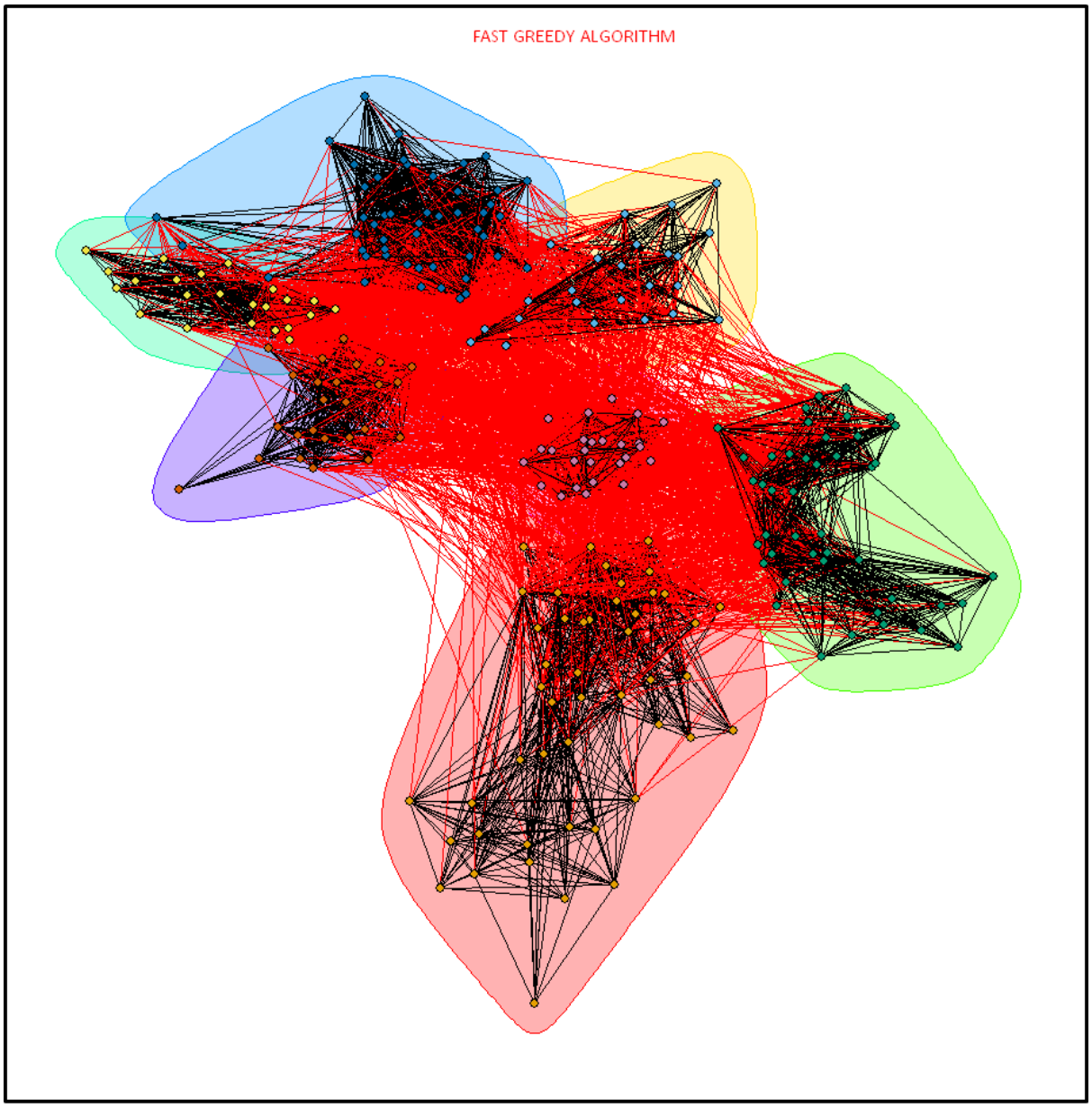
IDS 564
Lab3 Community Detection
Mohammad Umar- 672254237

