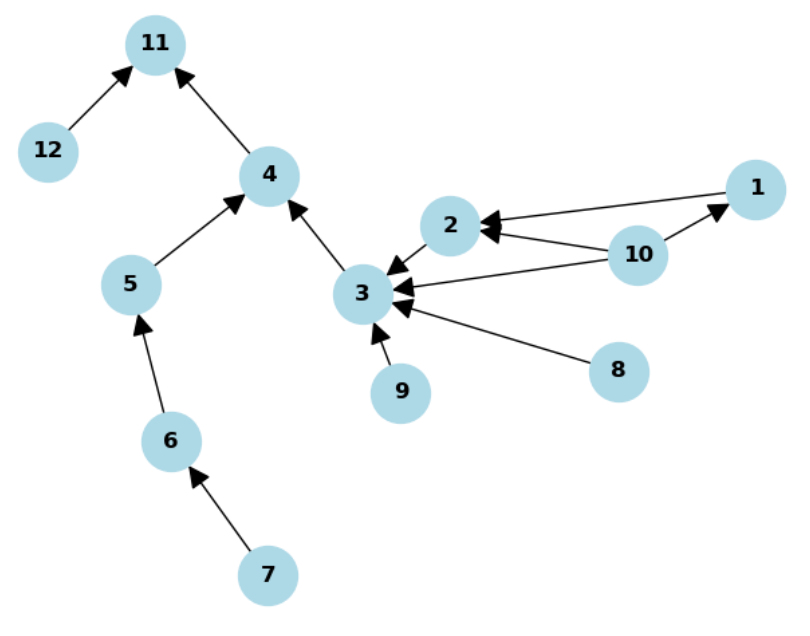
**CS5228 Tutorial 5 – Graphs**

**Q1: Centrality Measures on a Directed Graph**

The centrality of a node in a graph is a way of measuring its importance among all other nodes w.r.t. the graph structure. Figure 1 shows a directed graph G.

For a directed graph, we can compute closeness of a node v as the maximum length of paths from other nodes **to the node v**, and the betweenness is defined similarly to the undirected case.

