Fundamentals of Database Management

Social Network Analysis

Submitted to:

Prof. Ashok Harnal

Submitted By:

Dipu Kumar (025013)

Jatin Malhotra (025016)

Karthick P.M. (025052)

Political Blogs

Introduction:

A directed network of hyperlinks between weblogs on US politics, recorded in 2005 by Adamic and Glance.

Nodes: 1490

Edges: 19025, Directed Graph

Node "value" attributes indicate political leaning according to:

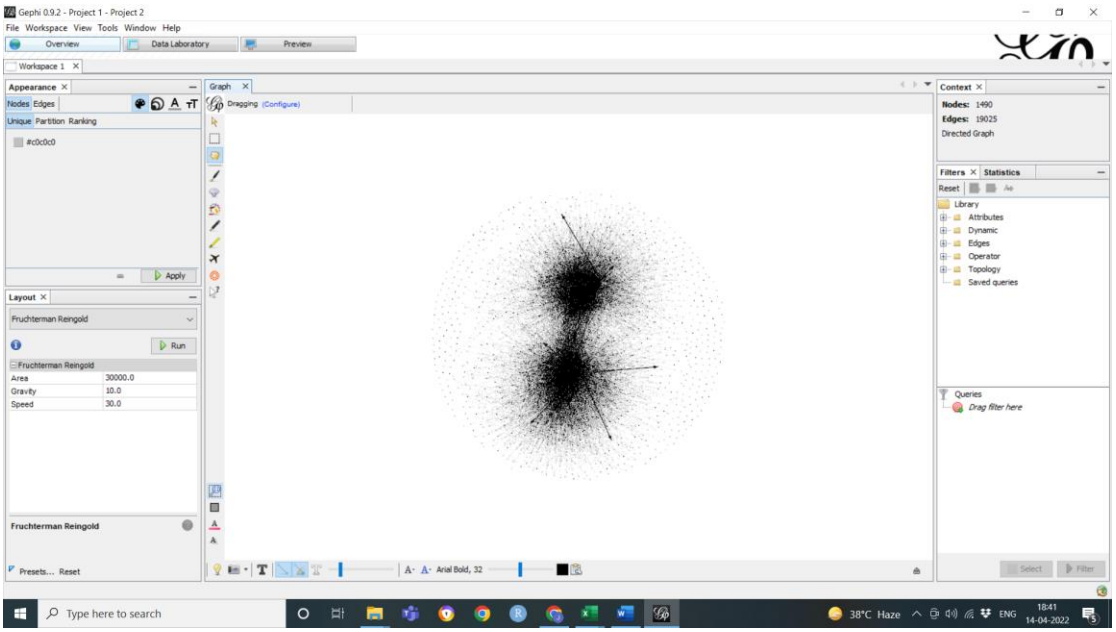
0 (left or liberal)

1 (right or conservative)

Data on political leaning comes from blog directories as indicated. Some blogs were labeled manually, based on incoming and outgoing links and posts around the time of the 2004 presidential election. Directory-derived labels are prone to error; manual labels are even more so. Links between blogs were automatically extracted from a crawl of the front page of the blog. This data was provided by Lada A. Adamic and Natalie Glance, "The political blogosphere and the 2004 US Election", in Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem (2005).

