Network Analysis : London Tube

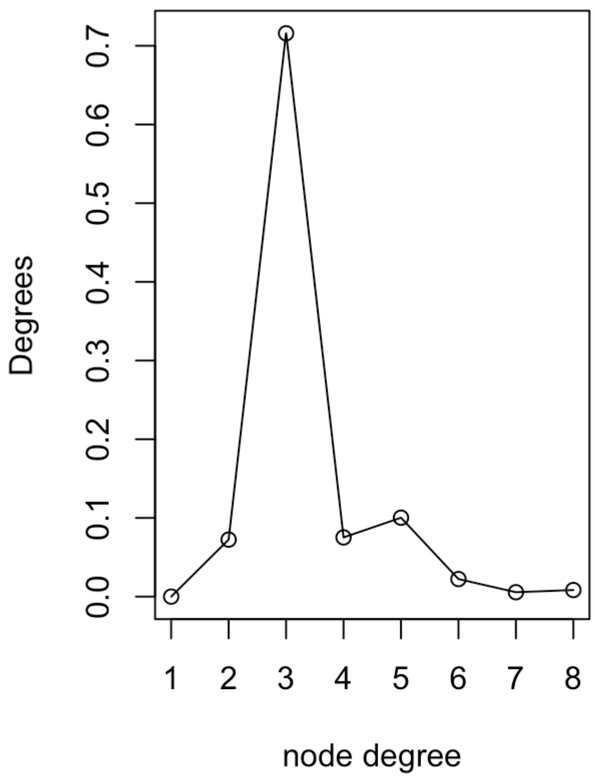
Kalyan and Vidal

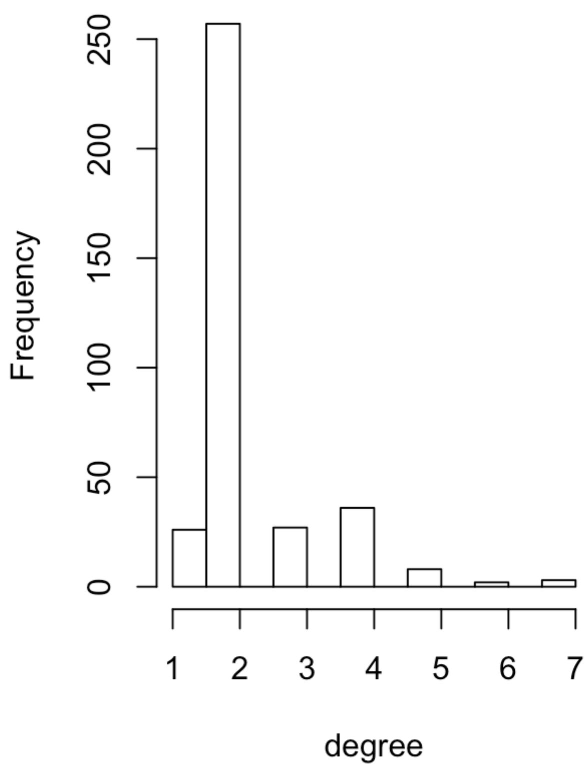
Introduction

The aim of this assignment is to analyse the `london tube graph`. We will be using `igraph` library to build our network. This data is a simple pairwise connection between metro stations. Our data is contains 419 observations. This parwise connection is a directed graph.

Analysis

Our analysis begun by trying to visulize the london tube graph. Our dataframe object had to be converted to graph data frame object. Below we observe that our degree distribution graph. We know that real world networks follow power distribution and we see some similarities.





According to our plots above we can conclude that our network more or less follows power law. This is a scale free network.

We know that real networks have a small diameter and high clustering co-efficient/ Transitivity. Transitivity measures the probability that the adjacent vertices of a vertex are connected. This is sometimes also called the clustering coefficient.

