Assignment - I

Social Media Analytics
CS G519

In

Master of Engineering (Software Systems)
By
Raksha Choudhary (2020H1120261P)

Chinmay Gosavi (2020H1120262P)

Under the supervision of
Dr. Poonam Goyal

Associate Professor
Department of Computer Science & Information Systems

Birla Institute of Technology and Science

Pilani, Pilani Campus, Rajasthan (India)

November, 2021

Public Repository Name - https://github.com/kirin-23/assignment

Graph Properties -

	No of nodes	No. of edges	Average Path lengths	Average Clustering Coeffecient
Karate Club Network	34	78	2.408199643	0.5706384782
Dolphins Social Network	62	159	3.356953993	0.2589582461
Jazz Musicians Network	199	2743	2.248007715	0.6138325373

Community Detection Algorithm Results -

	Karate Club Network		
	Number of Clusters	Modularity Score	Runtime Algorithm(ms)
Girvan Newman	2	0.3599605523	135.37
Modularity Based Clustering	3	0.3806706114	5.95
Spectral Clustering	2	0.233974359	13.363

Dolphins Social Network			
	Number of Clusters	Modularity Score	Runtime Algorithm(ms)
Girvan Newman	4	0.4580712788	621.827
Modularity Based Clustering	4	0.4954906847	12.414
Spectral Clustering	4	0.4643215896	22.894

| Jazz Musicians Network			
	Number of Clusters	Modularity Score	Runtime Algorithm(ms)
Girvan Newman	5	0.003625042846	268729.234
Modularity Based Clustering	4	0.439058079	245.578
Spectral Clustering	5	0.403974359	88.267