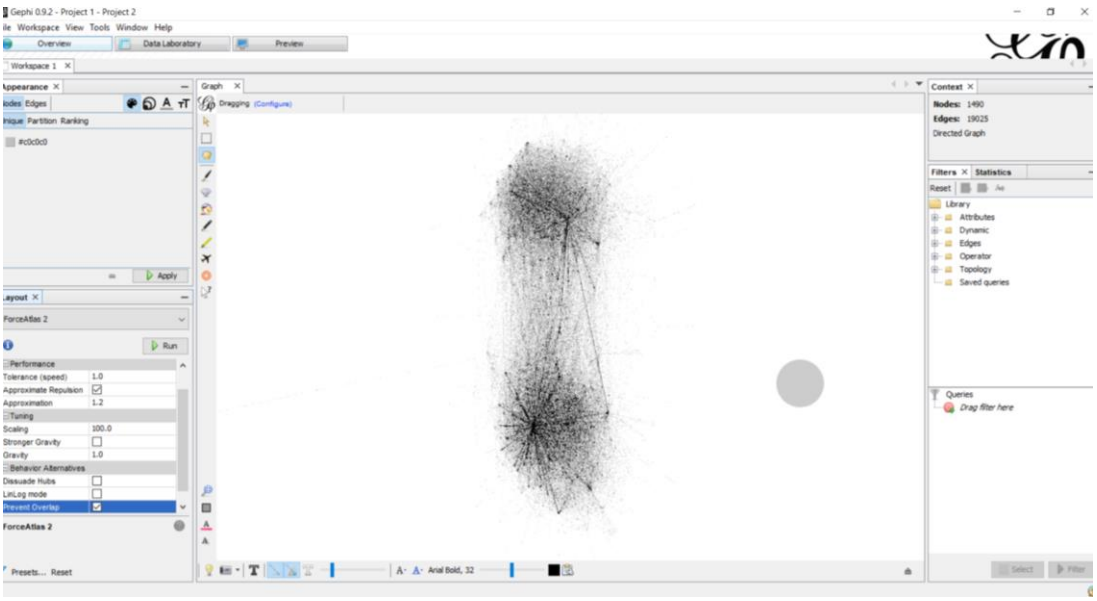
The procedure of social media analysis is as follows:

Fruchterman Reingold: The Fruchterman-Reingold Layout works well for many large social networks, though it may require some adjustment. It is an example of a force-directed algorithm, which uses an analogy of physical springs as edges that attract connected vertices toward each other and a competing repulsive force that pushes all vertices away from one another, whether they are connected or not [5, 7]. Area:30,000, Gravity:10, Speed:30.

