1 Question 1

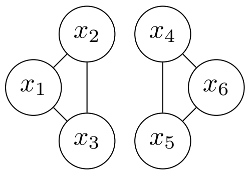
The maximum number of edges (resp. triangles) an undirected graph of n node without self-loops can have is reached for complete graphs. In this case, since all nodes are connected, the problem comes down to finding the number of ways we can select 2 (resp. 3 in the case of triangles) nodes out of the n nodes.

Therefore :

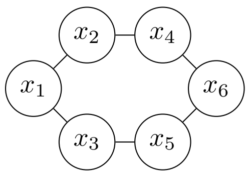
- Maximum # of edges = $\binom{n}{2} = \frac{n!}{2(n-2)!} = \frac{n(n-1)}{2}$
- Maximum # of triangles = $\binom{n}{3} = \frac{n(n-1)(n-2)}{6}$

2 Question 2

Having the same degree distribution is not a sufficient condition for two graphs to be isomorphic. For instance, let us study the two graphs in Figure 1. Both graphs have the same degree distribution : every node is connected to two others. However, graph a) is composed of two connected components whereas graph b) is only made of one connected component. Isomorphisms preserve some structural properties of graphs. In particular, there is no way to match vertices between the two graphs in a manner that preserves adjacency relationships (in graph a) there is no path between $\{x_1, x_2, x_3\}$ and $\{x_4, x_5, x_6\}$ but every node is connected in graph b)). Therefore, a) and b) are not isomorphic despite having the same degree distributions.



(a)



(b)

Figure 1: Two graphs with the same degree distribution (d=2)

3 Question 3

The global clustering coefficient C can be calculated as follows :

$$C = \frac{\text{number of closed triplets}}{\text{number of all triplets}}$$

Let us consider the cycle graphs C_n on n vertices.

- It is trivial that graphs with only one or two vertices are not concerned by such a coefficient (no possible triplet).
- C_3 is a simple triangle (closed triplet). Therefore it has a global clustering coefficient of 1 : $C_{C_3} = 1$
- $\forall n \geq 4$, C_n does not have any closed triplet (all vertices of degree 2 and forming a cycle) therefore, no matter how much triplets we consider we have $C_{C_n} = 0$

4 Question 4

We have :

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}([u_1]_i - [u_1]j)^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}[u_1]_i^2 - 2 \times \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}[u_1]_i[u_1]_j + \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}[u_1]_j^2 && (1)\\
&= \sum_{i=1}^{n} d_i[u_1]_i^2 - 2 \times \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}[u_1]_i[u_1]_j + \sum_{j=1}^{n} d_j[u_1]_j^2 && (2)\\
&= 2 \times \left(\sum_{i=1}^{n} d_i[u_1]_i^2 - \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}[u_1]_i[u_1]_j\right) && (3)\\
&= 2 \times (u_1^T.D.u_1 - u_1^T.A.u_1) && (4)\\
&= 2 \times (u_1^T.(D-A).u_1) && (5)\\
&= 2 \times u_1^T.D.(I - D^{-1}.A).u_1 && (6)\\
&= 2 \times u_1^T.L_{rw}.u_1 && (7)\\
&= 0 && (8)
\end{aligned}$$

We have $L_{rw}u_1 =$ smallest eigenvalue $= 0$ because L_{rw} is a positive semi definite matrix (all eigenvalue ≥ 0) with the sum of columns equal to 0. Therefore L_{rw} is of rank less than or equal to $n-1$ and has 0 as an eigenvalue. Finally we come to the result :

$$\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}([u_1]_i - [u_1]j)^2 = 0$$

5 Question 5

We compute the modularity of the two graphs :

5.1 Graph 1

- $n_c = 2$ (blue and orange)
- $m = 14$
- $l_{blue} = l_{orange} = 6$
- $d_{blue} = d_{orange} = 12$

