**DSM160: Network Science Coursework**

**London Railway Network**

**Coursework assignment 1: Network Science**

1) Is this network weighted or unweighted? Is it directed or undirected? Please justify. **[5 points]**

The network is unweighted and undirected. The railway network is undirected because it contains unordered pairs of vertices, the trains move in both directions simultaneously so are bi-directional. The network is unweighted as it does not take into account the time taken to get from one station to the other by including the length, this data is found in the file london\_transport\_multiplex.edges in the column name = ‘weight’. If this data formed part of the analysis data, it would be a weighted network.

(2) Import the data into Python using the pandas library and convert it into a networkx graph. **[2.5 points]**

(3) What is the number of nodes and links in the network? **[2.5 points]**

Number of nodes = 369

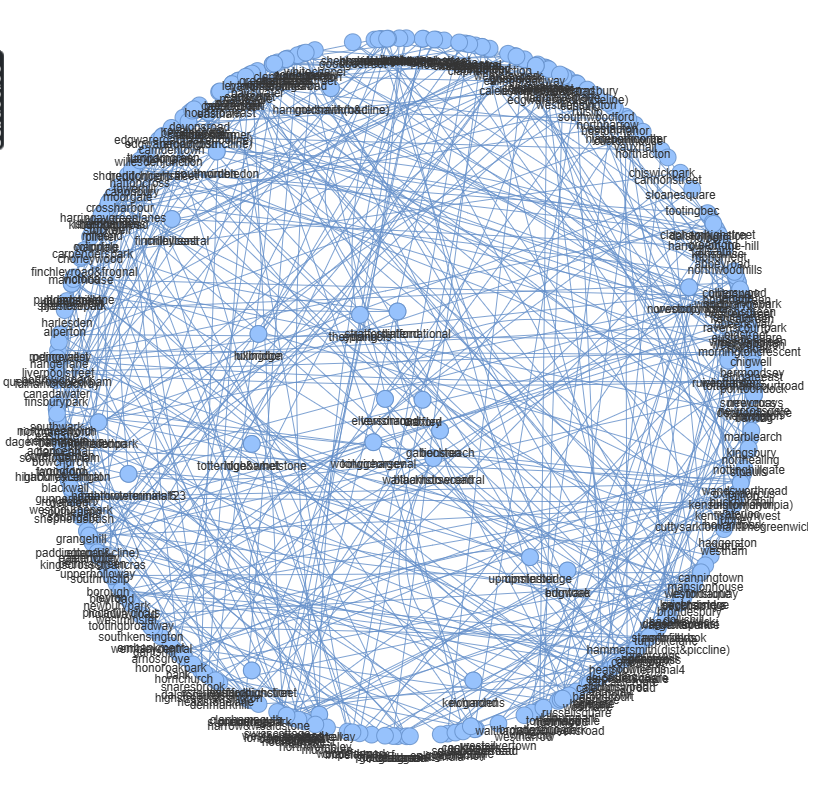
Number of links = 430

(4) Is the network sparse? Please justify. **[5 points]**

The network is sparse, the density is: 0.006333215506068104

The fewer the number of edges (links) in a network, the sparser the network will be. The number of links is smaller than the maximum possible number of links M.

(5) Plot the network using the library pyvis and include it as an image in your document. **[5 points]**



(6) Looking at your visualisation from item (5), what do you think would be the typical node degree of this network and why? **[5 points]**

Node degree: 2

Median = 2

Looking at the visual above it is evident that most nodes form around the edges and few are to be found in the centre. Therefore, the links are either between the centre to the edge or from one edge to another.

(7) What is the average degree of the network? How does it compare to your answer to item (6)? **[5 points]**

Average degree = 2.330623306233062

Average closeness centrality = 0.07700966107415387

Comparison between the node degree is: 0.3306

(8) Does the network have a core-periphery structure or a structure in which hubs are situated at the centre of star-like components? Please justify. **[5 points]**

The network has a core-periphery structure. The core contains important central stations, and the periphery has most of the connecting stations. The core stations are well connected as are the periphery ones are well connected to the core. There are sparse internal connections though.

(9) How many connected components does the network have? **[5 points]**

Number of connected components = 1

(10) What are the top 10 stations in terms of degree, closeness and betweenness centrality? **[5 points]**

By degree:

'bakerstreet',

'kingscrossstpancras',

'stratford',

'westham',

'bank',

'earlscourt',

'paddington',

'greenpark',

'oxfordcircus',

'canningtown']

By betweenness:

['bank',

'waterloo',

'kingscrossstpancras',

'greenpark',

'bakerstreet',

'euston',

'stratford',

'westminster',

'finchleyroad',

'liverpoolstreet']

By closeness:

['bank',

'waterloo',

'kingscrossstpancras',

'greenpark',

'bakerstreet',

'euston',

'stratford',

'westminster',

'finchleyroad',

'liverpoolstreet']

(11) Which stations are central according to all three centrality measures? **[5 points]**

Top 10 nodes by eigenvector centrality:

('oxfordcircus', 0.3968157978956346)

('greenpark', 0.3903624311661707)

('piccadillycircus', 0.30975629560797274)

('bondstreet', 0.28355136470871556)

('westminster', 0.22314485167327802)

('tottenhamcourtroad', 0.20996506946345467)

('bakerstreet', 0.2062259440657922)

('leicestersquare', 0.2059623772097717)

('waterloo', 0.18714423971611505)

('charingcross', 0.18061185439308267)

(12) Using the stations identified in item (10), which stations with high degree centrality are not betweenness-central? How do you interpret the role of such stations in the network? **[10 points]**

*(Hint: You may convert the list of stations according to each centrality metric into sets, using the set() function. Sets in Python have union, intersection and difference methods, allowing to combine sets in different ways)*

The stations with a high degree centrality that are not betweenness central are important in the transportation of information (or people) between important nodes. This occurs if almost all stations are common to the main route.

(13) What is the proportion of stations where, at least, two different (train) lines intersect? **[5 points]**

*(Hint: Think what the node’s degree needs to be to have an intersection between lines)*

The proportion of nodes with at least 2 train lines intersect is 0.92 (92%)

(14) Build and plot the node degree distribution for this network. How does it relate to your answer to item (6)? **[5 points]**

*(Hint: given the range of possible degree values, you do not need to transform the x-axis to log-scale)*

(15) How many steps does it take, on average, to go from one station to another using shortest paths? **[5 points]**

(16) What is the shortest path to go from the first to the second most central station by degree? What is the path length? **[5 points]**

(17) Amongst the top 5 stations by betweenness centrality, which are directly connected or have at most one intermediate station between them? **[10 points]**

(18) You want to perform some robustness tests on this transport network by simulating attacks which disrupt one station at a time. In particular, you want to compare the vulnerability of the network when nodes are removed one by one in decreasing order of (i) their degree, as opposed to (ii) their betweenness centrality. Which strategy – (i) or (ii) – will be faster in reducing to less than 80% the proportion of nodes in the (largest) connected component? Please justify. **[10 points]**

*(Hint: You may simulate each type of targeted attack removing one node at a time – rather than a proportion of nodes – and plot the results comparing (i) and (ii) using the log-scale for the x-axis)*