**Visualizations -
Karate Club Network**
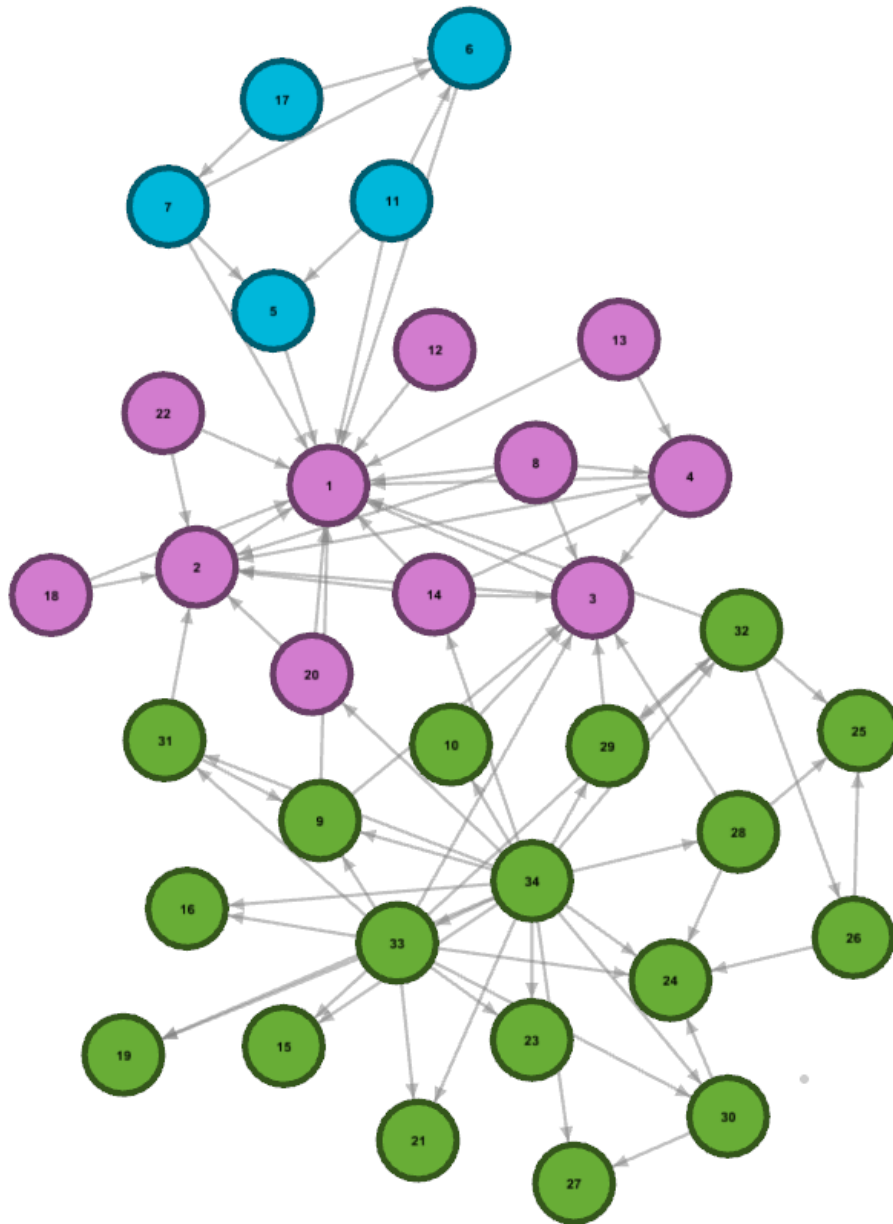
Layout - Force Atlas (Suitable for social networks) + Noverlap (Avoid node overlap)
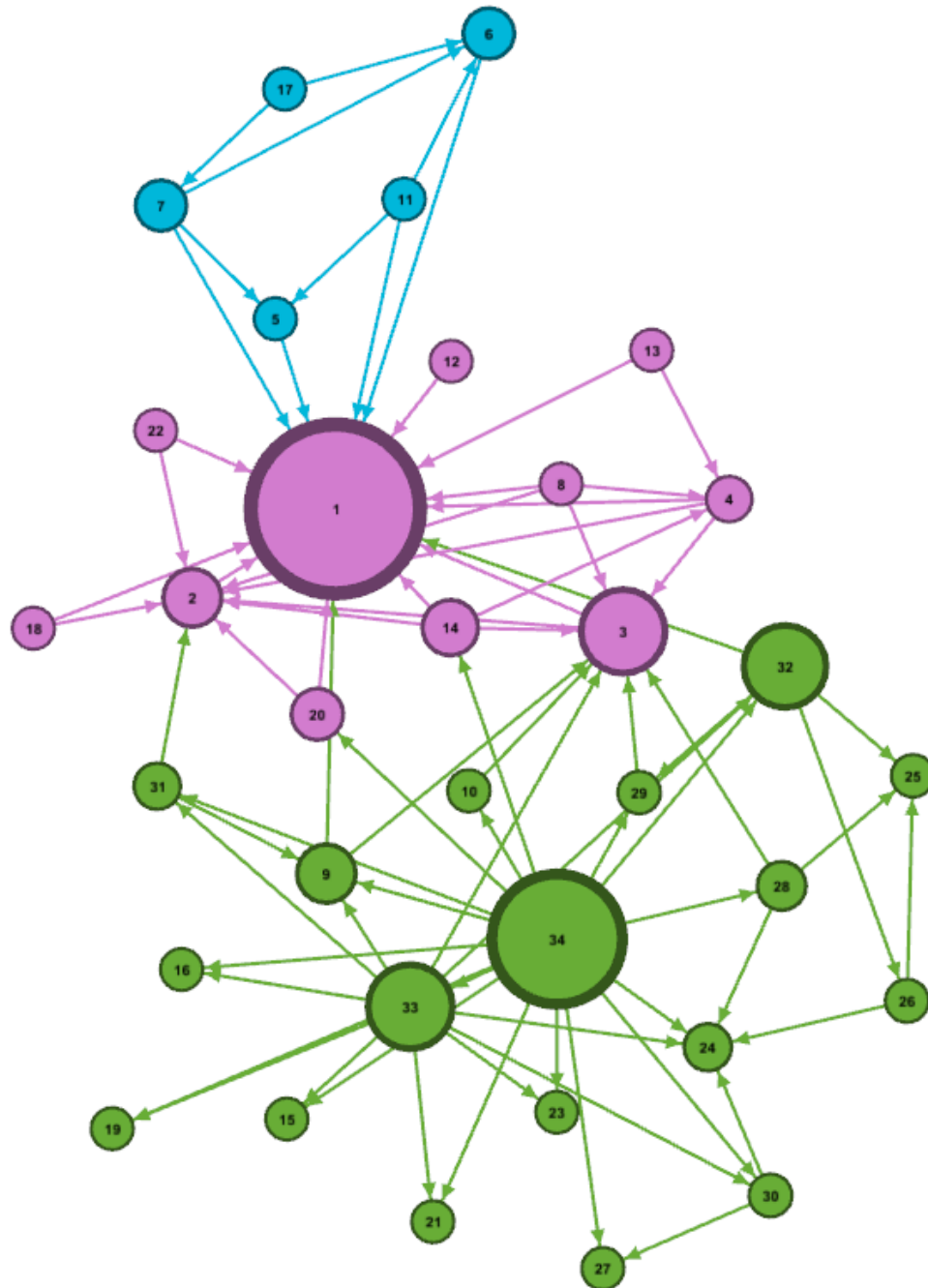**Description** - The nodes are colored as per communities.

