# 1 Question 1

Let G = (V, E) an undirected graph without self-loops with $|V|$ = n et $|E|$ = m. The maximum number of edges and triangles is reached when the graph is complete (every node is connected).

The maximum number of edges that G can contain is $\frac{n(n-1)}{2}$. This formula is derived from the fact that every node is connected to (n-1) other nodes (excluding itself), and divide by 2 to account for the fact that each edge is counted twice and the graph is undirected.

The maximum number of triangles that G can contain is $\binom{n}{3} = \frac{n(n-1)(n-2)}{6}$. This corresponds to the number of ways to take 3 nodes from a graph with n nodes, since all nodes are connected to each other.

# 2 Question 2

Two graphs with the same degree distribution are not necessarily isomorphic. In fact, here is a counter-example. The graphs G1 and G2 have the same degree distribution. Let denote the degree distribution $P_G(k)$ the fraction of nodes with degree k in the graph G :

$$P_G(k) = \begin{cases} \frac{1}{2} & \text{if k=1} \\ \frac{1}{3} & \text{if k=2} \\ \frac{1}{6} & \text{if k=3} \end{cases} \text{ for } G \in \{G_1, G_2\} \tag{1}$$

However, these graphs are clearly not isomorphic.

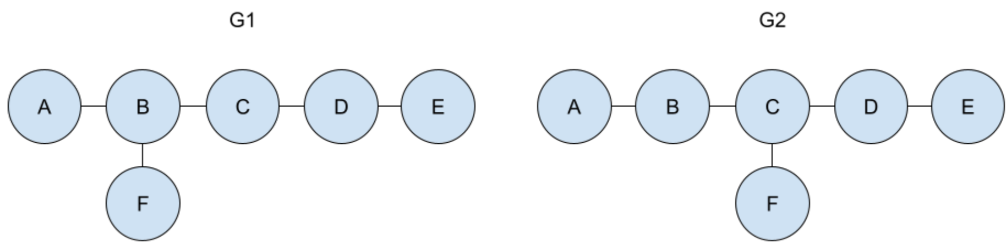


Figure 1: Two graphs with same degree distribution but non-isomorph

# 3 Question 3

Let $C_k$ a cycle of length k $\in \{3, 4, 5, ...\}$.

The global clustering coefficient is defined as the number of closed triplets divided by the total number of both triplets (open and closed).We note it $coef_{gc}$
$coef_{gc}(G) = \frac{\text{number of closed triplets}}{\text{number of closed triplets+number of open triplets}}$

If **k = 3**: there is only one closed triplet and zero open triplet. So, $coef_{gc}(C_k) = 1$

If **k** $>$ **3**: the graph no longer contains a closed triplet, as it is impossible to extract a triangle. So the number of closed triplets is zero. On the other hand, the number of open triplets is strictly positive (just take 3 consecutives nodes). Hence, $coef_{gc}(C_k) = 0$.

# 4 Question 4

Let $u_1 \in \mathbb{R}^n$ be the egeinvector associated with the smallest eigenvalue of $L_{rw}$. Let $[u_1]_i$ be the $i^{th}$ element of $u_1$ and A the adjacency matrix.The follwing quantity

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

Let's prove that result:

The random walk Laplacian is defined by $L_{rw} = I - D^{-1}A$ with D the degree matrix. We will show that the eigenvalues of this matrix are positive and that $u_1 = a\mathbb{1}$ with $a \in \mathbb{R}$ is an eigenvector associated with the smallest eigenvalue.

To show this result, we will show that the Laplacian $L_{sym} = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ is symmetric positive semi-definite (in particular its eigenvalues are positive or zero) and that $L_{rw}$ has the same eigenvalues as $L_{sym}$[1].

$L_{sym}$ is symmetric because I and $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ are symmetrics (A symmetric). Let's show that $L_{sym}$ is semi-positive, so $x \in \mathbb{R}^n$.

$$\begin{aligned}
x^T(I - D^{-1/2}AD^{-1/2})x &= x^Tx - x^TD^{-1/2}AD^{-1/2}x \\
&= x^Tx - z^TAz \; with \; z = \left[\frac{x_1}{\sqrt{d_1}}, ..., \frac{x_n}{\sqrt{d_n}}\right]^T \\
&= x^Tx - \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}z_iz_j = \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j}\frac{x_ix_j}{\sqrt{d_i}\sqrt{d_j}} \\
&= \frac{1}{2}\left(2 \times \sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j}\frac{x_ix_j}{\sqrt{d_i}\sqrt{d_j}}\right) = \frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j}\frac{x_ix_j}{\sqrt{d_i}\sqrt{d_j}} + \sum_{j=1}^{n} x_j^2\right)
\end{aligned}$$

But we have $\sum_{j=1}^{n} A_{i,j} = d_i$ so $\sum_{j=1}^{n} \frac{A_{i,j}}{d_i} = 1$

$$\begin{aligned}
x^T(I - D^{-1/2}AD^{-1/2})x &= \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i^2A_{i,j}\frac{1}{d_i} - 2\sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j}\frac{x_ix_j}{\sqrt{d_i}\sqrt{d_j}} + \sum_{i=1}^{n}\sum_{j=1}^{n} x_j^2A_{i,j}\frac{1}{d_j}\right) \text{ because } A_{ji} = A_{ij} \\
&= \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j}\left(x_i^2\frac{1}{d_i} - 2\frac{x_ix_j}{\sqrt{d_i}\sqrt{d_j}} + x_j^2\frac{1}{d_j}\right)\right) \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j}\left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}}\right)^2 \geq 0
\end{aligned}$$

Hence, $L_{sym} \geq 0$ so $0 \leq \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ its eigenvalues and $u_1, ..., u_n$ the associated eigenvectors.

