Bitcoin Alpha trust weighted signed network

(HW 1 report Task 1)

Bita Naserfarahmand

Student ID: 2021380075

Email: bita.nf@gmail.com

Computational sociology class

Content:

1)Background:

Homework 1 task 1:

Question:

Perform network analysis on one large-scale dataset:

- ***Crawl one real-world dataset or download one public dataset***
- ***Size of nodes: > 1000;***
- ***Size of edges: > 3000;***
- ***Quantify the centrality of nodes including degree, betweenness, closeness***
- ***Visualize the network topology***

Apparatus:

Gephi software

Dataset Information:

This is who-trusts-whom network of people who trade using Bitcoin on a platform called Bitcoin Alpha. Since Bitcoin users are anonymous, there is a need to maintain a record of users' reputation to prevent transactions with fraudulent and risky users. Members of Bitcoin Alpha rate other members in a scale of -10 (total distrust) to +10 (total trust) in steps of 1.

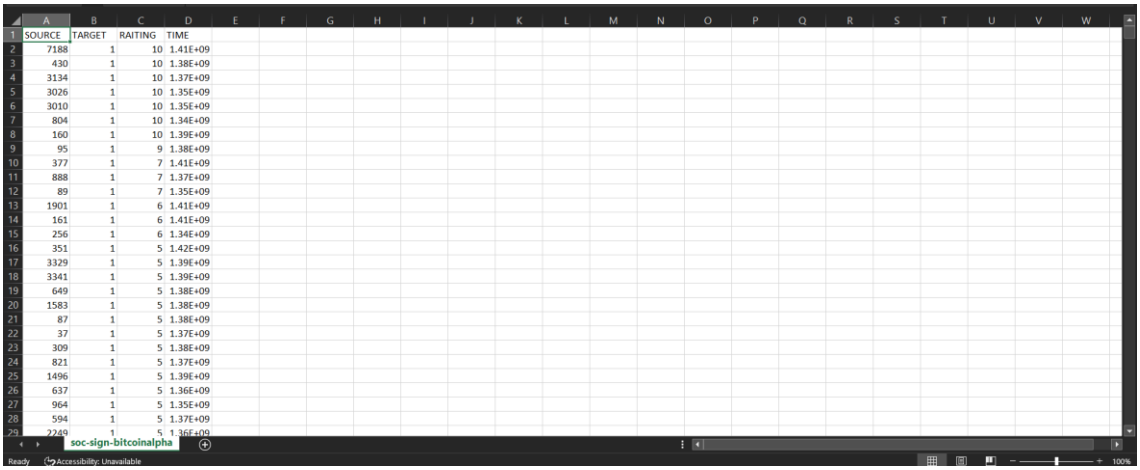
Nodes : 3783

Edges : 24186

Range of edge weight : -10 to +10

Percentage of positive edges : 93%

Here is s a brief look on how the dataset looks like in the excel format before being imported to Gefi:



SOURCE	TARGET	RAITING	TIME
7188	1	10	1.41E+09
430	1	10	1.38E+09
3134	1	10	1.37E+09
3026	1	10	1.35E+09
3010	1	10	1.35E+09
804	1	10	1.34E+09
160	1	10	1.39E+09
95	1	9	1.38E+09
377	1	7	1.41E+09
888	1	7	1.37E+09
89	1	7	1.35E+09
1901	1	6	1.41E+09
161	1	6	1.41E+09
256	1	6	1.34E+09
351	1	5	1.42E+09
3329	1	5	1.39E+09
3341	1	5	1.39E+09
649	1	5	1.38E+09
1583	1	5	1.38E+09
87	1	5	1.38E+09
37	1	5	1.37E+09
309	1	5	1.38E+09
821	1	5	1.37E+09
1496	1	5	1.39E+09
637	1	5	1.36E+09
964	1	5	1.35E+09
594	1	5	1.37E+09
2249	1	5	1.36E+09

The dataset has been divided into 4 columns, source, target, rating and time, as can be seen in the above picture.

SOURCE: node id of source, i.e., rater

TARGET: node id of target, i.e., ratee

RATING: the source's rating for the target, ranging from -10 to +10 in steps of 1

TIME: the time of the rating, measured as seconds since Epoch.

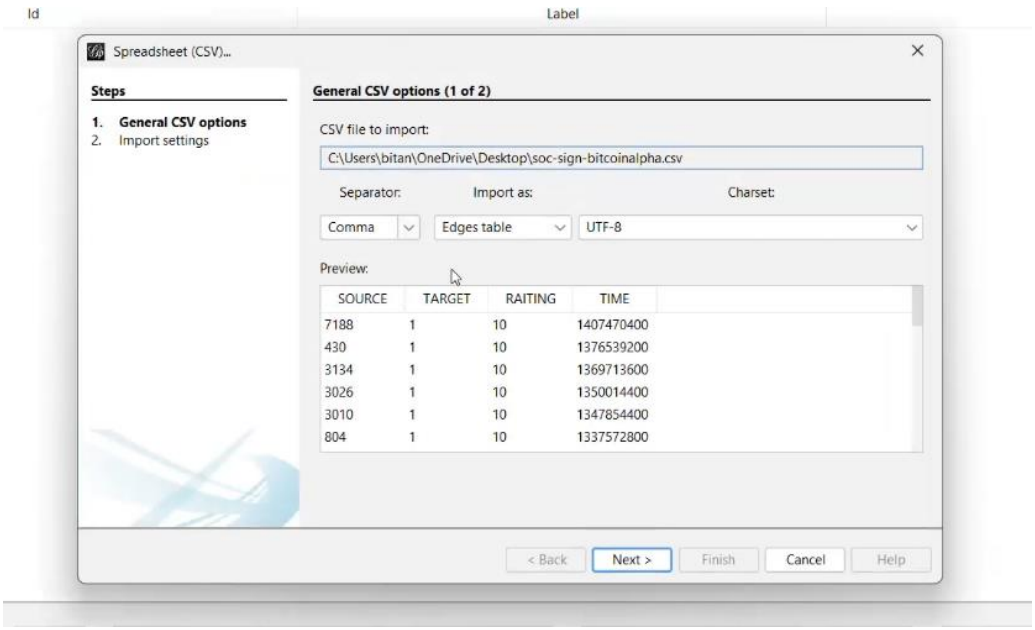
2)Goal:

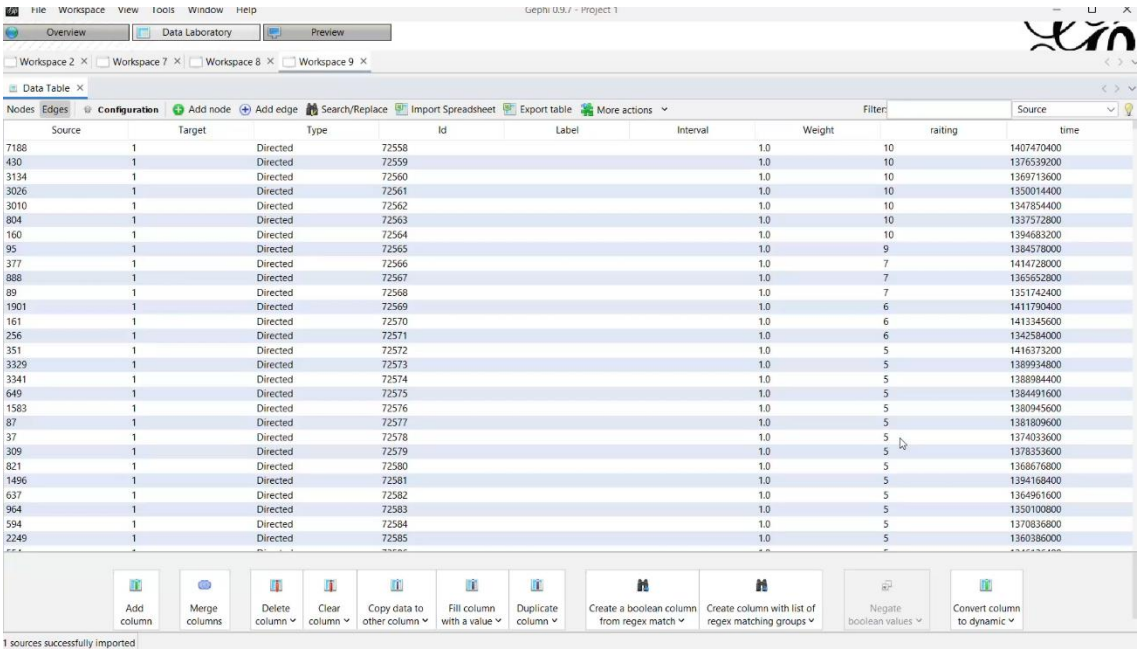
The goal of this project is to import the dataset into Gephi and visualize the data. This dataset is a who-trusts-whom network of people who trade using Bitcoin on a platform called Bitcoin Alpha. Since Bitcoin users are anonymous, there is a need to maintain a record of users' reputation to prevent transactions with fraudulent and risky users. So by the end of the project, one should be able to distinguish between a trustworthy and a fraud trader.
The main focus for this project is the big traders (big nodes) and to see whether or not they are trustworthy or not. For this purpose, a clear and easy graph must be created to make the analysis easy.

This report will show some changes being made to the graph step by step to reach a clear graph.
Unlike most datasets, this dataset focuses on edges instead of nodes, so the edges need to be distinguishable.
Since the question asked for closeness and in-betweenness calculations, this Report provides them however, these are attributes related to the nodes which we do not need.
At the end of this report some example analysis will be made on the biggest nodes in the graph to show how one can tell them apart.

3)Solutions:

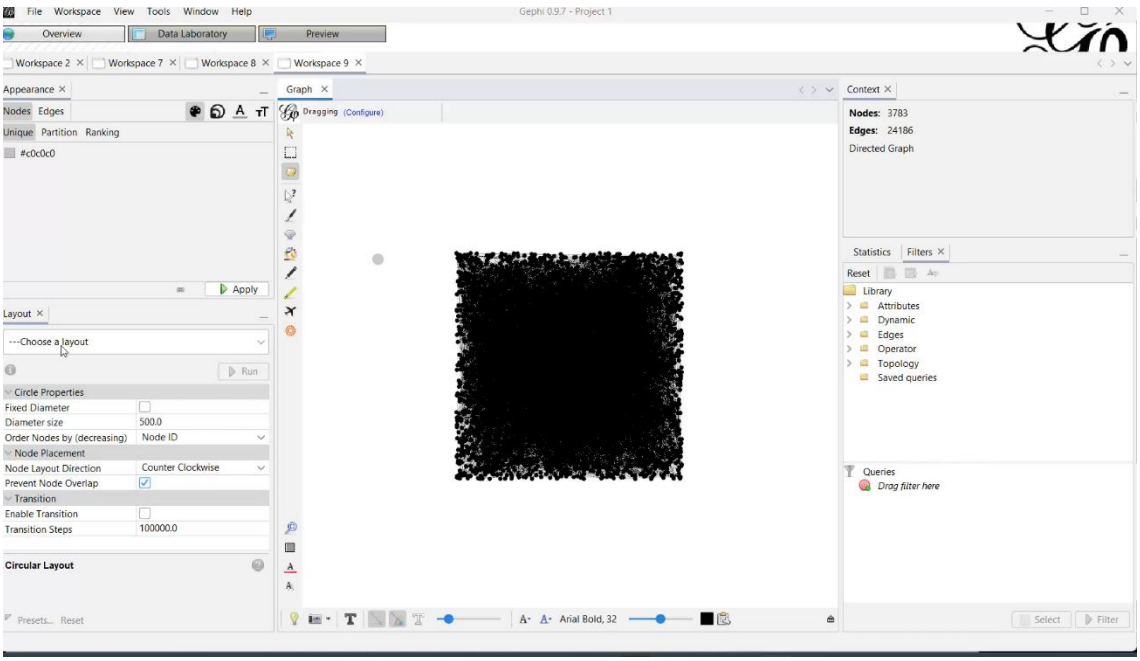
Importing the dataset to Gephi:



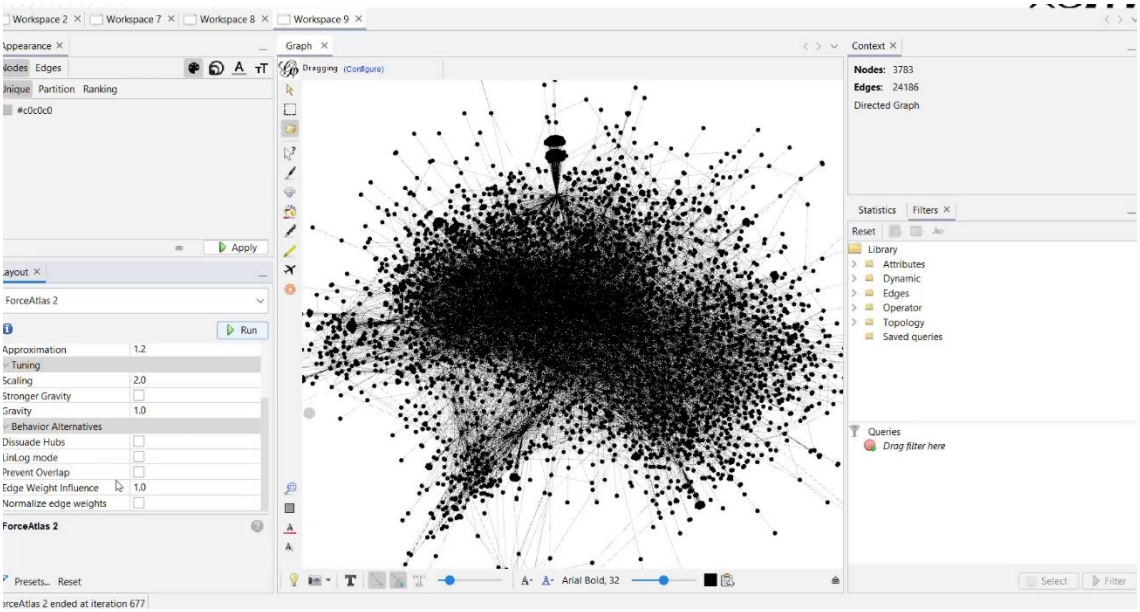


Our focus is the edges in our dataset. Shown above in the data library.

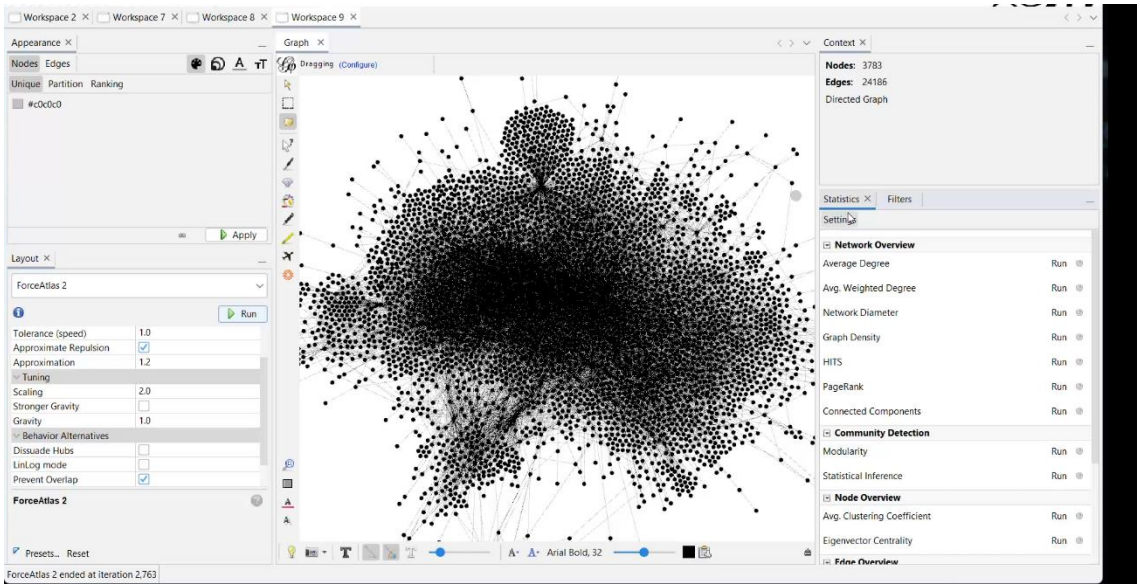
First look at the graph:



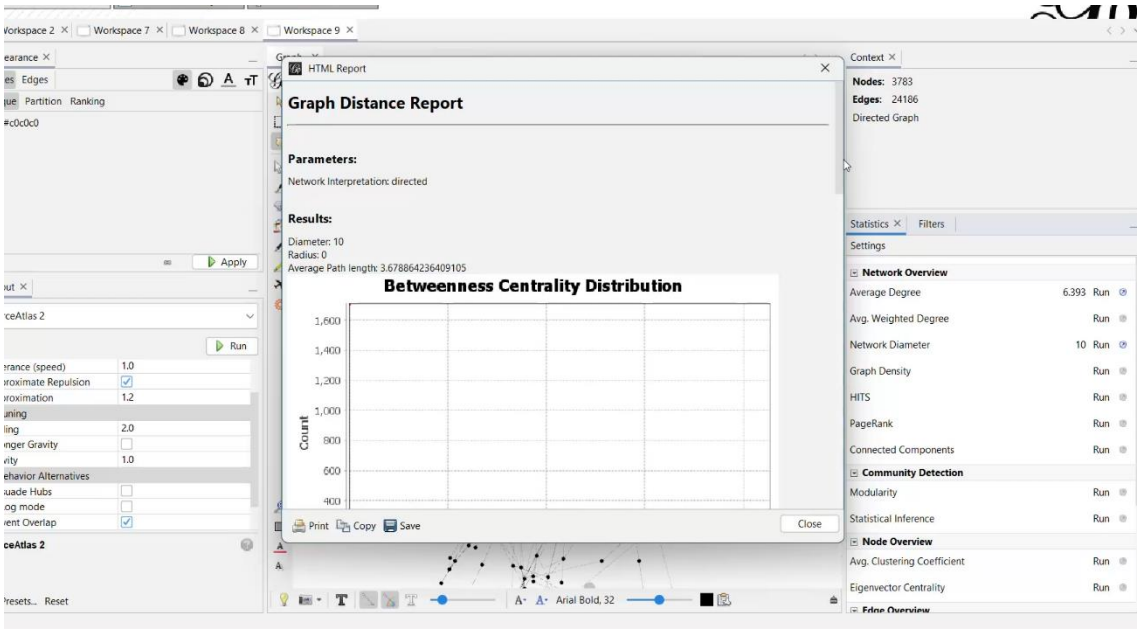
Using ForceAtlas2 layout:



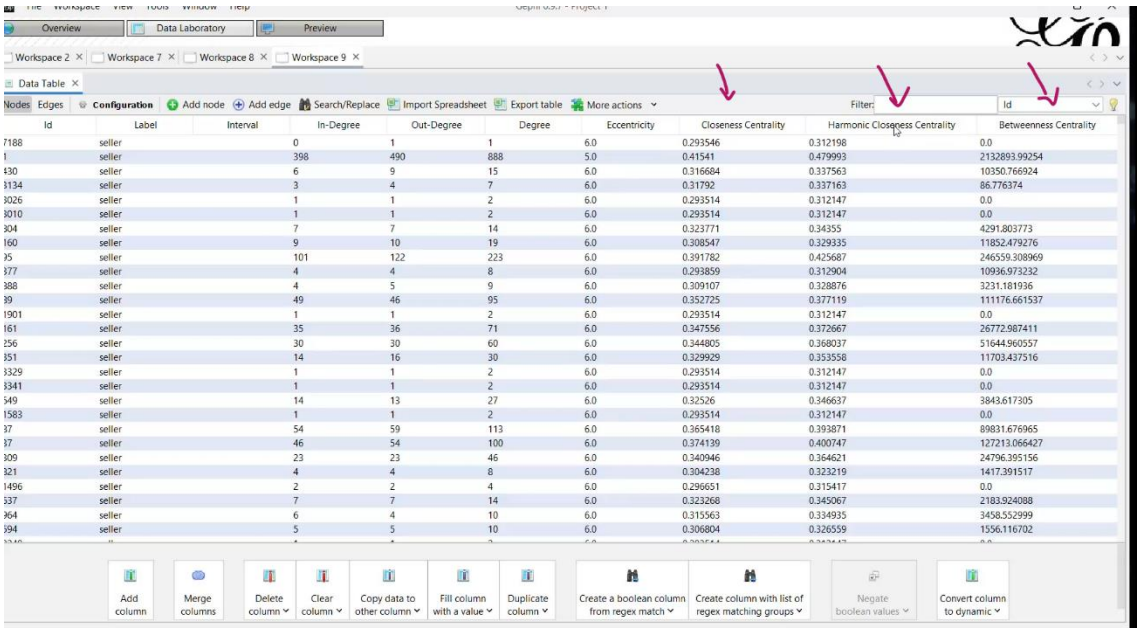
Prevent overlap:



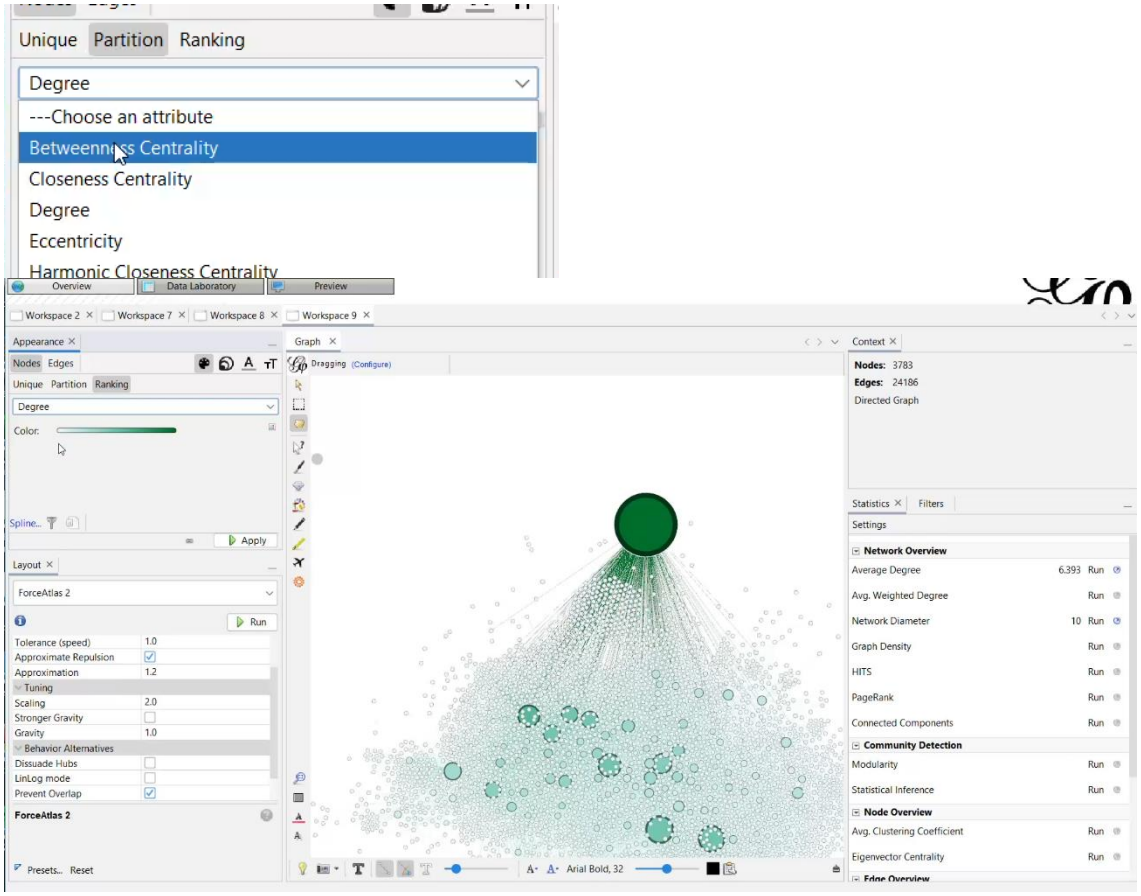
From statistics, run network diameter:



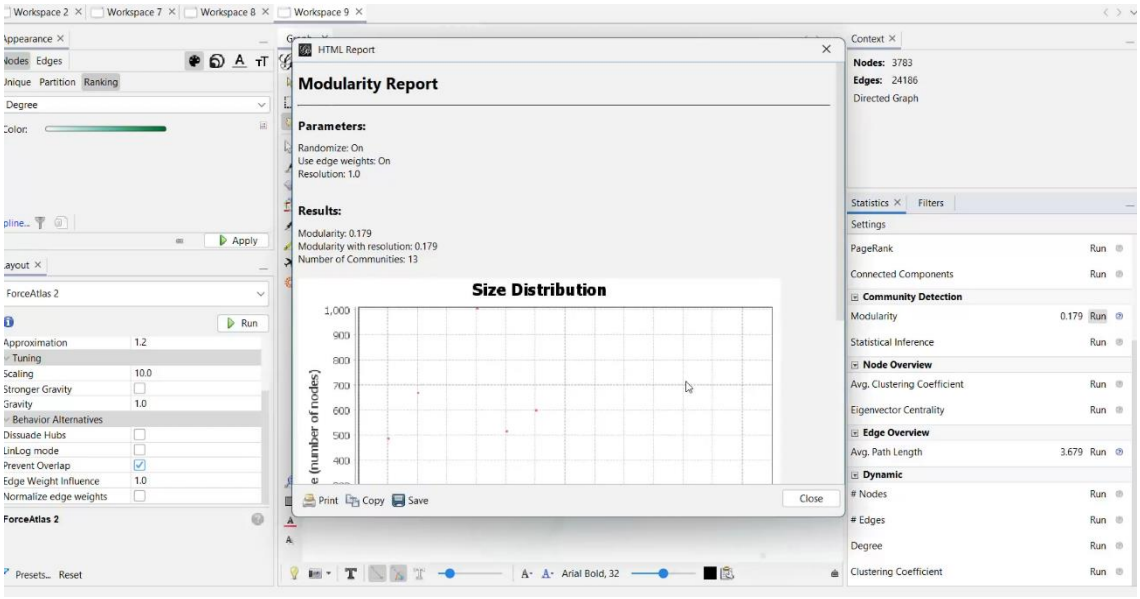
Now looking at the dataset it can be seen that the centrality of nodes including degree, betweenness, closeness has been calculated and added:



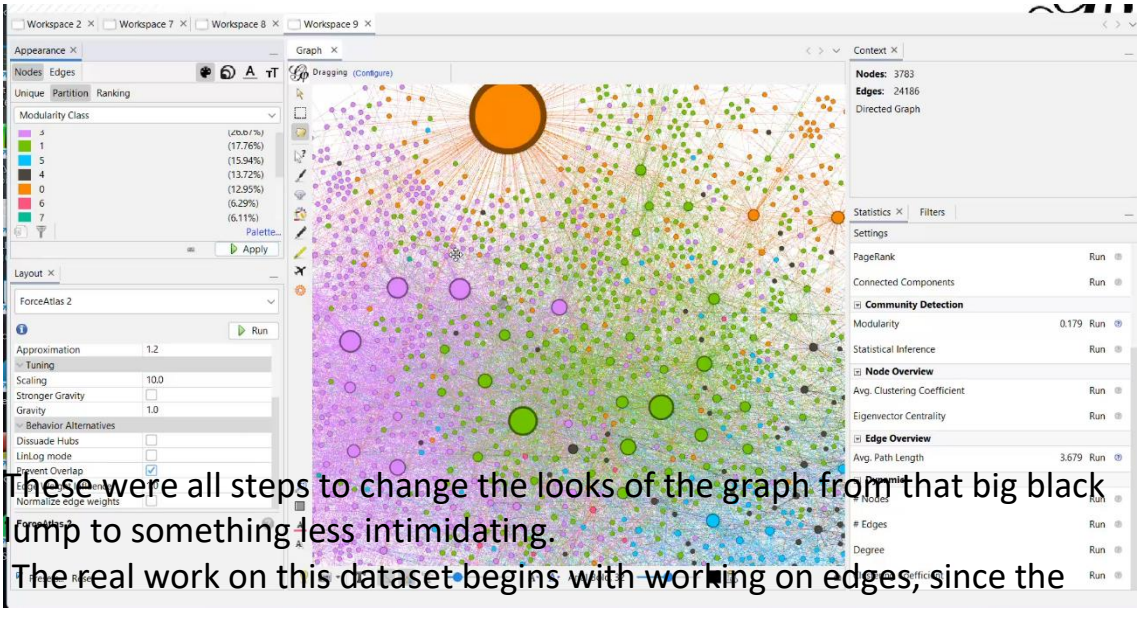
Changing the color of nodes based on betweenness centrality and make the scaling bigger so that the graph will expand (from 2.0 to 10.0):