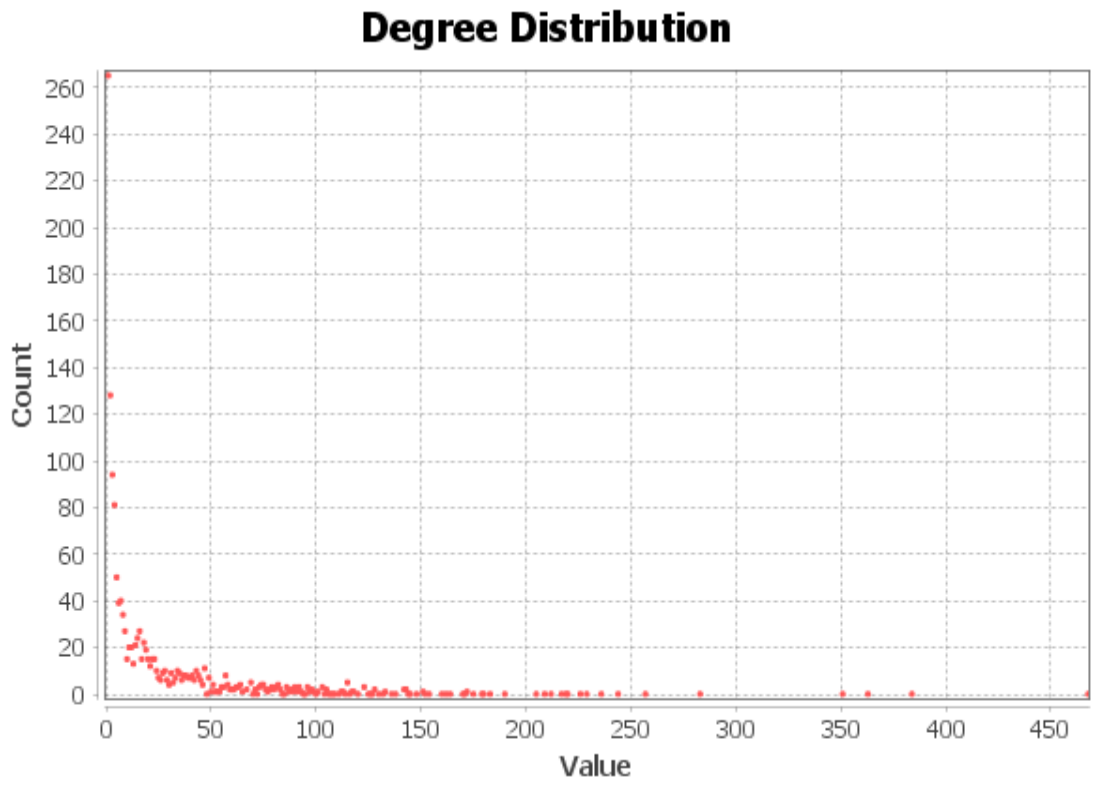


Force Atlas 2: ForceAtlas2 is a very fast layout algorithm for force-directed graphs. It's used to spatialize a weighted undirected graph in 2D (Edge weight defines the strength of the connection). Scaling. Control scale of the expansion of the graph.

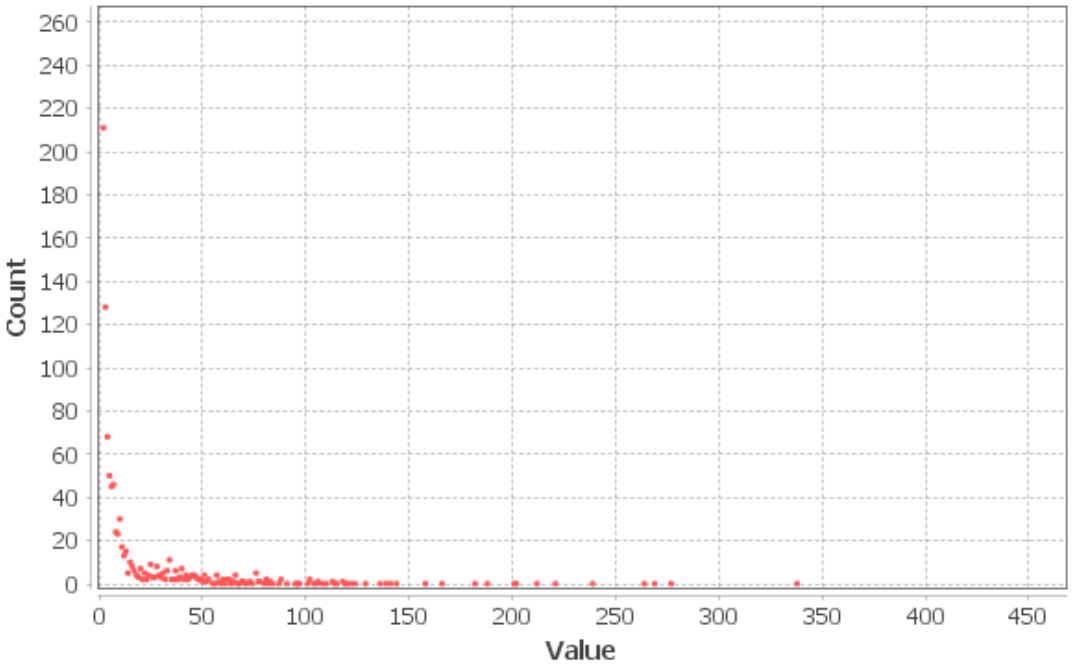
Prevent overlap. Prevent nodes from overlapping.