Layout - Force Atlas (Suitable for social networks) + Noverlap (Avoid node overlap)

Description - The nodes are colored as per communities. Edges are colored as per source network.
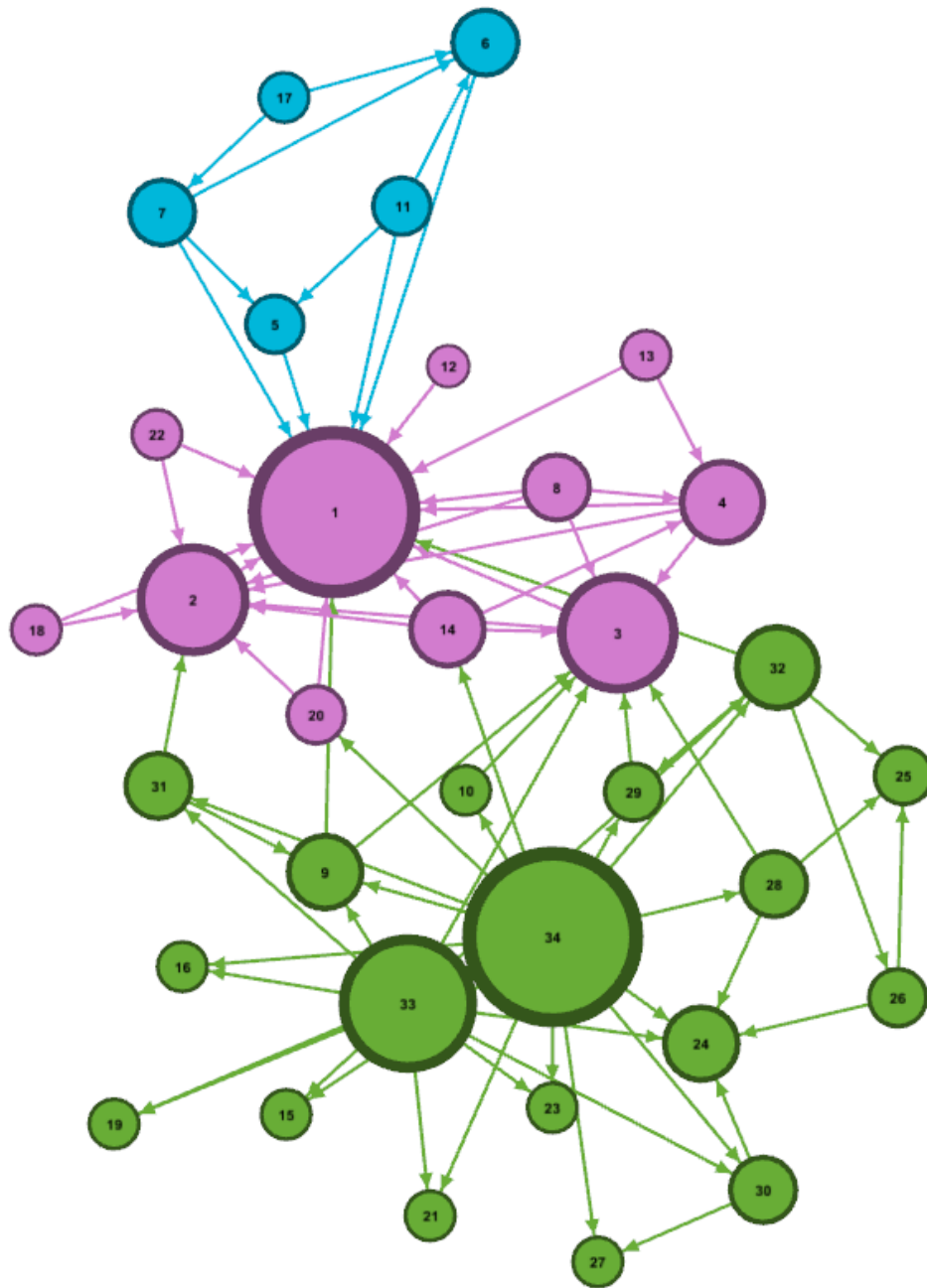
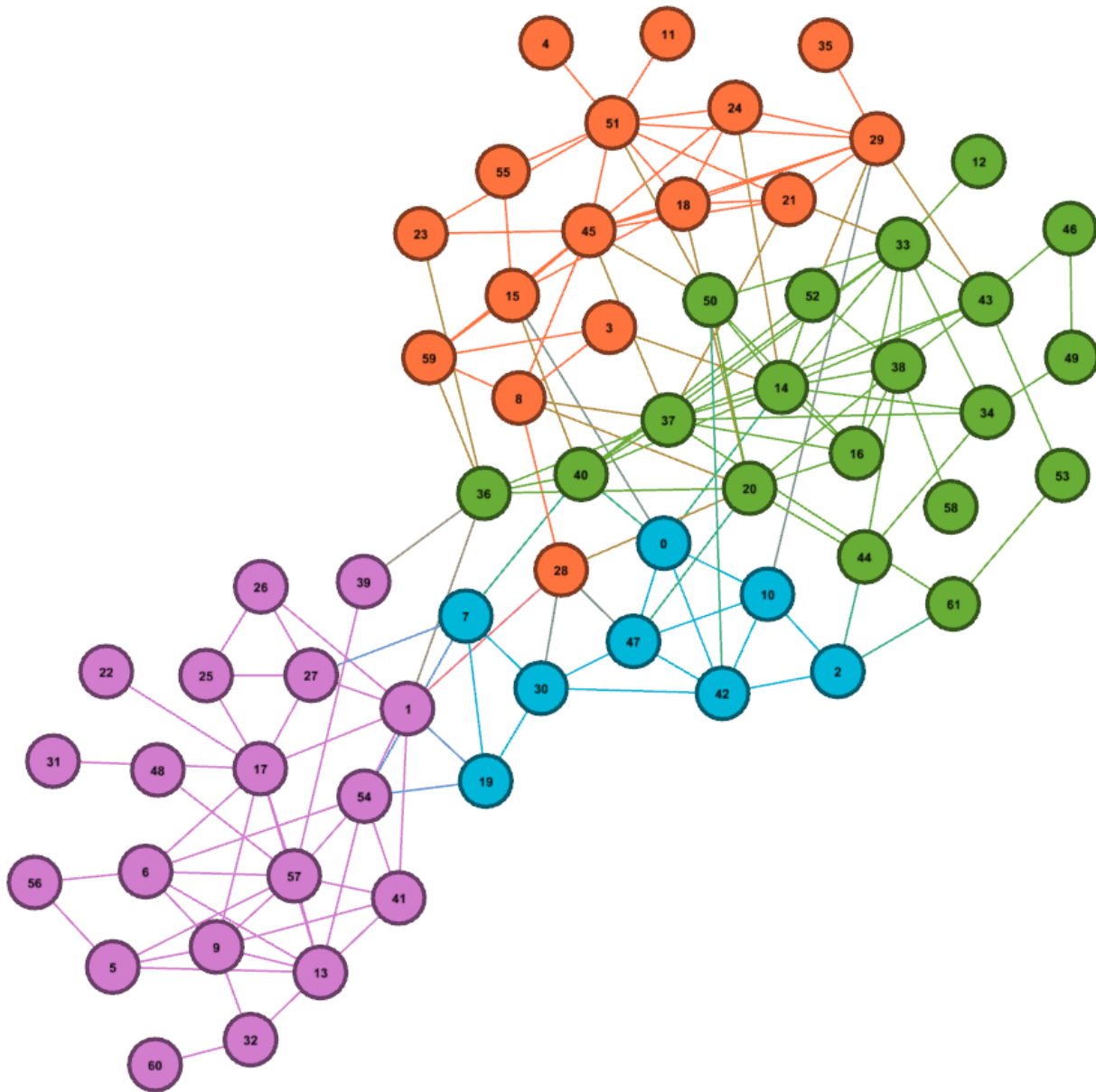**Layout** - Force Atlas (Suitable for social networks) + Noverlap (Avoid node overlap)

