Network Analysis Assignment

Name: **Sushil Sharma** PhD Cohort 2021

1. Load the data, transform the adjacency matrix into a `igraph` object, plot the network
    - Node size is determined by the degree of a node
    - Color is determined by the social contacts at a conference

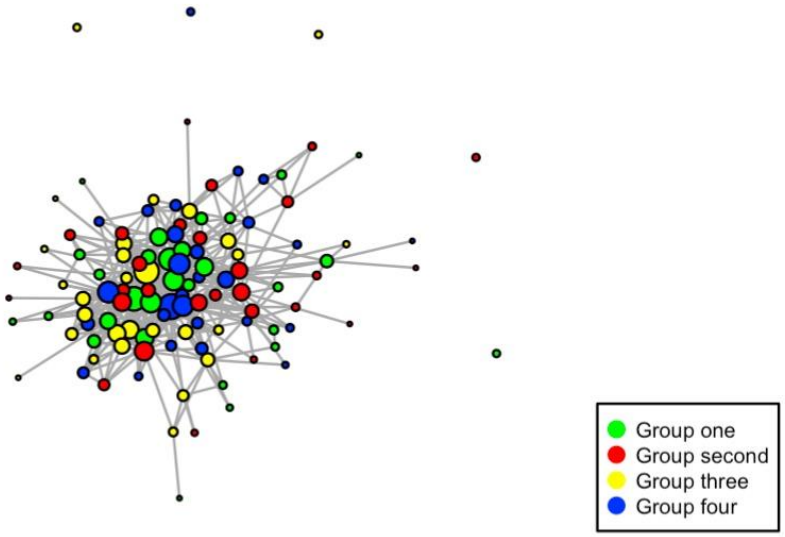

Fig 1: Plot the network

2. Calculate descriptive statistics of the network such as the number of nodes, number of edges, density, number of triangles, degrees and degree distribution.

The below table shows the descriptive statistics of the network:

Number of nodes	**113**
Number of edges	**441**
Number of possible edges	**6328**
Number of triangle	**480**
Density	**0.06969026**

This is a network subset of 113 nodes and 441 edges, For the degrees and degree distribution as we can see that below figures.

Plotting the degree distribution;

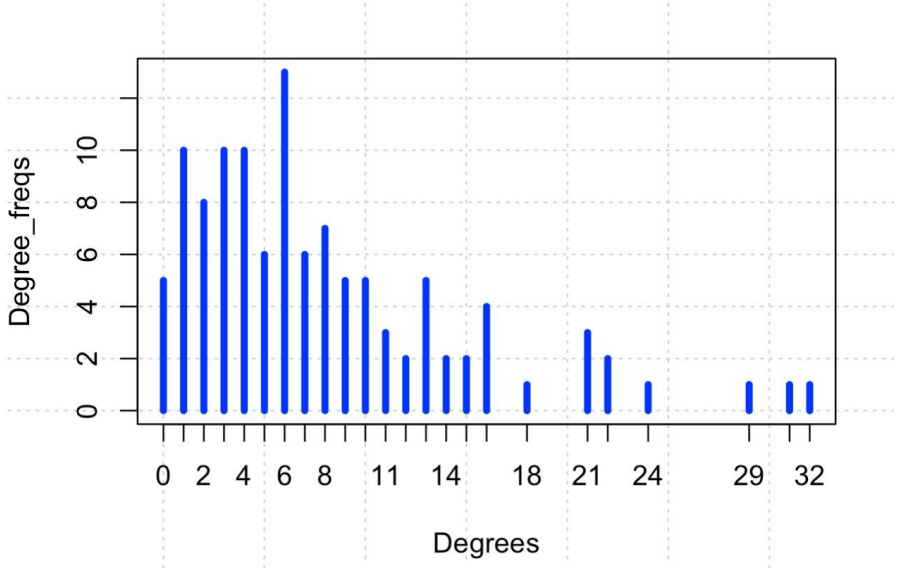

Fig 2: Degree distribution plot

3. Perform spectral clustering. Comment on whether the algorithm can find a reasonable clustering structure. Comment on whether some of the groups exhibit community structure. Firstly, spectral clustering is most often used for community finding in a network. In this Case calculate the eigenvalues of Y and decide the value of K, so I choose K = 4, then I select The smallest value of K such that $(K+1)^{th}$ Eigen value is much smaller than the K^{th} Eigenvalue (tradeoff between dimensions and variance explained). Usually K-means is run for some value of K, but the other values may be considered.

