Solal Danan
solal.danan@polytechnique.edu

Lab session # 4
ALTEGRAD 2024

11/26/24

1 Question 1

We are given a graph G with two disconnected components:

- The first component is a complete graph K_{100} with 100 vertices.
- The second component is a complete bipartite graph $K_{50,50}$ with 50 vertices in each partition.

Triangle: A triangle in a graph is a set of three vertices that are pairwise connected by edges.
Complete Graph (K_n): In q complete graph, every pair of vertices is directly connected by an edge.
Complete Bipartite Graph ($K_{m,n}$): A complete bipartite graph is a graph where the vertices are divided into two disjoint sets (or partitions), with edges connecting every vertex in one partition to every vertex in the other partition. No edges exist within a partition.

1. Total Number of Edges in G

Edges in K_{100}: A complete graph K_n has edges between every pair of its n vertices. The total number of edges in K_n is given by: $\binom{n}{2} = \frac{n(n-1)}{2}$.
For K_{100}, we substitute $n = 100$: Number of edges in $K_{100} = \binom{100}{2} = \frac{100\cdot 99}{2} = 4950$.

Edges in $K_{50,50}$: A complete bipartite graph $K_{m,n}$ has edges connecting every vertex in one partition to every vertex in the other partition. The total number of edges is $m \cdot n$.
For $K_{50,50}$, where $m = 50$ and $n = 50$: Number of edges in $K_{50,50} = 50 \cdot 50 = 2500$.

Total Number of Edges in G: Since G consists of two disconnected components, the total number of edges in G is the sum of the edges in K_{100} and $K_{50,50}$: Total edges in $G = 4950 + 2500 = 7450$.

2. Total Number of Triangles in G

Triangles in K_{100}: A triangle in K_{100} is formed by choosing any 3 vertices from its 100 vertices. The number of triangles is given by: $\binom{n}{3} = \frac{n(n-1)(n-2)}{6}$.
For K_{100}, we substitute $n = 100$: Number of triangles in $K_{100} = \binom{100}{3} = \frac{100\cdot 99\cdot 98}{6} = 161700$.

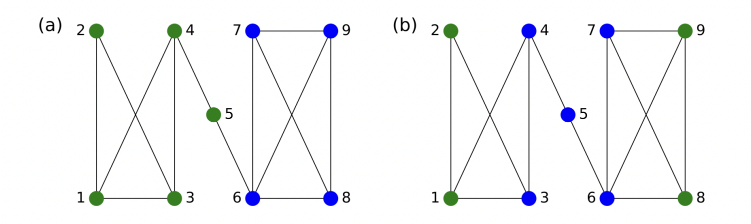
Triangles in $K_{50,50}$: A complete bipartite graph $K_{m,n}$ does not contain any triangles because there are no edges within a partition, and a triangle requires all three vertices to be pairwise connected. Thus: Number of triangles in $K_{50,50} = 0$

Total Number of Triangles in G: The total number of triangles in G is the sum of the triangles in K_{100} and $K_{50,50}$: Total triangles in $G = 161700 + 0 = 161700$.

2 Question 2

Modularity Q measures the quality of a clustering by comparing the observed number of edges within clusters to the expected number of edges in a random graph with the same degree distribution. The modularity formula is given as:

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



For figure 1.a:

- $m = 13$: Total number of edges in the graph.
- $l_{\text{green}} = 6$: Number of edges within the green cluster.
- $d_{\text{green}} = 13$: Sum of degrees of nodes in the green cluster.
- $l_{\text{blue}} = 6$ and $d_{\text{blue}} = 13$

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

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

Thus, the modularity of the clustering for Figure 1(a) is $Q = Q_{\text{green}} + Q_{\text{blue}} \approx 0.2115 + 0.2115 \approx 0.423$

For figure 1.b: $m = 13$, $l_{\text{green}} = 2$, $d_{\text{green}} = 11$, $l_{\text{blue}} = 4$, $d_{\text{blue}} = 15$

Green Cluster Contribution: $Q_{\text{green}} = \frac{l_{\text{green}}}{m} - \left(\frac{d_{\text{green}}}{2m}\right)^2 = \frac{2}{13} - \left(\frac{11}{26}\right)^2 \approx 0.1538 - 0.1791 \approx -0.0253$

