BARILLER Halvard
halvard.bariller@ens-paris-saclay.fr

1 Question 1

What is the maximum number of edges and the maximum number of triangles an undirected graph of n nodes without self-loops can have?

Let $G = (V, E)$ be an undirected graph of n nodes without self-loops. The maximum number of edges is obtained when each node is connected to every other node. Without counting the same edge several times, the total number of edges is thus :

$$|E|_{max} = \frac{(n-1)n}{2}.$$

The maximum number of triangles is obtained in the case where the graph is complete. In this case, the number of triangles is computed as the number of combinations of 3 nodes among n:

$$|T|_{max} = C(n, 3) = \frac{n!}{3!(n-3)!}.$$

2 Question 2

Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are isomorphic, $G_1 \cong G_2$, if there is a bijective mapping $f : V_1 \to V_2$ such that $(v_i, v_j) \in E_1 \iff (f(v_1), f(v_2)) \in E_2$. Let two graphs have identical degree distributions. Does this imply that the two graphs are isomorphic to each other? If not, give a counterexample.

In the following example, both graphs have the same degree distribution of $[3, 2, 2, 2, 1]$.

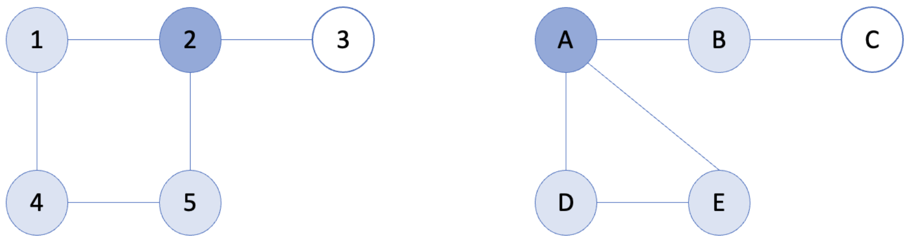

Figure 1: Non-Isomorphs Graphs (darker nodes as the degree is large)

For them to be isomorphic, we have the necessary and sufficient condition that there exists a bijective mapping f between the set of nodes preserving edges (as reminded in the question).
If such a mapping existed, then we would necessarily have the following:

$$\begin{aligned}
(1,2) \in E_1 &\iff (f(1), f(2)) \in E_2 \\
(2,5) \in E_1 &\iff (f(2), f(5)) \in E_2 \\
(5,4) \in E_1 &\iff (f(5), f(4)) \in E_2 \\
(4,1) \in E_1 &\iff (f(4), f(1)) \in E_2
\end{aligned}$$

However, that would imply that there would be a 4-cycle in the second graph, which is not the case. Therefore, we conclude that the graphs are not isomorphic, and thus that identical degree distributions is a necessary but not sufficient condition for graph isomorphism.

3 Question 3

Let C_n denote a cycle graph on n vertices, i.e., a graph containing a single cycle through all n nodes. What is the global clustering coefficient of graphs $C_3, C_4, C_5, \ldots$?

Let T_c be the number of closed triplets and T_o the number of open triplets. The global clustering coefficient formula is as follows:

$$CC_{global} = \frac{T_c}{T_c + T_o},$$

For C_3, $T_c = 3$ ($3 \times \#triangles$) and $T_o = 0$. The global clustering coefficient is thus of 1.
For $n > 3$, the global clustering coefficient of C_n is equal to 0. Indeed, in a cycle graph, no set of three nodes forms a triangle because each node is connected only to its two immediate neighbors. This makes sense because, in a cycle graph, while nodes are connected in a loop, there is no "tight-knittedness" beyond immediate neighbors.

4 Question 4

Let $\mathbf{u}_1 \in \mathbb{R}^n$ denote the eigenvector associated with the smallest eigenvalue of $\mathbf{L}_{rw}$ and let $[\mathbf{u}_1]_i$ denote the i-th element of $\mathbf{u}_1$. What is the output of the following expression $\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{A}_{ij}\left([\mathbf{u}_1]_i - [\mathbf{u}_1]_j\right)^2$?