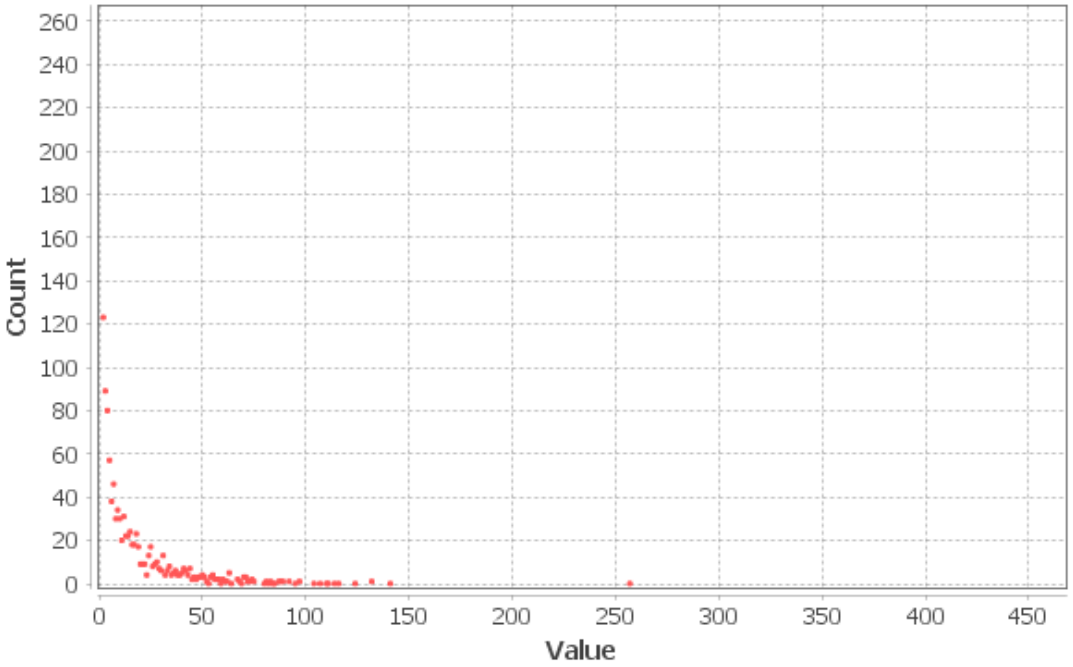
Degree + Modularity Partitions (without filter): In graph theory, the degree of a vertex is **the number of edges connecting it**. Degree centrality is the simplest centrality measure to compute. Recall that a node's degree is simply a count of how many social connections (i.e., edges) it has. The degree centrality for a node is simply its degree. A node with 10 social connections would have a degree centrality of 10. A node with 1 edge would have a degree centrality of 1. The average Degree of our node is 12.768.