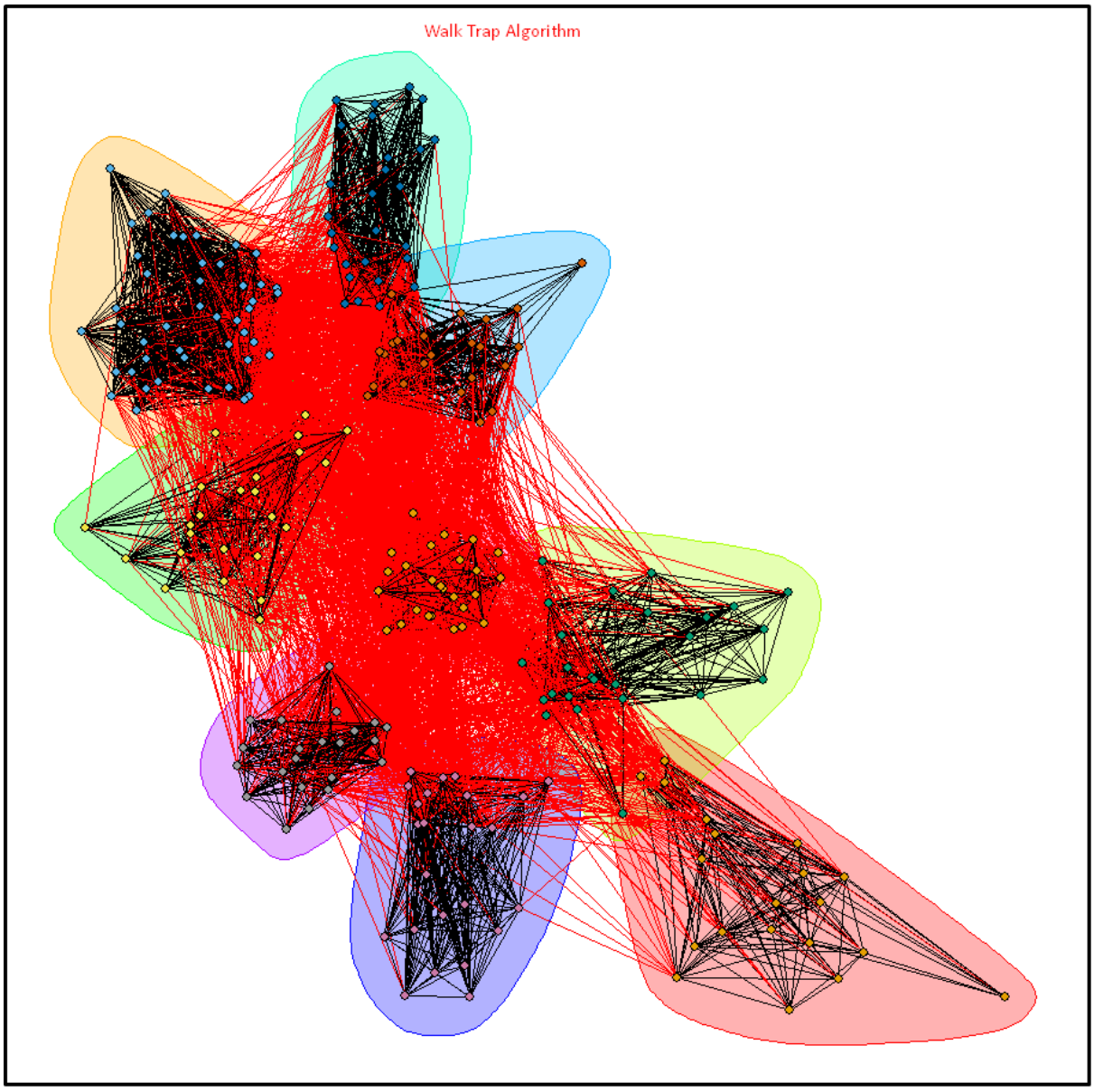
1) **Fast Greedy Algorithm Plot** source

This* algorithm is the Clauset-Newman-Moore algorithm. In this case the algorithm is agglomerative. At each step two groups merge. The merging is decided by optimising modularity. This is a fast algorithm but has the disadvantage of being a greedy algorithm.



