ARGHODEEP NANDI

2024EEZ8395

ELL 880 - Social Network Analysis – Fall 2024

Assignment 1

Part A- Dataset

Dataset link: https://nrvis.com/download/data/labeled/DD21.zip

Number of Nodes: 5748

Number of Edges: 14267

Undirected Graph

Steps for importing the data into Gephi Workspace.

1. Open Gephi and go to File.
2. Import Spreadsheet. Select the edges csv file.
3. Set Separator as Space and Import as Edges table. Verify the preview. Click on Next.
4. Select Time-Representation as Interval.
5. Select Undirected Graph and Don't Merge in Edges Merge Strategy.
6. Now for Nodes Label. Import Spreadsheet.
7. Set Separator as Space and Import as Nodes Table. Time Representation is interval.
8. The Imported columns are set as String type Data-type.
9. Select Undirected Graph and Don't Merge in Edges Merge Strategy. And append it to existing workspace.

Part B- Tools- Gephi

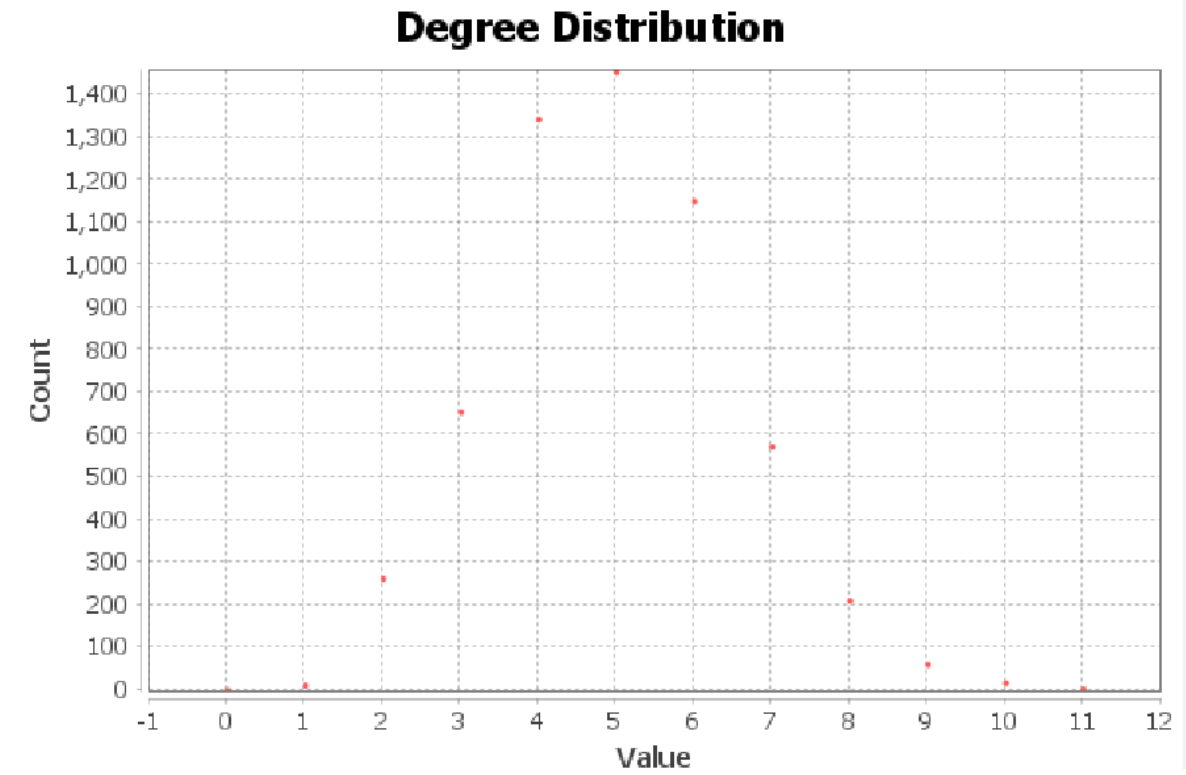
2. Degree Distribution

Statistics → Average Degree → Run

Degree Report

Results:

Average Degree: 4.963

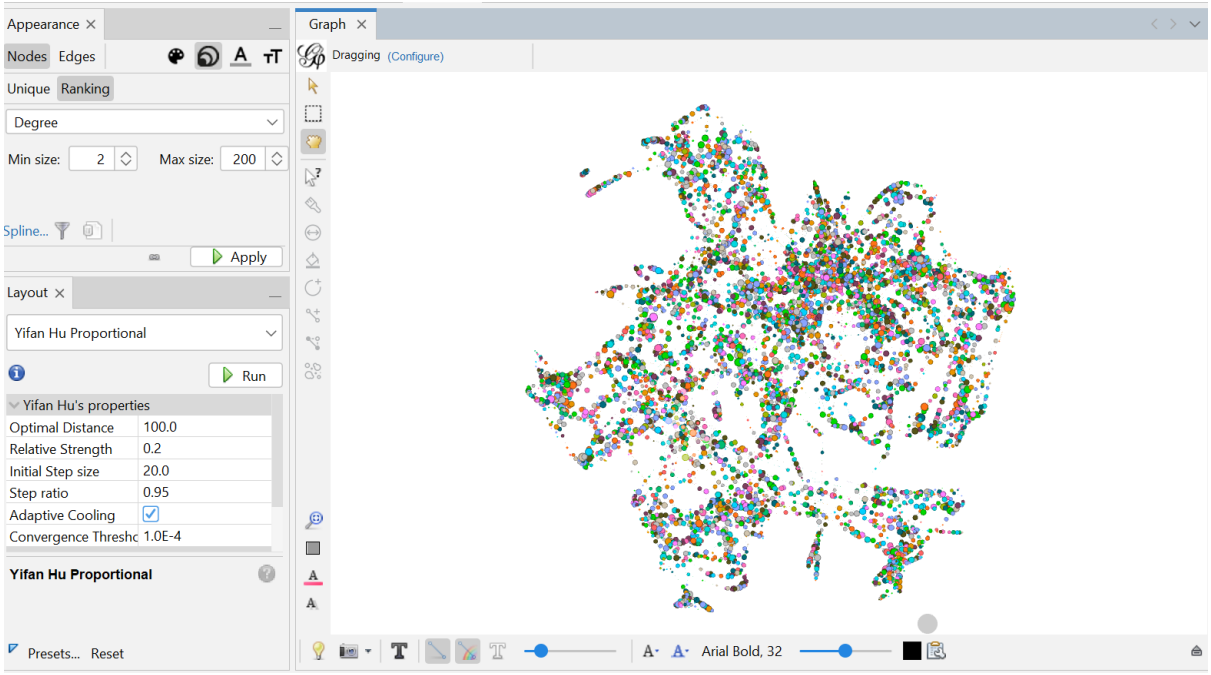


i: Assigning sizes to vertices based on their total degree.

Color Mode → Node → Partition → Node Label

Size Mode → Node → Ranking → Degree → Set Min and Max

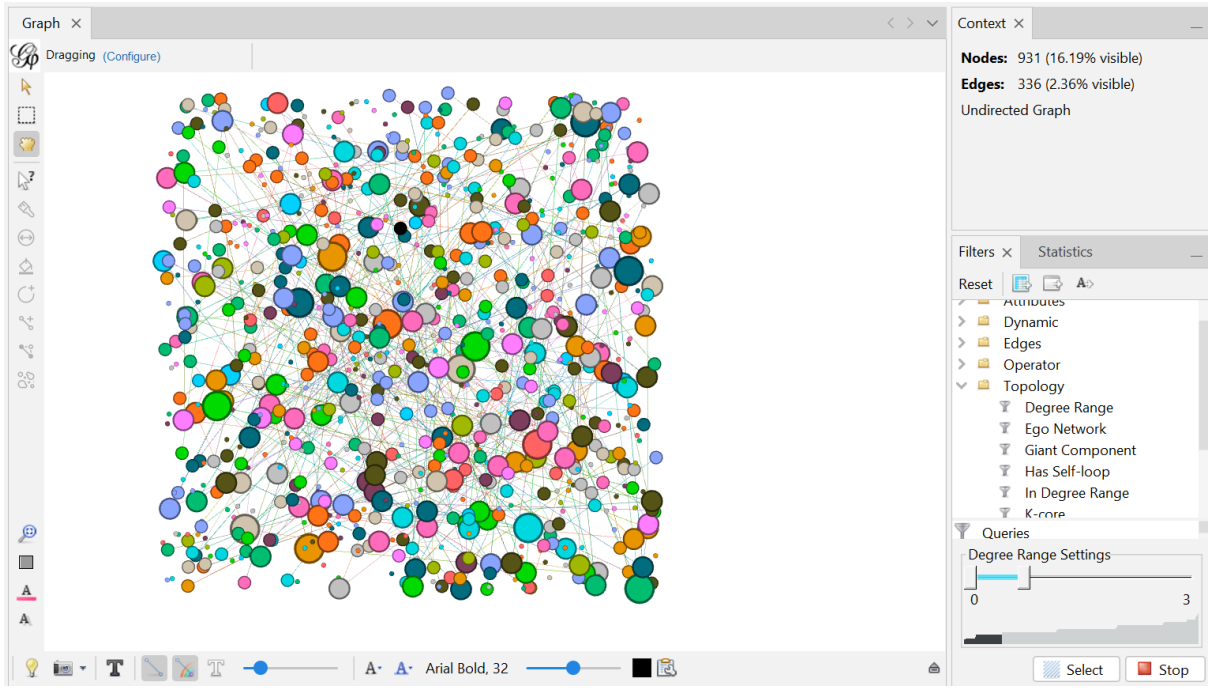
Layout → Yifan Hu Proportional → Default Parameters → Run