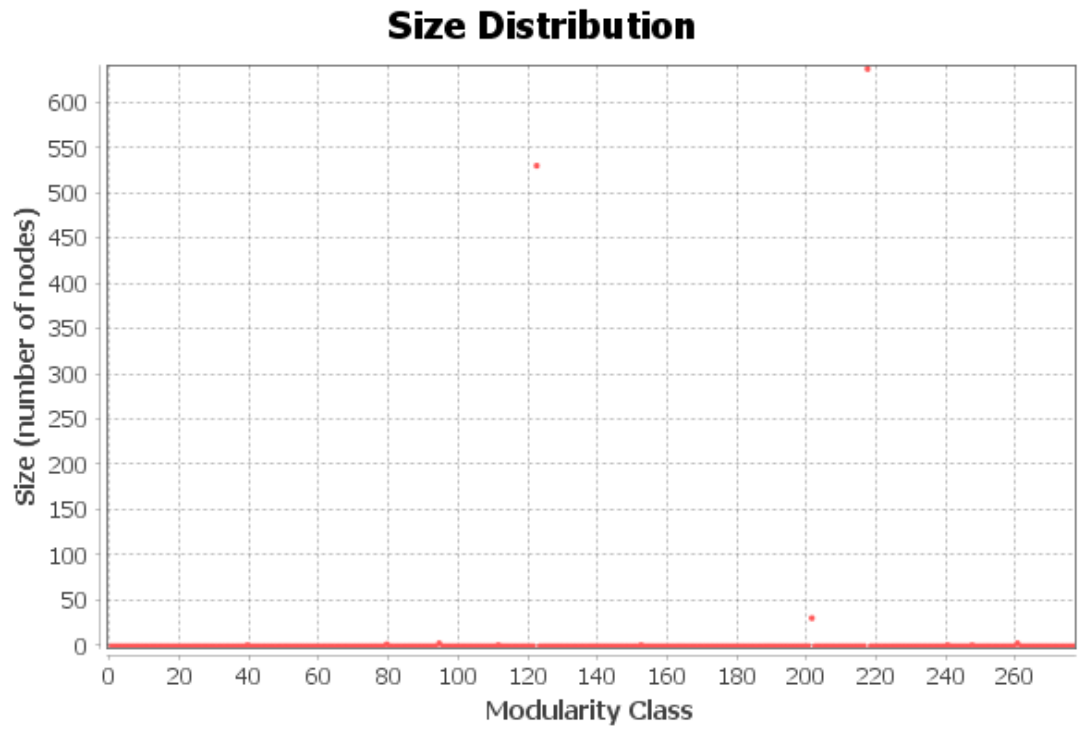
In-Degree Distribution



Out-Degree Distribution



Modularity is a system property that measures the degree to which densely connected compartments within a system can be decoupled into separate communities or clusters which interact more among themselves rather than in other communities. In a highly interconnected system with low levels of modularity, a shock to one compartment may cascade to another compartment and thus increase the risk of a system-wide collapse. Modularity = 0.427.



