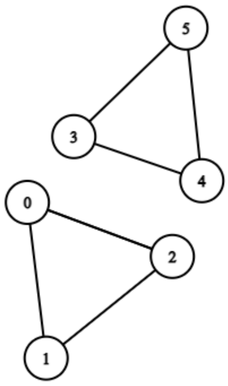
Gertner Victor
gertnervictor@gmail.com

1 Question 1

The maximum number of edge in a graph without any self loop having n nodes is the case where we have a complete graph. Thus, because all the nodes are connected we have $\frac{n(n-1)}{2}$ **edges at maximum.**
We can have the same logic for the number of triangles, the worst case is in a complete graph, and in this graph we can randomly take 3 nodes and create a triangle. **Thus the maximum number of triangle in a graph with n nodes is $\binom{n}{3}$.**

2 Question 2

Two graphs having the same degree distribution doesn't make them isomorphic. Here is an exemple :



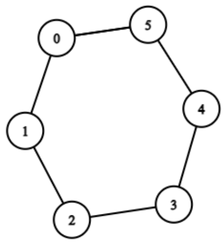


Figure 2: G'

Figure 1: G

Figure 3: Two graphs with same degree distribution but not isomorphic

We see here that each node in G and G' have a degree of 2.
However, we know that isomorphims between two algebric structures preserves the structures and its properties. Hence, if we have an equivalent relationship on a space T, and we have Φ an isomorphism from T into itself, we can apply the isomorphism to the equivalence relationship. Thus, we have for all a,b in T : a$\sim$b if and only if Φ(a) $\sim$ Φ(b). Thus, in our case we know that being in the same connected component can be seen as a equivalent relationship. Thus, if two graphs are isomorphic, they have the same number of connected components by what we just said. Using the contrapositive, we can conclude that if G and G' are not isomorphic.

3 Question 3

- **case n = 3 :** For C_3 we see that we only have one possibility for open triangle, and this one is close. **Hence, the global clustering coefficient for C_3 is 1.**
- **case n $\geq$ 4 :** If we take C_n, each node is only connected to it's two neighbours : right and left. Therefore there are no closed triangle in C_n, because the neighbours of the left neighbour don't contain the right neighbour and conversely. **Thus, the global clustering score is 0 for C_n**

4 Question 4 :

Befoire doing any computation, we can first note that :

$$\sum_i A_{i,j} = d_j$$

.
Thus we have :

$$\sum_{i,j} A_{i,j}([u_1]_i - [u_1]_j)^2$$

$$= \sum_{i,j} A_{i,j}[u_1]_i^2 + \sum_{i,j} A_{i,j}[u_1]_j^2 - 2\sum_{i,j} A_{i,j}[u_1]_i[u_1]_j$$

$$= \sum_{i} d_i[u_1]_i^2 + \sum_{j} d_j[u_1]_j^2 - 2\sum_{i,j} A_{i,j}[u_1]_i[u_1]_j$$

$$= 2\sum_{i} d_i[u_1]_i^2 - 2\sum_{i,j} A_{i,j}[u_1]_i[u_1]_j$$

$$= 2u_1^T D u_1 - u_1^T A u_1$$

$$= 2u_1^T D(I - D^{-1}A)u_1$$

$$= 2u_1^T D L_{rw} u_1$$

But we know that the eigenvector u1 is associated with the eigenvalue 0,ie $L_{rw}u_1 = 0$, because as we know the rank of $L_r w$ is n-1 at maximum and $L_r w$ is positive-semidefinite. Hence the smallest eigenvalue is 0. Thus we have :

$$\sum_{i,j} A_{i,j}([u_1]_i - [u_1]_j)^2 = 0$$

5 Question 5 :

First Graph : We take the two clusters as the blue cluster and the orange cluster respectively noted b and o. Here are the values of the parameters : $n_c = 2$, m=14, $l_b = 6$, $d_b = 14$, $l_o = 6$, $d_o = 14$
As we can see the two clusters have the same parameters. The calculation of Q becomes :

$$Q=2(\frac{6}{14} - (\frac{14}{28})^2) = 0.35$$

Second Graph : We take the two clusters as the blue cluster and the orange cluster respectively noted b and o. Here are the values of the parameters : $n_c = 2$, m=14, $l_b = 5$, $d_b = 17$, $l_o = 2$, $d_o = 11$

$$Q=(\frac{5}{14} - (\frac{17}{28})^2) + (\frac{2}{14} - (\frac{11}{28})^2) = -0.0229$$

One can comment that one could see this result only from the clustering, as we see the first one seems more logical by making compact clusters with nodes having low out of cluster edge, whereas the second cluster as a lot of nodes having out of cluster edge.

6 Question 6 :

We are going to use the shortest path kernel. Thus, we are going to first compute the vectors showing the feature map of the shortest paths for each of our graphs.

