1 Question 1

Let G be an undirected graph of n nodes without self-loops.

The maximum number of edges in G is $\frac{(n-1)n}{2}$. Indeed after selecting a first vertex, we can draw $n-1$ edges to the remaining vertexes. Then selecting a second vertex, it is already connected to the first one, so we can only draw $n-2$ edges to the remaining vertexes. And so on.

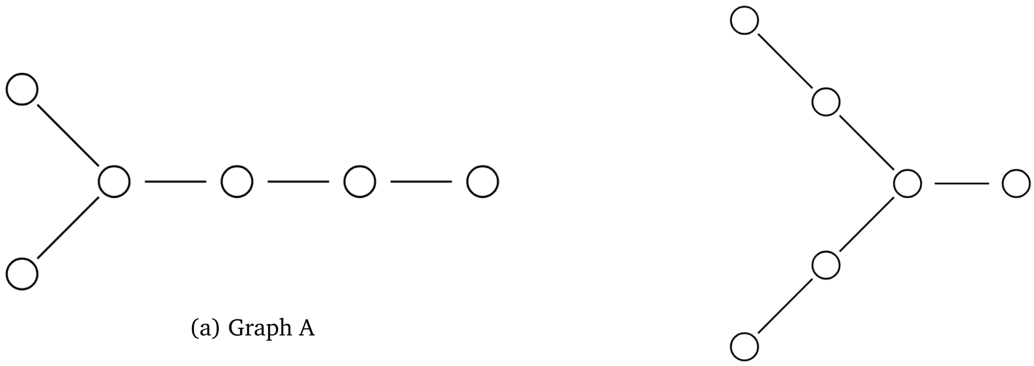
$$\sum_{i=1}^{n-1} i = \frac{(n-1)n}{2}$$

A triangle in G is obtained by choosing three vertexes among n without replacement. So the maximum number of triangles in G is simply :

$$\binom{n}{3}$$

2 Question 2

If two graphs have the same degree distributions, they are generally not isomorphic to each other. To prove this, here is a counter-example :



(a) Graph A

(b) Graph B

Figure 1: Two graphs with the same degree distribution but not isomorphic to each other

These two graphs have the same degree distribution (3, 2, 1). However they are not isomorphic to each other.

To prove this, let us first prove the following lemma : Let A and B be two graphs isomorphic to each other, and suppose there exists a path (a, b, c, d, e) in the first graph with no repeated vertex. Then there exists a path $(f(a), f(b), f(c), f(d), f(e))$ in the second graph with no repeated vertex.

Proof. Let f be the isomorphism between A and B. Since a, b, c, d and e are all different, we have that $f(a)$, $f(b)$, $f(c)$, $f(d)$ and $f(e)$ are all different.

- Since $a \leftrightarrow b$, we have that $f(a) \leftrightarrow f(b)$.
- Since $b \leftrightarrow c$, we have that $f(b) \leftrightarrow f(c)$.
- Since $c \leftrightarrow d$, we have that $f(c) \leftrightarrow f(d)$.
- Since $d \leftrightarrow e$, we have that $f(d) \leftrightarrow f(e)$.

Thus we have that $(f(a), f(b), f(c), f(d), f(e))$ is a path in B with no repeated vertex. $\square$

In our case, we see that the graph A exhibits two paths of length 5, while the graph B only exhibits one. If they were isomorphic to each other, two of the vertices of the graph A would be mapped to the same vertex in the graph B, since the two distinct paths in A would be mapped to the same path in B (by the lemma). This is a contradiction, so the two graphs are not isomorphic to each other.

3 Question 3

Let C_n be the cycle graph of n nodes.

- If $n = 3$, then C_n has exactly three closed triplets, each one centered on one of the three nodes. It has no open triplets.
- If $n > 3$, then C_n has no closed triplet. The number of open triplets is irrelevant in this case.

