1 Question 1

The number of edges in two components, a connected graph and a bipartite graph, is calculated as follows:

For the connected graph, the number edges is given by:

$$\sum_{i=1}^{n_1}(n_1 - i) = \frac{n_1(n_1 - 1)}{2}.$$

For the bipartite graph, the number of edges is calculated as:

$$n_2 \cdot n_2,$$

where n_2 represents the number of nodes in the bipartite component.

Given $n_1 = 100$ and $n_2 = 50$, we compute:

$$\frac{100 \cdot 99}{2} = 4950,$$

and

$$50 \cdot 50 = 2500.$$

Thus, the total number of pairs is:

$$4950 + 2500 = 7450.$$

In the case of triangles, we consider only the connected component, as a bipartite graph cannot form triangles. The number of triangles is determined by selecting 3 nodes from the n_1 nodes in the connected component. The number of ways to choose 3 nodes from n_1 nodes is given by:

$$\binom{n_1}{3} = \frac{n_1 \cdot (n_1 - 1) \cdot (n_1 - 2)}{6}.$$

For $n_1 = 100$, the calculation becomes:

$$\binom{100}{3} = \frac{100 \cdot 99 \cdot 98}{6}.$$

Simplifying this gives:

$$\binom{100}{3} = 161700.$$

Thus, the number of triangles that can be formed in the connected component is 161700.

2 Question 2

Question 2

To compute the modularity Q, we recall the formula:

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

where:

- m is the total number of edges in the graph,

- n_c is the number of communities,
- l_c is the number of edges within the community c,
- d_c is the sum of the degrees of the nodes that belong to community c.

Case (a):

- $n_c = 2$, $m = 13$,
- $l_{\text{blue}} = 6$, $l_{\text{green}} = 6$,
- $d_{\text{blue}} = 13$, $d_{\text{green}} = 13$.

Applying the formula:

$$Q = \left[\frac{6}{13} - \left(\frac{13}{26}\right)^2\right] + \left[\frac{6}{13} - \left(\frac{13}{26}\right)^2\right],$$

$$Q = \left[\frac{6}{13} - \frac{1}{4}\right] + \left[\frac{6}{13} - \frac{1}{4}\right].$$

Simplifying:

$$Q = 2 \times \left(\frac{6}{13} - \frac{1}{4}\right) = 2 \times (0.4615 - 0.25) = 2 \times 0.2115 = 0.423.$$

Thus, for case (a), $Q = 0.42$.

Case (b):

- $n_c = 2$, $m = 13$,
- $l_{\text{blue}} = 4$, $l_{\text{green}} = 2$,
- $d_{\text{blue}} = 15$, $d_{\text{green}} = 11$.

Applying the formula:

$$Q = \left[\frac{4}{13} - \left(\frac{15}{26}\right)^2\right] + \left[\frac{2}{13} - \left(\frac{11}{26}\right)^2\right],$$

$$Q = \left[\frac{4}{13} - \frac{225}{676}\right] + \left[\frac{2}{13} - \frac{121}{676}\right].$$

Simplifying:

$$Q = (0.3077 - 0.3333) + (0.1538 - 0.1791) = -0.0256 - 0.0253 = -0.0509.$$

Thus, for case (b), $Q = -0.050$.

Conclusion:
The modularity for graph (a) is $Q = 0.42$, while for graph (b), it is $Q = -0.050$. Therefore, graph (a) is better clustered than graph (b), which aligns with the visual inspection of the two graphs.

Question 3

To compute the similarities between graphs using the shortest path kernel, we first define the kernel vectors as column vectors:

$$P_4 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \quad C_4 = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}.$$

We compute the following dot products:

1. $P_4 \cdot P_4$:

$$P_4 \cdot P_4 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = (3 \cdot 3) + (2 \cdot 2) + (1 \cdot 1) = 9 + 4 + 1 = 14.$$

2. $P_4 \cdot C_4$:

$$P_4 \cdot C_4 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} = (3 \cdot 4) + (2 \cdot 2) + (1 \cdot 0) = 12 + 4 + 0 = 16.$$

3. $C_4 \cdot C_4$:

$$C_4 \cdot C_4 = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} = (4 \cdot 4) + (2 \cdot 2) + (0 \cdot 0) = 16 + 4 + 0 = 20.$$

Conclusion: - The self-similarity of P_4 is 14, and the self-similarity of C_4 is 20, indicating that C_4 is denser (has more internal structure) than P_4. - The similarity between P_4 and C_4 is 16, which is relatively high compared to the individual self-similarities. This suggests that P_4 and C_4 share significant structural similarities, despite differences in their specific configurations.

Question 4

The graphlet kernel $k(G, G') = f_G^T f_{G'}$ equals 0 when G and G' share no common graphlets of size 3.

Example: If G corresponds to graphlet 1 and G' to graphlet 2, their frequency vectors are:

$$f_G = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad f_{G'} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

The dot product is:

$$f_G^T f_{G'} = 0.$$