Let $\lambda_{min} = 0$ be the smallest eigenvalue of $\mathbf{L}_{rw}$. In the case of an undirected unweighted graph, the adjacency matrix is filled with 0's and 1's, and with D_{ii} the degree of the node i, we have that:

$$\sum_{i=1}^{n}\mathbf{A}_{ij} = \sum_{j=1}^{n}\mathbf{A}_{ij} = D_{ii}. \tag{1}$$

Moreover, following the given definition of the Laplacian matrix, we have that:

$$\mathbf{L}_{rw} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A} \iff \mathbf{D}\mathbf{L}_{rw} = \mathbf{D} - \mathbf{A}. \tag{2}$$

It then follows that:

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{A}_{ij}\left([\mathbf{u}_1]_i - [\mathbf{u}_1]_j\right)^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{A}_{ij}\left([\mathbf{u}_1]_i^2 - 2[\mathbf{u}_1]_i[\mathbf{u}_1]_j + [\mathbf{u}_1]_j^2\right) \\
&= \sum_{i=1}^{n}[\mathbf{u}_1]_i^2\sum_{j=1}^{n}\mathbf{A}_{ij} - 2\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{A}_{ij}[\mathbf{u}_1]_i[\mathbf{u}_1]_j + \sum_{j=1}^{n}[\mathbf{u}_1]_j^2\sum_{i=1}^{n}\mathbf{A}_{ij} \\
&\overset{(1)}{=} \sum_{i=1}^{n}[\mathbf{u}_1]_i^2\mathbf{D}_{ii} - 2\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{A}_{ij}[\mathbf{u}_1]_i[\mathbf{u}_1]_j + \sum_{j=1}^{n}[\mathbf{u}_1]_j^2\mathbf{D}_{jj} \\
&= 2\left(\sum_{i=1}^{n}[\mathbf{u}_1]_i^2\mathbf{D}_{ii} - \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{A}_{ij}[\mathbf{u}_1]_i[\mathbf{u}_1]_j\right) \\
&= 2\left(\langle\mathbf{D}\mathbf{u}_1, \mathbf{u}_1\rangle - \langle\mathbf{A}\mathbf{u}_1, \mathbf{u}_1\rangle\right) \\
&= 2\langle(\mathbf{D} - \mathbf{A})\mathbf{u}_1, \mathbf{u}_1\rangle \\
&\overset{(2)}{=} 2\langle\mathbf{D}\mathbf{L}_{rw}\mathbf{u}_1, \mathbf{u}_1\rangle \\
&= 2\underbrace{\lambda_{min}}_{=0}\langle\mathbf{D}\mathbf{u}_1, \mathbf{u}_1\rangle \\
&= 0.
\end{aligned}$$

5 Question 5

Compute (showing your calculations) the modularity of the clustering results shown in Figure 1. Note that different colors correspond to different clusters.

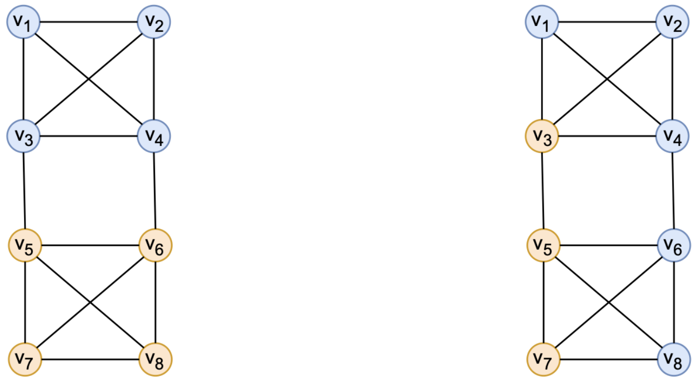

Figure 2: Two graphs (A left, B right) where nodes have been assigned to 2 clusters each. Cluster membership is indicated by node colour.

We recall that modularity is given by the following formula:

$$Q = \sum^{n_c} \left[\frac{l_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right], \text{ where } \begin{cases} n_c & \text{number of communities} \\ m = |E| & \text{total number of edges} \\ l_c & \text{number of edges within community } c \\ d_c & \text{sum of degrees of nodes in community } c \end{cases}$$

For graph A, we have the following:

$$\begin{cases} n_c &= 2 \\ m &= 14 \\ l_c &= [6, 6] \\ d_c &= [14, 14] \end{cases}$$

Therefore, we have that:

$$Q_A = \left[\frac{6}{14} - \left(\frac{14}{2 \times 14} \right)^2 \right] + \left[\frac{6}{14} - \left(\frac{14}{2 \times 14} \right)^2 \right] = 0.35714.$$

For graph B, we have the following:

$$\begin{cases} n_c &= 2 \\ m &= 14 \\ l_c &= [5, 2] \\ d_c &= [17, 11] \end{cases}$$

Therefore, we have that:

$$Q_A = \left[\frac{5}{14} - \left(\frac{17}{2 \times 14} \right)^2 \right] + \left[\frac{2}{14} - \left(\frac{11}{2 \times 14} \right)^2 \right] = -0.02296.$$

The obtained results are thus aligned with our expectations regarding the modularity score: a higher score is testament of a better community structure, which appears clearly in our case for the graph A.