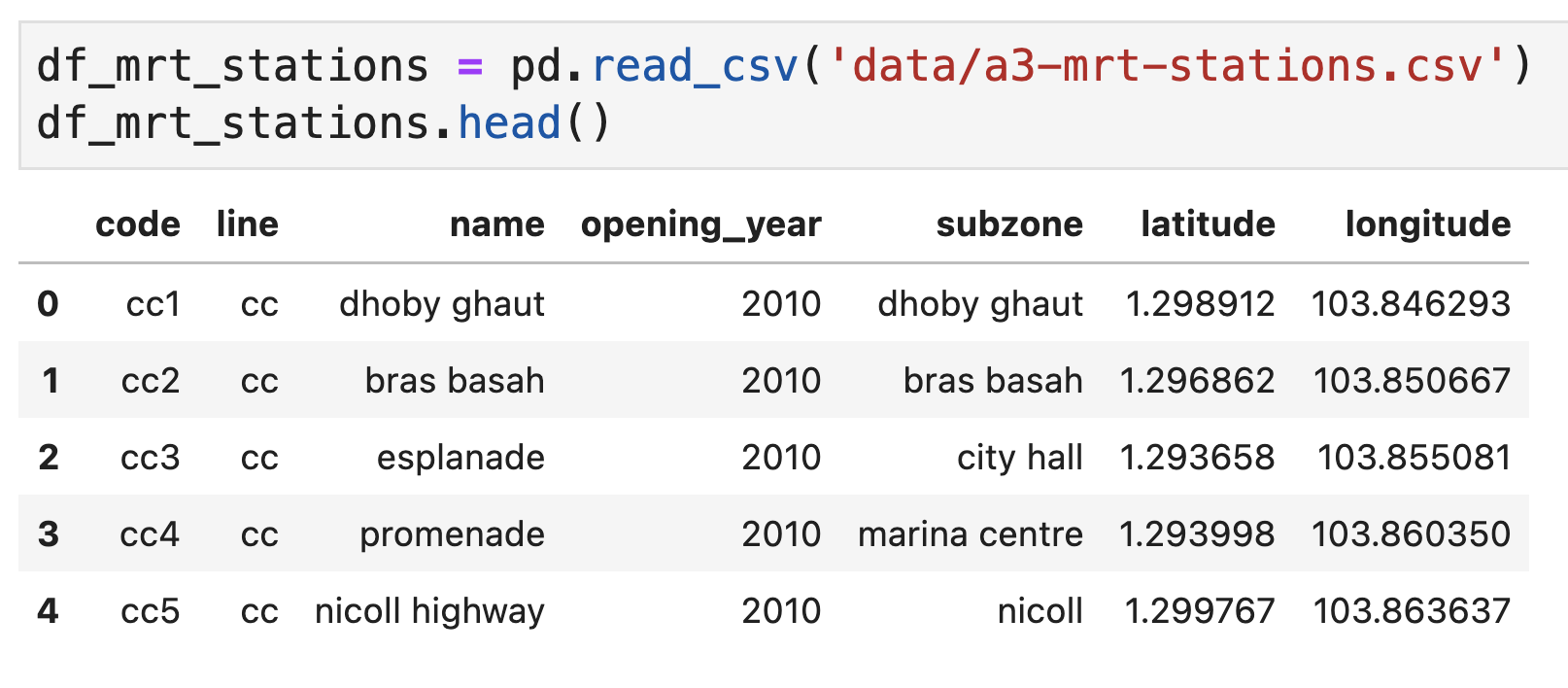


Simply by eye-balling graph G try to identify the nodes with the highest score according to the 4 centrality measures:

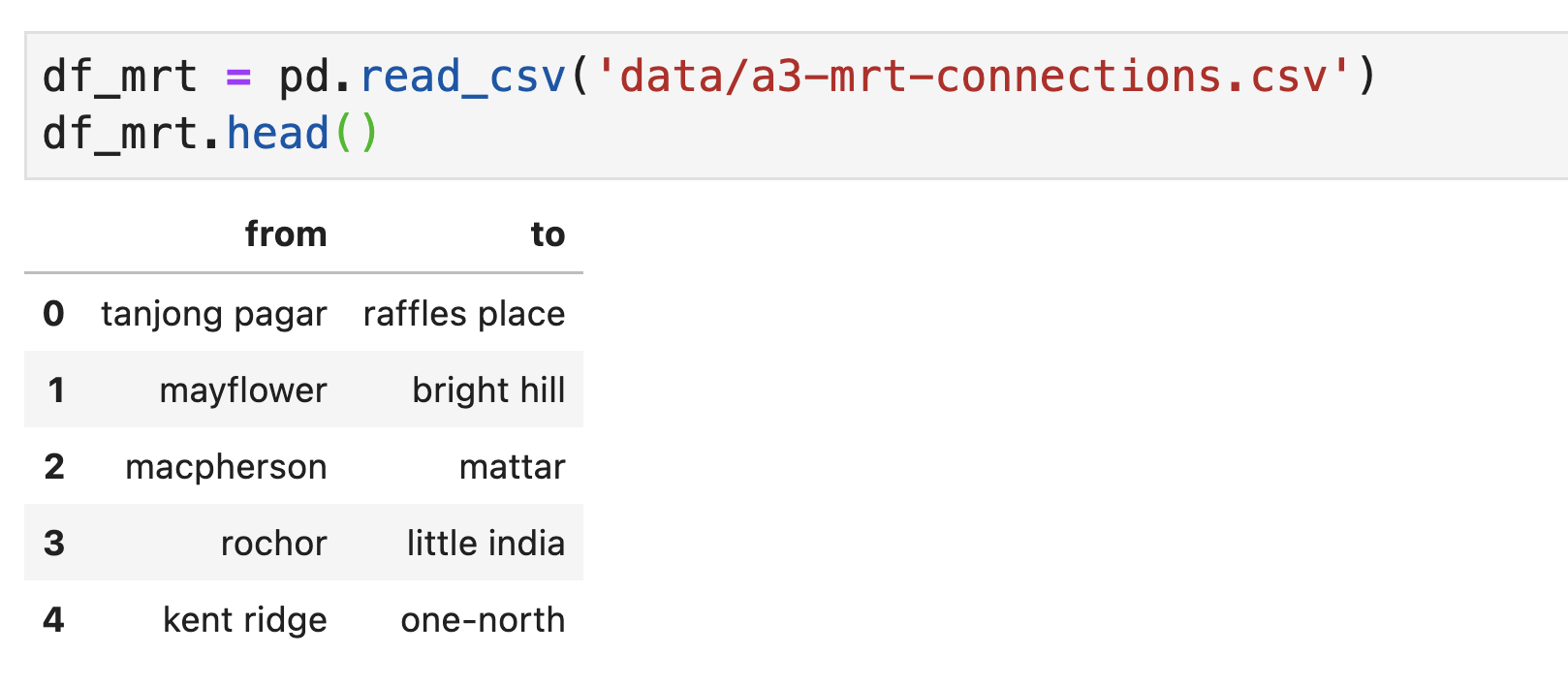
* Out-degree
* In-degree
* Closeness
* Betweenness
* Out-degree: 10 (the only node with more than 1 outgoing edge)
* In-degree: 3 (the node with most in-coming edges)
* Closeness: 4 (Merging point of 2 subgraphs, [4,5,6,7] and [1,2,3,8,9,10]), making it the easiest to reach on average)
* Betweenness: 3 (benefits from its high in-degree and central position in the graph)