We have :

$$\begin{aligned}
Q_1 &= \sum^{n_c}\left[\frac{l_c}{m} - \left(\frac{d_c}{2m}\right)^2\right] && (9)\\
&= \left[\frac{l_{blue}}{m} - \left(\frac{d_{blue}}{2m}\right)^2\right] + \left[\frac{l_{orange}}{m} - \left(\frac{d_{orange}}{2m}\right)^2\right] && (10)\\
&= 2 \times \left[\frac{6}{14} - \left(\frac{14}{2\times 14}\right)^2\right] && (11)\\
&\approx 0.3571 && (12)
\end{aligned}$$

5.2 Graph 2

- $n_c = 2$ (blue and orange)
- $m = 14$
- $l_{blue} = 5$ and $l_{orange} = 2$

- $d_{blue} = 17$ and $d_{orange} = 11$

$$Q_2 = \sum^{n_c} \left[\frac{l_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right] \tag{13}$$

$$= \left[\frac{l_{blue}}{m} - \left(\frac{d_{blue}}{2m} \right)^2 \right] + \left[\frac{l_{orange}}{m} - \left(\frac{d_{orange}}{2m} \right)^2 \right] \tag{14}$$

$$= \left[\frac{5}{14} - \left(\frac{17}{2 \times 14} \right)^2 \right] + \left[\frac{2}{14} - \left(\frac{11}{2 \times 14} \right)^2 \right] \tag{15}$$

$$\approx -0.0229 \tag{16}$$

6 Question 6

We wish to calculate the shortest path kernel for the pairs (P_4, P_4), (S_4, S_4) and (P_4, S_4) where P_4 is the path graph of length 4 and S_4 is the star graph of length 4. Let us first draw the FLoyd transformations of P_4 and S_4.

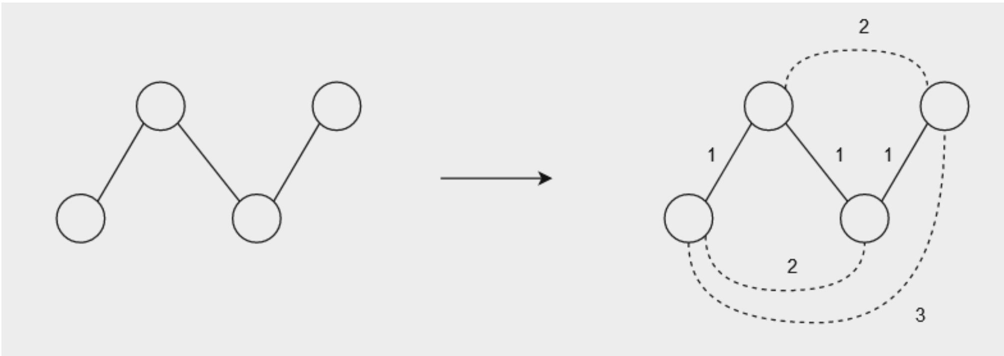


Figure 2: Floyd Transformation of P_4

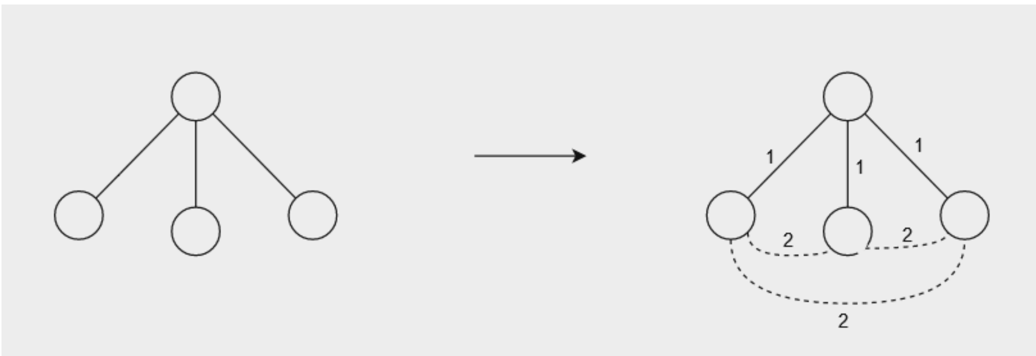


Figure 3: Floyd Transformation of S_4

In the Floyd transformation of P_4 we have :