Note: For more details about the eigenvalue and K ← 4 see the code in R

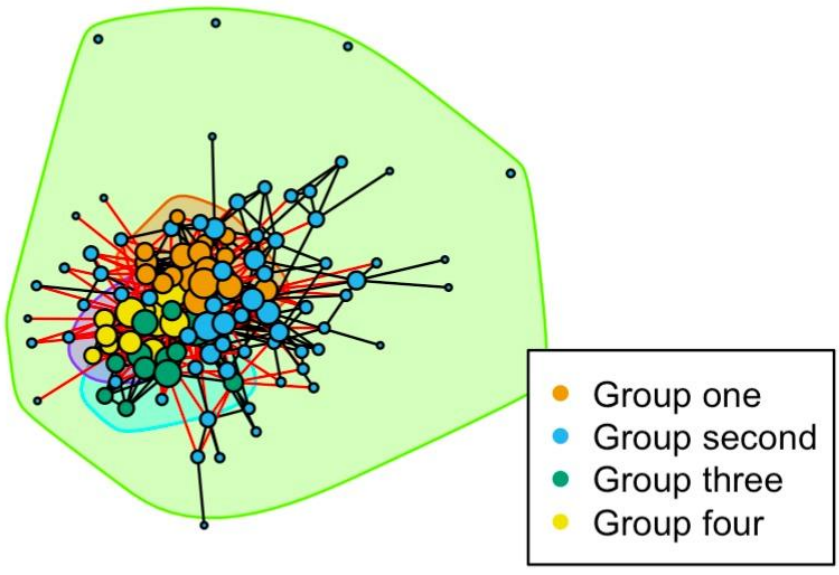

Fig 3: Spectral Clustering Plot

The figure above reveals that group two (blue nodes) and group one (orange nodes) have the most common nodes, indicating a community pattern between the two groups.