**Q2: Centrality Measures on MRT Map**

We have a dataset of MRT stations, which are nodes:



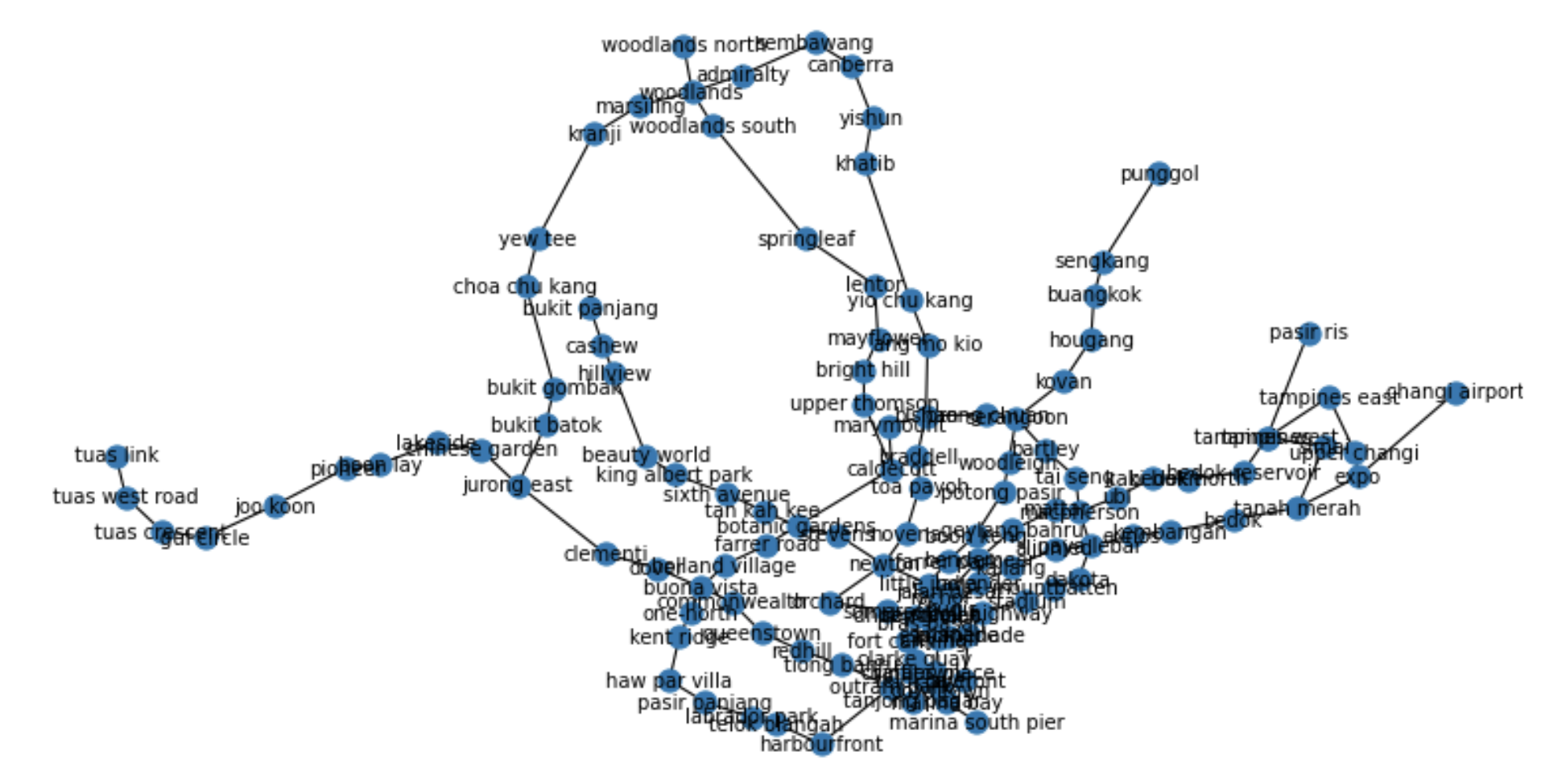
We also have the connections between them, which are edges:



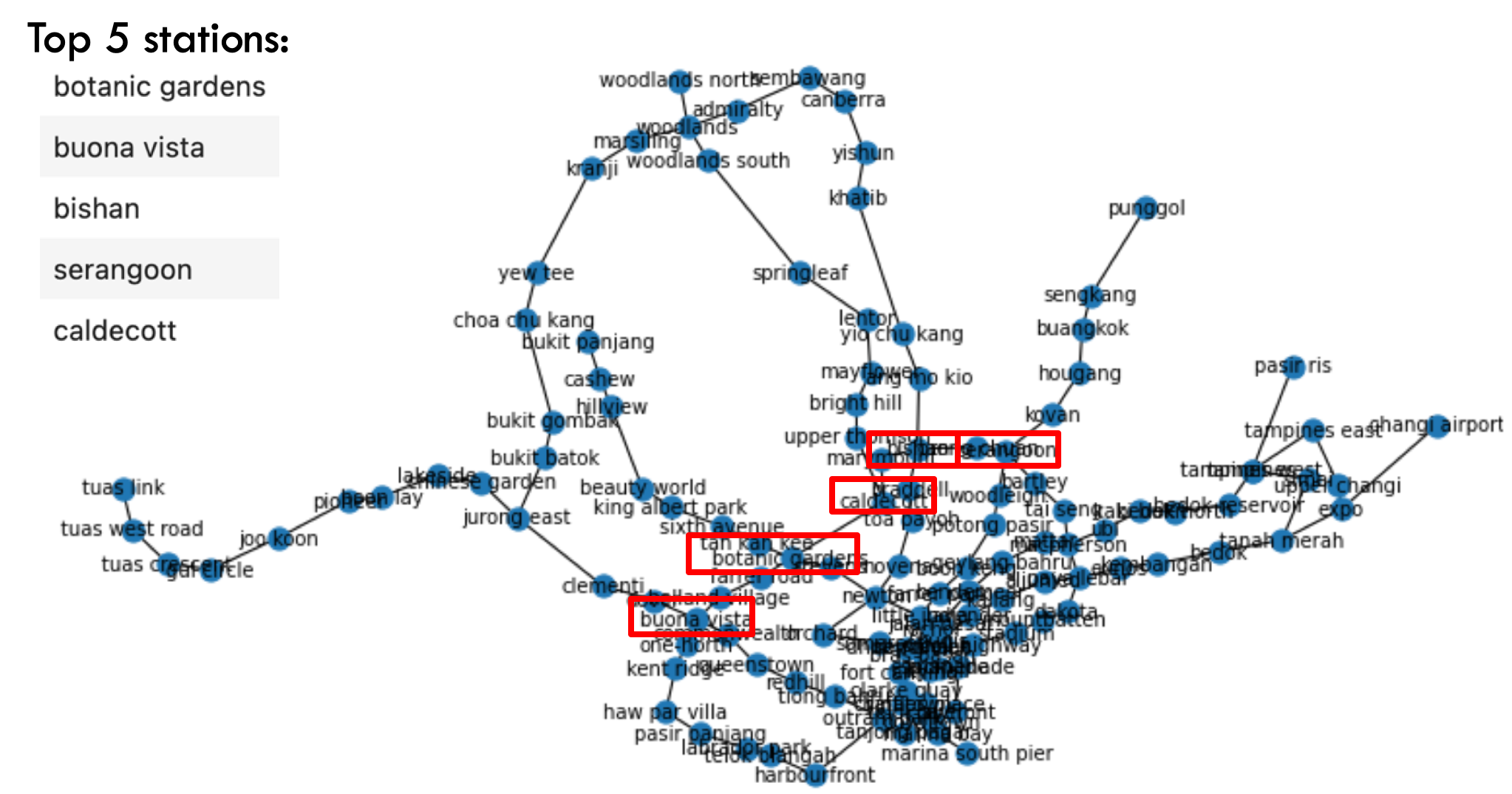
Converting this to a networkx graph:



Finally, we can visualize the graph (using the “spring\_layout” layout in networkx), producing the following result:

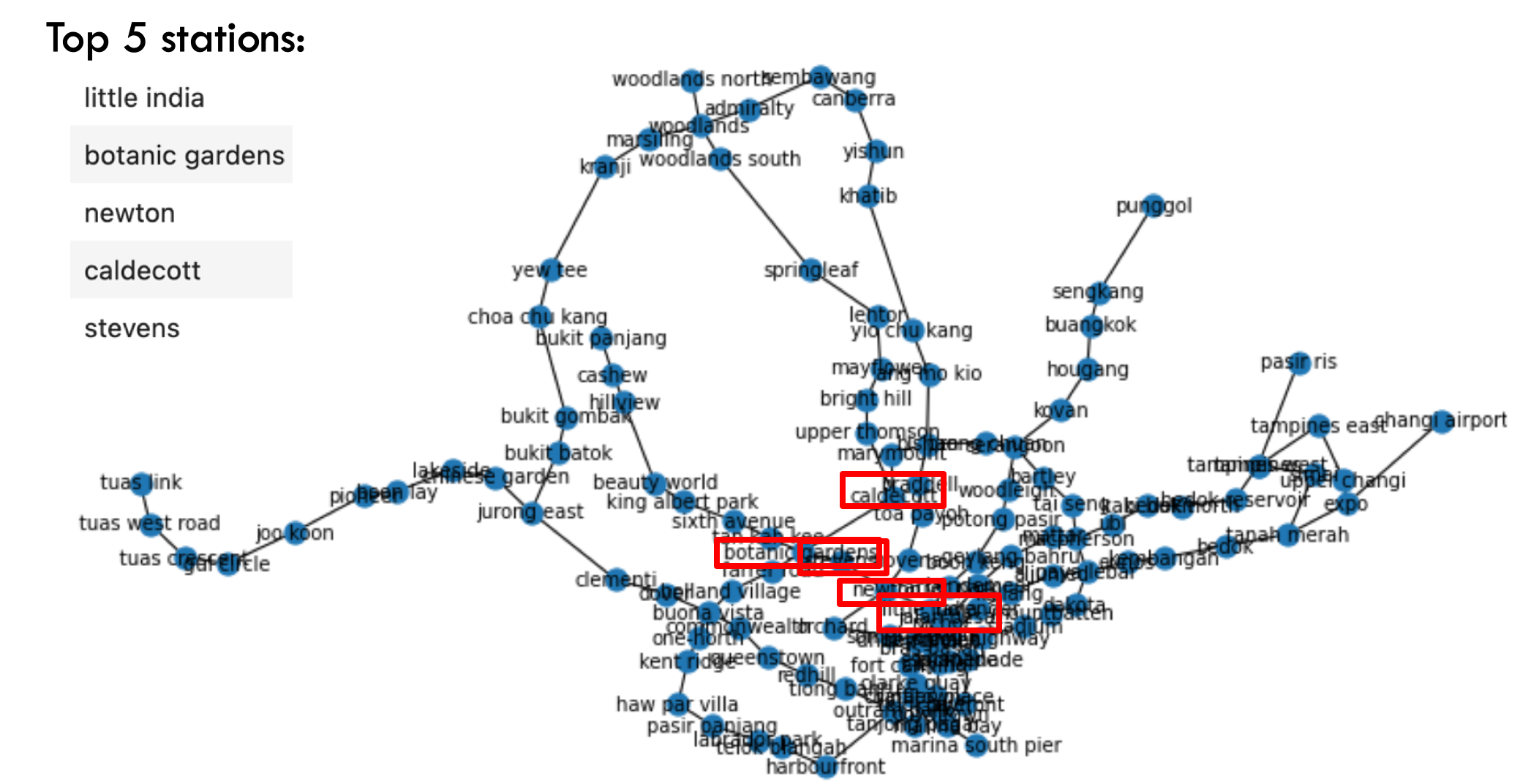


**2a)** We have 3 centrality measures: degree, centrality, betweenness. For one measure, the top 5 highest centrality stations are as follows. Which measure is it?