Thus the global clustering coefficient of C_n is :

$$\begin{cases} 1 & \text{if } n = 3 \\ 0 & \text{if } n > 3 \end{cases}$$

4 Question 4

Our goal is to prove the following result :

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

We denote by λ_1 the eigenvalue associated with the eigenvector u_1.
Let $f \in \mathbb{R}^n$, similarly to [1], we compute $f^T L f$:

$$\begin{aligned} f^T L f &= f^T f - f^T D^{-1} A f \\ &= \sum_{i=1}^{n} f_i^2 - \sum_{i=1}^{n} \sum_{j=1}^{n} f_i D_{ij}^{-1} A_{ij} f_j \\ &= \frac{1}{2} \left(\sum_{i=1}^{n} f_i^2 - 2 \sum_{i=1}^{n} \sum_{j=1}^{n} f_i D_{ij}^{-1} A_{ij} f_j + \sum_{i=1}^{n} f_i^2 \right) \\ &= \frac{1}{2} \sum_{i,j=1}^{n} D_{ij}^{-1} A_{ij} (f_i - f_j)^2 \\ &\geq 0 \end{aligned}$$

Thus for any f and λ such that $Lf = \lambda f$, we have that $f^T L f = \lambda f^T f = \lambda ||f||^2 \geq 0$. All eigenvalues of L are non-negative.

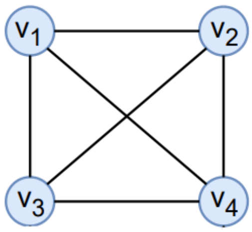
Let us consider the vector $\mathbb{1}$, we easily have that $L\mathbb{1} = \mathbb{0}$ (by construction of D and A), so $\lambda_1 = 0$ and $u_1 = \mathbb{1}$.

This suffices to prove (1) as all terms in the sum are null.

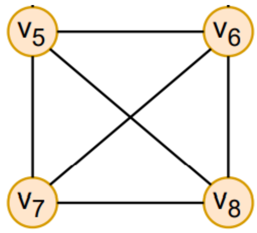
5 Question 5

For the left graph :

We denote the cluster containing the nodes $v_1, \ldots, v_4$ as cluster 1 and the cluster containing the nodes $v_5, \ldots, v_8$ as cluster 2.



(a) Cluster 1



(b) Cluster 2

Figure 2: The two communities of the left graph

Based on 2 and the initial graph, we compute the following :

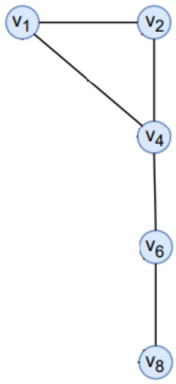
- $m = 14$
- $l_1 = 6$ and $l_2 = 6$
- $d_1 = 14$ and $d_2 = 14$

Thus we have that :

$$Q = \sum_{i=1}^{2} \left[\frac{l_i}{m} - \left(\frac{d_i}{2m} \right)^2 \right] = \frac{5}{14} \approx 0.36$$

For the right graph :

We denote the cluster containing the nodes v_1, v_2, v_4, v_6, v_8 as cluster 1 and the cluster containing the nodes v_3, v_5, v_7 as cluster 2.



(a) Cluster 1

(b) Cluster 2

Figure 3: The two communities of the left graph

Based on 3 and the initial graph, we compute the following :

- $m = 14$
- $l_1 = 5$ and $l_2 = 2$
- $d_1 = 17$ and $d_2 = 11$

Thus we have that :

$$Q = \sum_{i=1}^{2} \left[\frac{l_i}{m} - \left(\frac{d_i}{2m} \right)^2 \right] = \frac{-9}{392} \approx -0.02$$

Conclusion :

In this example the left graph has a higher modularity than the right graph. This is consistent with the fact that the communities in the left graph are more cohesive than the communities in the right graph.

6 Question 6

We have that:

- $\phi(P_4) = [3, 2, 1, 0, \ldots, 0]$
- $\phi(S_4) = [4, 6, 0, \ldots, 0]$

Thus :

- $k(P_4, P_4) = 6$
- $k(P_4, S_4) = 5$
- $k(S_4, S_4) = 10$

7 Question 7

In the graphlet case, a kernel value equal to 0 means that no subgraph of size 3 is shared between the two graphs. Two graphs with a kernel value of 0 must be very dissimilar. We give different examples of graphs with a kernel value of 0 :

- A complete graph of any size with a cycle of more than 3 edges.
- A complete graph with a binary tree.
- Basically any graph without a triangle; with a complete graph.
- If we restrain ourselves to small graphs, we can easily build graphs that only have a certain type of graphlet. As the graphs grow, it becomes more and more likely to find a graphlet of a certain type, so the kernel value will probably not be 0.

References

[1] Ulrike Von Luxburg. A tutorial on spectral clustering. 2007.