Question 1

The graph G consists of two connected components:

- A complete graph K_{100} with 100 vertices, which has $\frac{100 \times 99}{2} = 4950$ edges.

- A complete bipartite graph $K_{50,50}$ with two sets of 50 vertices, which has $50 \times 50 = 2500$ edges.

Thus, the total number of edges in G is $4950 + 2500 = 7450$.

For the number of triangles:

- The complete graph K_{100} has $\binom{100}{3} = 161700$ triangles.

- The complete bipartite graph $K_{50,50}$ has no triangles.

Therefore, the total number of triangles in G is 161700.

Question 2

The formula is:

$$Q = \sum_{c=1}^{nc} \left(\frac{l_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right)$$

- For clustering A:
$$\text{mod}_1 = \left(\frac{6}{13} - \left(\frac{13}{2 \times 13} \right)^2 \right) + \left(\frac{6}{13} - \left(\frac{13}{2 \times 13} \right)^2 \right) = 0.423$$

- For clustering B:
$$\text{mod}_2 = \left(\frac{2}{13} - \left(\frac{11}{2 \times 13} \right)^2 \right) + \left(\frac{4}{13} - \left(\frac{15}{2 \times 13} \right)^2 \right) = 0.050$$

The modularity for clustering B is very poor and close to random as the modularity is close to 0 whereas the clustering A is more relevant with a modularity closer to 1.

Question 3

Here are the feature vectors for C4 and P4 based on the frequencies of their shortest paths:

- C4: [0, 4, 2, 2]

- P4: [0, 3, 2, 1]

$$\text{Kernel}(C4, C4) = \text{dot}([0, 4, 2, 2], [0, 4, 2, 2]) = 24$$

$$\text{Kernel}(C4, P4) = \text{dot}([0, 4, 2, 2], [0, 3, 2, 1]) = 18$$

$$\text{Kernel}(P4, P4) = \text{dot}([0, 3, 2, 1], [0, 3, 2, 1]) = 14$$

The kernel value for (C4, C4) is the highest, suggesting that the two cycle graphs are the most similar in terms of their shortest path distributions.(C4, P4) has a lower kernel value, indicating that the cycle and path graphs are less similar. (P4, P4) has the lowest kernel value, as two path graphs are less structurally similar compared to a cycle and a path.

Question 4

A kernel value of 0 means that the two graphs are dissimilar according to the kernel, in this case, the graphlet kernel. Specifically, the graphlet kernel measures the similarity between two graphs based on the number of matching subgraphs (graphlets of size 3) they share. If the kernel value is 0, this indicates that the two graphs do not share any common graphlets of size 3.

Let's consider the following graphs:

- G: A Path graph (P3) with 3 nodes, connected as a straight line (1-2-3).

- G: A Star graph with nodes without edges (1, 2, 3).

In this case:

- The feature vector for G would be: '[0, 0, 1]'.

- The feature vector for G' would be: '[1, 0, 0]'.

Therefore, the dot product of these feature vectors would be 0.