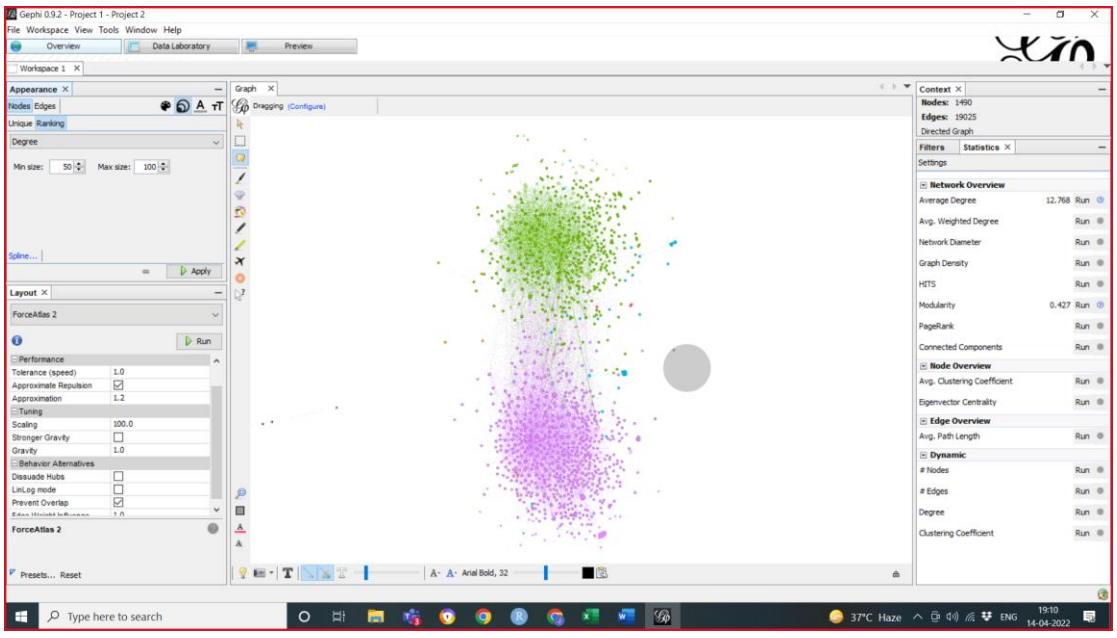
Modularity Report

Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

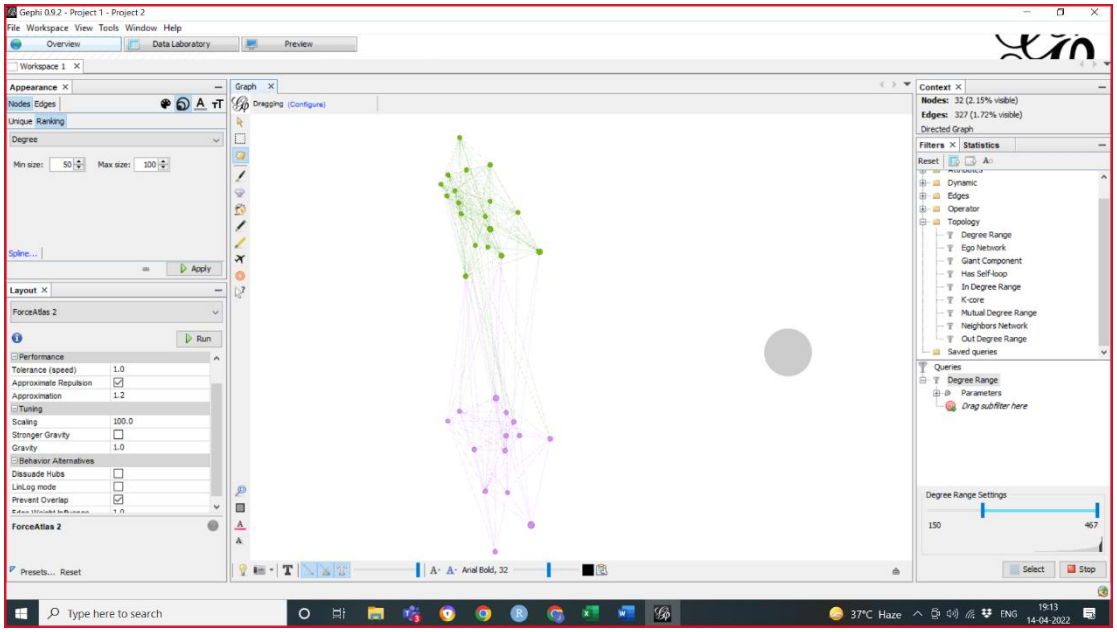
Results:

Modularity: 0.426
Modularity with resolution: 0.426
Number of Communities: 278



Topology → Degree Range (with Filter)

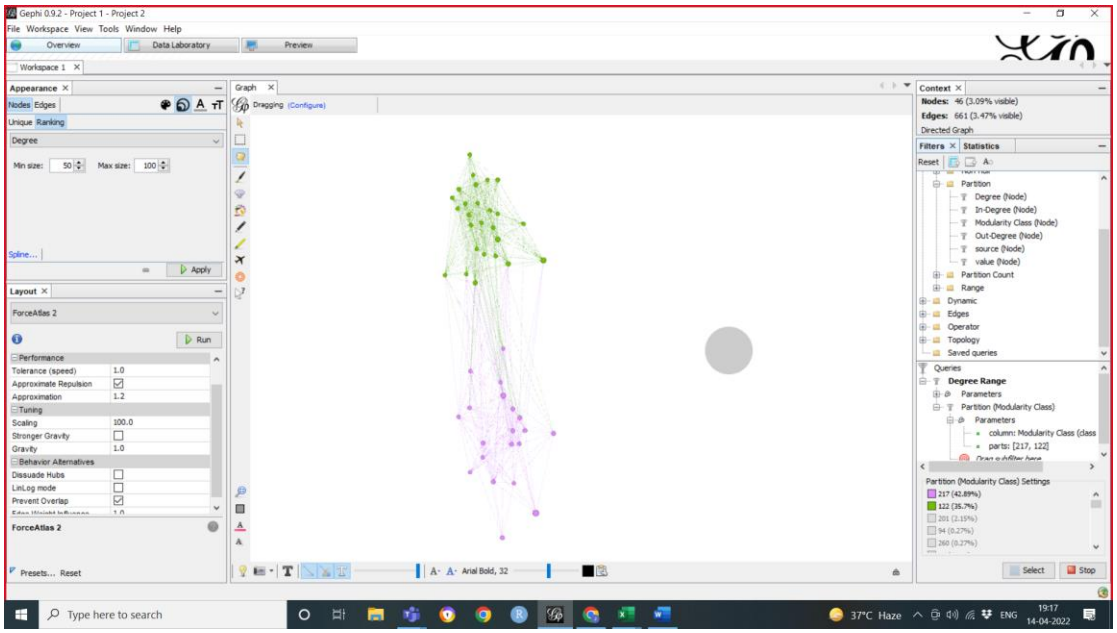
Here we selected the nodes whose degree was more than 150 and less than 467.