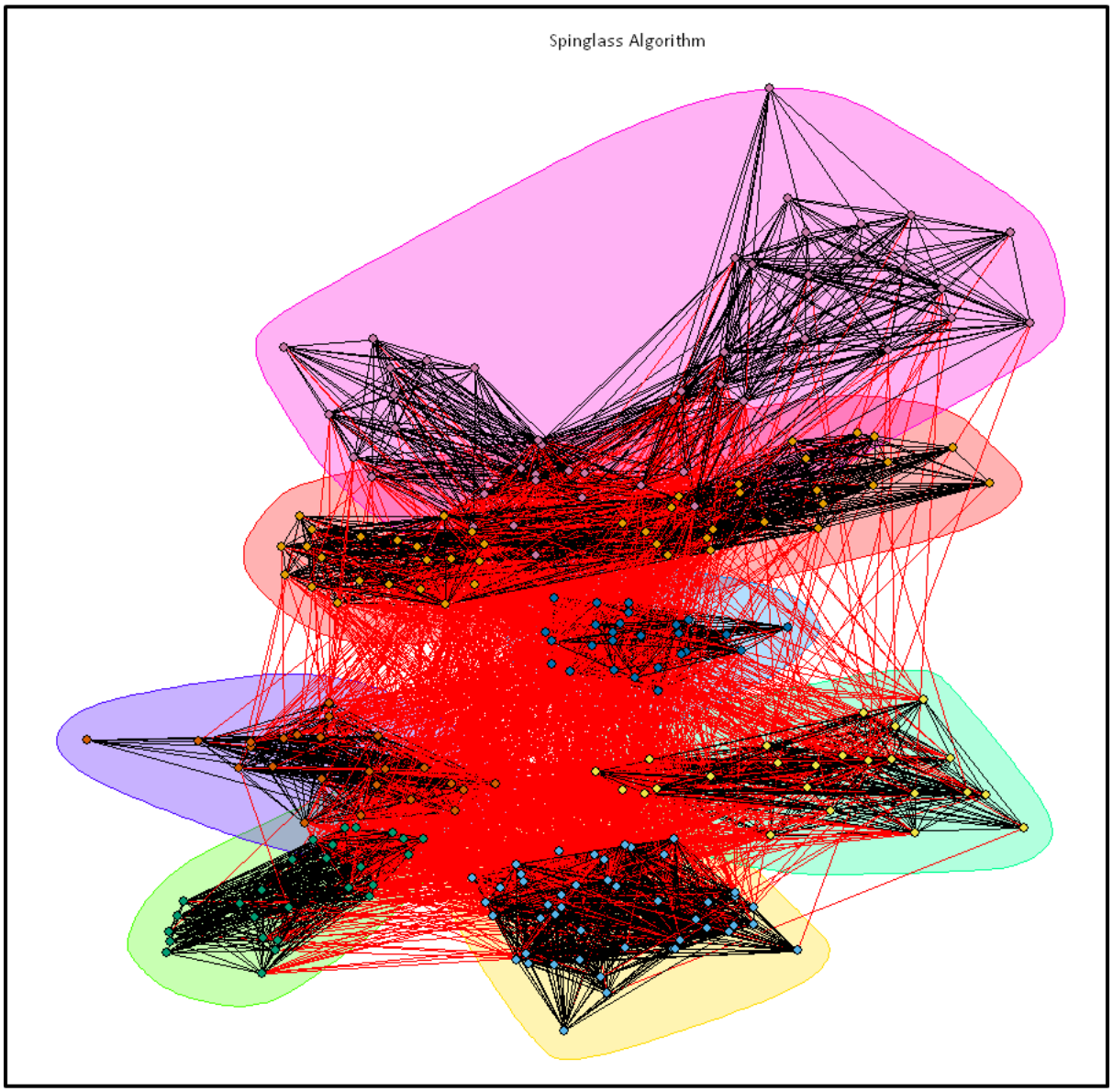
2) WalkTrap Algorithm

This algorithm finds densely connected subgraphs by performing random walks. The idea is that random walks will tend to stay inside communities instead of jumping to other communities.