Description - The nodes are colored as per communities. Edges are colored as per source network. The node size is as per degree centrality. Correctly reflects that the largest nodes are 1 and 34 who were the instructor and administrator respectively are important.

Layout - Force Atlas (Suitable for social networks) + Noverlap (Avoid node overlap)
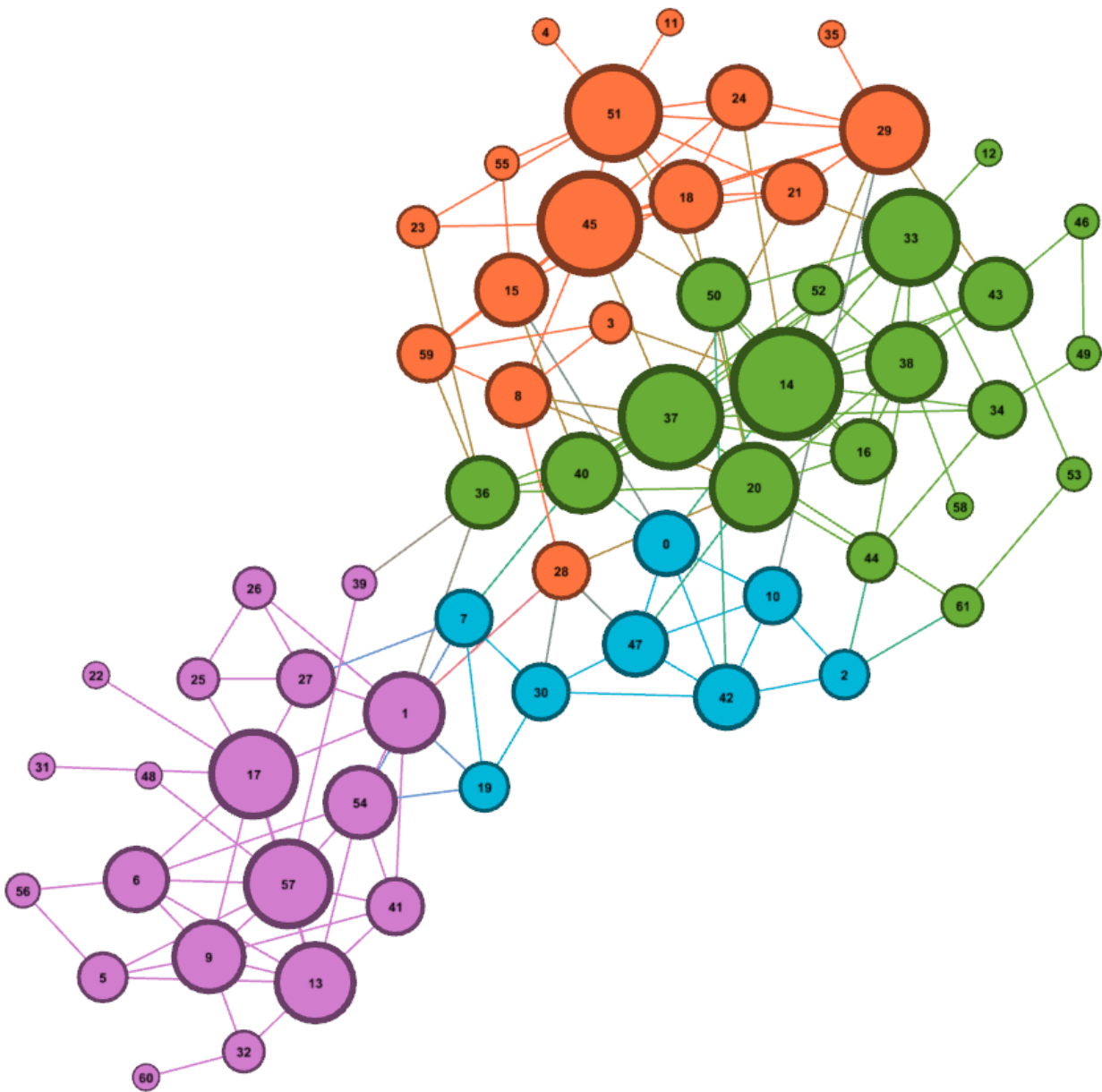
Description - The nodes are colored as per communities. Edges are colored as per source network. The node size is as per betweenness centrality. Correctly reflects that the largest nodes are 1 and 34 who were the instructor and administrator respectively are important.