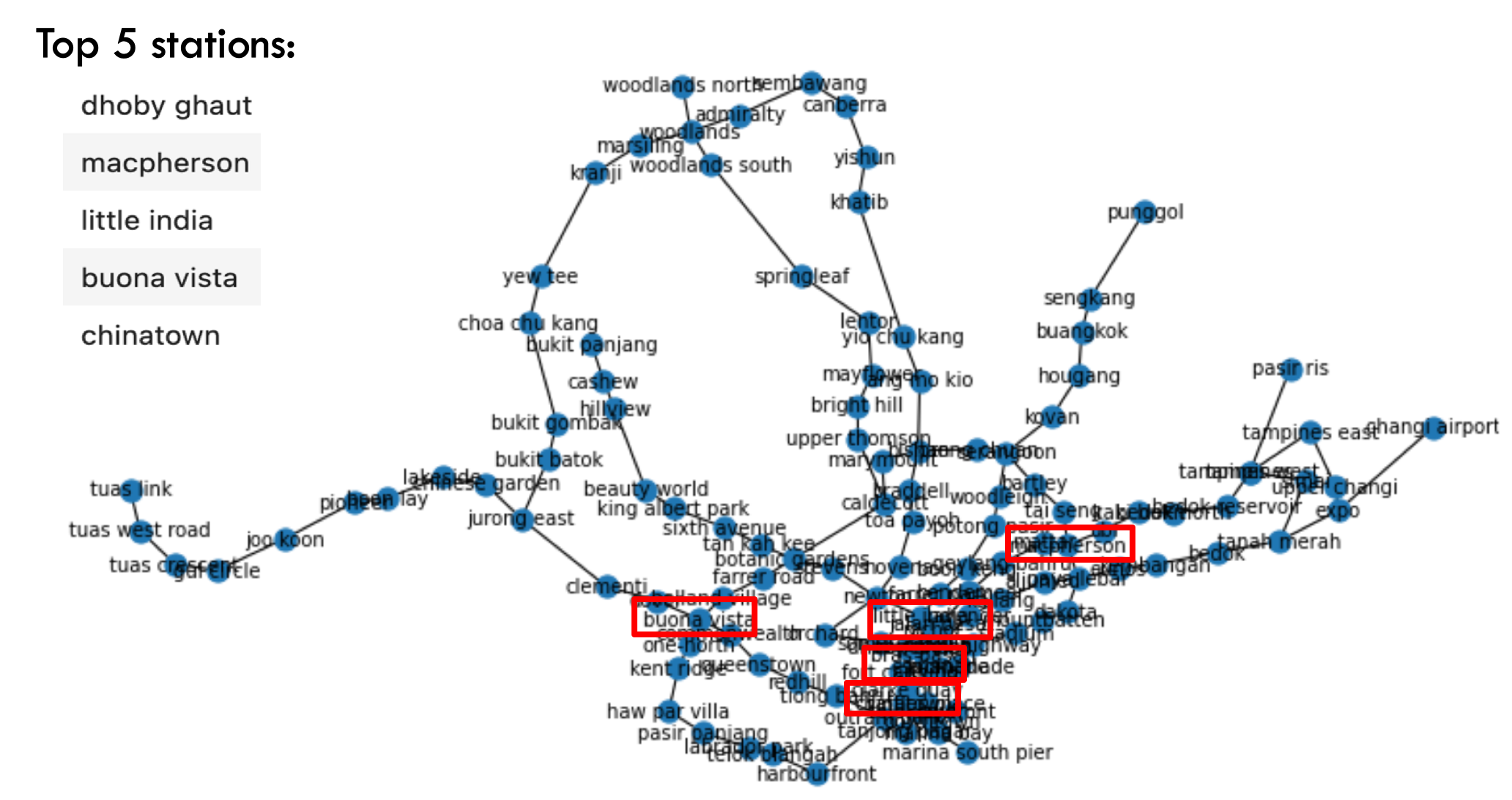
Betweenness. The nodes tend to be near the center and also ‘bridge nodes’ connecting regions of the graph, thus lying on many shortest paths.

**2b)** For the next centrality measure, the top 5 highest centrality stations are as follows. Which centrality measure is it?



Closeness. The nodes are all in the central region of the graph, thus having shorter distances to every other node.

**2c)** Finally, the last centrality measure (which can be determined by elimination) is:



Degree. The nodes have degree at least 4.

**2d)** For each of closeness and betweenness, come up with 1 concrete application scenario utilizing that centrality measure on the MRT graph, and discuss why that centrality measure is suitable.

Closeness: MRT stations with the highest closeness are those that on average can be reached the fastest from any other MRT station. For example, if we wanted to build a **hospital (or other similar facility)**, it makes sense to locate it on an MRT station with high closeness, so that it can be reached quickly from any other station.