6 Question 6

Let P_n denote a path graph on n vertices and S_n denote a star graph on n vertices. Calculate the shortest path kernel for the pairs $(P_4, P_4), (P_4, S_4)$ and (S_4, S_4).

The feature map ϕ associated to the shortest path kernel returns a vector containing the number of shortest paths of length n for each possible length in a graph.
Depending on the implementation, the feature map can be computed slightly differently: indeed, in the given code, by using the *dict(nx.shortest_path_length())* function, the first key-value pair associated to each primary key of the dictionary contains the length of the shortest path to itself, which is zero. Since the nested *for* loops going through the nodes are allowing the iterators to be equal to the same node, the first entry

of the vector returned by the feature map is the number of shortest paths of length zero, i.e. the number of nodes in the graph. However, in the example given for a path graph P_3, it seems that the first element in the feature map output is the number of shortest path of length 1. In the following, the kernels will be calculated omitting the number of shortest paths of length zero (the results including this element as in the script can be found by adding the product of the number of nodes of each graph to the below kernel values).

For the path graph P_4, we have $\phi(P_4) = [6, 4, 2]$, meaning that there are 6 shortest paths of length 1, 4 shortest paths of length 2 and 2 shortest paths of length 3. Thus the kernel value is:

$$k(P_4, P_4) = \langle \phi(P_4), \phi(P_4) \rangle = 56.$$

For the star graph S_4, we have $\phi(S_4) = [8, 12, 0]$, meaning that there are 8 shortest paths of length 1, 12 shortest paths of length 2 and no path of length 3. Thus the kernel value is:

$$k(S_4, S_4) = \langle \phi(S_4), \phi(S_4) \rangle = 208.$$

The kernel of P_4 and S_4 is computed as follows:

$$k(P_4, S_4) = \langle \phi(P_4), \phi(S_4) \rangle = 96.$$

7 Question 7

Let k denote the graphlet kernel that decomposes graphs into graphlets of size 3. Let also G, G' denote two graphs and suppose that $k(G, G') = f_G^\top f'_G = 0$. What does a kernel value equal to 0 mean? Give an example of two graphs G, G' for which $k(G, G') = 0$ holds.

Let G, G' be two graphs. We have that:

$$\begin{aligned} k(G, G') = f_G^\top f'_G = 0 &\iff \sum_{i=1}^{4} f_G^{(i)} f_{G'}^{(i)} = 0 \\ &\iff \forall i \in \{1, 2, 3, 4\},\ f_G^{(i)} f_{G'}^{(i)} = 0. \end{aligned}$$

This means that for each graphlet, there is a graph between G and G' for which there does not exist a subgraph of size 3 isomorphic to the considered graphlet. As a result, one can expect the graphs to be completely dissimilar with respect to the subgraphs structured considered by using this kernel. The more straightforward example of such case is if the considered graphs are equal to different graphlets.

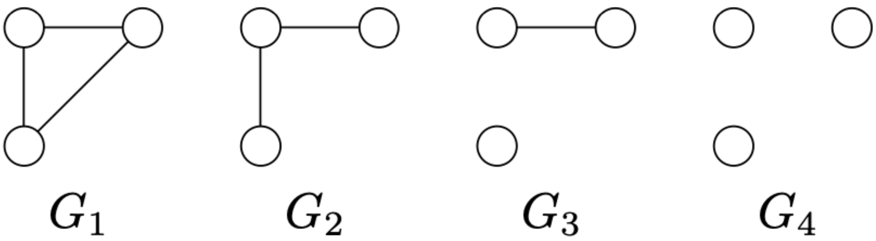

Figure 3: Set of the graphlets of size 3.

In the above figure, $\forall i, j \in \{1, 2, 3, 4\}$ such that $i \neq j$, we have that $k(G_i, G_j) = 0$.
As another example, the same result can be obtained with the two below graphs.

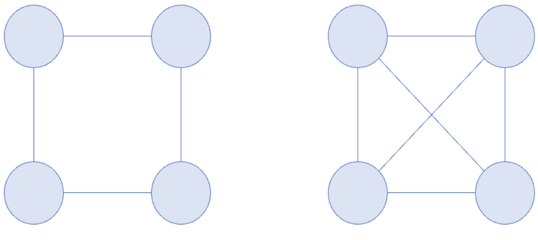

Figure 4: Graphs with null graphlet kernel value.

Indeed, each subgraph of three nodes in the left graph is isomorphic to G_2 (cf. Figure 3) whereas each subgraph of three nodes in the right graph is isomorphic to G_1. As a result, the kernel value between those two is equal to zero.