Blue Cluster Contribution: $Q_{\text{blue}} = \frac{l_{\text{blue}}}{m} - \left(\frac{d_{\text{blue}}}{2m}\right)^2 = \frac{4}{13} - \left(\frac{15}{26}\right)^2 \approx 0.3077 - 0.3330 \approx -0.0253$
Thus, the modularity of the clustering for Figure 1(b) is: $Q = Q_{\text{green}} + Q_{\text{blue}} \approx -0.0253 + -0.0253 \approx -0.0506$

3 Question 3

To compute the shortest path kernel for the pairs (C_4, C_4), (C_4, P_4), and (P_4, P_4), we calculate the feature vectors for each graph as follows:

Feature Vectors

- $\phi(C_4) = [4, 4, 0, \dots]$:
  - 4 shortest paths of length 1.
  - 4 shortest paths of length 2.
- $\phi(P_4) = [3, 2, 1, 0, \dots]$:
  - 3 shortest paths of length 1.
  - 2 shortest paths of length 2.
  - 1 shortest path of length 3.

Kernel Computations

The shortest path kernel is defined as:

$$k(G_1, G_2) = \sum_{d=1}^{\infty} \phi_d(G_1) \cdot \phi_d(G_2)$$

1. (C_4, C_4):
   $k(C_4, C_4) = \phi_1(C_4) \cdot \phi_1(C_4) + \phi_2(C_4) \cdot \phi_2(C_4) = 4 \cdot 4 + 4 \cdot 4 = 16 + 16 = 32$
2. (C_4, P_4):
   $k(C_4, P_4) = \phi_1(C_4) \cdot \phi_1(P_4) + \phi_2(C_4) \cdot \phi_2(P_4) = 4 \cdot 3 + 4 \cdot 2 = 12 + 8 = 20$
3. (P_4, P_4):
   $k(P_4, P_4) = \phi_1(P_4) \cdot \phi_1(P_4) + \phi_2(P_4) \cdot \phi_2(P_4) + \phi_3(P_4) \cdot \phi_3(P_4) = 3 \cdot 3 + 2 \cdot 2 + 1 \cdot 1 = 9 + 4 + 1 = 14$

4 Question 4

In the graphlet kernel, $k(G, G') = f_G^\top f_{G'}$ measures the similarity between two graphs G and G' based on the occurrences of graphlets (size-3 subgraphs). A value $k(G, G') = 0$ means G and G' share **no common graphlets of size 3**.

Example:

Let G be a path graph: $A \longleftrightarrow B \longleftrightarrow C$, and G' be a triangle graph: $A \longleftrightarrow B \longleftrightarrow C \longleftrightarrow A$.

Graphlets of Size 3

For $k = 3$, the possible graphlets are:

- G_1: Disconnected graph (3 isolated nodes).
- G_2: Single edge (2 connected nodes with 1 isolated node).
- G_3: Path graph (3 nodes connected in a line).
- G_4: Triangle (3 fully connected nodes).

Subsets of 3 Nodes

For G ($A \longleftrightarrow B \longleftrightarrow C$), the only subset of size 3 is the graph itself:

$$\text{Subset: } \{A, B, C\}, \quad \text{Induced Subgraph: } A \longleftrightarrow B \longleftrightarrow C$$

This is isomorphic to G_3 (path graph).
For G' ($A \longleftrightarrow B \longleftrightarrow C \longleftrightarrow A$), the only subset of size 3 is the graph itself:

$$\text{Subset: } \{A, B, C\}, \quad \text{Induced Subgraph: } A \longleftrightarrow B \longleftrightarrow C \longleftrightarrow A$$

This is isomorphic to G_4 (triangle).

Feature Vectors:

- For G: $f_G = [0, 0, 1, 0]$.
- For G': $f_{G'} = [0, 0, 0, 1]$.

Kernel Computation:

The graphlet kernel is computed as the dot product of the feature vectors:

$$k(G, G') = f_G^\top f_{G'} = [0, 0, 1, 0] \cdot [0, 0, 0, 1] = 0$$

Conclusion:

The kernel value $k(G, G') = 0$ indicates that G (path graph) and G' (triangle graph) share no common graphlets of size 3. The feature vector for G reflects one path graphlet, while G''s feature vector reflects one triangle graphlet, leading to no overlap.