Let's show that $L_{rw}$ has the same eigenvalues as $L_{sym}$. Let $u_k$ be an eigenvector of $L_{sym}$.

$$\begin{aligned}
L_{sym}u_k &= \lambda_k u_k \\
u_k - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}u_k &= \lambda_k u_k \\
D^{-\frac{1}{2}}u_k - D^{-1}AD^{-\frac{1}{2}}u_k &= \lambda_k D^{-\frac{1}{2}}u_k \\
(I_n - D^{-1}A)D^{-\frac{1}{2}}u_k &= \lambda_k D^{-\frac{1}{2}}u_k \\
(l_{rw})D^{-\frac{1}{2}}u_k &= \lambda_k D^{-\frac{1}{2}}u_k
\end{aligned}$$

So $\lambda_k$ is an eigenvalue of $L_{rw}$ associated with the vector $D^{-\frac{1}{2}}u_k$. Hence the eigenvalues of $L_{rw}$ and $L_{sym}$ are the same. In particular, they are positive.

But $L_{rw}\mathbb{1} = \mathbb{1} - D^{-1}A\mathbb{1}$ and $(A\mathbf{1})_i = \sum_{j=1}^{n} A_{ij}$ then $\left(D^{-1}A\mathbb{1}\right)_i = \frac{1}{d_i} \times \sum_{j=1}^{n} A_{ij} = 1$ (as shown earlier), so 0 is an eigenvalue of $L_{rw}$. The eigenvalues of $L_{rw}$ are positive, therefore $\lambda_1 = 0$ is the smallest egeinvalue

[1]We use $L_{sym}$ because $L_{rw}$ is generally not symmetric.

associated with the vector $u_1 = a\mathbb{1}\ with\ a \in \mathbb{R}$.

In conclusion, $\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\left([u_1]_i - [u_1]_j\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\left(a - a\right)^2 = 0$.

# 5 Question 5

Modularity M is defined by:

$$M = \sum^{n_c}\left[\frac{l_c}{m} - \left(\frac{d_c}{2m}\right)^2\right] \tag{3}$$

Modularity provides an evaluation measure for assessing the quality of clustering. At first glance, if I had to choose one of the two partitions, I would go for the first, as we can identify two cliques of 4 nodes. So, we would like to group these nodes in the same clusters. Thus, we expect a higher modularity score for graph 1 than graph 2.

Both graphs have two clusters: blue nodes and yellow nodes.

## 5.1 Graph 1

The graph 1 has m = 14 edges :

| | Cluster 1 | Cluster 2 |
|---|---|---|
| $l_c$ | 6 | 6 |
| $d_c$ | 14 | 14 |

So, M = $\frac{l_1}{m} - \left(\frac{d_1}{2m}\right)^2 + \frac{l_2}{m} - \left(\frac{d_2}{2m}\right)^2 = \left(\frac{6}{14} - \left(\frac{14}{2\times 14}\right)^2\right) \times 2 = \frac{5}{14} \simeq 0.357$

## 5.2 Graph 2

Graph 2 also has m = 14 edges.

| | Cluster 1 | Cluster 2 |
|---|---|---|
| $l_c$ | 5 | 2 |
| $d_c$ | 17 | 11 |

So, M = $\frac{l_1}{m} - \left(\frac{d_1}{2m}\right)^2 + \frac{l_2}{m} - \left(\frac{d_2}{2m}\right)^2 = \left(\frac{5}{14} - \left(\frac{17}{2\times 14}\right)^2\right) + \left(\frac{2}{14} - \left(\frac{11}{2\times 14}\right)^2\right)$ = -$\frac{9}{392} \simeq -0.023$

Hence, M($G_1$) $>$ M($G_2$). The results are in line with our expectations.

# 6 Question 6

Let $\phi$ be the function that associates with a graph the vector that counts the number of shortest paths of length k in the graph.

We have $\phi(P4)$ = [3, 2, 1] and $\phi(S4)$ = [3, 3, 0]. Let $k_{sp}$ be the shortest path kernel defined by

$$k_{sp}(G, G') = \phi(G)^T\phi(G') \tag{4}$$

The result is:

$k_{sp}(P_4, P_4) = 9 + 4 + 1 = 14$
$k_{sp}(S4, S4) = 9 + 9 = 18$
$k_{sp}(P4, S4) = 9 + 6 = 15$

# 7 Question 7

Let k denote the graphlet kernel that decomposes graphs into graphlets of size 3 and G, G' two graphs. The fact that k(G, G') $= f_G^T f_{G'} = 0$ means that there is no subgraph of 3 nodes in G and subgraph of 3 nodes in G' that are both isomorphic to a same graphlet.

Consider the set of graphlets of size 3: G1, G2, G3, G4 (figure 4 of the poly.). In graph G, we can only extract sub-graphs with 3 nodes which are isomorphic to $G_1$. For the graph G', we can only extract subgraphs with 3 nodes that are isomorphic to $G_2$. The result is:

$$\begin{cases} f_G = [4, 0, 0, 0] \\ f_{G'} = [0, 4, 0, 0] \end{cases} \Rightarrow f_G^T f_{G'} = 0 \tag{5}$$

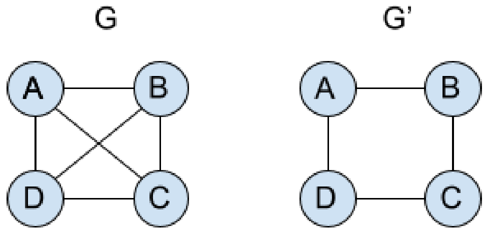


Figure 2: Two graphs with k(G, G')=0