Rosalie MILLNER
rosalie@millner.fr

1 Question 1

What is the number of edges in graph G consisting of two connected components, where one of the connected components is a complete graph on 100 vertices and the other connected component is a complete bipartite graph with 50 vertices in each partition set? What is the number of triangles in G?

Let's divide the problem into sub-problems. We can first start by calculating the number of edges there is in each of the two connected components of G separately.

The first one, that we will call G_1 is a complete graph on $n = 100$ vertices. By definition, a complete graph is a graph in which every vertex is connected to every other vertex, in the sense that there are edges connecting every vertex to every other vertex in the graph. Its number of edges is given by:

$$\begin{aligned}\text{Number of vertices in } G_1 &= \sum_{k=1}^{n-1} k \\ &= \frac{n(n-1)}{2} \\ &= \frac{100 \times 99}{2} \\ &= 4950\end{aligned}$$

The intuition behind the formula is the following: suppose one wants to connect n vertices to each other. Taking a first vertex, we must connect it to the $(n-1)$ other ones, creating $(n-1)$ edges. For a second one, we do the same but without the edge we have just created connecting it to the first vertex, we thus create $(n-2)$ edges, and so one until we get to the last vertex that has already been connected to every other one through this process.

The second connected component, G_2, is a complete bipartite graph with $n_1 =$ 50 vertices in the first partition set and $n_2 =$ 50 vertices in the second partition set. By definition, such a graph is defined as being one with an edge between every pair of vertices that is from both different sets. There are no edges within one partition set. The number of edges here is given by:

$$\begin{aligned}\text{Number of vertices in } G_2 &= n_1 \times n_2 \\ &= 50 \times 50 \\ &= 2500\end{aligned}$$

The intuition behind this formula is straightforward: every vertex from every partition set is connected once to every vertex from the other partition set.

Finally, we obtain the the total number of edges in the graph G by summing the number of edges in G_1 and G_2:

$$\begin{aligned}\text{Total number of vertices in } G &= 4950 + 2500 \\ &= 7450\end{aligned}$$

Now, let's calculate the number of triangles that are in G. This boils down to finding how many triangles there are in G_1 and in G_2 and then summing the results.

Since G_1 is a complete graph, finding the number of triangles boils down to calculating how many different sets made of 3 vertices there are in G_1, as they are all connected to each other by edges. This is given by:

$$\begin{aligned}
\text{Number of triangles in } G_1 &= \binom{n}{3} \\
&= \binom{100}{3} \\
&= \frac{100!}{3!\,(100-3)!} \\
&= \frac{100 \times 99 \times 98}{3 \times 2 \times 1} \\
&= 161700
\end{aligned}$$

As for G_2, it is a bipartite graph, meaning it is impossible to have an edge within one partition set. This makes it impossible to find triangles in such a graph. We thus have: Number of triangles in $G_2 = 0$.

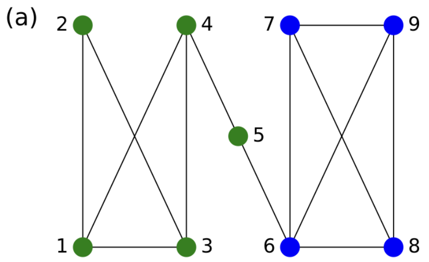
Therefore, the total number of triangles in G is equal to the number of triangles in G_1, which is: 161700.

2 Question 2

Compute (showing your calculations) the modularity of the clustering results shown in Figure 1. Note that different colors correspond to different clusters.

Let us consider that the nodes represented in green belong to cluster 1 and the ones represented in blue belong to cluster 2.

Let's begin with figure (a).

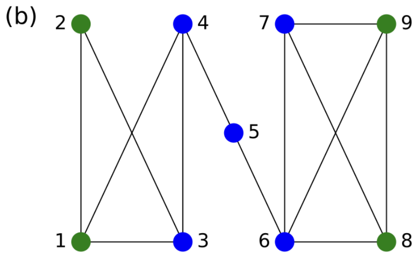


We have:
$m = 13$
$n_c = 2$
$l_1 = 6$
$d_1 = 13$
$l_2 = 6$
$d_2 = 13$

We can thus apply the formula with these values:

$$\begin{aligned}
Q_a &= \sum^{n_c} \left[\frac{l_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right] \\
&= \frac{l_1}{m} - \left(\frac{d_1}{2m} \right)^2 + \frac{l_2}{m} - \left(\frac{d_2}{2m} \right)^2 \\
&= \left(\frac{6}{13} - \left(\frac{13}{26} \right)^2 \right) \times 2 \\
&= \frac{11}{26} \approx 0.42
\end{aligned}$$

We do the same for figure (b):



We have:
$m = 13$
$n_c = 2$
$l_1 = 2$
$d_1 = 11$
$l_2 = 4$
$d_2 = 15$

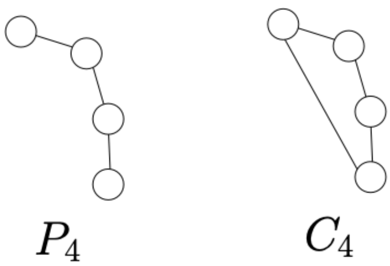
This gives us:

$$\begin{aligned} Q_b &= \frac{l_1}{m} - \left(\frac{d_1}{2m}\right)^2 + \frac{l_2}{m} - \left(\frac{d_2}{2m}\right)^2 \\ &= \frac{2}{13} - \left(\frac{11}{26}\right)^2 + \frac{4}{13} - \left(\frac{15}{26}\right)^2 \\ &= -\frac{17}{338} \approx -0.05 \end{aligned}$$

We obtain $Q_a > Q_b$. Therefore, according to the modularity metric, the clustering from figure (a) has a higher quality than the clustering from figure (b). This is coherent with the intuition when we visually examine the figures.

3 Question 3

Let P_n denote a path graph on n vertices and C_n denote a cycle graph on n vertices. Calculate the shortest path kernel for the pairs (C_4, C_4), (C_4, P_4) and (P_4, P_4).



Following the notations from the Lab handout, we have:

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

and

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

There is no path of length 4 in C_4 because the two extremities of any chain of length 4 would be the same node, which contradicts the definition of a path.

By applying dot products, we obtain:

$$k(C_4, C_4) = \phi(C_4)^T \phi(C_4) \ = \ 4^2 + 4^2 + 4^2 = 48$$

$$k(C_4, P_4) = \phi(C_4)^T \phi(P_4) \ = \ 4 \times 3 + 4 \times 2 + 4 \times 1 = 24$$

$$k(P_4, P_4) = \phi(P_4)^T \phi(P_4) \ = \ 3^2 + 2^2 + 1^2 = 14$$

4 Question 4

Let k denote the graphlet kernel that decomposes graphs into graphlets of size 3. Let also G, G' denote two graphs and suppose that $k(G, G') = f_G^T \ f'_G = 0$. What does a kernel value equal to 0 mean? Give an example of two graphs G, G' for which $k(G, G') = 0$ holds.

A kernel value equal to zero indicates that there is no common graphlet between the two graphs, which is coherent with the fact that k is used as a measure of the similarity between 2 graphs.

A simple example of G and G' such that $k(G, G') = 0$ would be to take any two of the graphlets we have used to build the graphlet kernel. For instance, if G is the graph from the figure bellow, we would have $k(G, G') = 0$.