By changing the degree range of the network, we can see that there is a significant change in the graph as we have opted for the directed graph.

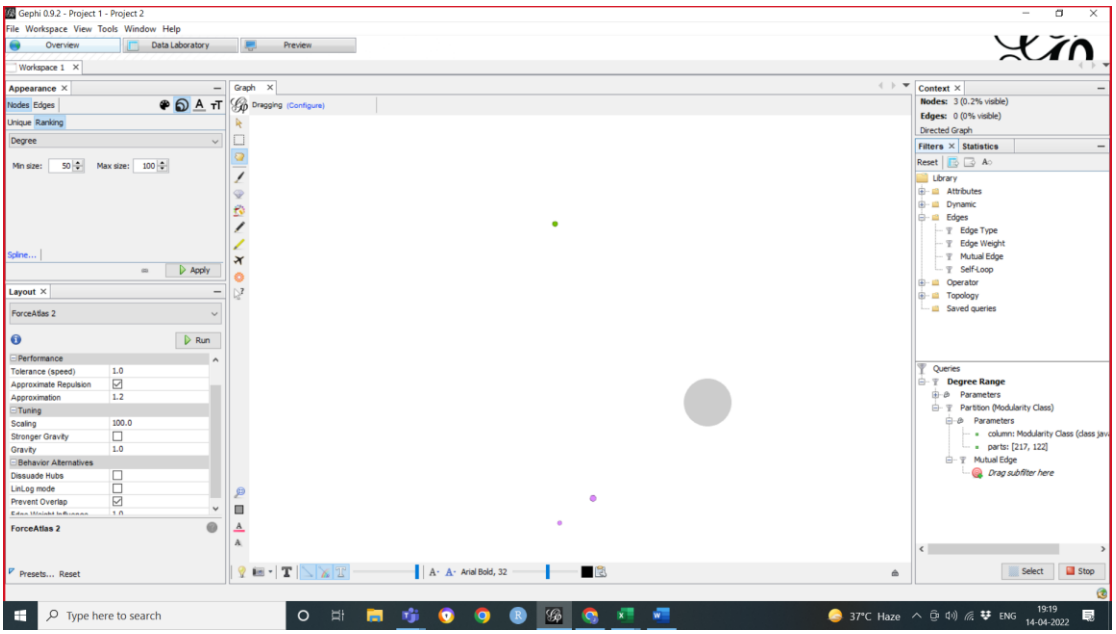
Attributes → Modularity Class

Partitioning was done based on Modularity, the top 2 communities were selected, i.e., the communities with Degree above 130 and forming the major communities were selected.