Dolphins Social Network

Layout - Force Atlas (Suitable for social networks) + Noverlap (Avoid node overlap)

Description - The nodes are colored as per communities. Edges are colored as per source network.

Layout - Force Atlas (Suitable for social networks) + Noverlap (Avoid node overlap)

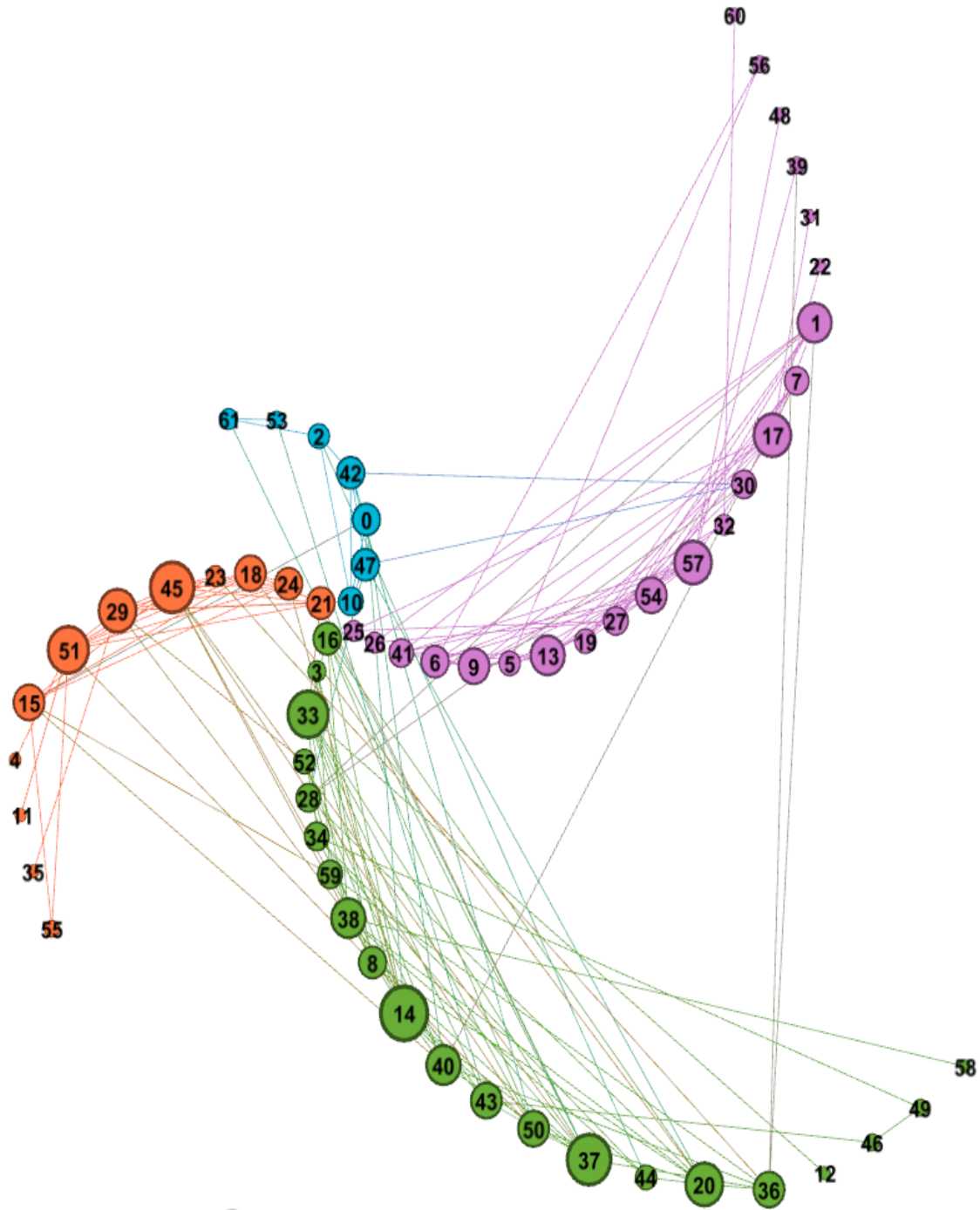**Description** - The nodes are colored as per communities. Edges are colored as per source network. The node size is as per degree centrality.