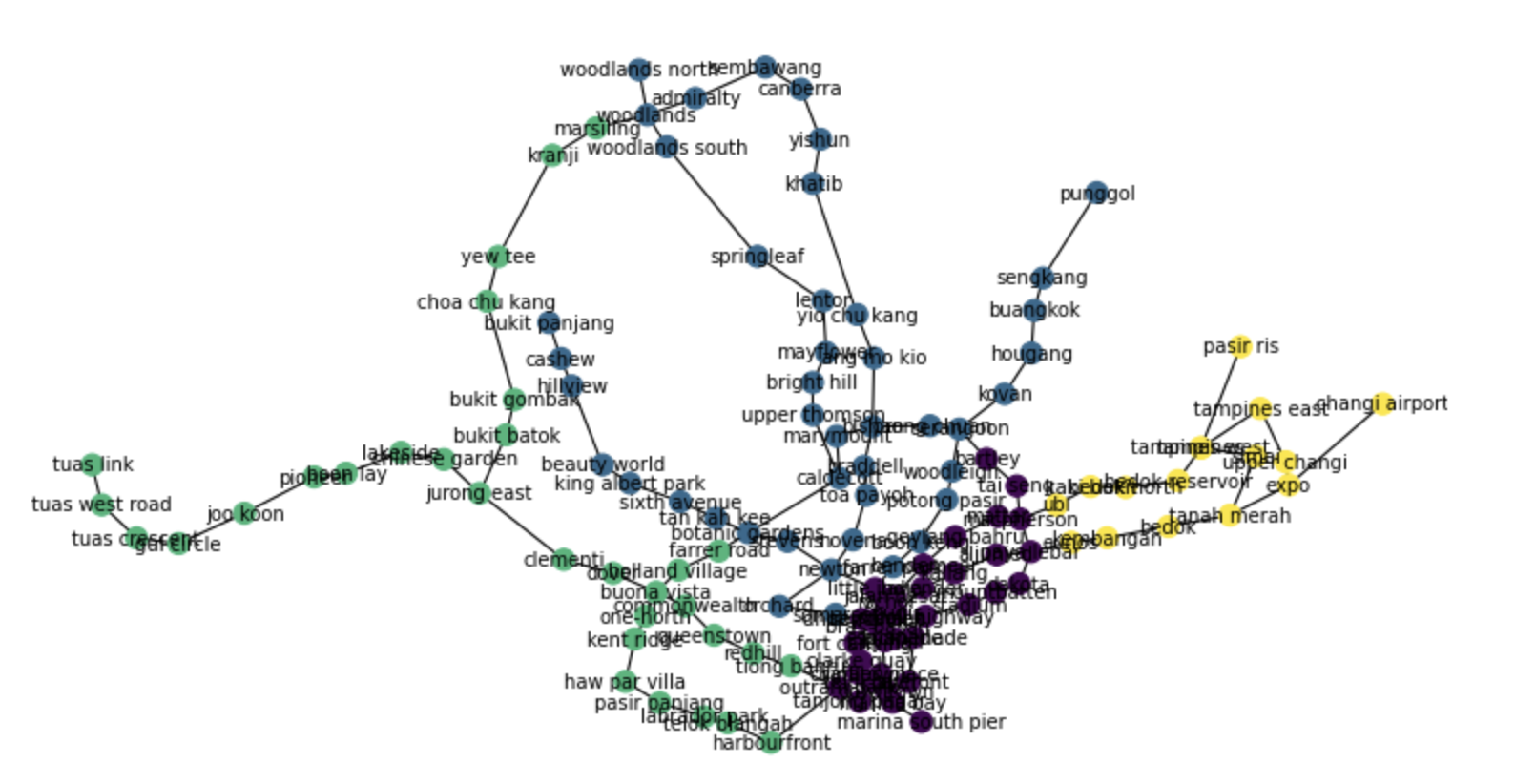
Betweenness: MRT stations with high betweenness are those that get passed most often when travelling from an origin to a destination MRT station (along shortest paths). As such, **interruptions at these stations should be particularly avoided / quickly fixed.**

**2e)** Most likely, the suitability of the centrality measures in the previous question’s scenarios are likely to be over-simplifications. For each of your scenarios, discuss one practical limitation of using that centrality measure in the scenarios you have come up with.

Closeness: this only considers the case of a single hospital / facility.

Betweenness: assumes that each rider has a randomly sampled origin and destination, which is a definite oversimplification. Weighting based on ridership statistics would be more accurate.

**2f)** We can run the Girvan-Newman algorithm to obtain “communities” of the nodes in this MRT map. The Girvan-Newman Algorithm iteratively removes a minimum set of edges until the graph breaks into 2 components. Edge with the highest Edge Betweenness Centrality are removed first. The resulting communities are as follows:

****

Do you think these communities make sense? We can see some rather sparse and large clusters (e.g. green) and some tightly connected and smaller clusters (e.g. black). Why do you think this is the case?

Generally, regions in the outskirts have less shortest paths going through them, so their betweenness score is lower, and are less likely to be cut. This results in smaller communities in the dense central region, and larger communities in the outskirts.