1 Question 1

The number of edges in an unoriented complete graph on 100 vertices is $\binom{100}{2}$.
The number of edges in an unoriented complete bipartite graph with 50 vertices in each partition set is 50×50.
And since there are no edges between the two connected components of G, the number of edges in G is:
$\binom{100}{2} + 50 \times 50 = \frac{100 \times 99}{2} + 50 \times 50 = \boxed{7450 \text{ edges}}$.

A triangle in G is given by 3 points which belong to the same connected component.
For the connected component which is a complete graph on 100 vertices, there are $\binom{100}{3}$ triangles.
For the connected component which is a complete bipartite graph, given three distinct vertices, they do not form the triangle: two of the vertices necessarily belong to the same partition, and there is therefore no edge between them.
So there are $\boxed{\binom{100}{3} = 161,700 \text{ triangles in } G}$.

2 Question 2

- Graph 1.(a): Here, $m = |E| = 13$, $n_c = 2$.
  For the green cluster, $l_c = 6$, and $d_c = 3 + 2 + 3 + 3 + 2 = 13$.
  For the blue cluster, $l_c = 6$, and $d_c = 4 + 3 + 3 + 3 = 13$.
  Therefore $Q = \left(\frac{6}{13} - \left(\frac{13}{2 \times 13}\right)^2\right) + \left(\frac{6}{13} - \left(\frac{13}{2 \times 13}\right)^2\right) = \boxed{\frac{11}{26}} \approx 0.423$.
- Graph 1.(b): Here, $m = |E| = 13$, $n_c = 2$.
  For the green cluster, $l_c = 2$, and $d_c = 3 + 2 + 3 + 3 = 11$.
  For the blue cluster, $l_c = 4$, and $d_c = 3 + 3 + 2 + 4 + 3 = 15$.
  Therefore $Q = \left(\frac{2}{13} - \left(\frac{11}{2 \times 13}\right)^2\right) + \left(\frac{4}{13} - \left(\frac{15}{2 \times 13}\right)^2\right) = \boxed{\frac{-17}{338}} \approx -0.050$.

So the modularity of clustering (a) is higher than the modularity of clustering (b). This was expected, since the clustering in (b) is obviously not good.

3 Question 3

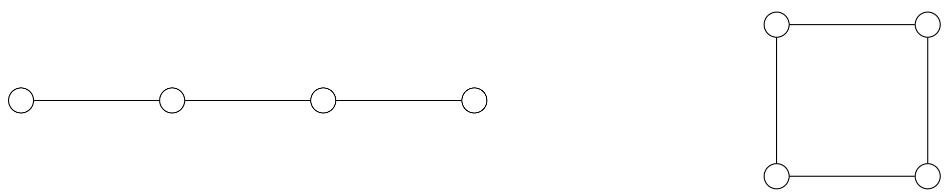

Figure 1: P_4 (left) and C_4 (right)

C_4 and P_4 can be seen in Fig. 1.

We have $\phi(C_4) = [4, 2, 0, 0]$ (between two opposite vertices, there are 2 shortest paths of length 2, but only one of them counts in $\phi(C_4)$, as can be seen in [1]), and $\phi(P_4) = [3, 2, 1, 0]$.

In the code, paths of longer 0 are counted.

Then, the shortest path kernel is obtained by computing the scalar products:

- For (C_4, C_4): $\phi(C_4)^T\phi(C_4) = 20$.
- For (C_4, P_4): $\phi(C_4)^T\phi(P_4) = 16$.
- For (P_4, P_4): $\phi(P_4)^T\phi(P_4) = 14$.

4 Question 4

A kernel value equal to 0 means that G **and** G' **have no graphlet in common**: for each $i \in \{1, 2, 3, 4\}$, we have $(f_G)_i = 0$ or $(f_{G'})_i = 0$.
We can consider the example in Fig. 2.

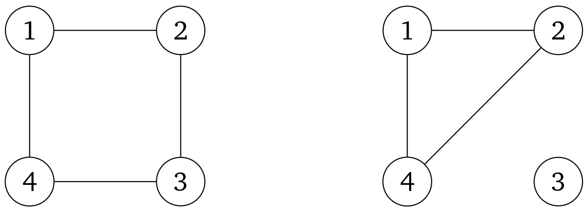

Figure 2: G and G' such that $k(G, G') = 0$

Indeed, we have $f_G = (0, 4, 0, 0)^T$ and $f_{G'} = (1, 0, 3, 0)^T$, and therefore $k(G, G') = f_G^T f_{G'} = 0$.

References

[1] K.M. Borgwardt and H.P. Kriegel. Shortest-path kernels on graphs. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*, pages 8 pp.–, 2005.