1 Question 1

As the graph G is composed of two connected components, we just have to count the number of edges in each connected component:

- The first component is a complete graph with 100 vertices, therefore there are $\sum_{k=0}^{100-1} k = \frac{99*100}{2} = 4950$ edges in that component.
- In the second component, there are two partition set, with 50 nodes each which are linked to all the nodes of the other partition set. Therefore there are $50 * 50 = 2500$ edges in that component. (It can be seen as a linear layer of neural network without bias)

So, in total there are **7450** edges in the graph.

Similarly, we count the number of triangles in each component:

- A triangle is just a triplet, therefore, counting the number of triangles is equivalent to counting the number of triplets in a set of 100 nodes: this is exactly $\binom{100}{3} = \frac{100*99*98}{6} = 161,700$ triangles.
- In the complete bipartite component, since there are no edges between the nodes from the same partition set, there are no triangles.

So, in total there are **161,700** triangles in G.

2 Question 2

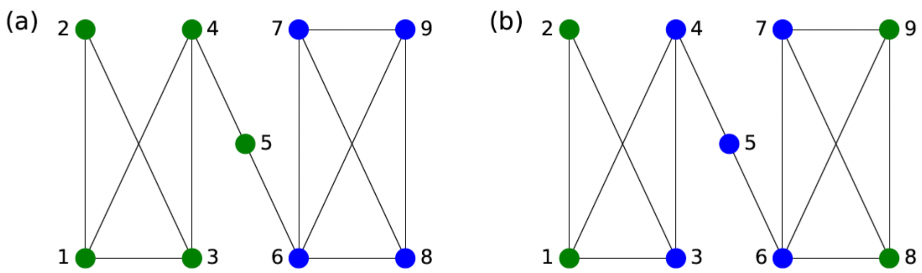

Figure 1: Two graphs where nodes have been assigned to 2 clusters. Cluster membership is indicated by node color.

In both graph (a) and (b), $m = 13$, $n_c = 2$,

- For the graph (a), $l_{green} = 6$, $l_{blue} = 6$, $d_{green} = 13$ and $d_{blue} = 13$. Then,

$$Q = (\frac{6}{13} - (\frac{13}{26})^2) + (\frac{6}{13} - (\frac{13}{26})^2) = \frac{11}{26} \approx 0.42$$

- For the graph (b), $l_{green} = 2$, $l_{blue} = 4$, $d_{green} = 11$ and $d_{blue} = 15$. Then,

$$Q = (\frac{2}{13} - (\frac{11}{26})^2) + (\frac{4}{13} - (\frac{15}{26})^2) = -\frac{17}{338} \approx -0.05$$

This results seems to be coherent with the natural idea that the clustering from graph (a) is better than the one from graph (b).

3 Question 3

If we note k the shortest path kernel, then there exists a function ϕ such that: $k(G_1, G_2) = \langle \phi(G_1), \phi(G_2) \rangle$. The function ϕ counts the frequency of shortest path distance in the graph i.e. $\phi(G)[i]$ counts the number of shortest paths of distance $i+1$ in the graph.

Therefore, $\phi(P_4) = [3, 2, 1]$ and $\phi(C_4) = [4, 4, 0]$.

Finally, $k(C_4, C_4) = 32$, $k(C_4, P_4) = 20$ and $k(P_4, P_4) = 14$.

4 Question 4

As the elements of the vectors $f_G, f_{G'}$ are all non negative, if the kernel value : $k(G, G') = f_G^T f_{G'} = 0$, then it means that there are no types of graphlets of size 3 that both graphs G and G' share.

If we take $G_1 = C_3$, a complete graph with 3 nodes and $G_2 = P_3$, a path graph with 3 nodes. Then, $f_G = [1, 0, 0, 0]$ and $f'_G = [0, 1, 0, 0]$, therefore $k(G, G') = f_G^T f_{G'} = 0$, as both graphs don't share a common 3-node subgraph.