- 3 edges with label 1
- 2 edges with label 2
- 1 edge with label 3

In the Floyd transformation of S_4 we have :

- 3 edges with label 1
- 3 edges with label 2

Given the Floyd-transformed graphs $S_1 = (V_1, E_1)$ and $S_2 = (V_2, E_2)$ of G_1 and G_2, the shortest path kernel is defined as:

$$k(G_1, G_2) = \sum_{e_1 \in E_1} \sum_{e_2 \in E_2} k_{edge}(e_1, e_2)$$

Where k_{edge} is a kernel on edges : $k_{edges} = \begin{cases} 1 & \text{if } l(e_1) = l(e_2) \\ 0 & \text{else} \end{cases}$ with $l(e)$ the label of edge e

Therefore, we have the following kernel values :

- $k(P_4, P_4) = 3 \times 3 + 2 \times 2 + 1 = 14$
- $k(P_4, S_4) = 3 \times 3 + 3 \times 2 = 15$
- $k(S_4, S_4) = 3 \times 3 + 3 \times 3 = 18$

7 Question 7

A kernal value of zero means that the vectors f_G and $f_{G'}$ are orthogonal. It indicates a high disimilarity between the two graphs with respect to the specific structural patterns being captured by the graphlet kernel (in this case graphlets with 3 nodes).
In the following figures, we consider two graphs G and G' and we study their decompositions in terms of graphlets of size three which will allow us to compute f_G and $f_{G'}$.

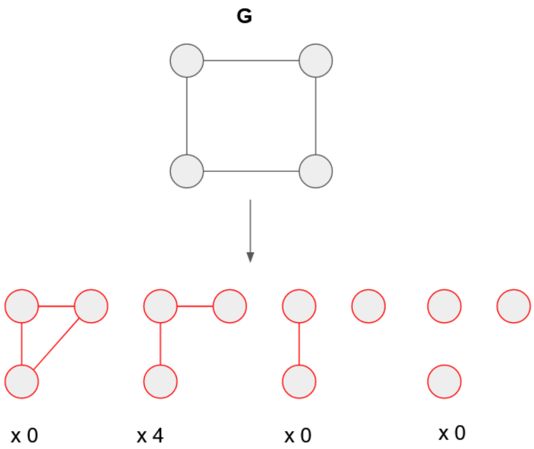


Figure 4: Graph G and its decompositon into size 3 graphlets

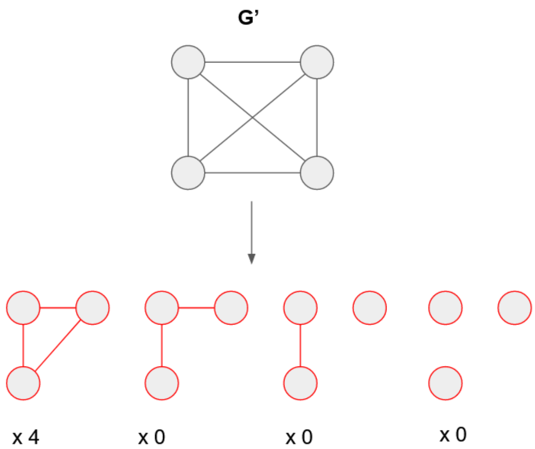


Figure 5: Graph G' and its decompositon into size 3 graphlets

According to those decompositions we get the two following vectors :

- $f_G = \{0, 4, 0, 0\}$
- $f_{G'} = \{4, 0, 0, 0\}$

Therefore, we have the graphlet kernel :

$$k(G, G') = f_G^T f_{G'} = 0$$