Julien Delavande
juliendelavande@gmail.com

1 Question 1: Graph with Two Connected Components

1.1 1. Number of Edges in Graph G

The graph G consists of two connected components:

- **First Component:** A complete graph K_{100} on 100 vertices. The number of edges in K_{100} is given by:

$$\text{Number of edges in } K_{100} = \binom{100}{2} = \frac{100 \cdot (100-1)}{2} = 4950$$

- **Second Component:** A complete bipartite graph $K_{50,50}$ with 50 vertices in each partition. The number of edges in $K_{50,50}$ is given by:

$$\text{Number of edges in } K_{50,50} = 50 \cdot 50 = 2500$$

- **Total Edges:** The total number of edges in G is:

$$4950 + 2500 = 7450$$

1.2 2. Number of Triangles in Graph G

- **First Component:** The complete graph K_{100} contains triangles. The number of triangles in K_{100} is given by:

$$\text{Number of triangles in } K_{100} = \binom{100}{3} = \frac{100 \cdot 99 \cdot 98}{6} = 161700$$

- **Second Component:** The bipartite graph $K_{50,50}$ does not contain any triangles, as triangles cannot be formed in a bipartite graph.

- **Total Triangles:** The total number of triangles in G is:

$$161700 + 0 = 161700$$

1.3 Final Answer

- Total number of edges in G: **7450**.
- Total number of triangles in G: **161700**.

Question 2: Modularity Calculation

The modularity Q is calculated using the formula:

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

where:

- m: Total number of edges in the graph.
- l_c: Number of edges within cluster c.
- d_c: Sum of degrees of the nodes in cluster c.
- n_c: Total number of clusters.

—

(a) Partition 1

Step 1: Total number of edges (m) From the graph in (a), the total number of edges is:

$$m = 13$$

Step 2: Cluster-specific values For the two clusters (Green and Blue):

1. **Green cluster**:
   - Nodes: $\{1, 2, 3, 4, 5\}$
   - $l_{\text{green}} = 6$ (edges within the green cluster).
   - $d_{\text{green}} = 3 + 2 + 3 + 3 + 2 = 13$ (sum of degrees of the nodes in the green cluster).
2. **Blue cluster**:
   - Nodes: $\{6, 7, 8, 9\}$
   - $l_{\text{blue}} = 6$ (edges within the blue cluster).
   - $d_{\text{blue}} = 4 + 3 + 3 + 3 = 13$ (sum of degrees of the nodes in the blue cluster).

Step 3: Modularity Calculation

Green cluster (C_1):

$$Q_{\text{green}} = \frac{l_{\text{green}}}{m} - \left(\frac{d_{\text{green}}}{2m}\right)^2 = \frac{6}{13} - \left(\frac{13}{26}\right)^2$$
$$Q_{\text{green}} = 0.4615 - 0.25 = 0.2115$$

Blue cluster (C_2):

$$Q_{\text{blue}} = \frac{l_{\text{blue}}}{m} - \left(\frac{d_{\text{blue}}}{2m}\right)^2 = \frac{6}{13} - \left(\frac{13}{26}\right)^2$$
$$Q_{\text{blue}} = 0.4615 - 0.25 = 0.2115$$

Total Modularity:

$$Q = Q_{\text{green}} + Q_{\text{blue}} = 0.2115 + 0.2115 = 0.423$$

Final Answer:

$$Q = 0.423$$

(b) Partition 2

Step 1: Cluster-specific values For the two clusters (Green and Blue):

1. **Green cluster**:
   - Nodes: $\{1, 2, 8, 9\}$
   - $l_{\text{green}} = 2$ (edges within the green cluster).
   - $d_{\text{green}} = 3 + 2 + 3 + 3 = 11$.
2. **Blue cluster**:
   - Nodes: $\{3, 4, 5, 6, 7\}$
   - $l_{\text{blue}} = 4$.
   - $d_{\text{blue}} = 3 + 3 + 2 + 4 + 3 = 15$.

