Question 1

Let define G_1 and G_2 as the two connected components described in the question. Define N_1 and N_2 as $N_1 = 100$ and $N_2 = 50$.

Number of edges in G

In G_1, the number of edges is $\binom{N_1}{2} = \frac{N_1(N-1)}{2} = 4950$.
In G_2, the number of edges is $N_2 \times N_2$ as G_2 is a bipartie graph where each partition set has N_2 vertices. Thus, the number of edges in G_2 is $N_2 \times N_2 = 2500$. Finally, the total number of edges in G is $4950 + 2500 = 7450$.

Number of triangles in G

In G_1, the number of triangles is $\binom{N_1}{3} = \frac{N_1(N_1-1)(N_1-2)}{3\times 2} = 161700$.
In G_2, the number of triangles is 0 because it is a complete bipartite graph.
Finally, the total number of triangles in G is 161700.

Question 2

For subfigure (a):
In the proposed graph, we can observe two different communities (clusters), the green one with nodes $\{1, 2, 3, 4, 5\}$ and the blue one with nodes $\{6, 7, 8, 9\}$. We thus have $n_c = 2$. To compute the modularity metric, we need to compute the number of edges inside each community (l_1 and l_2), the total number of edges of the whole graph (m) and the sum of the degrees of each nodes inside a community(d_1 and d_2).

We observe that:

- $m = 13$
- $l_1 = 6$ and $l_2 = 6$
- $d_1 = 13$ and $d_2 = 13$

Thus, the modularity metric is:

$$\begin{aligned}
Q &= \left(\frac{l_1}{m} - \left(\frac{d_1}{2m}\right)^2\right) + \left(\frac{l_1}{m} - \left(\frac{d_1}{2m}\right)^2\right) \\
&= \left(\frac{6}{13} - \left(\frac{13}{26}\right)^2\right) + \left(\frac{6}{13} - \left(\frac{13}{26}\right)^2\right) \\
&= \frac{6}{13} - \frac{1}{4} + \frac{6}{13} - \frac{1}{4} \\
&= \frac{12}{13} - \frac{1}{2} \\
&= \frac{24}{26} - \frac{13}{26} \\
&= \frac{11}{26} \in [-1, 1].
\end{aligned}$$

For subfigure (b):
Using the same method as for subfigure (a), we have:

- $m = 13$
- $l_1 = 2$ and $l_2 = 4$

- $d_1 = 11$ and $d_2 = 15$

Thus, the modularity metric is:

$$\begin{aligned}
Q &= \left(\frac{2}{13} - \left(\frac{11}{26}\right)^2\right) + \left(\frac{4}{13} - \left(\frac{15}{26}\right)^2\right) \\
&= \frac{2}{13} - \frac{121}{676} + \frac{4}{13} - \frac{225}{676} \\
&= \frac{6}{13} - \frac{346}{676} \\
&= \frac{6}{13} - \frac{173}{338} \\
&= \frac{156}{338} - \frac{173}{338} \\
&= -\frac{17}{338} \in [-1, 1].
\end{aligned}$$

Question 3

For $n = 4$, let P_4 be:

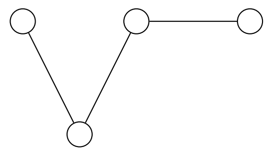


Figure 1: P_4

and C_4 be:

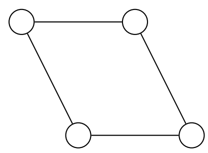


Figure 2: C_4

for C_4:

- 4 paths of length 1
- 2 paths of length 2
- 0 paths of length 3

for P_4:

- 3 paths of length 1
- 2 paths of length 2
- 1 paths of length 3

To compute the shortest path kernel between 2 graphs, we need to compute the inner product between the two feature vectors.. Doing though for all lengths, we get the following kernels:

$$k(P_4, C_4) = 3 \times 4 + 2 \times 2 + 1 \times 0 = 16.$$

$$k(C_4, C_4) = 4 \times 4 + 2 \times 2 = 20.$$

$$k(P_4, P_4) = 3 \times 3 + 2 \times 2 + 1 \times 1 = 14.$$

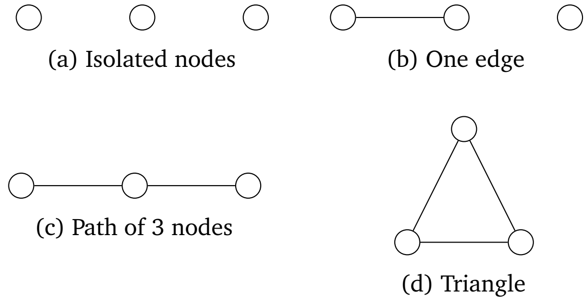
Question 4

Let G, $G^{'}$ be two undirected graphs. Let $\mathcal{G} = \{graphlet(1), \dots, graphlet(N_3)\}$, with $N_3 = 4$. With the notations of the question and the article [1], let f_G (respectively $f_{G'}$) be a vector of length N_3, whose i^{th} component is the number of occurence of $graphlet(i)$ in G (respectively in $G^{'}$). The graphlet kernel is defined as

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

If $k(G, G^{'}) = 0$, it means that $f_G^\top f_{G'} = 0$, so f_G and $f_{G'}$ are orthogonal. As $\forall i \in \{1, \dots, N_3\}, \forall G, f_G^i \geq 0$ (denoting f_G^i as the i^{th} component of vector f_G), the only possibility for k to be equal to 0 is that $\forall i, f_G^i f_{G'}^i = 0$. This implies that there is no similar graphlets between the compared graphs (they don't have any common subgraphs structure).
However, it is possible that a 3-size graphlet kernel can lead to a zero value, rather that a 4-size graphlet kernel can capture similarities (between the same graphs). This can happens because of a graph complexity.

Example of two graphs G and $G^{'}$ which $k(G, G^{'}) = 0$:
For a 3-size kernel: the 4 different graphlets are represented in figure 3.

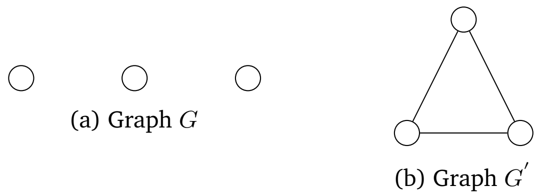


(a) Isolated nodes (b) One edge

(c) Path of 3 nodes (d) Triangle

Figure 3: The four different graphlets of size 3

Let's choose two undirected graphs G and $G^{'}$:



(a) Graph G (b) Graph $G^{'}$

Figure 4

Let's compute the graphlet kernel between G and $G^{'}$:

$$\begin{aligned} k(G, G^{'}) &= f_G^\top f_{G'} \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \\ &= 0. \end{aligned}$$

References

[1] Nino Shervashidze, SVN Vishwanathan, Tobias Petri, Kurt Mehlhorn, and Karsten Borgwardt. Efficient graphlet kernels for large graph comparison. In David van Dyk and Max Welling, editors, *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 488–495, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 16–18 Apr 2009. PMLR.