Question 1

First, let's study the number of edges in each connected components.

- The first connected component is a complete graph on 100 vertices, it hence has $99 + 98 + ... + 2 + 1$ edges (as there is no inner loop) which equals to $\frac{(99*100)}{2}$ = **4950 edges**.
- The second connected component is a complete bipartite graph with 50 vertices in each partition set which means that there is $50 * 50$ =**2500 edges** because each of the 50 nodes of a partition is linked to the 50 nodes of the other partition.

The total number of edges is therefore **7450**.
Now let's compute the number of triangle in each connected components.

- The number of triangles in the first connected component is related to the binomial coefficient and is equal to $\binom{100}{3}$ = **161700 triangles** as we look for triangles that are composed of 3 nodes among a graph that is composed of 100 nodes.
- The second connected component however **does not have any triangle** because each node of each partition is connected to each node of the other partition but as nodes within each partition are not connected together, there cannot exist any triangle.

Therefore, the number of triangle in the graph is **161700**.

Question 2

First, let's recall the formula of the modularity :

$$Q = \sum^{n_c} \left[\frac{l_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right]$$

With m the total number of edges, n_c the number of clusters, l_c the number of edges within cluster c and d_c the sum of the degrees of the nodes within cluster c.

Now, we compute the modularity for graph a. For this graph we have $m = 13$, $l_1 = 6$ and $l_2 = 6$ (both clusters have the same amount of edges), by counting the node degree in ascending node index, we get $d_1 = 3 + 2 + 3 + 3 + 1 = 12$ and $d_2 = 3 + 3 + 3 + 3 = 12$.
The modularity is then $Q_a = 2 * (6/13 - (12/26)^2) \approx \mathbf{0.49}$.
For graph b we have $m = 13$, $l_1 = 2$ and $l_2 = 4$, by counting the node degree in ascending node index, we get $d_1 = 1 + 1 + 1 + 1 = 4$ and $d_2 = 1 + 2 + 2 + 2 + 1 = 8$.
The modularity is therefore $Q_b = (2/13 - (4/26)^2) + (4/13 - (8/26)^2) \approx \mathbf{0.34}$.
We can observe that as expected, the modularity of graph a is larger than graph b because the clustering seems to separate better the two parts of the graph.

Question 3

Let's compute first $\phi_{sp}(C_4)$ and $\phi_{sp}(P_4)$. For the cyclic graph $\phi_{sp}(C_4)$, there are 4 shortest path of distance 1 corresponding to the neighboring nodes and 4 shortest path of distance 2 corresponding to the opposite nodes of the square. As there are not any longer distance shortest path, we have $\phi_{sp}(C_4) = [4, 4, 0, ..., 0]$.

Now for the path graph P_4, there are 4 shortest path of distance 1 corresponding to the path between each neighboring nodes, there are 2 shortest path of distance 2 (one between node 1 and node 3 and one between node 2 and node 4), and 1 shortest path of distance 3 between node 1 and node 4. Hence, we have $\phi_{sp}(P_4) = [3, 2, 1, 0, ..., 0]$.

Let's now compute the value of the shortest path kernel for each pairs :

- $k_{sp}(C_4, C_4) = 4 * 4 + 4 * 4 = \mathbf{32}$
- $k_{sp}(C_4, P_4) = 4 * 3 + 4 * 2 + 0 * 1 = \mathbf{20}$
- $k_{sp}(P_4, P_4) = 3 * 3 + 2 * 2 + 1 * 1 = \mathbf{14}$

Question 4

Two graphs G and G' such that $k(G, G') = 0$ means that graph G and graph G' do not have any common graphlet of size 3, so no 3-nodes subgraph of graph G and no 3-nodes subraph of graph G' can be isomorphic to the same graphlet.

Now, let's show an example where $k(G, G') = 0$. Let G be a cycle graph of 4 nodes and G' be a graph with 4 isolated nodes. Any 3-nodes subgraph of graph G will be a path graph corresponding to G_2 (thus not isomorphic to G_1, G_3, G_4) and any 3-nodes subgraph of graph G' will be a graph with 3 isolated nodes which corresponds to G_4 (thus not isomorphic to G_1, G_2, G_3). Hence in this case we will have $k(G, G') = 0$ because f_G will have a non-zero value only at index 2 and $f_{G'}$ will have non-zero value value only at index 4.