Layout - Force Atlas (Suitable for social networks) + Noverlap (Avoid node overlap)
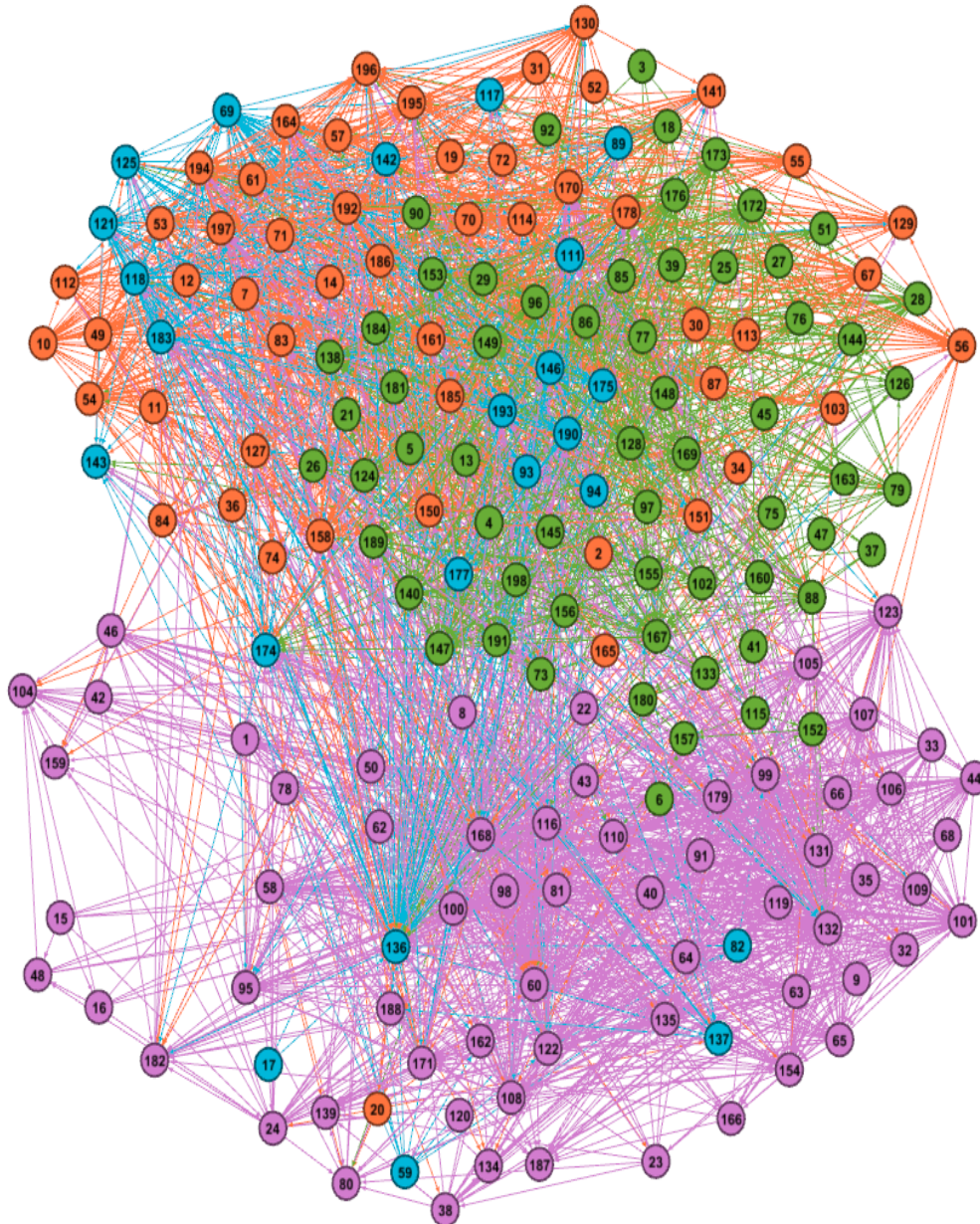
Description - The nodes are colored as per communities. Edges are colored as per source network. The node size is as per betweenness centrality.