Run community detection in the statistics section to get the modularity report to see different communities which was not a necessary step but it makes the graph more pleasing to the eyes:



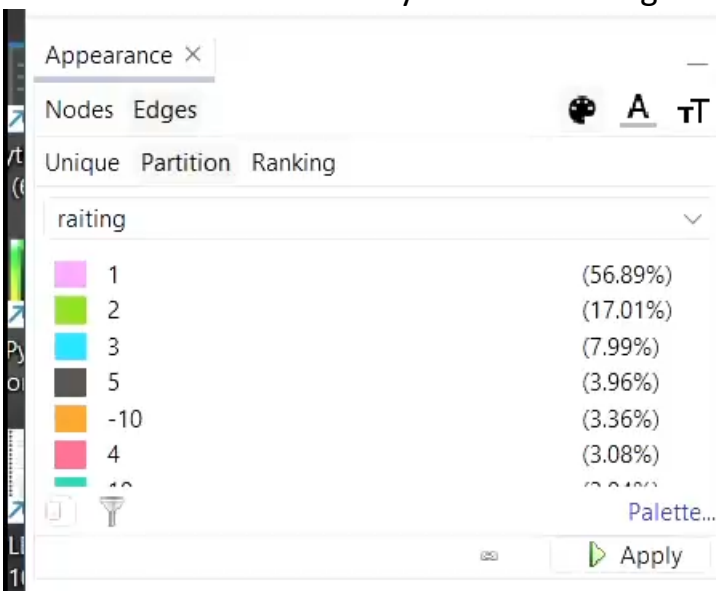
This is how the graph looks when the class color is a applied to the nodes:



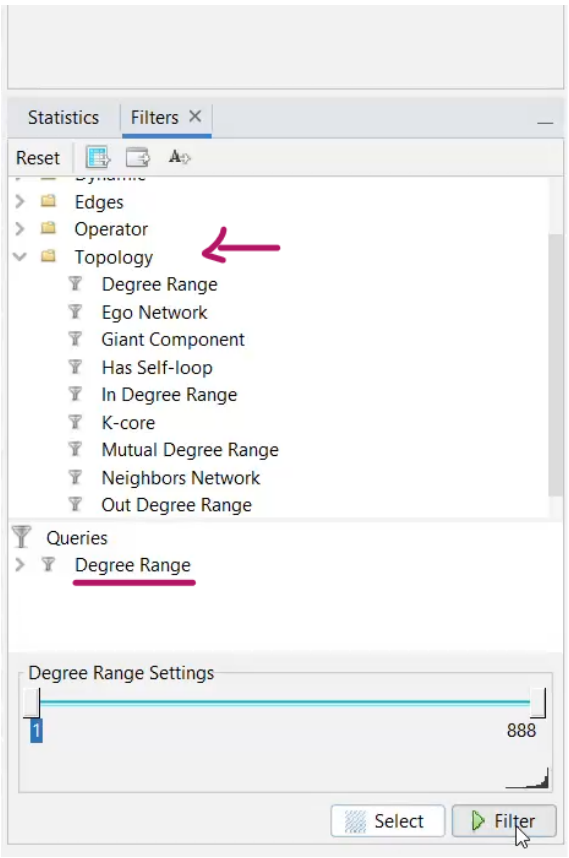
These were all steps to change the looks of the graph from that big black lump to something less intimidating.

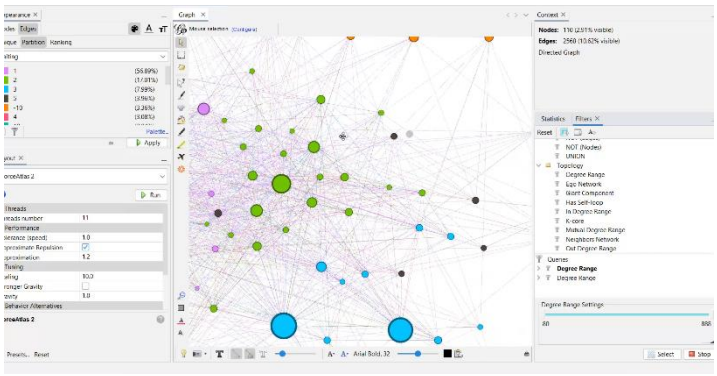
The real work on this dataset begins with working on edges, since the dataset has weighted edges, the best choice was to color code the rating of the edges. Edges have weights from -10 to +10, which is the rating of

the traders.
For example, from the picture below it can be seen that 56.89% of the edges in the graph have 1 rating, so not many are satisfied with trading with the nodes that have been rated 1. Which is shown by purple in the graph.
Orange means -10 which is the worst rating and green shows 10 which is the highest rating. By looking at the color and the number of the edges going to the nodes, which represent the traders, one can see how many people were satisfied with their service. By choosing the biggest nodes (which shows the ones with the most buyers) we choose the most popular traders and can see of they are fraud or legit.