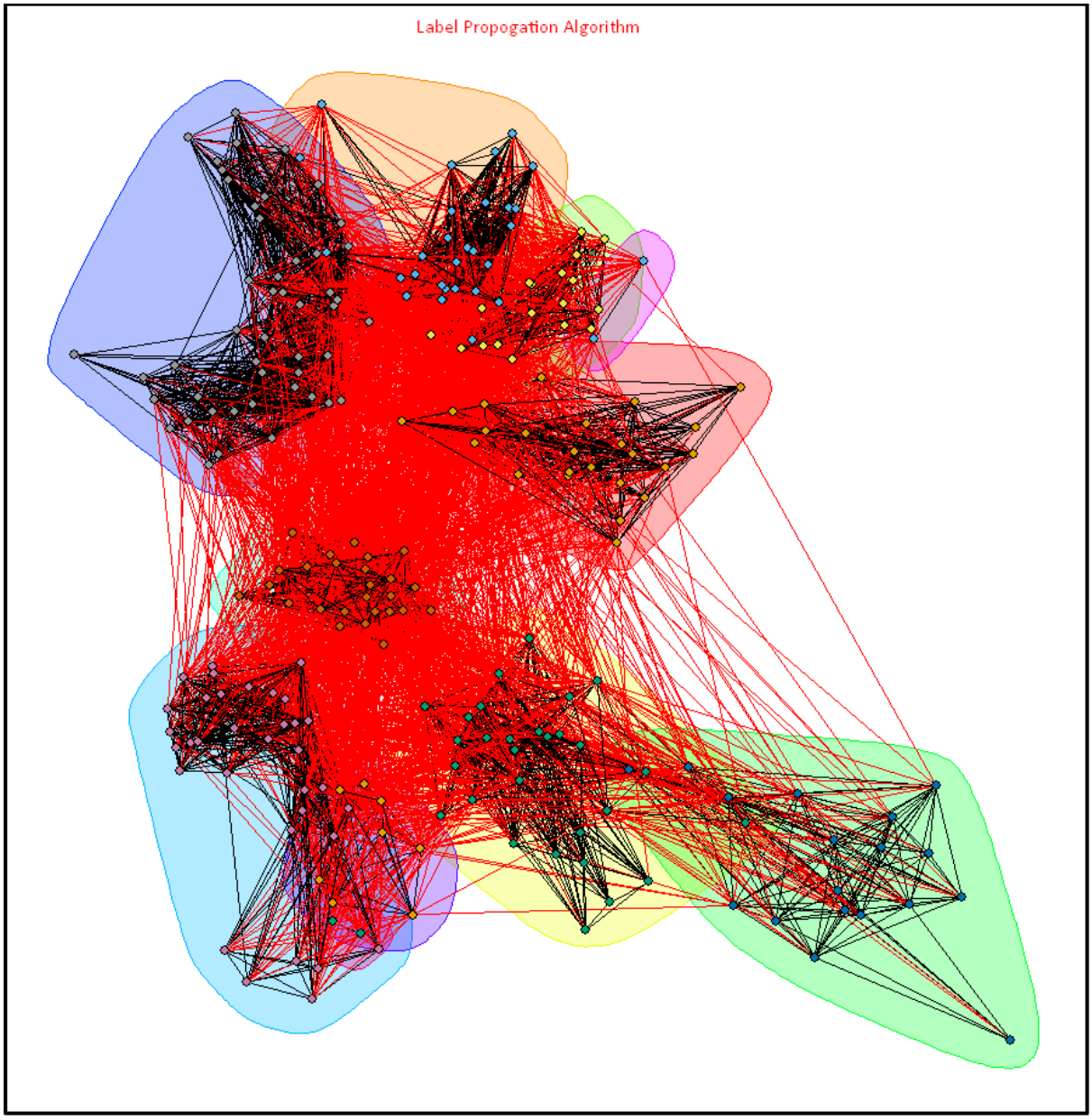
3) Spinglass Algorithm

This algorithm uses as spin-glass model and simulated annealing to find the communities inside a network