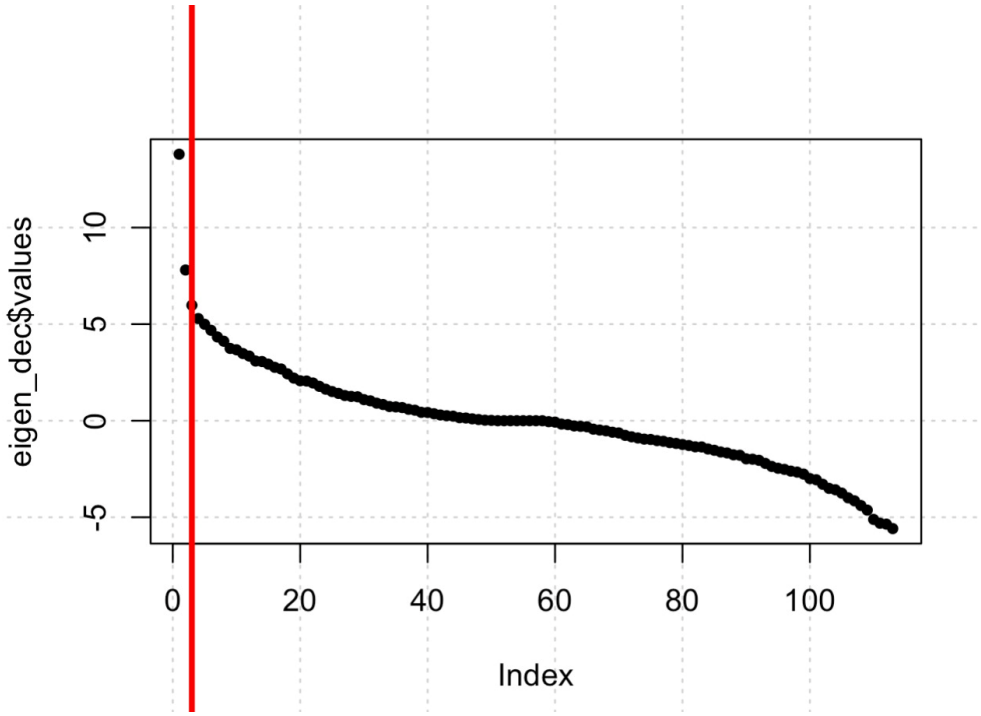

Fig 4: Eigenvalues Plot

4. Use the variational expectation-maximization algorithm to fit a stochastic blockmodel to the data. Describe the connectivity behavior for each of the group found, discussing whether they exhibit community structure. Compare the optimal partition with the one obtained through spectral clustering

   In general, two nodes in the same group are said stochastically equivalent, because they have the same probability of connecting to any other node. In other words, a SBM highlights groups of nodes that have the same behavior. The variational EM technique with the Bernoulli probability distribution was performed with the blockmodel package in R.
   The probability that an edge appears between the node in group one and a node in group second is **0.1852.**

Group one	0.1124
Group two	**0.1861**
Group three	0.0858
Group four	0.0337

As we can clearly see that group 2 has a strong community structure and group 4 do not.

Compare the optimal partition with the one obtained through SBM Model.

Group one	**12.704**
Group two	21.030
Group three	9.7009
Group four	3.8165

The table shows how many nodes has conference community, group 2 which is **21.030** but the expected number of edges for a node in group 1 which is **12.704**

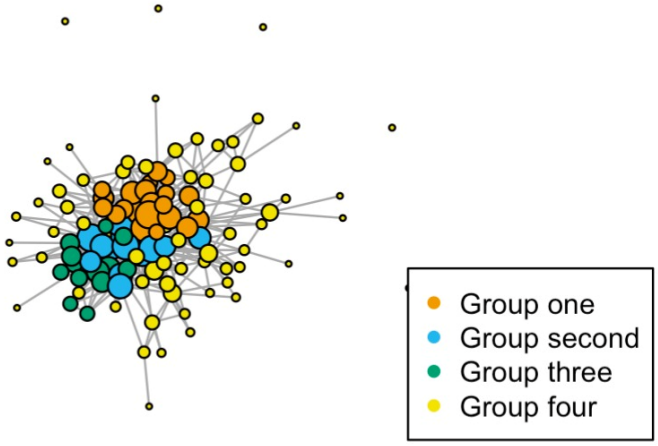

Fig 5: Cluster graph

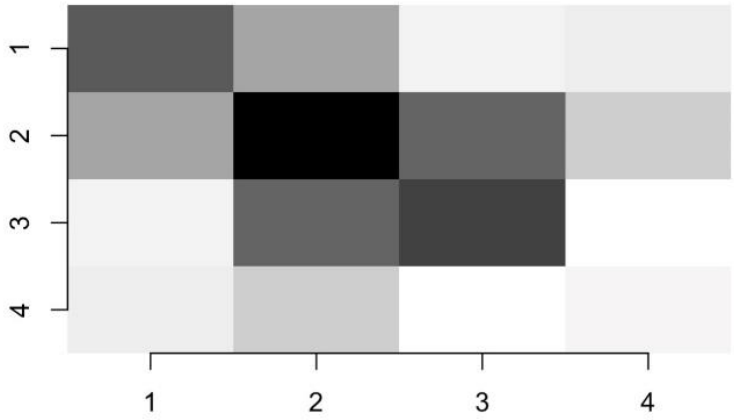

Fig 6: Community behavior between group

- The strongest community pattern was found between Group two and Group three
- The largest number of nodes Group two
- The highest number of nodes Group four

Please locate the code included as Assignment1.R in your email.