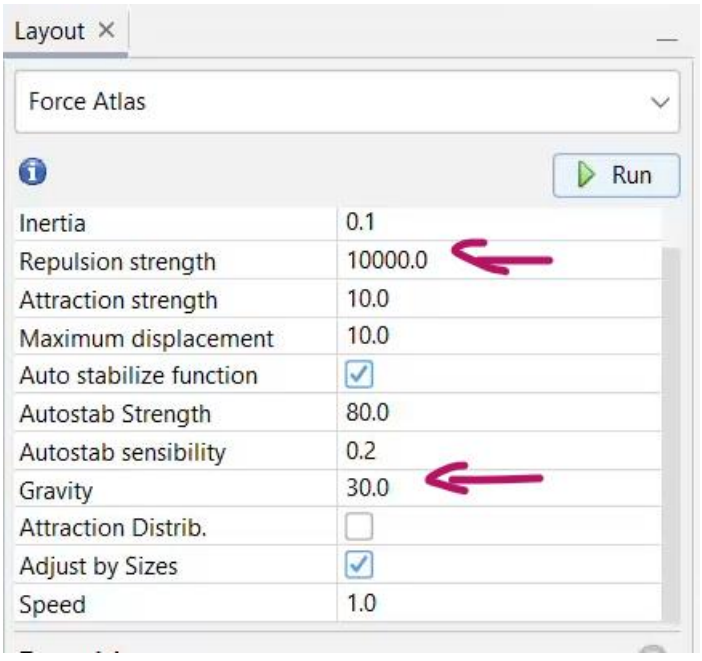


However, the number of edges is too high to see clearly. So in this next step by running topology and choosing degree range, some nodes can be deleted. For example nodes that have 80 edges or less.

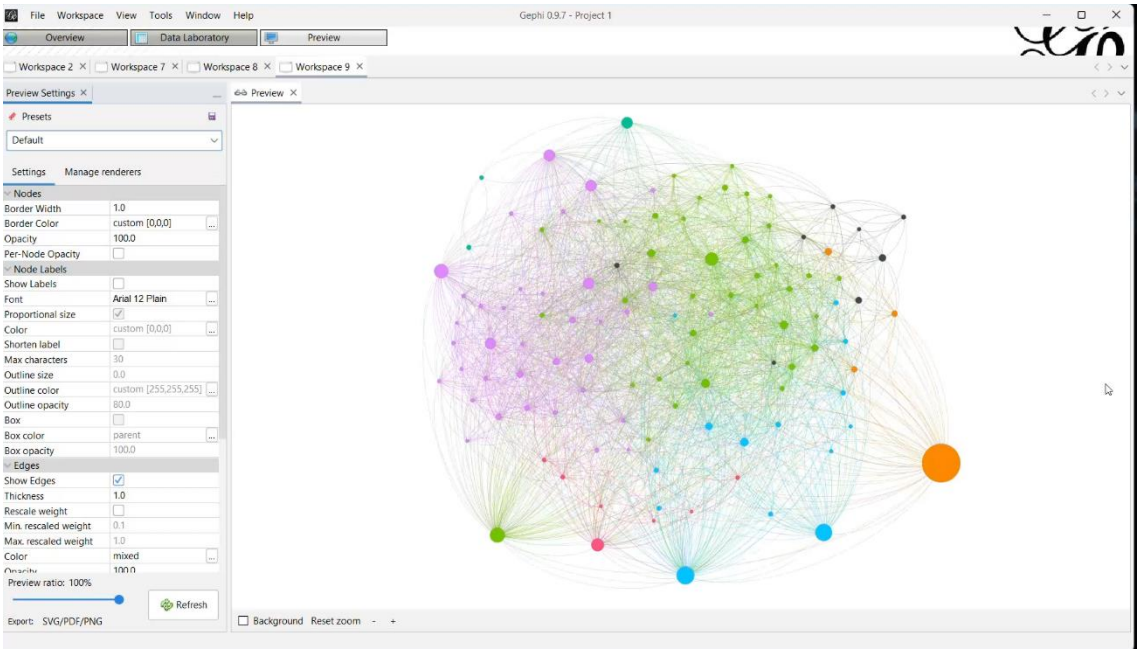




Change to ForceAtlas with characteristics shown below:



At the end we can look at the preview:



4)Findings:

There were many ways that this graph could be analyzed but in this project the focus was to look at the biggest traders and see if they are a fraud or can be trusted.

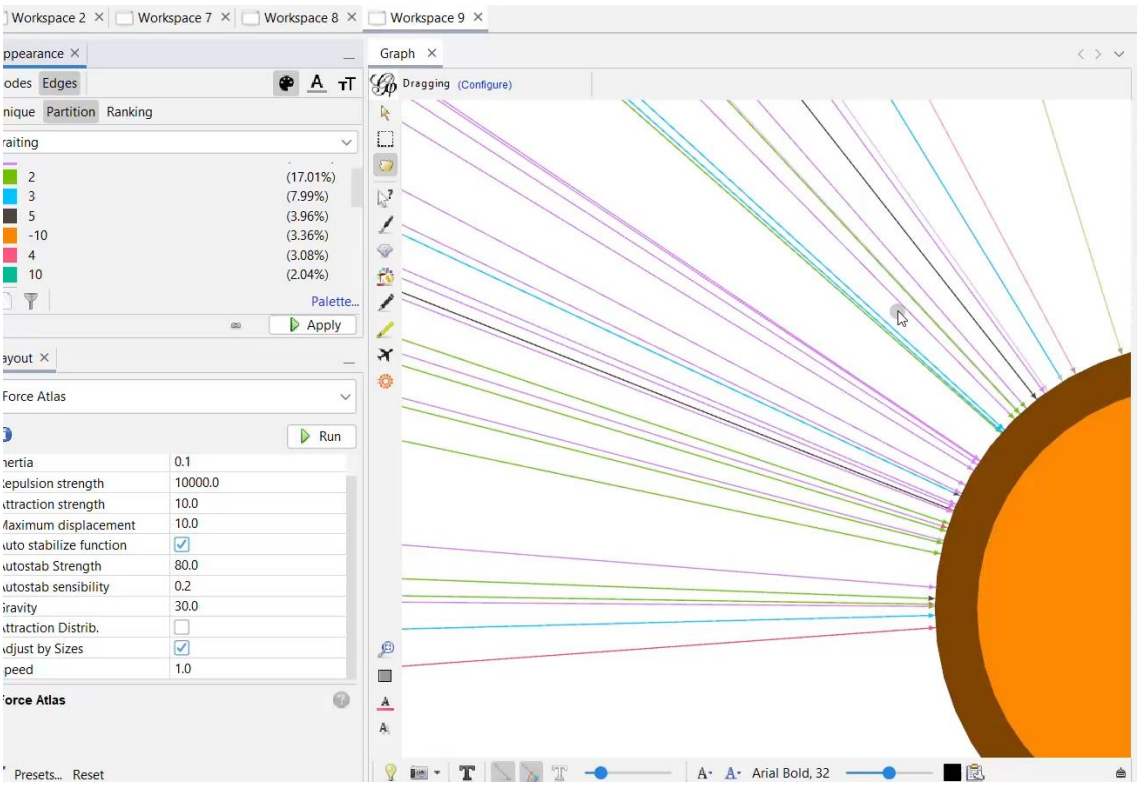
Now that the graph is a bit organized, analyzing the biggest nodes which represent the most popular traders and sellers is possible.