4) Label Propagation Algorithm

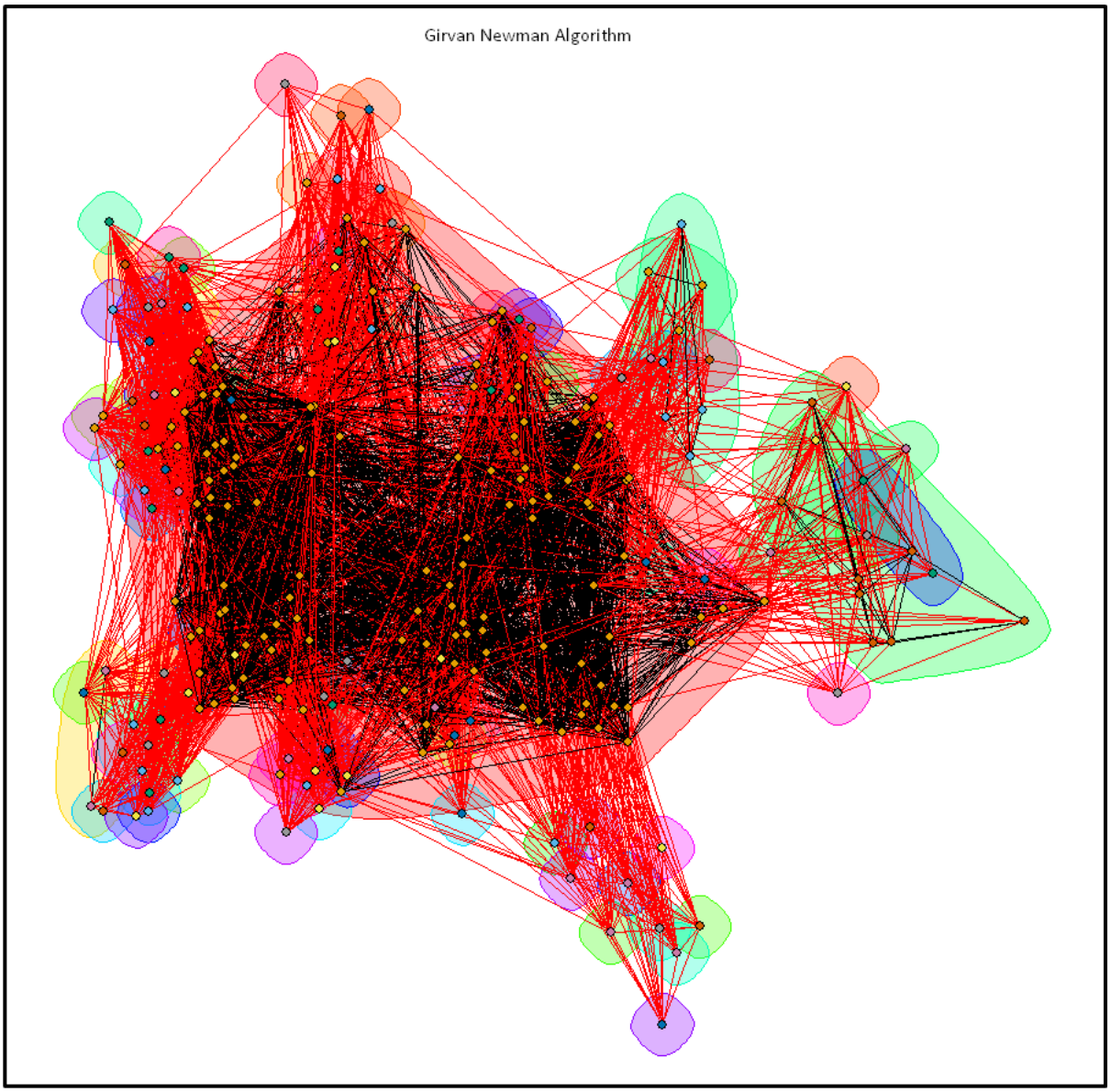
It works by labelling the vertices with unique labels and then updating the labels by majority voting in the neighbourhood of the vertex.



