SOCIAL NETWORK ANALYSIS

Github Social Network

INTRODUCTION

A large social network of GitHub developers which was collected from the public API in June 2019. Nodes are developers who have starred at least 10 repositories and edges are mutual follower relationships between them. The vertex features are extracted based on the location, repositories starred, employer and e-mail address. The task related to the graph is binary node classification - one has to predict whether the GitHub user is a web or a machine learning developer.

Dataset Link

Gephi software was used for the visualization and analysis of this dataset.

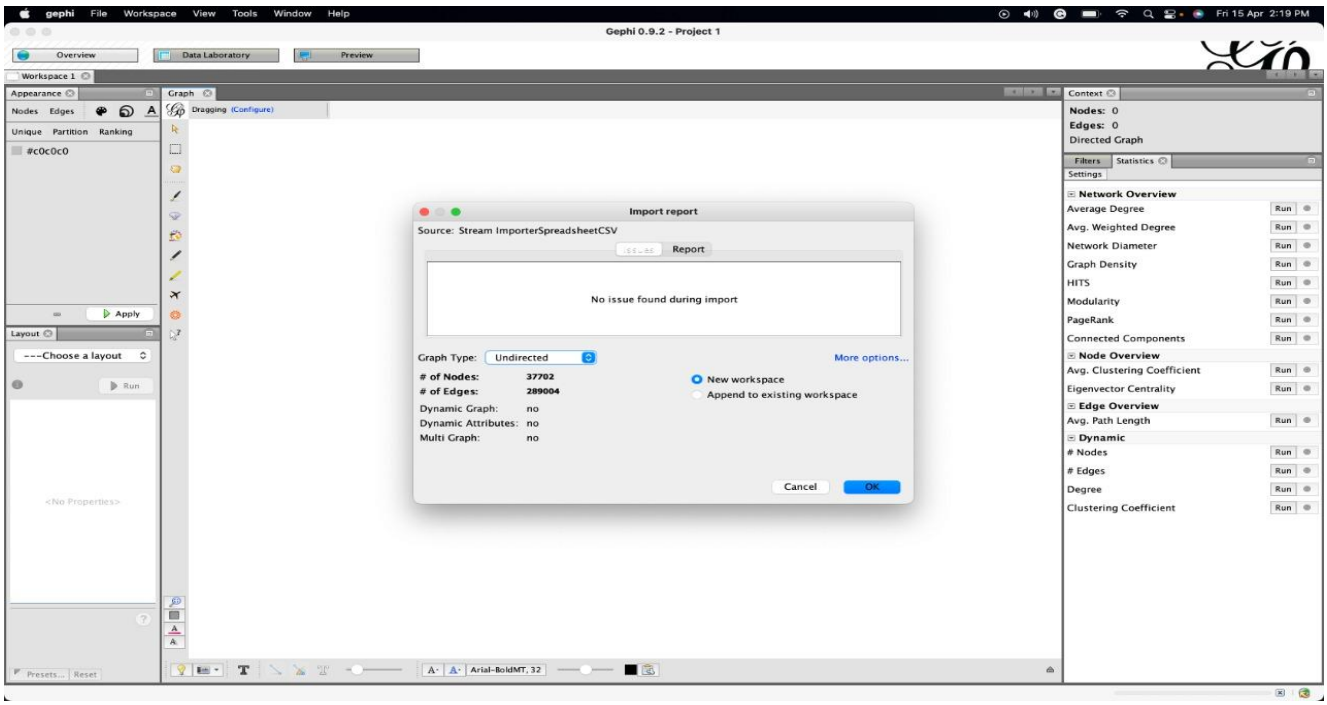
The following steps show the procedure followed for the analysis and interpretation.

Step 1:

The .gml file for the initial data cluster was imported into Gephi. The number of nodes and edges were mentioned in the following window.

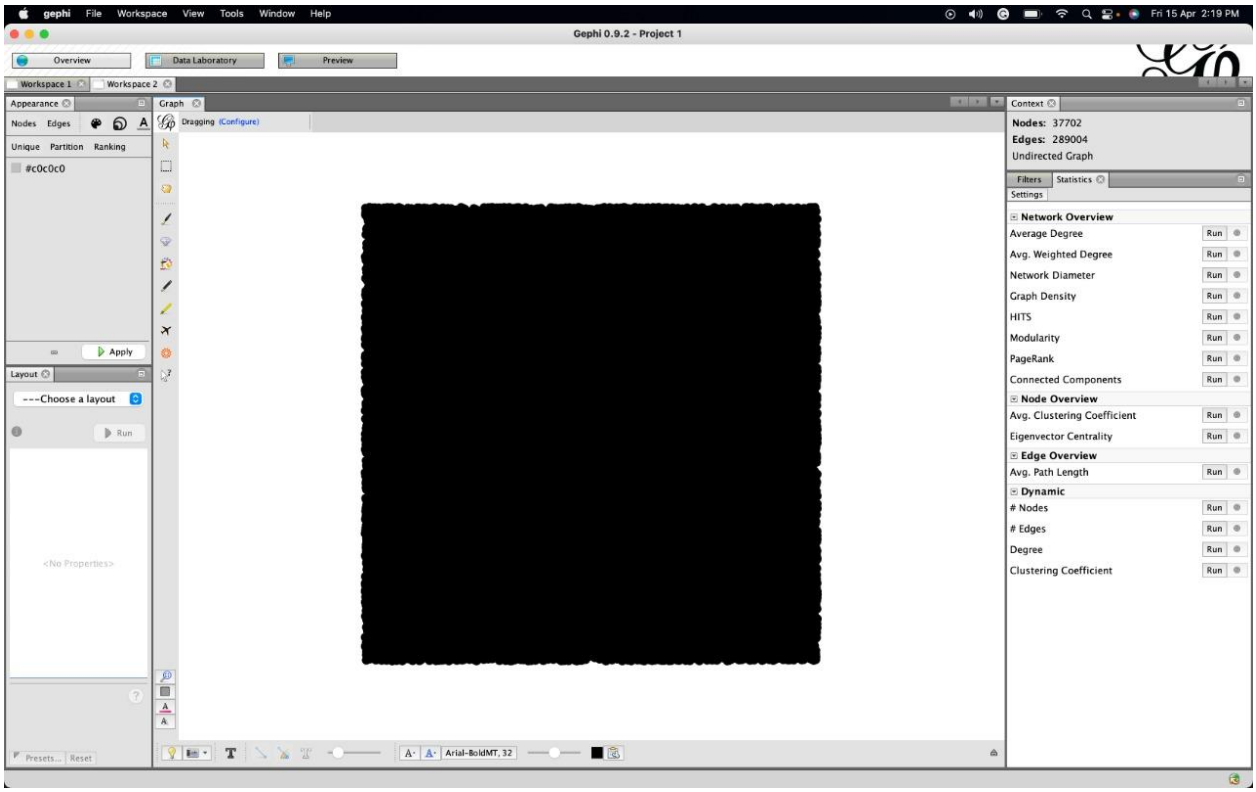
Number of Nodes- 37702

Number of Edges - 289004



Step 2:

Initially, the data cluster looked like a square with multiple nodes and edges through which no interpretation and analysis could be done.



Step 3:

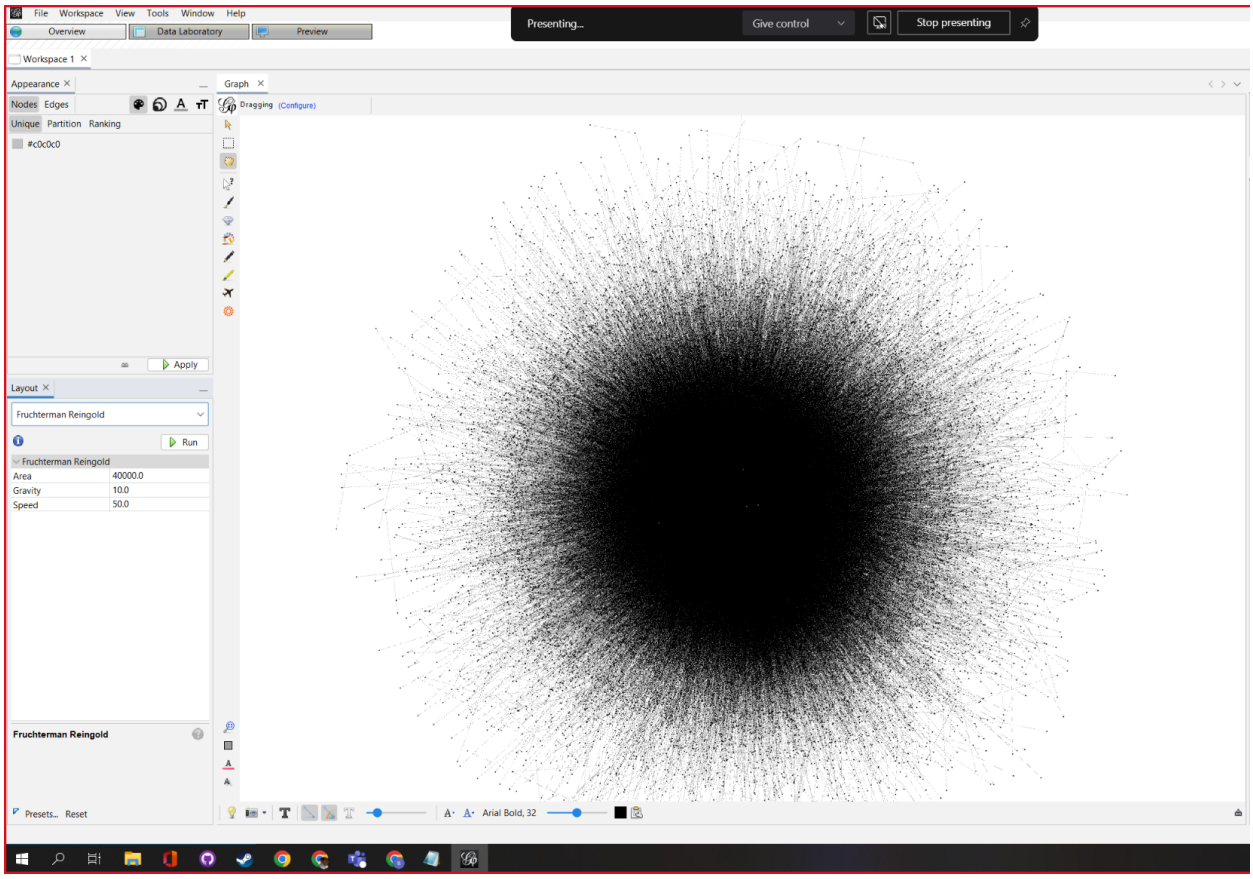
In this step, Fuchterman Reingold was selected in the Layout option. The parameter were set as-