Step 2: Modularity Calculation The modularity formula is:

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

where $n_c = 2$, $m = 13$, and the cluster-specific values are:

$$l_{\text{green}} = 2, \quad d_{\text{green}} = 11, \quad l_{\text{blue}} = 4, \quad d_{\text{blue}} = 15$$

Green cluster (C_1):

$$Q_{\text{green}} = \frac{l_{\text{green}}}{m} - \left(\frac{d_{\text{green}}}{2m}\right)^2 = \frac{2}{13} - \left(\frac{11}{26}\right)^2$$

$$Q_{\text{green}} = 0.1538 - (0.4231)^2 = 0.1538 - 0.1790 = -0.0252$$

Blue cluster (C_2):

$$Q_{\text{blue}} = \frac{l_{\text{blue}}}{m} - \left(\frac{d_{\text{blue}}}{2m}\right)^2 = \frac{4}{13} - \left(\frac{15}{26}\right)^2$$

$$Q_{\text{blue}} = 0.3077 - (0.5769)^2 = 0.3077 - 0.3330 = -0.0253$$

Total Modularity:

$$Q = Q_{\text{green}} + Q_{\text{blue}} = -0.0252 + (-0.0253) = -0.0505$$

Final Answer:

$$Q = -0.0505$$

Question 3: Shortest Path Kernel Calculation

The shortest path kernel $k(G, G')$ for two graphs G and G' is defined as:

$$k(G, G') = \phi(G) \cdot \phi(G') = \sum_d \phi_d(G) \cdot \phi_d(G')$$

where $\phi_d(G)$ is the frequency of shortest paths of distance d in graph G.

(1) Pair (C_4, C_4)

- For C_4 (a cycle graph with 4 nodes):

$$\phi(C_4) = [4, 4, 0, \ldots]$$

- 4 shortest paths of distance 1. - 4 shortest paths of distance 2.

$$k(C_4, C_4) = 4 \cdot 4 + 4 \cdot 4 = 16 + 16 = 32$$

(2) Pair (C_4, P_4)

- For P_4 (a path graph with 4 nodes):

$$\phi(P_4) = [3, 2, 1, 0, \ldots]$$

- 3 shortest paths of distance 1. - 2 shortest paths of distance 2. - 1 shortest path of distance 3.

$$k(C_4, P_4) = 4 \cdot 3 + 4 \cdot 2 + 0 \cdot 1 = 12 + 8 + 0 = 20$$

—

(3) Pair (P_4, P_4)

$$k(P_4, P_4) = 3 \cdot 3 + 2 \cdot 2 + 1 \cdot 1 = 9 + 4 + 1 = 14$$

—

Final Answers:

- $k(C_4, C_4) = 32$
- $k(C_4, P_4) = 20$
- $k(P_4, P_4) = 14$

Final Answer

The modularity values for the two partitions are:

1. Partition (a): $Q = 0.423$
2. Partition (b): $Q = -0.0505$

Partition (a) has a much better modularity score than Partition (b).

Question 4: Graphlet Kernel Value Equal to 0

Answer to Question 4

If $k(G, G') = 0$, it implies that the feature vectors f_G and $f_{G'}$ are orthogonal. In other words, the two graphs do not share any common graphlets of size 3. This occurs when every graphlet of size 3 present in G is absent in G', and vice versa.

Example

Let G and G' be two graphs defined as follows:

- G consists of a single cycle of three nodes (a triangle graph).
- G' consists of three nodes connected in a path (a line graph of three nodes).

Explanation:

- In G, the only graphlet of size 3 is a triangle.
- In G', the only graphlet of size 3 is a path.

Since G and G' do not share any common graphlets of size 3, their feature vectors f_G and $f_{G'}$ are orthogonal, resulting in:

$$k(G, G') = f_G^\top f_{G'} = 0.$$

Thus, $k(G, G') = 0$ holds for these two graphs.

References