Layout - Radial Layout (To easily visualize the connections between the network as well as the connections outside the community)
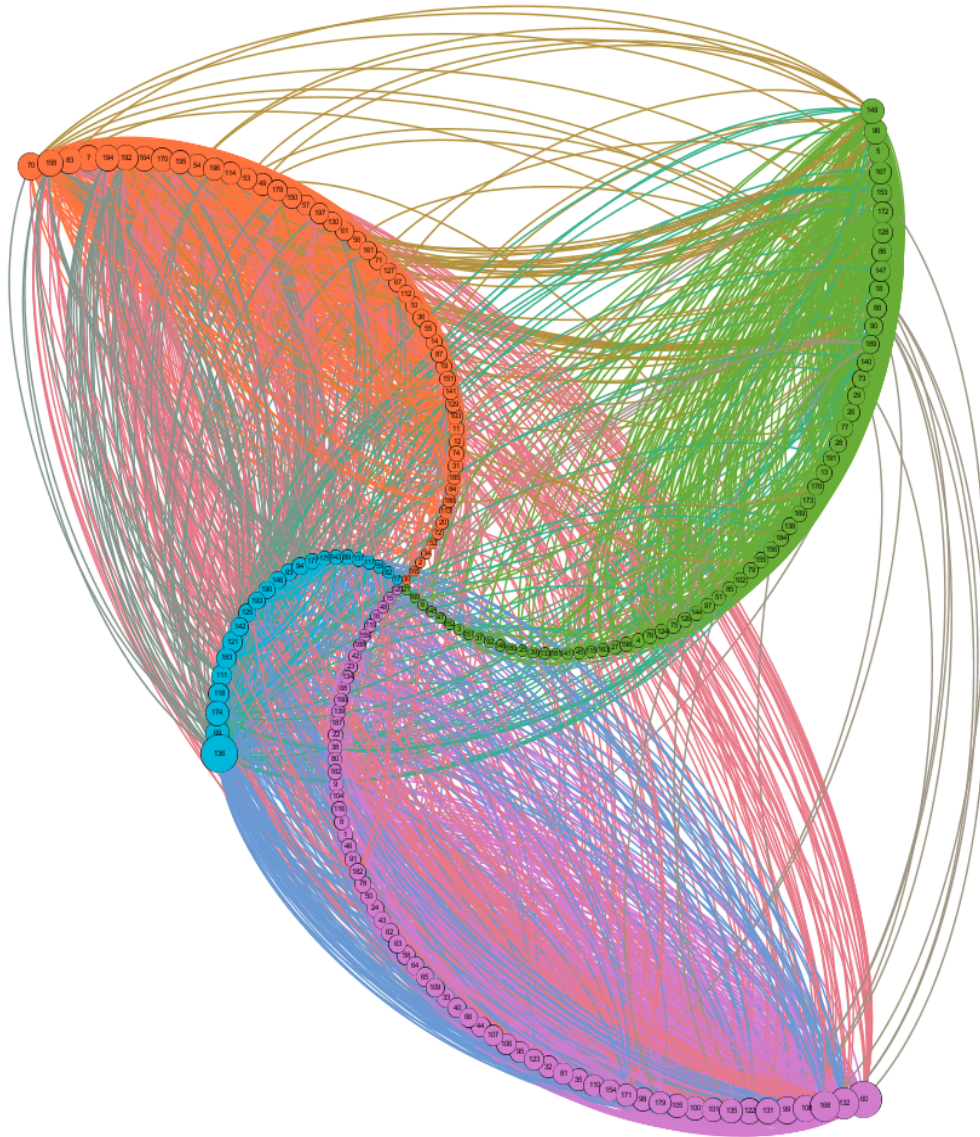
Description - The nodes are colored as per communities. Edges are colored as per source network. The node size is as per degree centrality and ordered as per clustering coefficient.
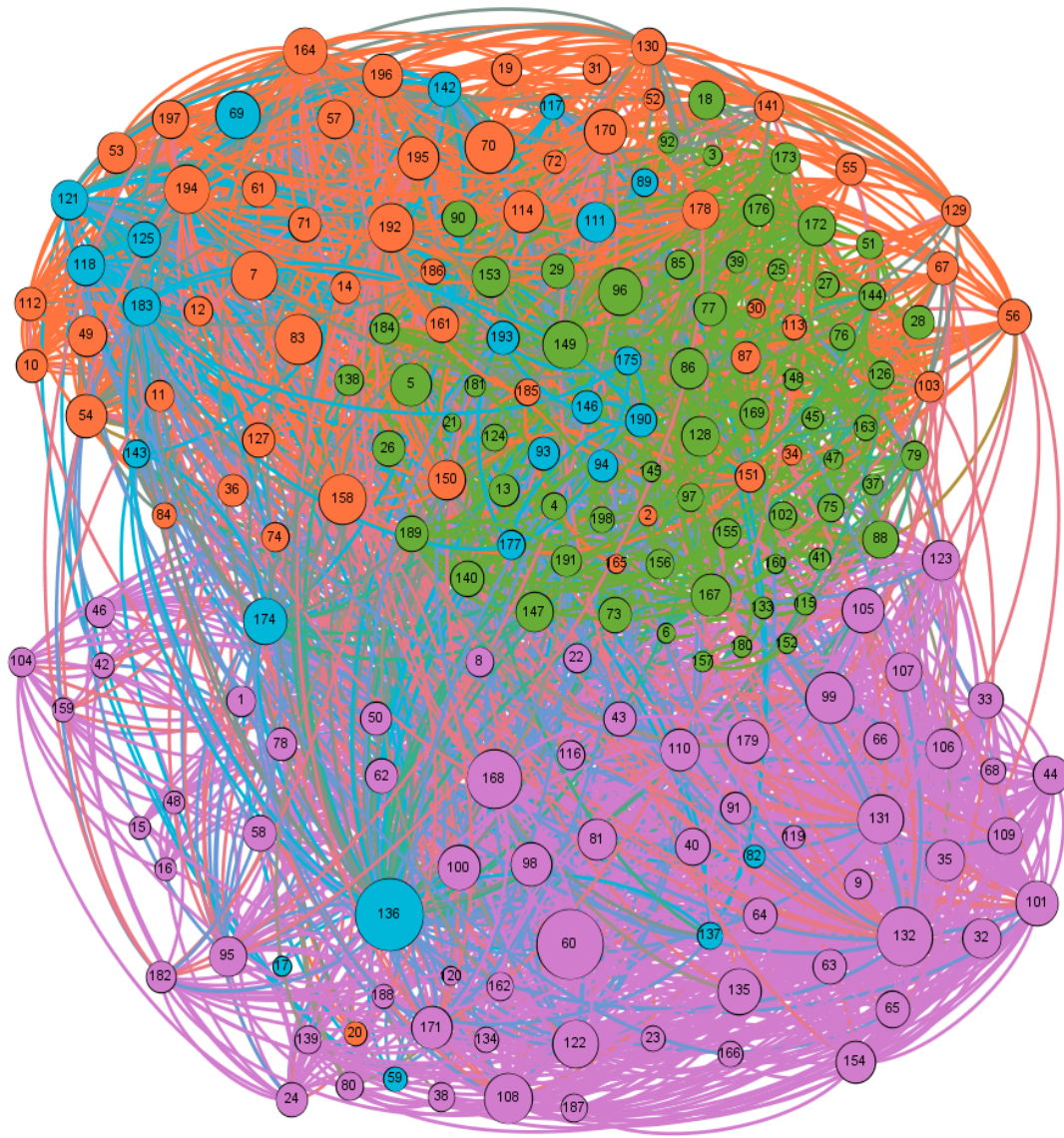
Jazz Musicians Network