Area- 40000

Speed- 50

Gravity- 10

The clusters rearranged themselves in a manner as shown in the picture below-



Step 4:

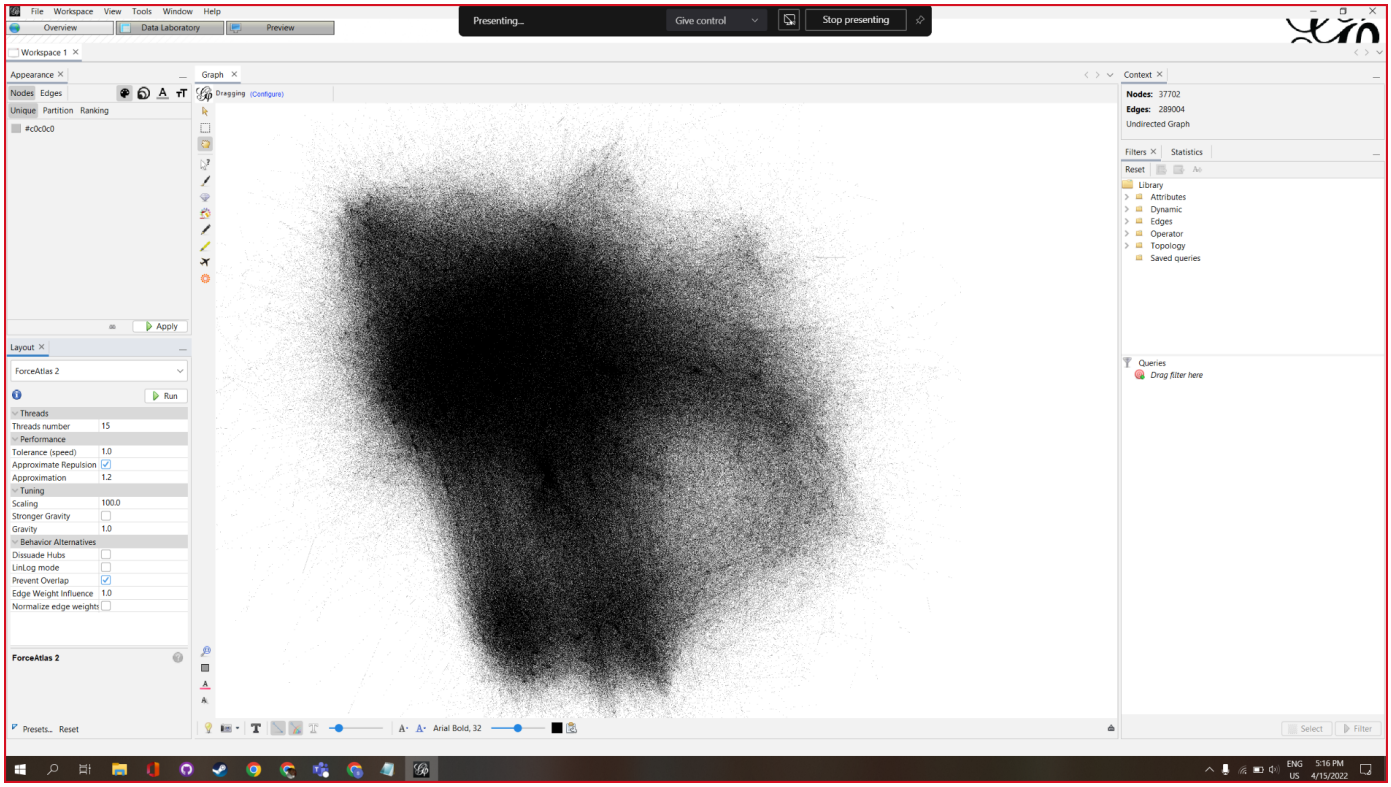
In this step, Force Atlas 2 was selected in the Layout option. The parameter were set as-

Tolerance Speed- 1.0

Scaling- 100

Prevent Overlap- ✅

The clusters rearranged themselves in a manner as shown in the picture below- (run for 5 mins)



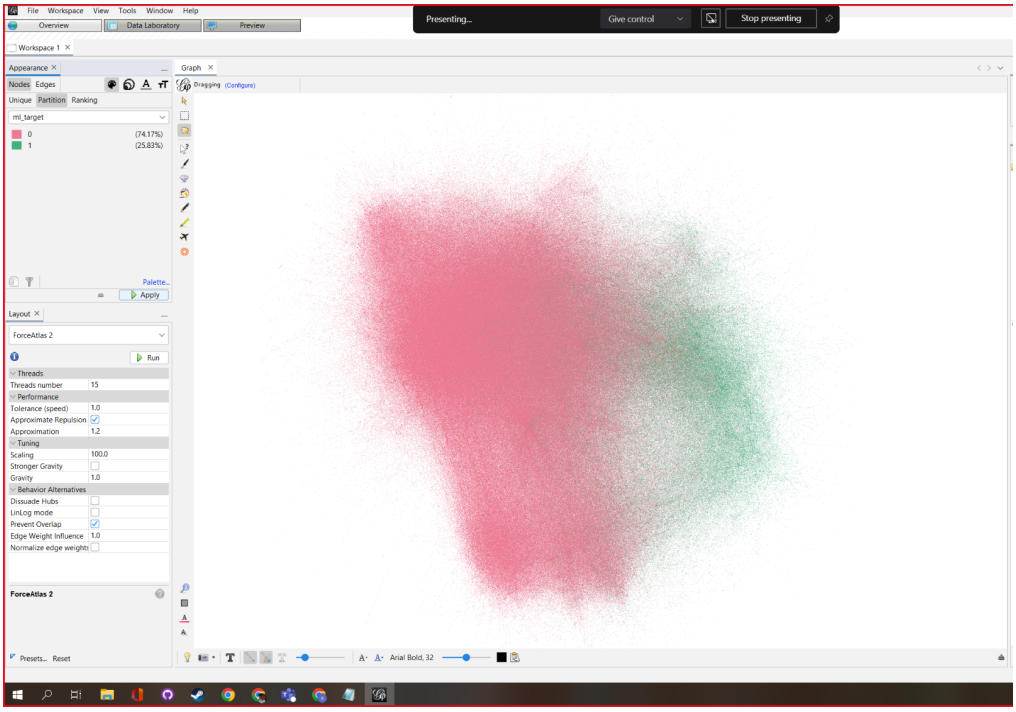
Step 5:

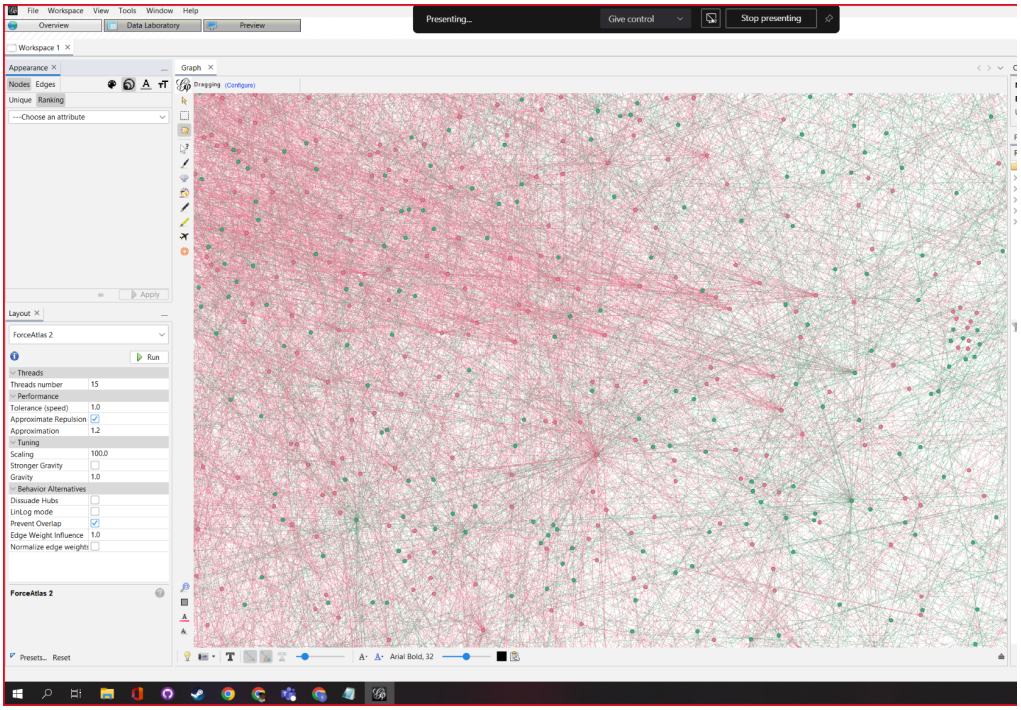
In this step, we tried to differentiate between the github users who are machine learning developers or not.

Nodes – Color Palette – Partition – “ml target” – Apply

Observation- Pink Cluster- Non- Machine Learning Developer

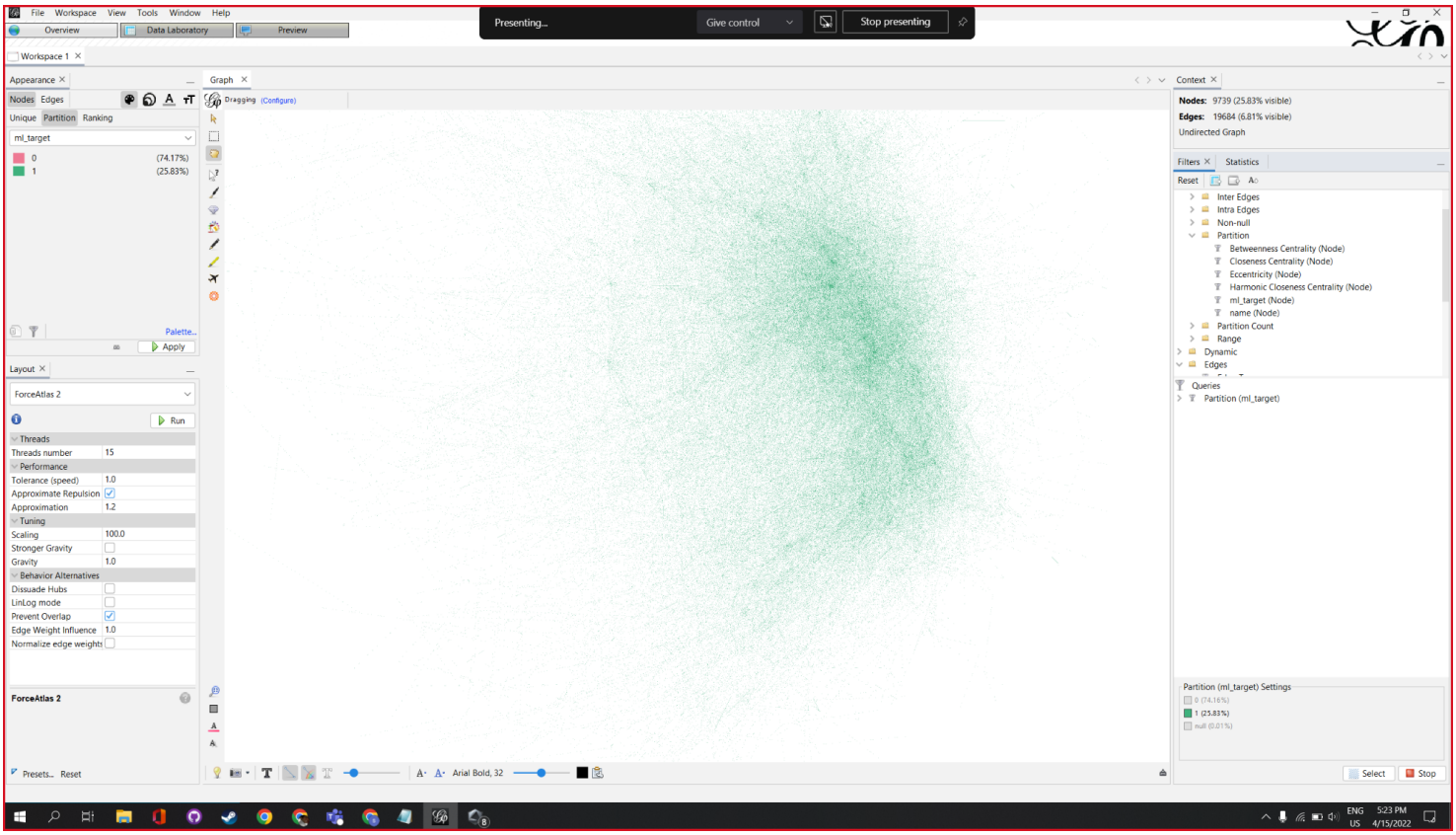
Green Cluster- Machine Learning Developer





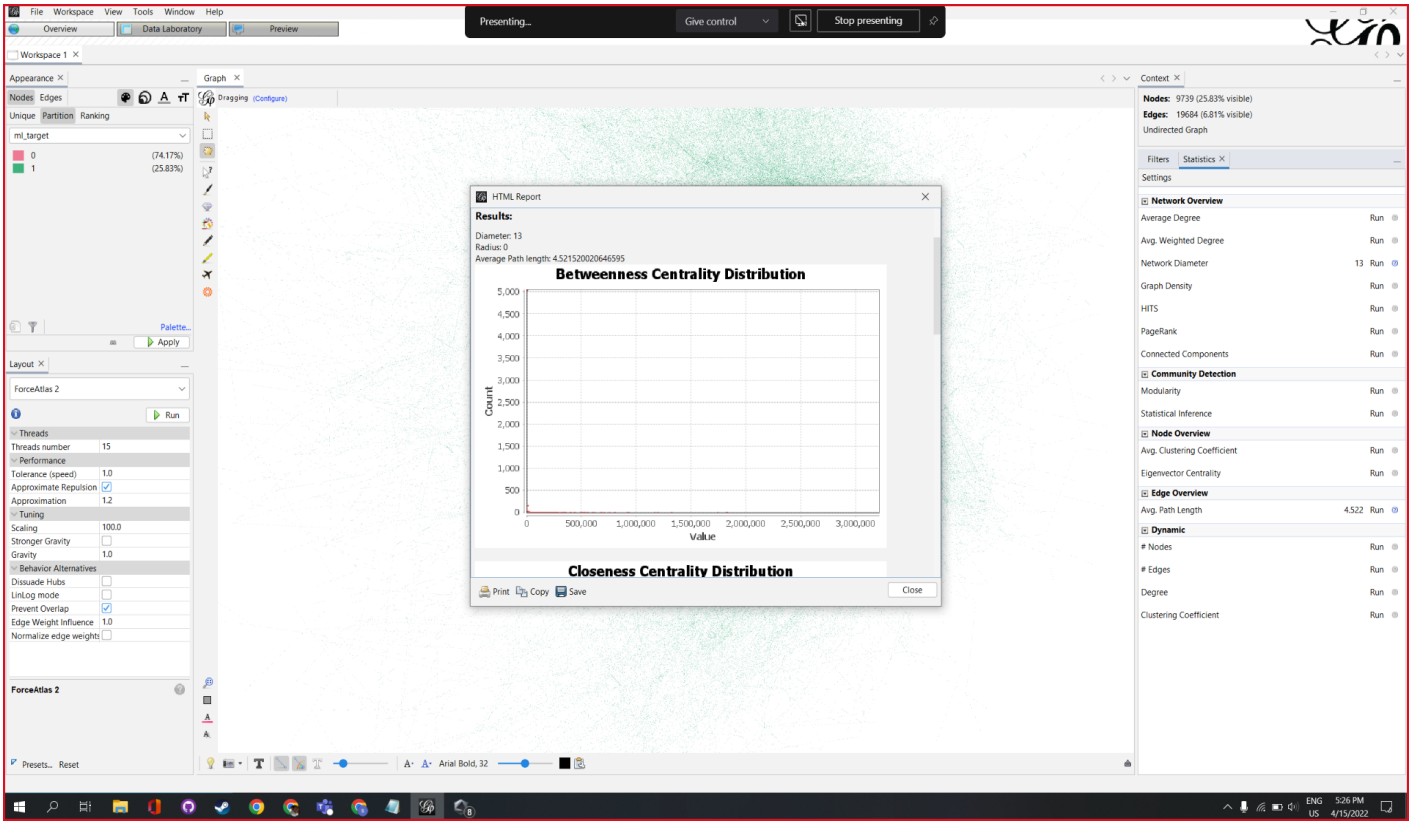
Step 6:

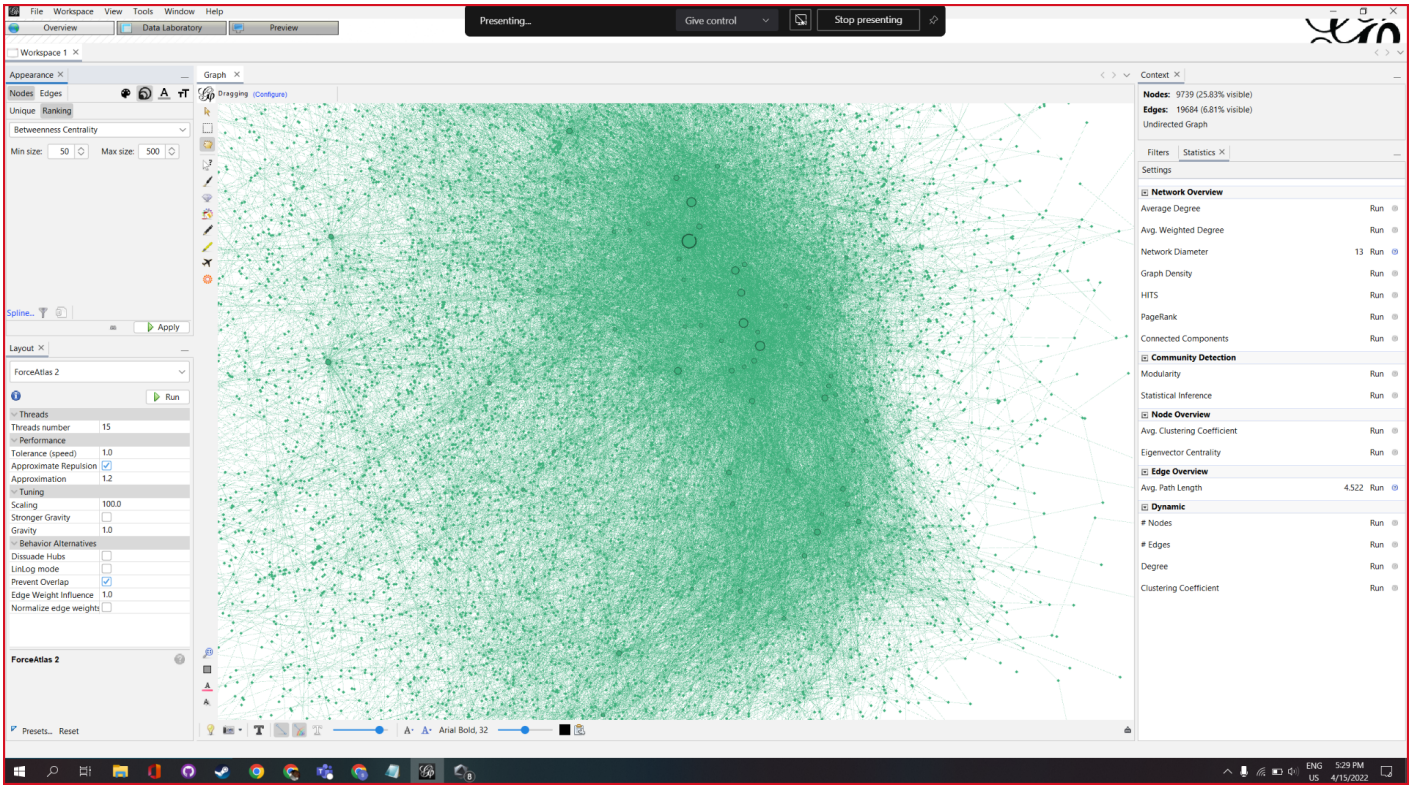
In this step, we applied a filter to only view the green cluster, i.e., the Machine Learning developers.



Step 7:

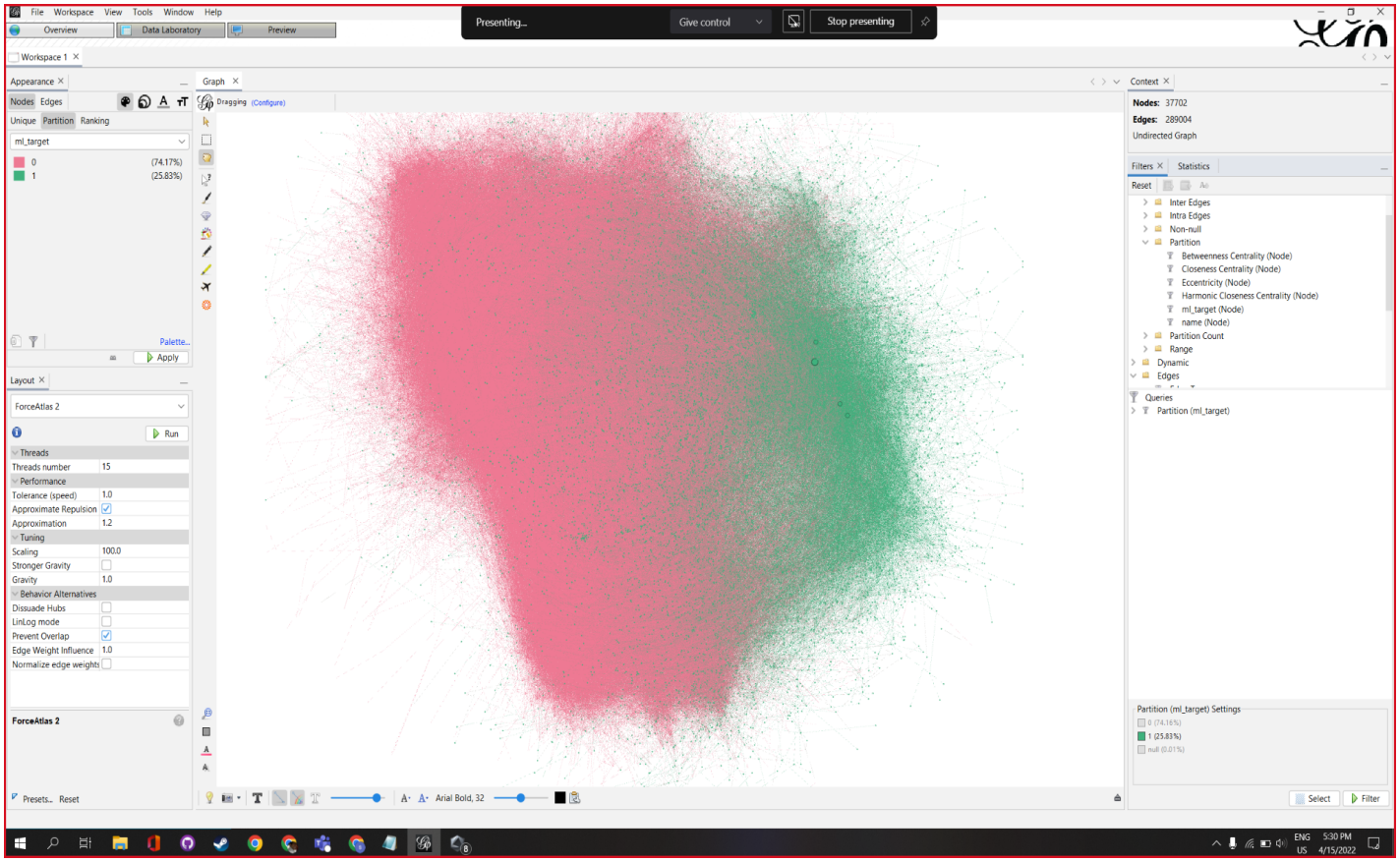
In this step, we calculated the centrality measure, and the centrality measure in size. Also, made the filter and then calculated





Step 8:

Removed filter on “ml_target” to view both, Machine Learning developers and Non-Machine Learning developers.



