Assignment CRT2021 Network Analys

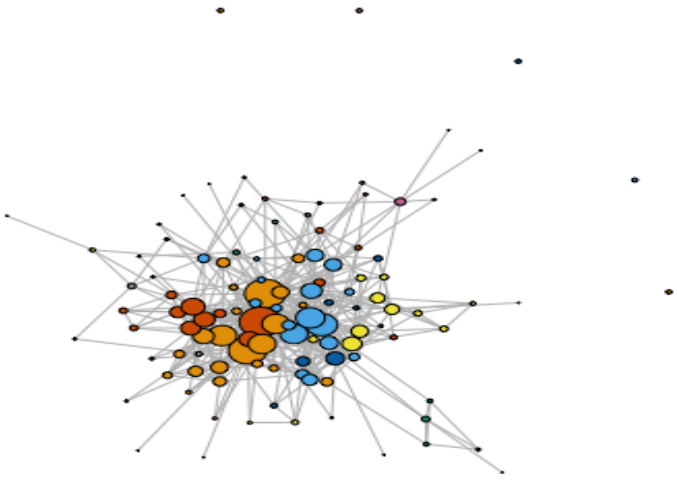
Aryan Singh (aryan.singh@ul.ie)

January 30, 2022

1. Load the data, transform the adjacency matrix into an igraph object, plot the network.

Plotted network in figure 1, node size is based on degree of a node and color is based on betweeness criteria.

Figure 1: Network plot

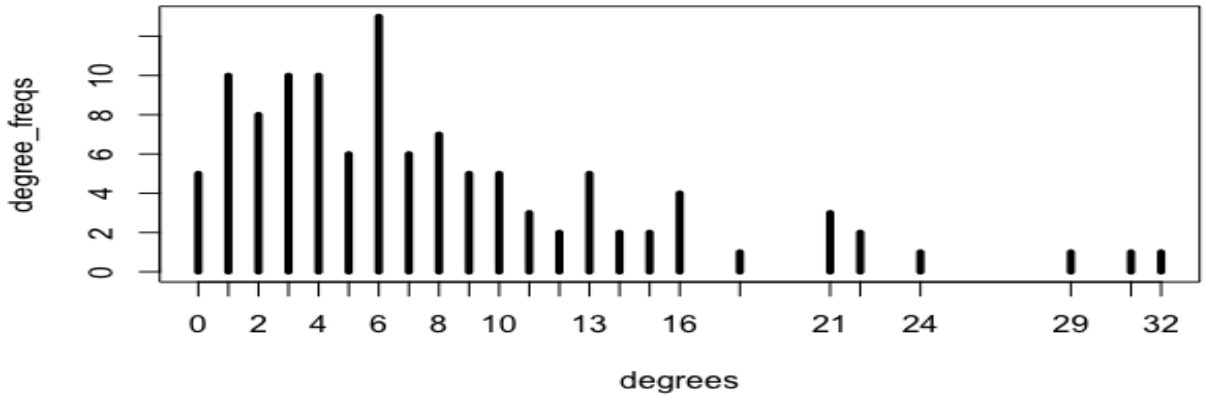


2. Descriptive statistics of the network such as the number of nodes, number of edges, density, number of triangles, degrees and degree distribution.

Table 1: Descriptive statistics of given network

Descriptive statistics of network	value
Nodes	113
Edges (dyad mutual)	441
Possible edges	6328
Triangles (triad census type 300)	480
Density	0.069
Number of components	6
Connectedness	0.913