Layout - OpenOrd (Suitable for networks with large number of edges) + Noverlap (Avoid node overlap)

Description - The nodes are colored as per communities. Edges are colored as per source network.

Layout - Radial Layout (To easily visualize the connections between the network as well as the connections outside the community)

Description - The nodes are colored as per communities. Edges are colored as per source network. The node size is as per degree centrality and ordered as per clustering coefficient.

Layout - OpenOrd (Suitable for networks with large number of edges) + Noverlap (Avoid node overlap)

Description - The nodes are colored as per communities. Edges are colored as per source network and node size as per degree.

Algorithm Comparison -
- **Efficiency -**
  - **Girvan-Newman performs slower** compared to the other algorithms, especially slower as the number of edges increases in the graph. This is reflected in the results in the huge difference between the run-time on the Jazz Network graph and the other two graphs.
  - As per the results we can observe for computational speed, **Modularity Maximization outperforms the other algorithms** except for the Jazz network which has a high amount of edges.
- **Quality -**
  - Gephi tool utilizes **Louvain method for community detection**, thus finding communities based on modularity maximization.
  - For comparing the results we'll have to compare with the results obtained in the modularity-maximization code. We've not compared with the results from other methods as Spectral clustering will provide balanced communities and Girvan-Newman based on betweenness of edges.
  - Consider the **number of communities as a factor**, the number of communities by the code and as per the visualization are the same for all the networks.
  - In Karate Club Graph, the results match with most of the nodes classified in the same community as per the visualization. Accuracy **achieved is ~90%** with 3 nodes misplaced in result comptared visualization.

Representative Network of Karate Club Network for Girvan-Newman Method
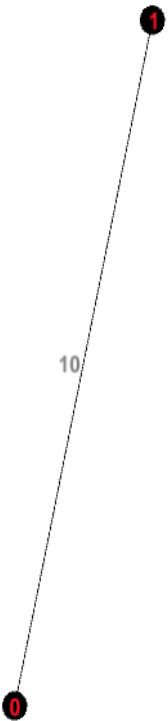
Representative Network of Dolphins Social Network for Girvan-Newman Method