Step 9:

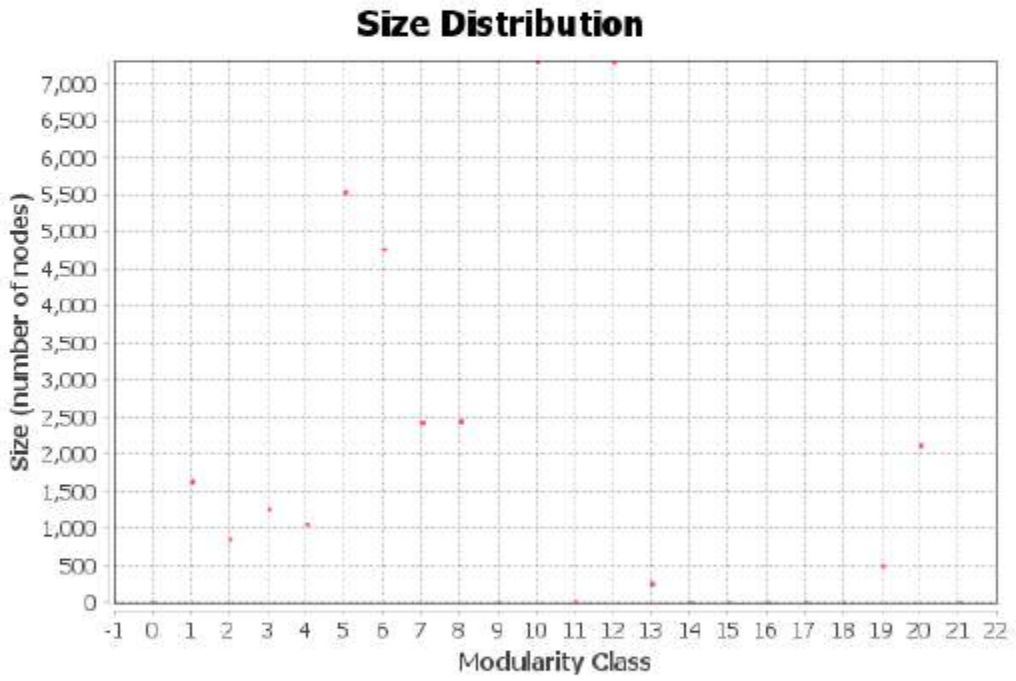
Modularity Report

Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

Modularity: 0.458
Modularity with resolution: 0.458
Number of Communities: 22



Algorithm:

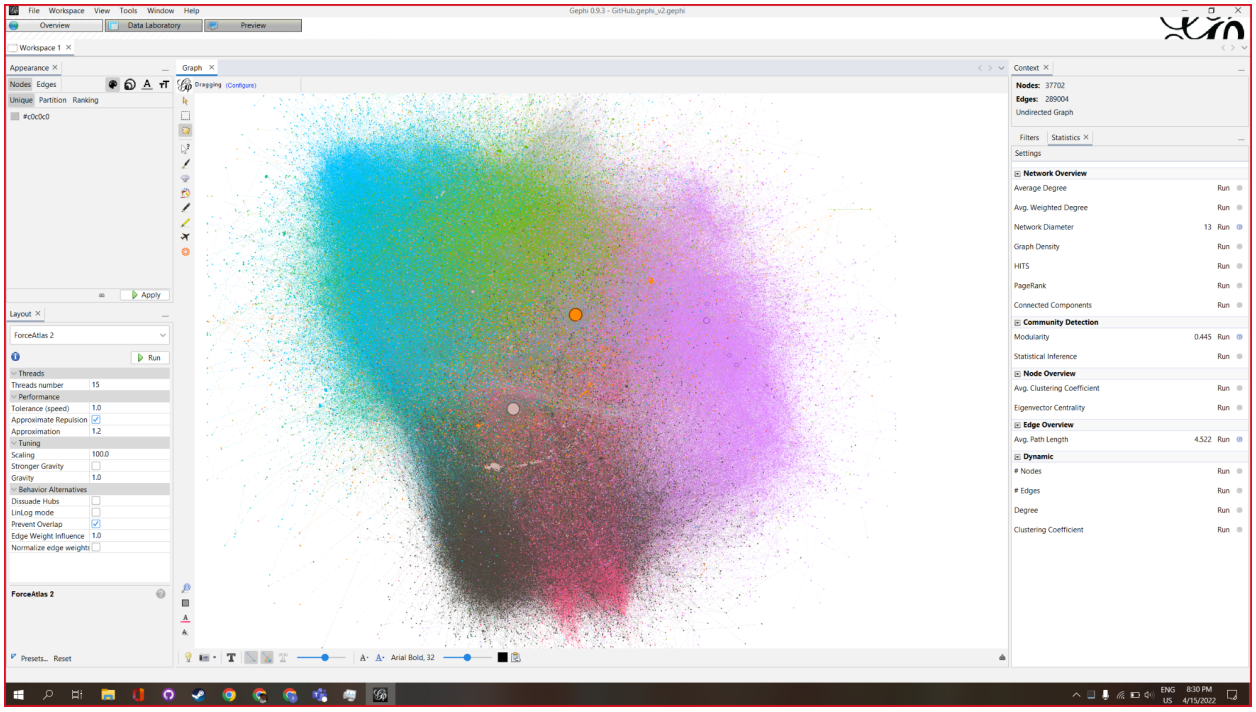
Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, *Fast unfolding of communities in large networks*, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

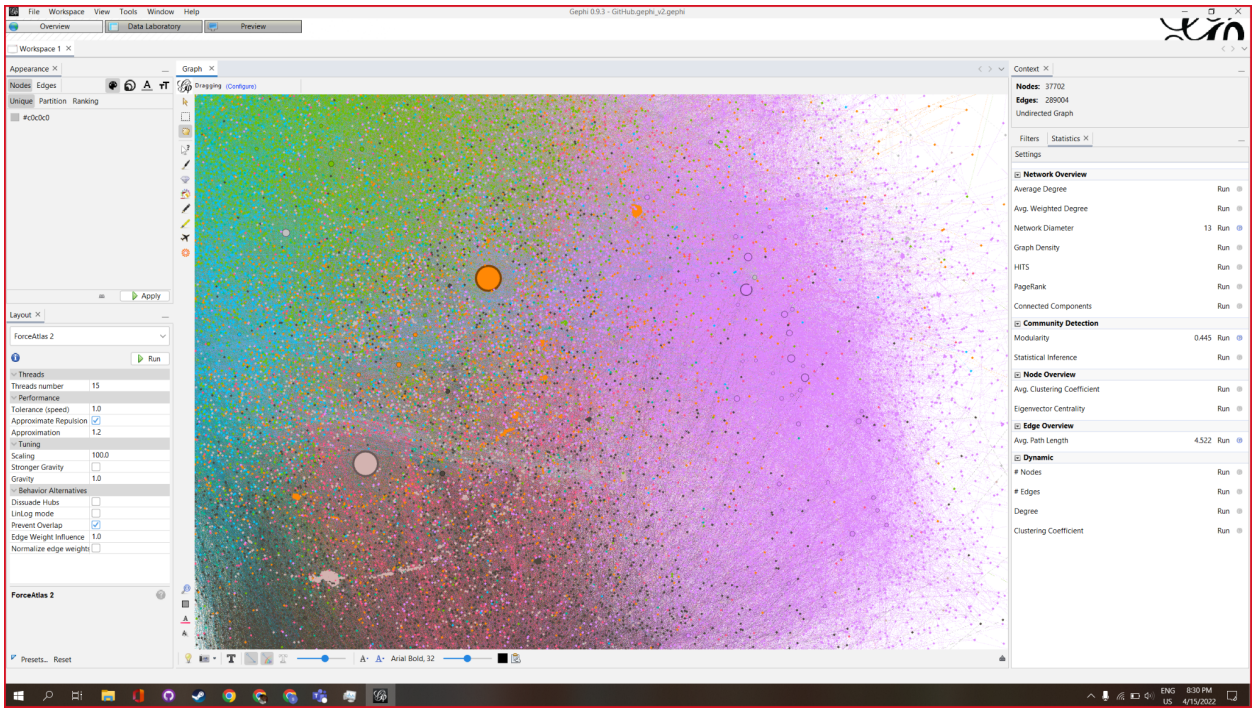
Resolution:

R. Lambiotte, J.-C. Delvenne, M. Barahona *Laplacian Dynamics and Multiscale Modular Structure in Networks 2009*

Step 10:

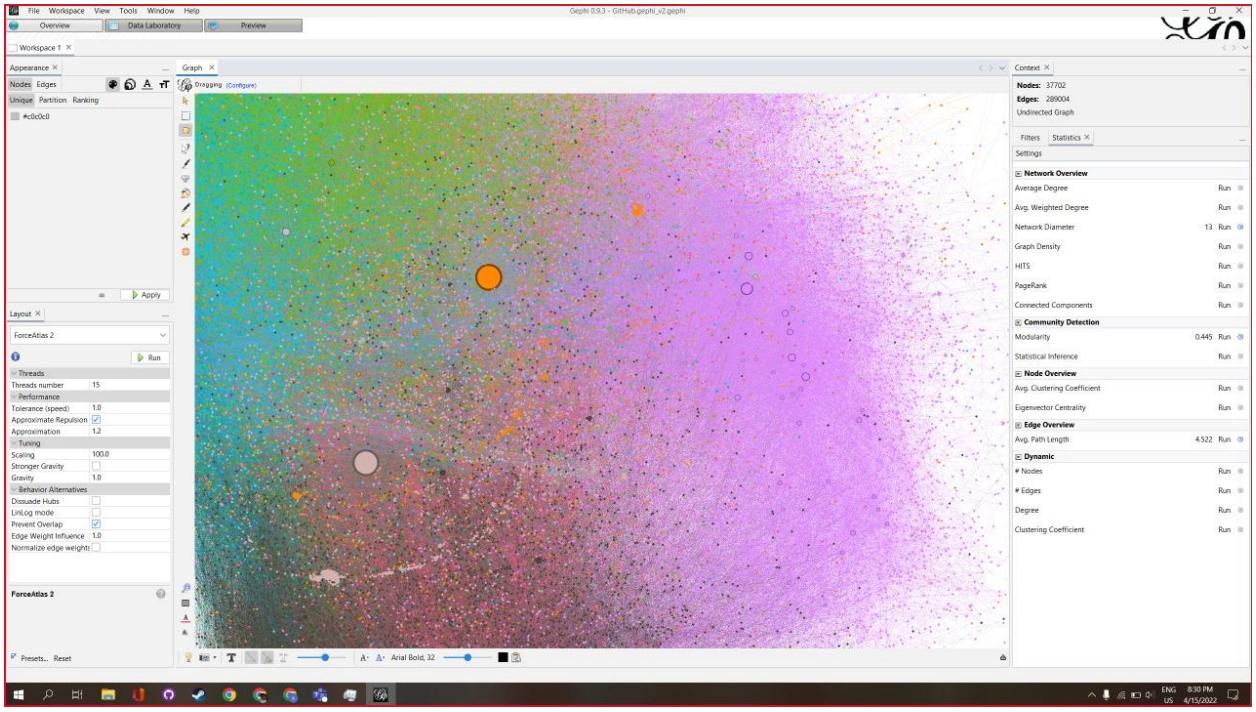
Colour as Modularity classes, size as betweenness of ML nodes

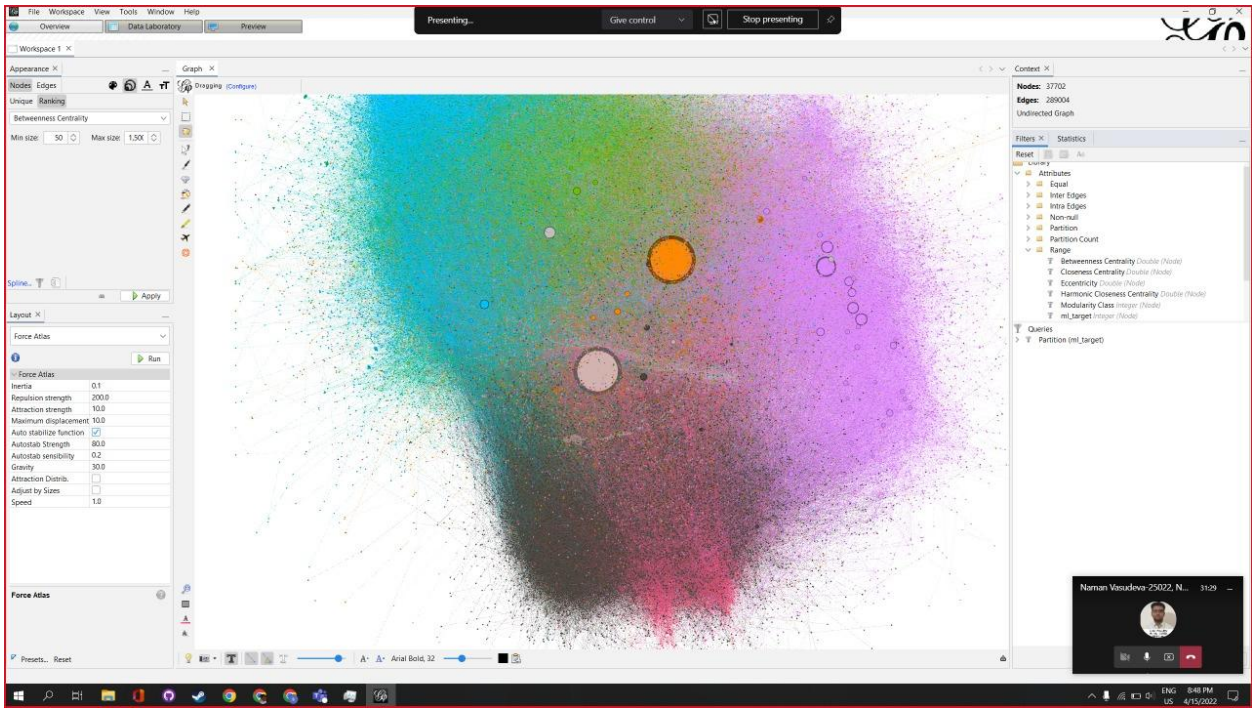




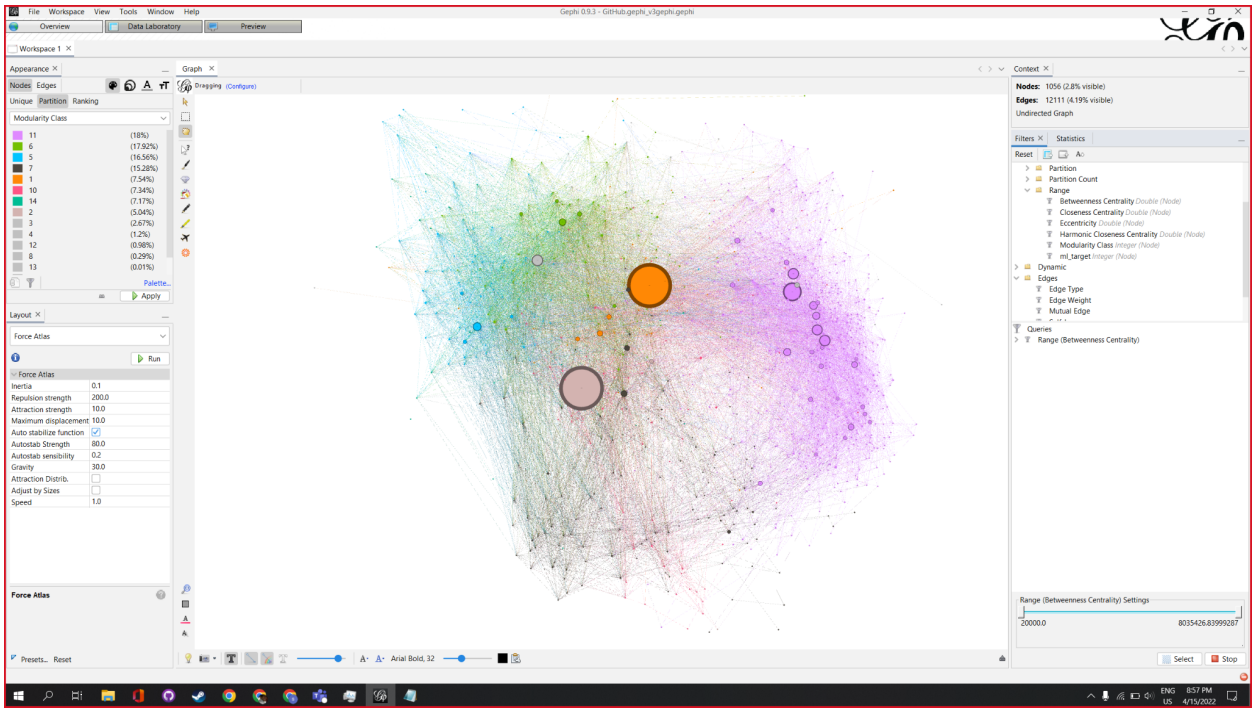
Step 11:

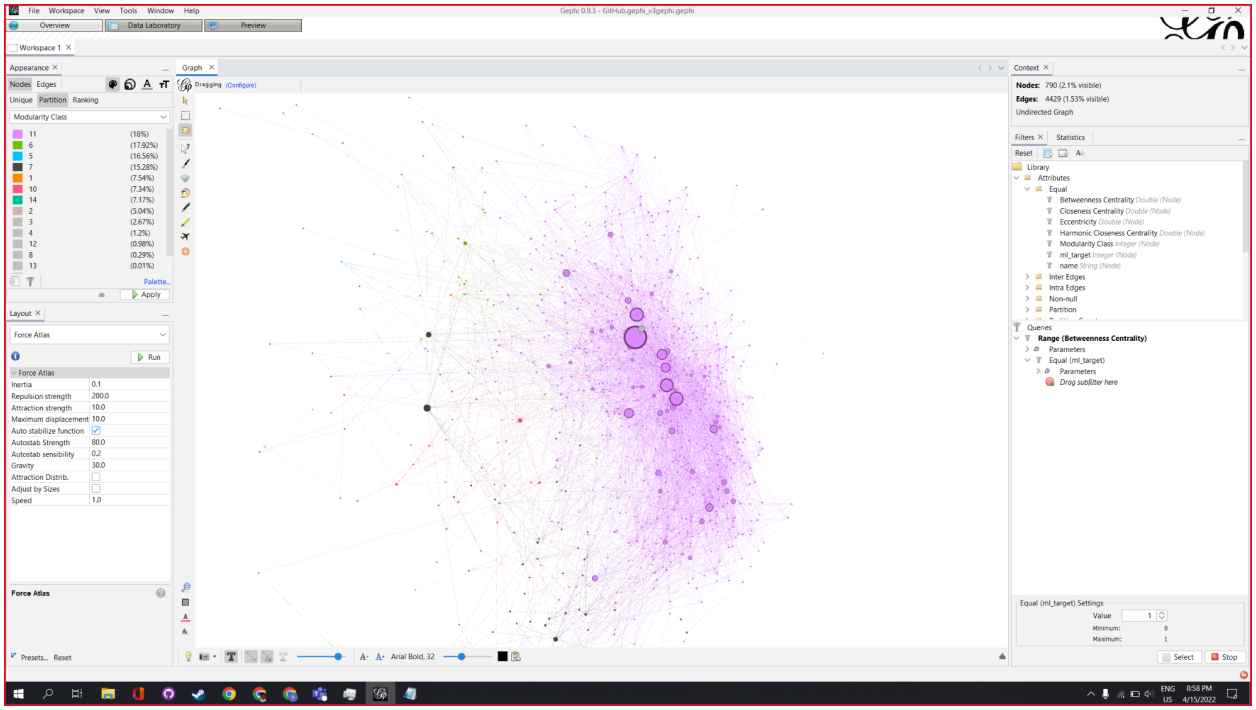
Calculated betweenness of the whole data set in this step.