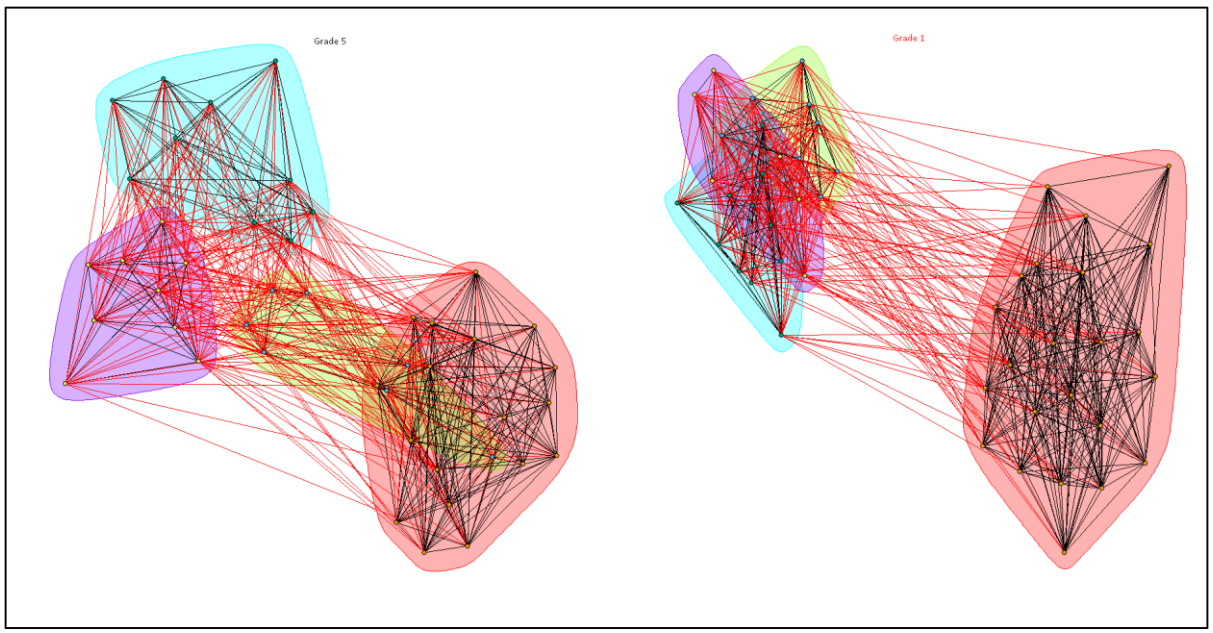
5) **Girvan Newman Algorithm**[Source]

The edge betweenness score of an edge measures the number of shortest paths through it. The idea of the edge betweenness based community structure detection is that it is likely that edges connecting separate structures have high edge betweenness as all the shortest paths from one module to another must travel through them.

This is important for our setting specifically as we know the data has clear partitions into sections and further into genders. Deleting the high betweenness edge would make intuitive sense as it would be like removing the friendship between students in 2 different sections and between males and females



Difference between Grade 1 and Grade 5:



It clearly shows that the nodes for grade 5 are more segregated. Which means that kids in 5th grade start getting an idea about gender and hence make closer friendships with their own kind

Girvan Newman algorithm shows that community number 42 has in common 1 teacher and 3 students from section 5B so we should dive deep into what is happening there. All the other communities suggested do not have in common teachers and children

Most of the p-values from girvan newman algorithm are 1