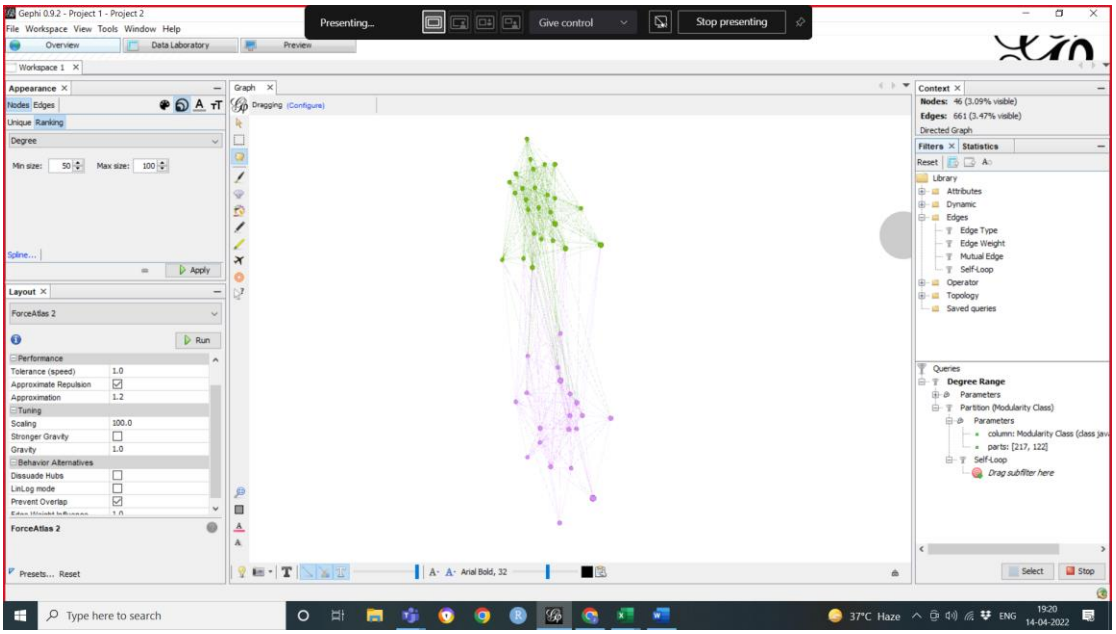
Edges → Mutual Edges

Mutual Edges (Filter) keeps edges that have a multiple/reciprocal edge (Only for Directed edges). Here we can see that only 3 nodes have multiple edges.



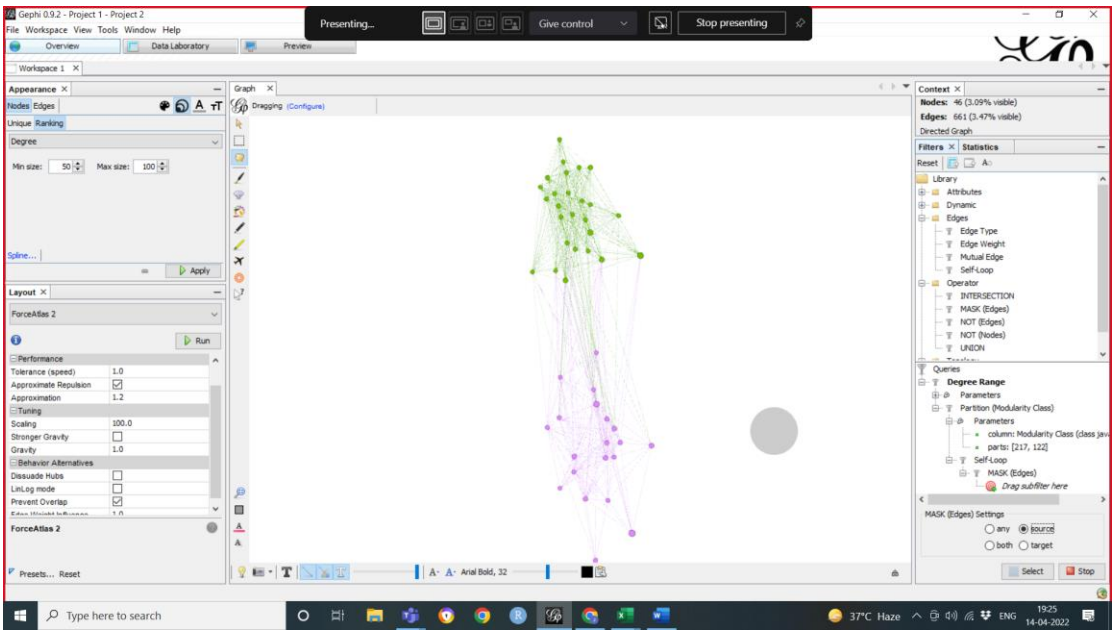
Edges → Self-Loop

This filter removes all the self-loops that exist in the network. After proper comparison, we can see the difference between networks.



Operator → Edge (MASK)

This filter returns the complete graph with only edges from the node filter sub-query. For instance, keep edges with source node degree>5. We can see a slight difference in both of the networks formed.



Conclusion:

The last image suggests that the dataset has many blogs(items) which are interrelated to each other and hyperlinks between weblogs on US politics (recorded in 2005 by Adamic and Glance) are having a strong relationship with each other. There are 278 different communities and filtering them suggests the top two communities are bonded strongly enough.