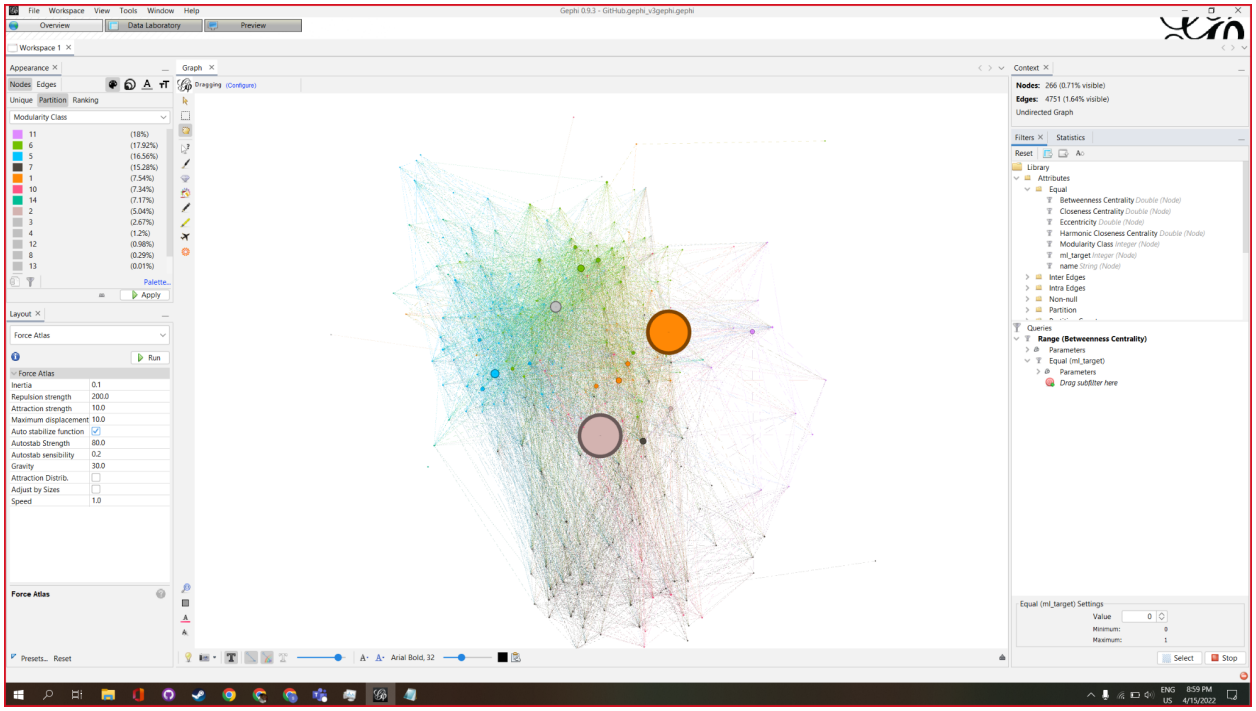


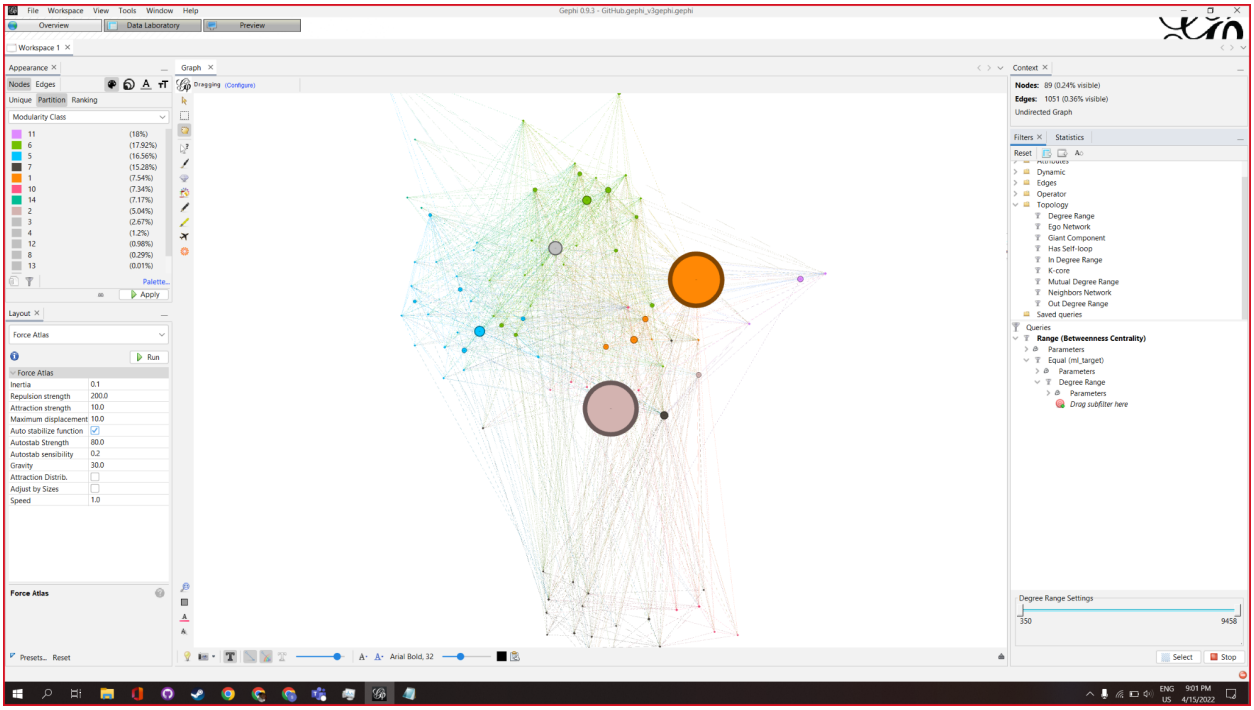


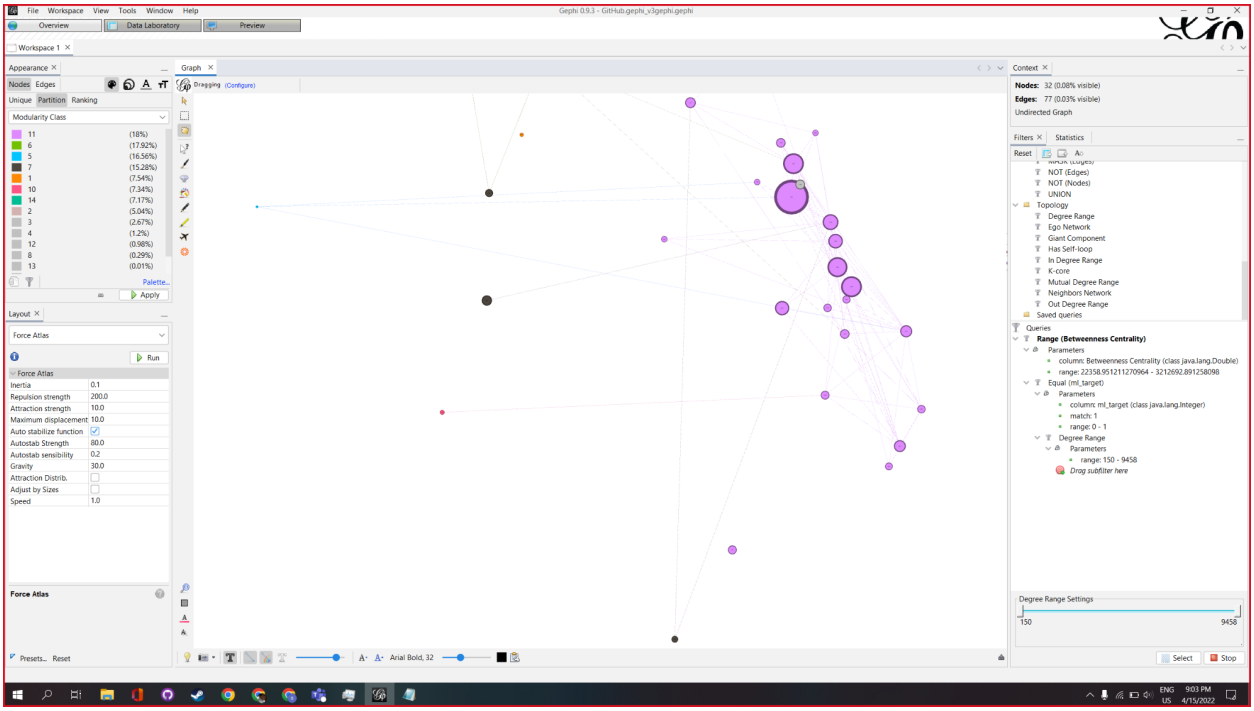
Filters to explore more




File Workspace View Tools Window Help
Gephi 0.9.3 - GitHub.gephi_v3gephi.gephi
Overview
Data Laboratory
Preview
Workspace 1
Appearance
Nodes Edges
Unique Partition Ranking
Modularity Class
11 (18%)
6 (17.92%)
5 (16.56%)
7 (15.28%)
1 (7.54%)
10 (7.34%)
14 (7.17%)
2 (5.04%)
3 (2.67%)
4 (1.2%)
12 (0.98%)
8 (0.29%)
13 (0.01%)
Palette...
Apply
Layout
Force Atlas
Run
Force Atlas
Inertia 0.1
Repulsion strength 200.0
Attraction strength 10.0
Maximum displacement 10.0
Auto stabilize function
Autostab Strength 80.0
Autostab sensibility 0.2
Gravity 30.0
Attraction Distrib.
Adjust by Sizes
Speed 1.0
Force Atlas
Presets... Reset
Graph
Dragging (Configure)
Arial Bold, 32
Context
Nodes: 790 (2.1% visible)
Edges: 4429 (1.53% visible)
Undirected Graph
Filters
Statistics
Reset
Library
Attributes
Equal
Betweenness Centrality Double (Node)
Closeness Centrality Double (Node)
Eccentricity Double (Node)
Harmonic Closeness Centrality Double (Node)
Modularity Class Integer (Node)
ml_target Integer (Node)
name String (Node)
Inter Edges
Intra Edges
Non-null
Partition
Queries
Range (Betweenness Centrality)
Parameters
Equal (ml_target)
Parameters
Drag subfilter here
Equal (ml_target) Settings
Value 1
Minimum: 0
Maximum: 1
Select
Stop
ENG US
8:58 PM
4/15/2022


File Workspace View Tools Window Help
Gephi 0.9.3 - GitHub.gephi_v3gephi.gephi
Overview
Data Laboratory
Preview
Workspace 1
Appearance
Nodes Edges
Unique Partition Ranking
Modularity Class
11 (18%)
6 (17.92%)
5 (16.56%)
7 (15.28%)
1 (7.54%)
10 (7.34%)
14 (7.17%)
2 (5.04%)
3 (2.67%)
4 (1.2%)
12 (0.98%)
8 (0.29%)
13 (0.01%)
Palette...
Apply
Layout
Force Atlas
Run
Force Atlas
Inertia 0.1
Repulsion strength 200.0
Attraction strength 10.0
Maximum displacement 10.0
Auto stabilize function
Autostab Strength 80.0
Autostab sensibility 0.2
Gravity 30.0
Attraction Distrib.
Adjust by Sizes
Speed 1.0
Force Atlas
Presets... Reset
Graph
Dragging (Configure)
Arial Bold, 32
Context
Nodes: 266 (0.71% visible)
Edges: 4751 (1.64% visible)
Undirected Graph
Filters
Statistics
Reset
Library
Attributes
Equal
Betweenness Centrality Double (Node)
Closeness Centrality Double (Node)
Eccentricity Double (Node)
Harmonic Closeness Centrality Double (Node)
Modularity Class Integer (Node)
ml_target Integer (Node)
name String (Node)
Inter Edges
Intra Edges
Non-null
Partition
Queries
Range (Betweenness Centrality)
Parameters
Equal (ml_target)
Parameters
Drag subfilter here
Equal (ml_target) Settings
Value 0
Minimum: 0
Maximum: 1
Select
Stop
ENG US
8:59 PM
4/15/2022


File Workspace View Tools Window Help
Gephi 0.9.3 - GitHub.gephi_v3gephi.gephi
Overview
Data Laboratory
Preview
Workspace 1
Appearance
Nodes Edges
Unique Partition Ranking
Modularity Class
11 (18%)
6 (17.92%)
5 (16.56%)
7 (15.28%)
1 (7.54%)
10 (7.34%)
14 (7.17%)
2 (5.04%)
3 (2.67%)
4 (1.2%)
12 (0.98%)
8 (0.29%)
13 (0.01%)
Palette...
Apply
Layout
Force Atlas
Run
Force Atlas
Inertia 0.1
Repulsion strength 200.0
Attraction strength 10.0
Maximum displacement 10.0
Auto stabilize function
Autostab Strength 80.0
Autostab sensibility 0.2
Gravity 30.0
Attraction Distrib.
Adjust by Sizes
Speed 1.0
Force Atlas
Presets... Reset
Graph
Dragging (Configure)
Arial Bold, 32
Context