In this section 4 node will be analyzed:

Node1:

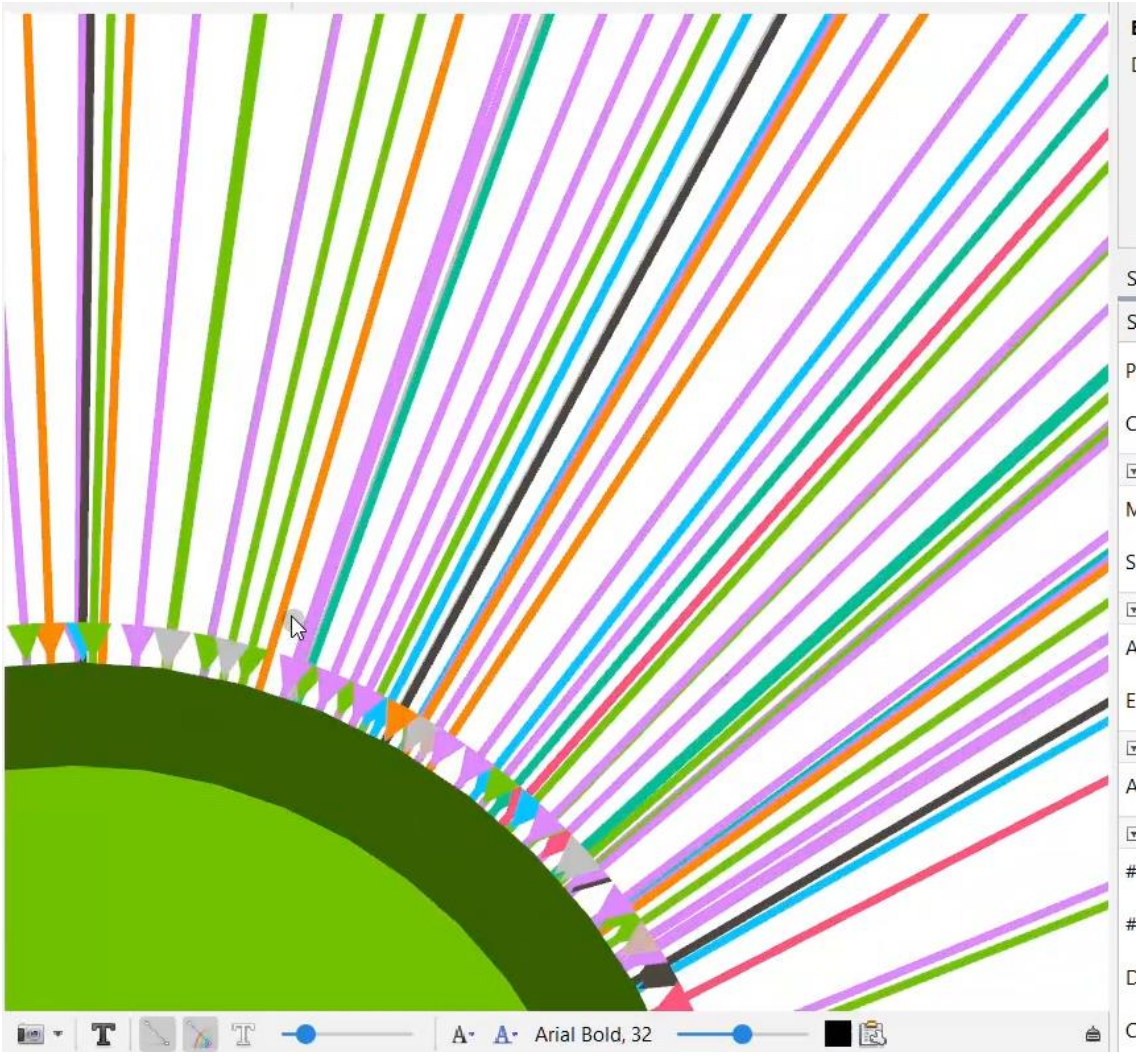
The largest node was dragged further so that we could take a closer look at its edges.

Not many orange edges (-10) going in this node. most buyers are somewhat satisfied with their trade. So, this node does not represent a fraud trader.


Workspace 2
Workspace 7
Workspace 8
Workspace 9
Edges
Partition
Ranking
2 (17.01%)
3 (7.99%)
5 (3.96%)
-10 (3.36%)
4 (3.08%)
10 (2.04%)
Palette...
Apply
Force Atlas
Run
0.1
10000.0
10.0
10.0
80.0
0.2
30.0
1.0
Presets...
Reset
Graph
Dragging
(Configure)
Arial Bold, 32

Node2:

mostly gave the rating 1. Most people are happy about this one but there is still orange which shows complete dissatisfaction but mostly positive so this node is most likely not a fraud.



Node3:

This node has the most orange edges, so a lot of people gave -10. This trader is not to be trusted as is most likely a fraud.