$\phi(P_4) = [3, 2, 1]$: Three shortest path of size 3, two of size 2 and one of size 1.
$\phi(S_4) = [3, 3, 0]$: Three shortest path of size 3 and 3 of size 2.

So we now can compute the shortest path kernel :

$$k(P_4, P_4) = \phi(P_4) * \phi(P_4)^T = [3, 2, 1] * [3, 2, 1]^T = 3 * 3 + 2 * 2 + 1 = 14$$
$$k(P_4, S_4) = \phi(P_4) * \phi(S_4)^T = [3, 2, 1] * [3, 3, 0]^T = 3 * 3 + 2 * 3 + 1 * 0 = 15$$
$$k(S_4, S_4) = \phi(S_4) * \phi(S_4)^T = [3, 3, 0] * [3, 3, 0]^T = 3 * 3 + 3 * 3 = 18$$

We can see that we don't have the same value, thus P_4 and S_4 are two separable class of graphs.

7 Question 7

Let G and G' be two graphs.
Let $f_G = (g_1, g_2, g_3, g_4)$ and $f_{G'} = (g'_1, g'_2, g'_3, g'_4)$ being the results of the sampling of graphlets, having g_i and g'_i being the number of time the i-th graphlet of size 3 was sampled in respectfully G and G'.
we have :

$$\text{k(G,G')} = f_G^T f_{G'}$$
$$= \Sigma_{i=1}^{n} g_i * g'_i$$

Hence :

$$k(G, G') = 0$$
$\Leftrightarrow g_i = 0$ or $g'_i = 0$ for all i in [1,4] since g_i and g'_i are positive.

A sufficient condition to have this for two graphs G and G' is that from all the possible subgraphs of size 3 for each of those graphs if a type of graphlet appears in the subgraphs of G it cannot appear in the subgraphs of G'.
We can find exemples of this :

Trivial Exemple : If we take two different graphlet of size 3, it's obvious their graphlet kernel will be 0.

More complex exemple : Here we are going to take a fully connected grah : however we sample the nodes, the subgraph of size 3 will always the size 3 graphlet that is fully connected. And we are taking another graph that no matter how we sample the 3 nodes we never find the 3 graphlet that is fully connected.

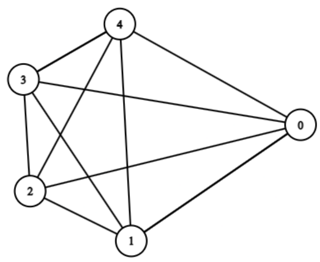


Figure 4: G1

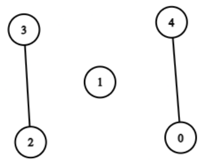


Figure 5: G2

Figure 6: The two graphs we are studying for this question

As we can see, the only graphlets of size 3 in all the subgraphs of size 3 of G1 are :

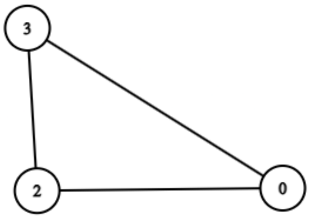


Figure 7: The only possible subgraph of size 3 for G1

While the possible subgraphs of size 3 for G2 are :

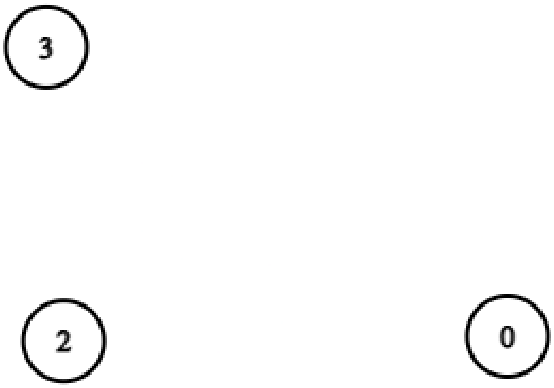


Figure 8:

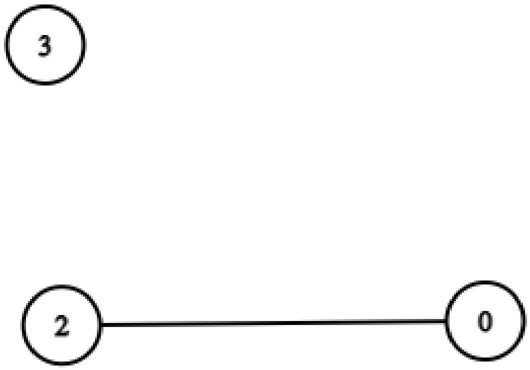


Figure 9:

Figure 10: Possible subgraphs of size 3 for G2

So as we can see, it's not the same. Thus the graphlet kernel between G1 and G2 will be 0.