	Values
diameter	37.0000000

	Values
clustring_coefficient	0.0418994
average_path_length	13.1594620
total_vertices	359.0000000
total_edges	419.0000000
avg_degree	0.1250000

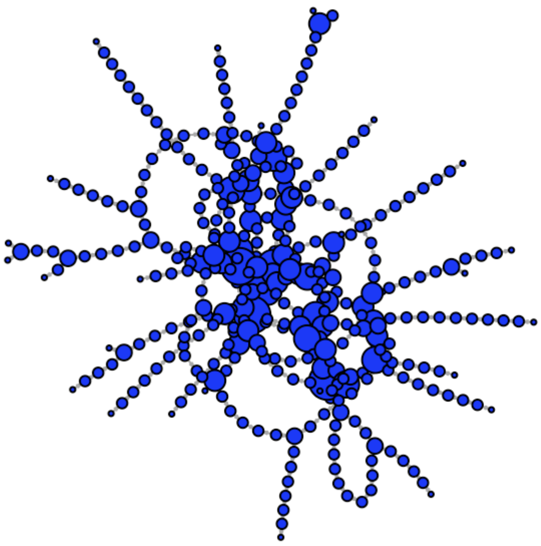
```
## [1] "phat estimation 0.0065"
```

```
## [1] "ER diameter estimation -2.82928001127435"
```

The value of `clustring_coefficient` `C` is `0.0418994` for our network. Estimated value of phat is `0.041`. We know that if `C >> phat` (probability to connect 2 nodes) then we can say that clustering coefficient is high. Hence this network cannot be an `ER` network as out network has high clustering co-efficient. We also `ER` networks have small diameter. According to our estimated diameter we get a negative value or roughly value of `log n` that is around `5`.

Graph Structure

Plotting the graph structure below we see a strucutre that a realword starion metro would be like. For example stations far out will be less connected that stations in the city.



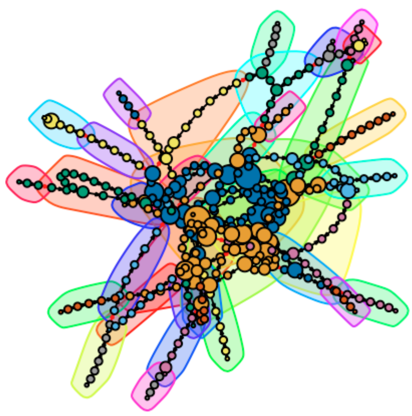
Communities

Below we show the visualization of communities formed using the `Walktap Community` and `Edge Betweenness Community`.

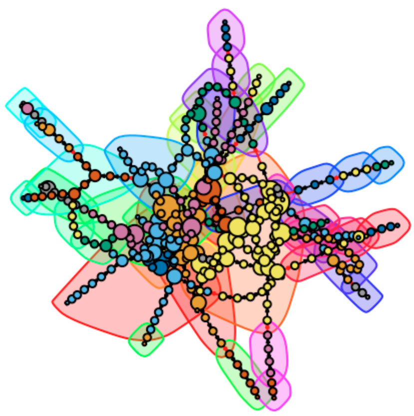
Walktrap Community is an approach based on random walks. The general idea is that if you perform random walks on the graph, then the walks are more likely to stay within the same community because there are only a few edges that lead outside a given community.

Edge Betweenness Community is a hierarchical decomposition process where edges are removed in the decreasing order of their edge betweenness scores (i.e. the number of shortest paths that pass through a given edge).

walktrap community



edge betweenness community



We will also analysie Modularity, It is one measure of the structure of networks or graphs. It was designed to measure the strength of division of a network into modules (also called groups, clusters or communities). We also calculate the total communities observed in nodes as Membership.

	walktrap	edge_betweenness
Modularity	0.7791508	0.7462392
Membership	36.0000000	45.0000000

Difficulties

We had issues with python and igraph basically installing the cairo library. Hence we had to shift to networkx. However we realized that networkx did not have good plotting libraries and finally used R with igraph.

Conclusion

We observed that walktrap community produced less clusters and better modularity compared to edge betweenness community.

- We have observed high clustering coefficient as seen in real networks.
- Our network is scale free and follows power law.
- We can also say that our network has hubs since it follows a power law (Nodes with high degree). This asserts our assumtion out a tube network which has lots of connected stations.
- Low value diameter value.
- Based on values of modularity and membership we can say communities represent close neighbouring metro station. We see a larger node size if the station is well connected.