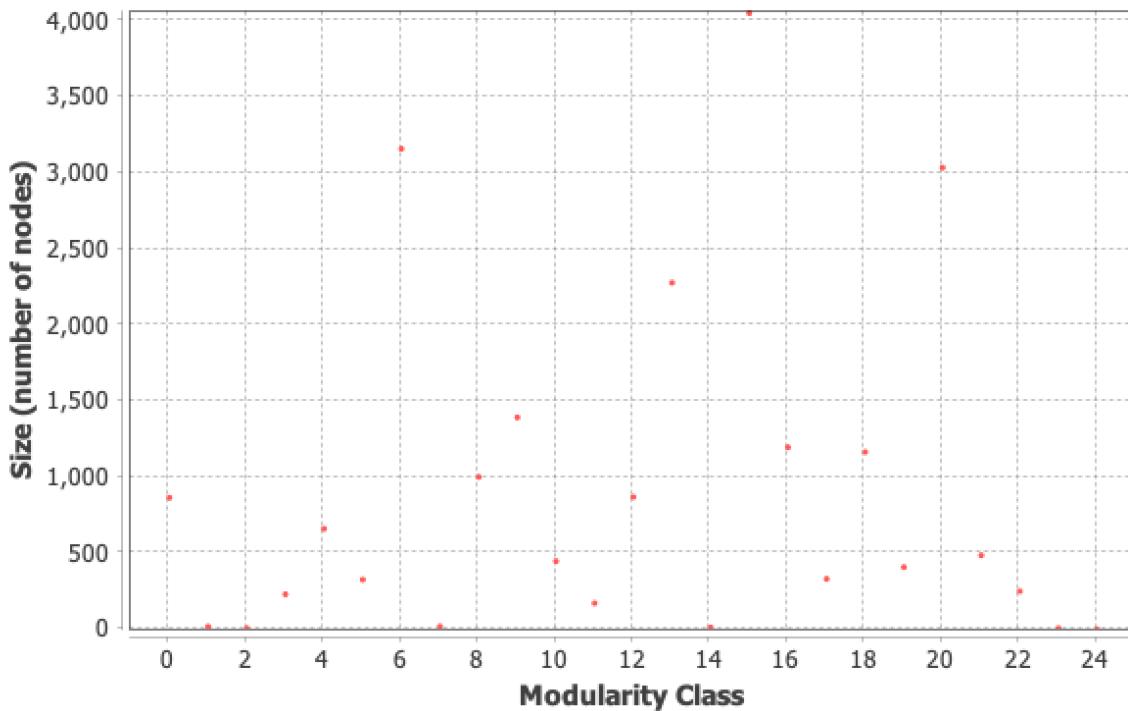
Results:

Modularity: 0.776

Modularity with resolution: 1.702

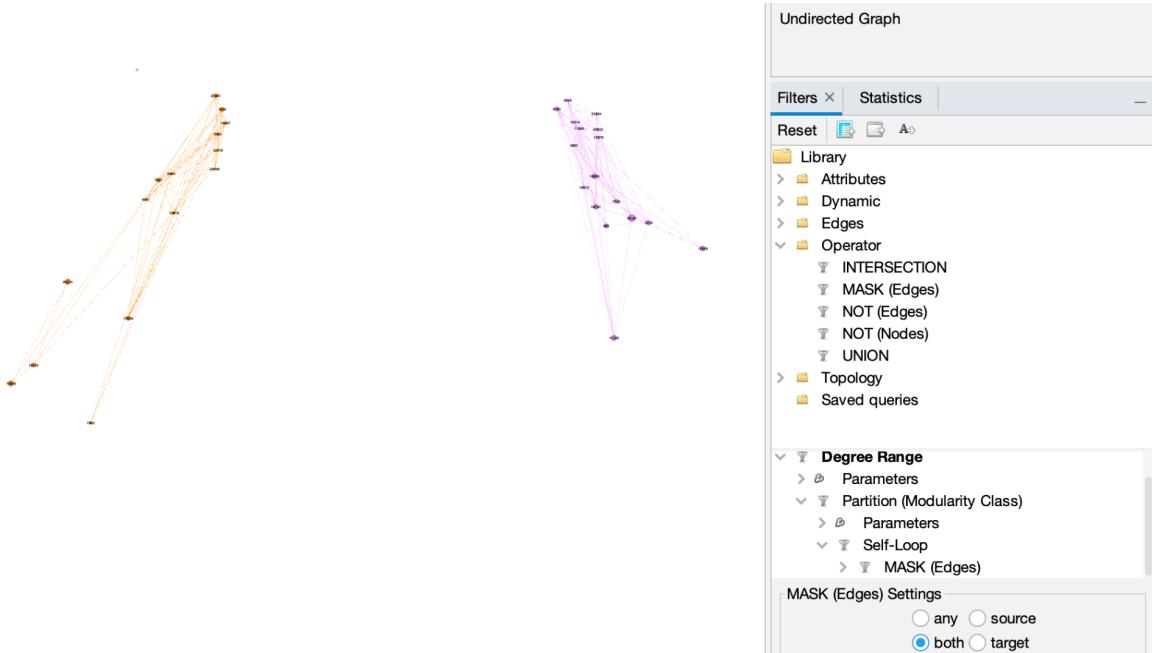
Number of Communities: 25

Size Distribution



Q3: Examine relationship of nodes within and outside communities

The dataset contains 22,472 nodes and there are 44 communities and the relationship within the communities are very strong and there are some communities which do not have any relationship to other communities



Q4: Any further insights that you may draw by analyzing the network

The most important insight from the analysis is that the two top nodes does not have any relation between them