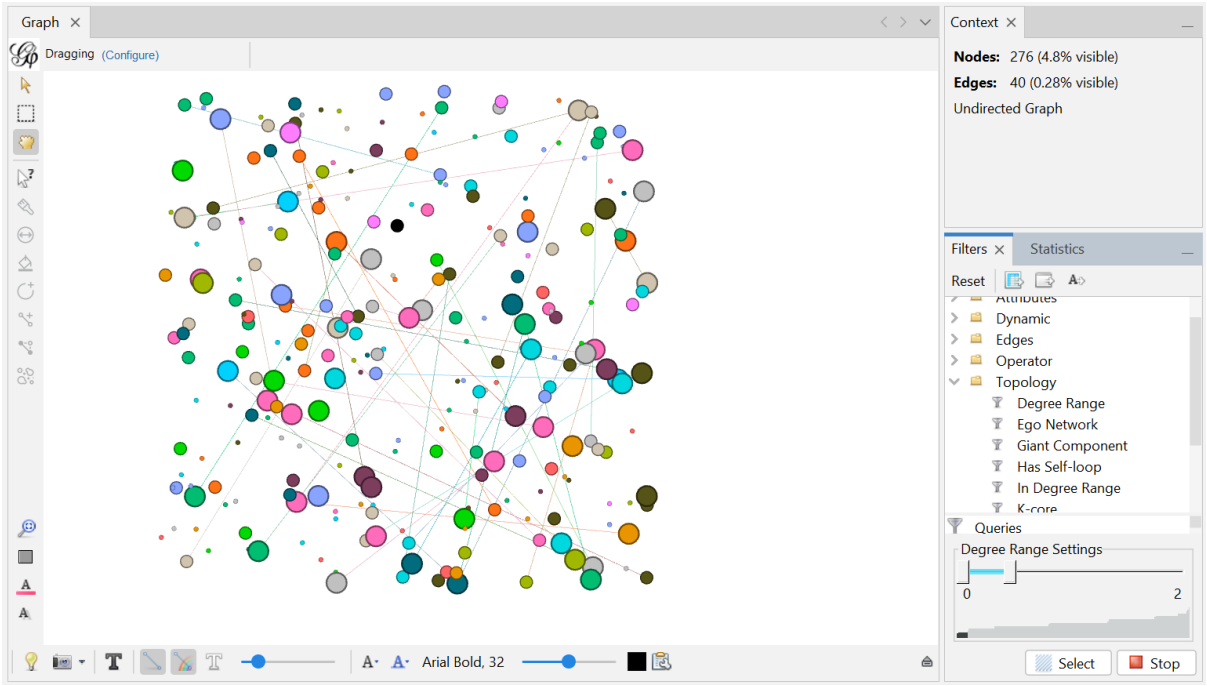


3. Filter the network by degree such that only the:

Filters → Topology → Degree Range → Degree Range Settings → Filter

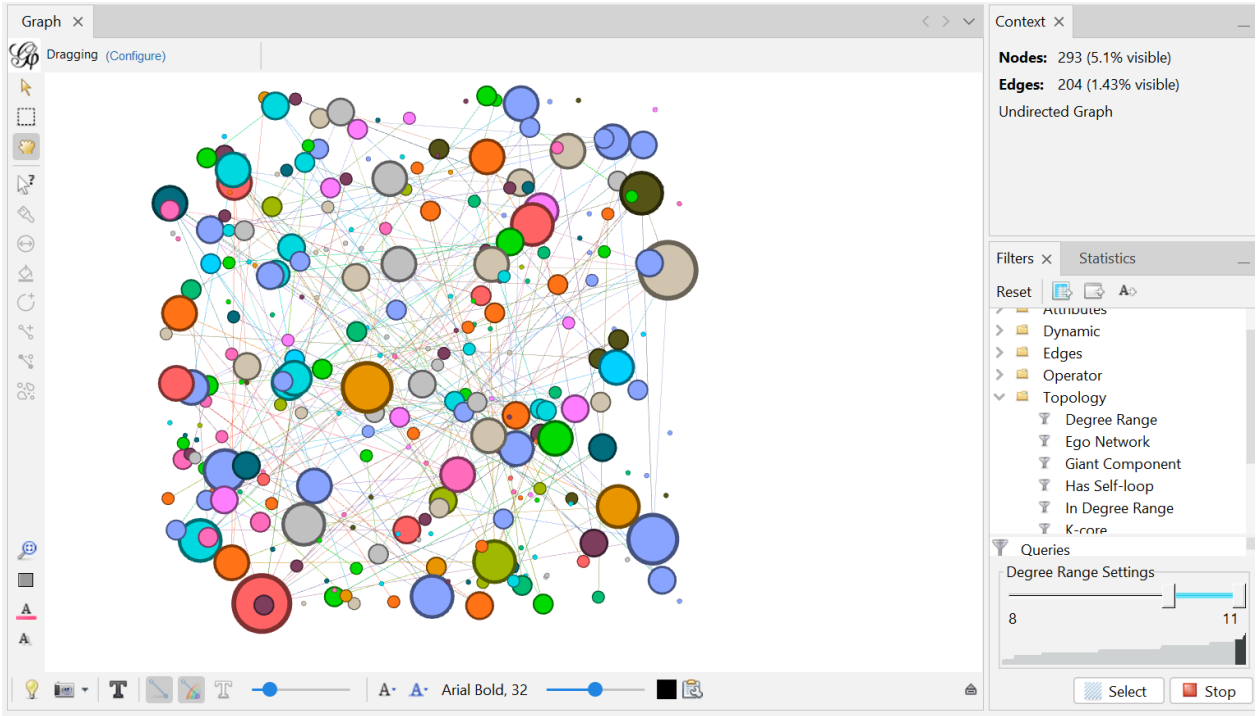
i. Bottom 10% of nodes and the connection among them are visible.





In my graph exact bottom 10% was not possible so I have attached with max degree 3 (16.19%) and max degree 2(4.8%).

ii. Top 5% of nodes and the connections among them are visible.



4. Find:

i. All the connected components of the network.

