1 Question 1

We can compute separately the number of edges for each component and then sum them up. For the first component, since it is a complete graph we have:

$$n_1 = \frac{n_{vertices} * (n_{vertices} - 1)}{2} = \frac{100 * 99}{2} = 4950$$

For the second component, since it is a complete bipartite graph we have:

$$n_2 = n_{vertices1} * n_{vertices2} = 50 * 50 = 2500$$

So in totol we have:

$$n_{total} = n_1 + n_2 = 7450$$

Since in a bipartite graph there is no triangles (the vertices from the same partition are not connected) and giving that in a complete graph, the number of triangles is simply the number of ways to choose 3 vertices which is :

$$C_{100}^3 = 161700$$

2 Question 2

The maximum possible value for transitivity is equal to 1, which is achievable if the number of open triplets is 0. Taking this into account, the condition we are looking for is that all connected components should be complete.

3 Question 3

In a connected graph, if we take the vector equal to one in all it's components, then we have a null vector after applying the Laplacian matrix. Which means that the smallest eigenvalue in this case is 0 and the associated vector is the vector equal one in all its components. Adding this vector will only add a 1 in all the rows of U, since k-means is a distance based algorithm, this won't change the results, hence it is a good practice to not take this eigenvalue into account.

4 Question 4

Since K-means is a stochastic algorithm, then the output of the clustering algorithm is stochastic.

5 Question 5

Denote by 1 the green cluster, and by 2 the blue one:
(a) :

$$m = 13; n_c = 2; l_c1 = 6; l_c2 = 6; d_c1 = d_c2 = 13, thus : Q = 0.423$$

(b) :

$$m = 13; n_c = 2; l_c1 = 2; l_c2 = 4; d_c1 = 11; d_c2 = 16, thus : Q = -0.096$$

6 Question 6

Giving that :

$$K(G_1, G_2) = < \phi(G_1) | \phi(G_2) >$$

and that

$$\phi(C4) = [4, 4, 4, 0]; \phi(P4) = [3, 2, 1, 0]$$

so:

$$K(C4, C4) = 48; K(C4, P4) = 24; K(P4, P4) = 14$$