Figure 2: degree distribution plot



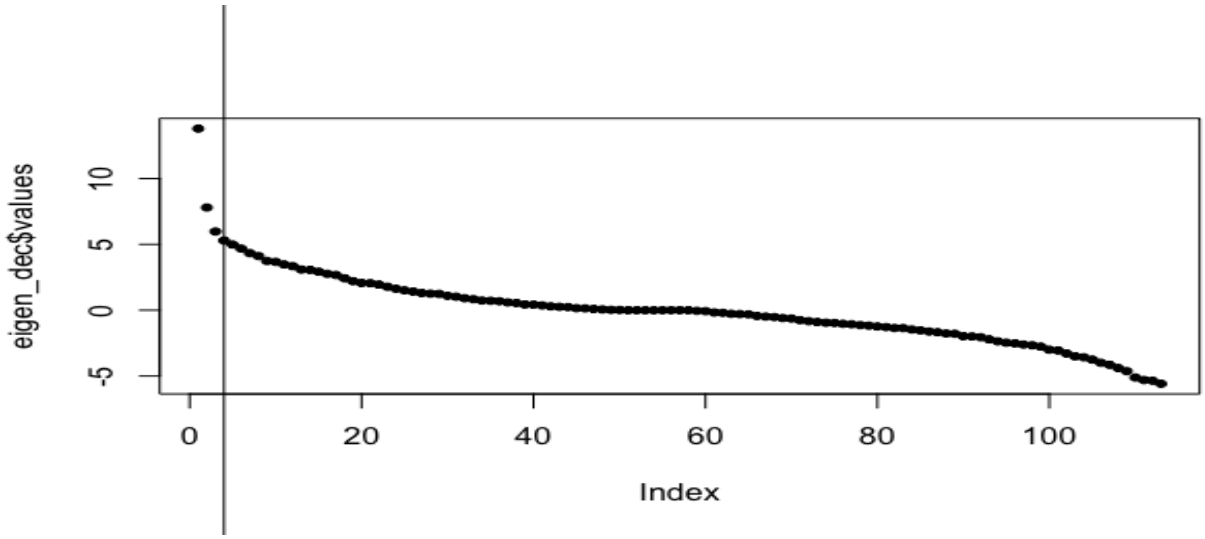
3. Perform spectral clustering. Comment on whether the algorithm can find a reasonable clustering structure. Comment on whether some of the groups exhibit community structure.

Plot shown in figure3, we can see four clusters assigned to nodes based on eigen value decomposition of adjacency matrix, Group1 (with blue nodes) and Group 2 (with yellow nodes) have the most number of edges between them, which reflects a community pattern between the two groups. Number of nodes in group 1 is 71 and in group 2 is 25 with 49 edges between both the groups.

Figure 3: spectral clustering plot

For optimal number of cluster value K, we used the plot shown in figure 2, K value 4 has been chosen based on point where there exist the largest eigengap [1] $|\lambda_k - \lambda_{k+1}|$.

Figure 4: eigen value plot



4. Variational expectation-maximisation algorithm to fit a stochastic blockmodel to the Data. Discuss whether they exhibit community structure and compare it with spectral clustering.

The Stochastic BlockModel (SBM) is a finite mixture model for network data, Blockmodels package in R has been used, which uses variational EM algorithm,with Bernoulli probability distribution. When we extract the best model according to ICL values,we get K = 4 as the best number of clusters.

some stats (K = 4)	group1	group2	group3	group4
Number of nodes	21	12	13	67
Expected degree	0.11	0.18	0.08	0.033
Expected number of edges	12.7	21.03	9.7	3.8

Table 3: community pattern between groups

probability of edge formation	group1	group2	group3	group4
group1	-	0.18	0.033	0.044
group2	0.18	-	0.313	0.033
group3	0.033	0.313	-	0.034
group4	0.044	0.033	0.034	-

References

[1] Leslie Lamport, *A Tutorial on Spectral Clustering.* Ulrike von Luxburg 2007.

From Table 3 we can see that group 2 and group 3 shows strong sign of community.

The main difference between clustering using spatial clustering and SBM model is that with spatial clustering, the community behaviour is observed between two groups with the largest number of nodes whereas, with SBM model we can see that though the group 4 (coloured yellow) has the highest number of nodes but the stronger community pattern is observed between group 2 and group 3.

Figure 5: Heat-map of community behavior between groups

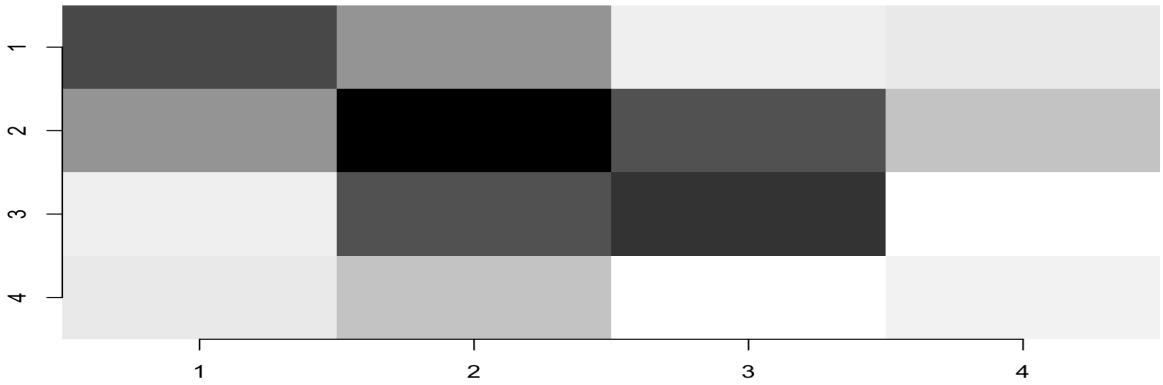


Figure 6: Plot of clustered graph using SBM