Nodes: 89 (0.24% visible)
Edges: 1051 (0.36% visible)
Undirected Graph
Filters
Statistics
Reset
Dynamic
Edges
Operator
Topology
Degree Range
Ego Network
Giant Component
Has Self-loop
In Degree Range
K-core
Mutual Degree Range
Neighbors Network
Out Degree Range
Saved queries
Queries
Range (Betweenness Centrality)
Parameters
Equal (ml_target)
Parameters
Degree Range
Parameters
Drag subfilter here
Degree Range Settings
350
9458
Select
Stop
ENG US
9:01 PM
4/15/2022


File Workspace View Tools Window Help
Gephi 0.9.3 - GitHub.gephi_v3gephi.gephi
Overview
Data Laboratory
Preview
Workspace 1
Appearance
Nodes Edges
Unique Partition Ranking
Modularity Class
11 (18%)
6 (17.92%)
5 (16.56%)
7 (15.28%)
1 (7.54%)
10 (7.34%)
14 (7.17%)
2 (5.04%)
3 (2.67%)
4 (1.2%)
12 (0.98%)
8 (0.29%)
13 (0.01%)
Palette...
Apply
Layout
Force Atlas
Run
Force Atlas
Inertia 0.1
Repulsion strength 200.0
Attraction strength 10.0
Maximum displacement 10.0
Auto stabilize function
Autostab Strength 80.0
Autostab sensibility 0.2
Gravity 30.0
Attraction Distrib.
Adjust by Sizes
Speed 1.0
Force Atlas
Presets... Reset
Graph
Dragging (Configure)
Arial Bold, 32
Context
Nodes: 32 (0.08% visible)
Edges: 77 (0.03% visible)
Undirected Graph
Filters
Statistics
Reset
NOT (Edges)
NOT (Nodes)
UNION
Topology
Degree Range
Ego Network
Giant Component
Has Self-loop
In Degree Range
K-core
Mutual Degree Range
Neighbors Network
Out Degree Range
Saved queries
Queries
Range (Betweenness Centrality)
Parameters
columns: Betweenness Centrality (class java.lang.Double)
range: 22358.951211270964 - 3212692.891258098
Equal (ml_target)
Parameters
column: ml_target (class java.lang.Integer)
match: 1
range: 0 - 1
Degree Range
Parameters
range: 150 - 9458
Drag subfilter here
Degree Range Settings
150
9458
Select
Stop
ENG US
9:03 PM
4/15/2022

Questions

	QUESTIONS	ANSWERS
a.	**Who are important entities from different points of view.**	**31890 , 27803** , 19222 , 14954, 20193 , 20528 , 29023 , 16631 Based on the highest degrees of nodes, the above mentioned entities can be regarded as important. Nodes 31890 and 27803 are of the most importance
b.	**How many communities exist within the network?** **Examine the characteristics of each community.** **Why is a community different from other communities?**	A total of 1**5 communities** exist within this network. Communities 12,8,13 and 9 are least related to other communities. The community of machine learning developers who only do machine learning is community 11 Community 6 is possibly a community which consists of developers focused on machine learning and another topic which is common among them.
c.	**Examine the relationship of nodes within and outside communities.**	The dataset contains a list of all of the links, where a link represents mutual follow relationship between developers. Possibly the people using ML need help or ideas in ML coding for which they are following other developers. People who generally post ML codes are also being visited by other developers. This shows that the relevance of ML is increasing in other fields as well
d.	**Any further insights that you may draw by analyzing the network**	As the need for machine learning and data science has increased, it is consistent with the dataset that people from non coding background and non ML background are venturing into ML.