Statistics → Connected Components → Run

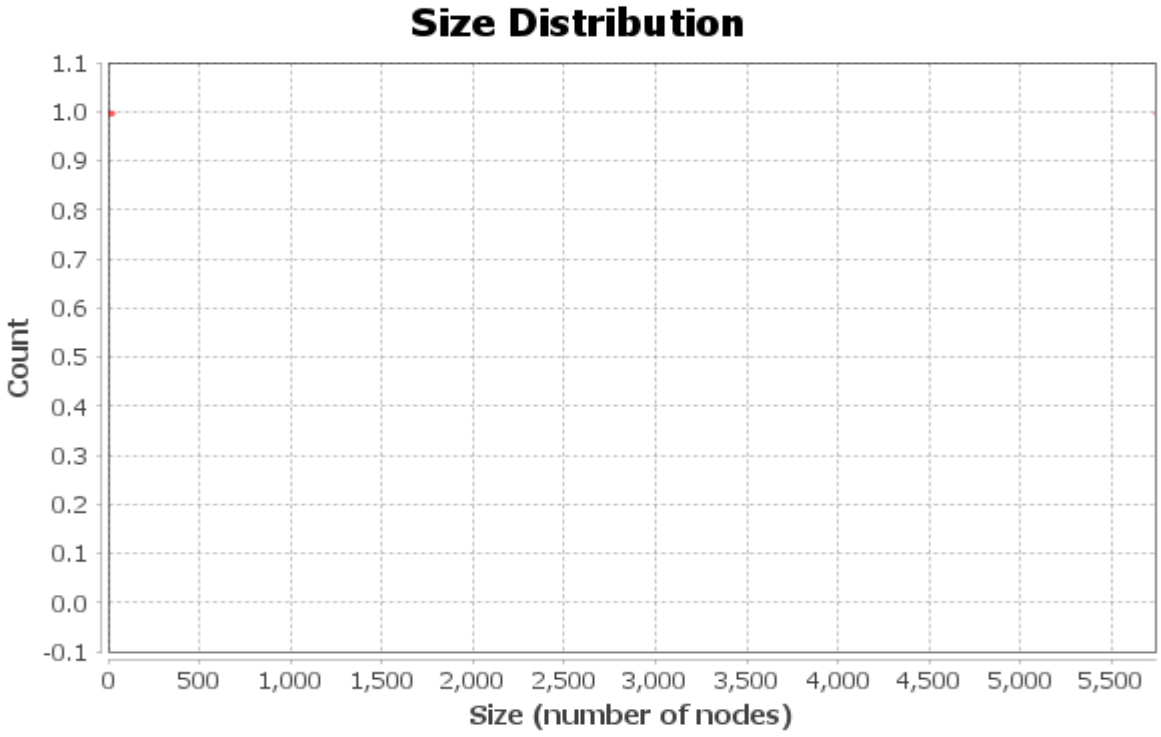
Connected Components Report

Parameters:

Network Interpretation: undirected

Results:

Number of Weakly Connected Components: 2



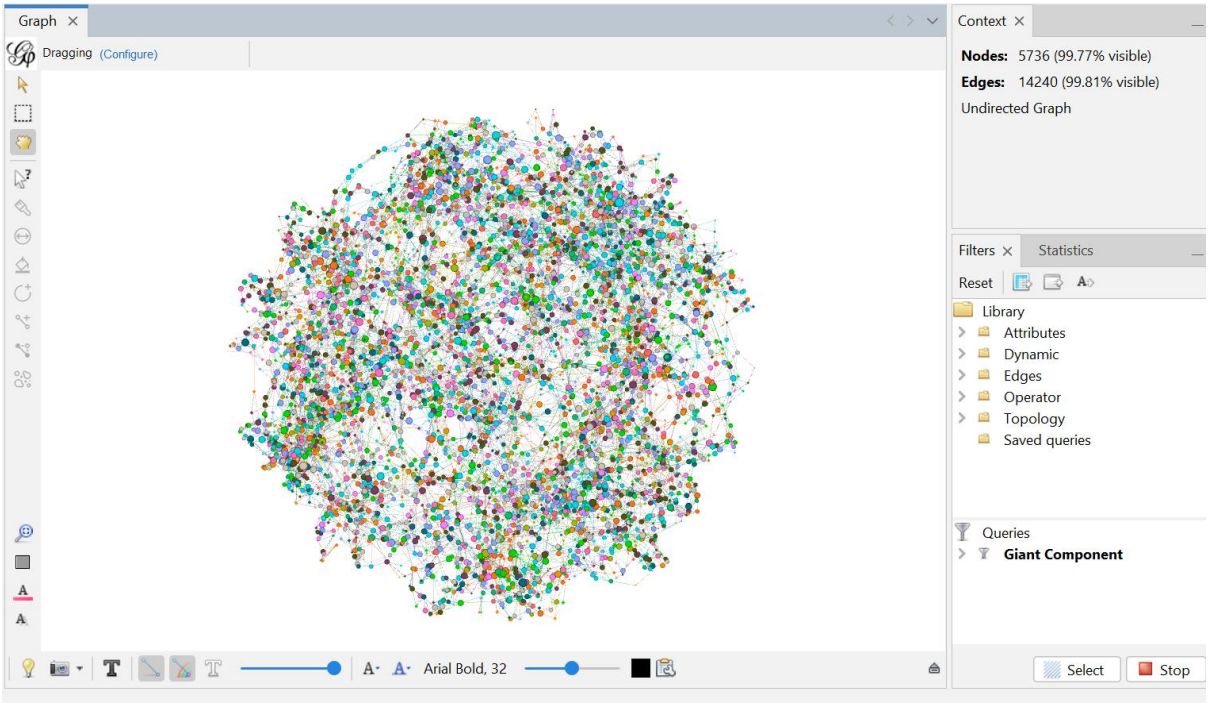
Algorithm:

Robert Tarjan, *Depth-First Search and Linear Graph Algorithms*, in SIAM Journal on Computing 1 (2): 146–160 (1972)

The Giant Component consists of 5736 (99.77%) nodes (see next answer). Hence could not fit into the above graph. The other component is too small.

ii. The size of the giant component of the network.

Filters → Topology → Giant Component → Filter



Part C- Libraries

Google Collab link:
https://colab.research.google.com/drive/1Xta_1IX13f8u6pmmtPUC9MCtxWd0Kj75