

Question 1 :

Let G be a cycle graph that consists of n nodes, i.e., a graph containing a single cycle through all n nodes. We randomly remove two edges from G (with uniform probability). What is the number of connected components of the emerging graph?

The number of connected components of the emerging graph is 2. As we can see in 1, removing two edges from G always result in two connected components.

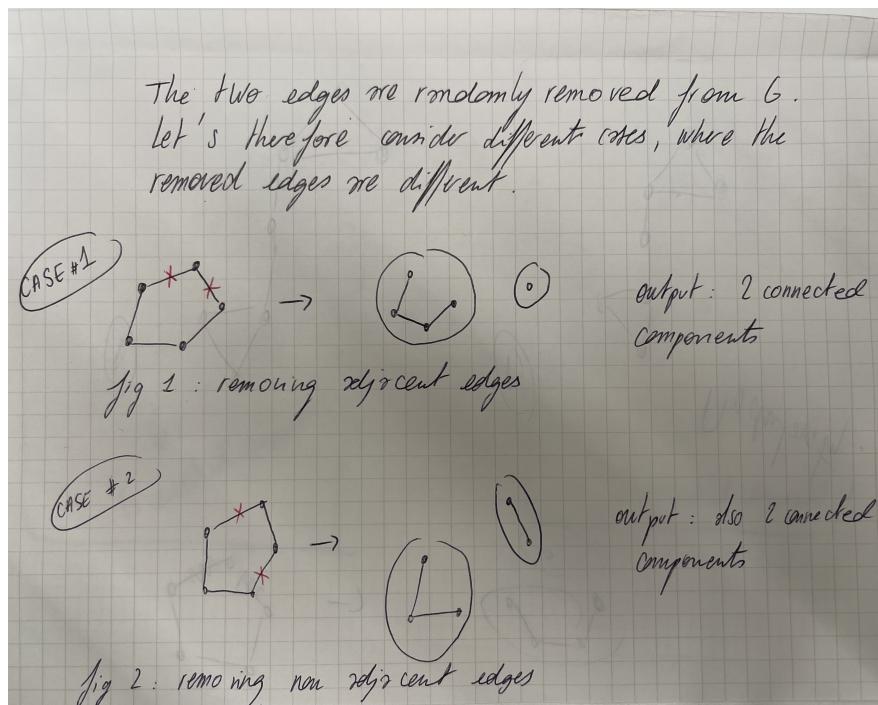


Figure 1: Connected components cycle graph

Question 2 :

Isomorphism implies the preservation of the vertex degrees. If two graphs have identical degree distributions, it however does not necessarily imply that they are isomorphic to each other.
The following figure shows that we can find a counter-example to show that :

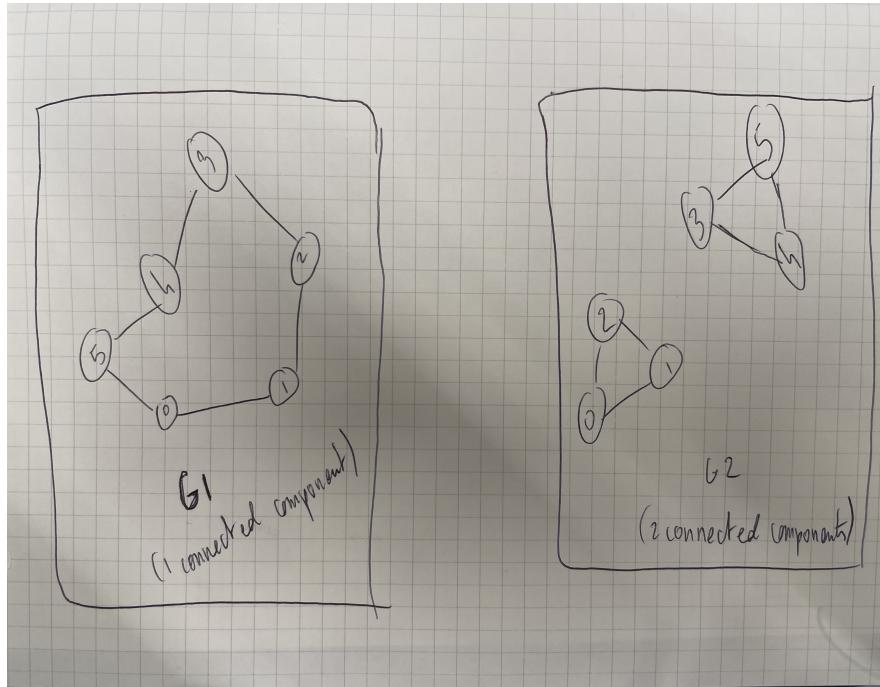


Figure 2: Representation of G_1 and G_2 , two graphs with identical degree distribution

In this representation, G_1 has one number of connected components (and n -cycles = 1), and G_2 has two number of connected components (and n -cycles = 2). Therefore, no bijective mapping between G_1 and G_2 can be applied. In this case, although having identical degree distribution, G_1 and G_2 are not isomorphic.

Question 3 :

Let G be an undirected graph (without self-loops) consisting of n nodes and $\frac{n(n-1)}{2} - 1$ edges. What is the global clustering coefficient of that graph?

The global clustering coefficient of a graph is given by :

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

Here, we consider a graph without any self loops. For a node with n adjacent nodes, the clustering coefficient is given by $\frac{N}{n*(n-1)}$, with N the number of edges between these nodes. As we have n nodes here, we can re-write the clustering coefficient as follow :

$$\frac{N}{(n-1)*(n-2)}$$

$$\text{with } N = \frac{n(n-1)}{2} - 1$$

Therefore, in this case, the undirected graph (without self-loops) consisting of n nodes and $\frac{n(n-1)}{2} - 1$ edges can be represented by a full graph where $n - 2$ nodes are connected to each others (i.e. one edge has been removed from a complete graph). To re-write it, the number of edges in a graph of n nodes is :

$$\binom{n}{2} = \frac{n(n-1)}{2}$$

Moreover, the number of triplets is given by the number of triads in this graph, and can be denoted as : $\binom{n}{3}$. Since two nodes has been removed, the number of closed triplets in our graph is given by :

$$\binom{n}{3} - \binom{n-2}{1}$$

The glocal cluster coefficient can therefore be re-written as :

$$GCC = \frac{\binom{n}{3} - \binom{n-2}{1}}{\binom{n}{3}}$$

Question 5 :

Given a graph G is the output of the spectral clustering deterministic or stochastic?

Let's first recall the pseudo code for the spectral clustering algorithm, given is this TP (see 3) :

Algorithm 1 Spectral Clustering

Input: Graph $G = (V, E)$ and parameter k

Output: Clusters C_1, C_2, \dots, C_k (i.e., cluster assignments of each node of the graph)

- 1: Let A be the adjacency matrix of the graph
 - 2: Compute the Laplacian matrix $L_{rw} = I - D^{-1}A$. Matrix D corresponds to the diagonal degree matrix of graph G (i.e., degree of each node v (= number of neighbors) in the main diagonal)
 - 3: Apply eigenvalue decomposition to the Laplacian matrix L_{rw} and compute the eigenvectors that correspond to d smallest eigenvalues. Let $U = [u_1|u_2|\dots|u_d] \in \mathbb{R}^{m \times d}$ be the matrix containing these eigenvectors as columns
 - 4: For $i = 1, \dots, m$, let $y_i \in \mathbb{R}^d$ be the vector corresponding to the i -th row of U . Apply k -means to the points $(y_i)_{i=1,\dots,m}$ (i.e., the rows of U) and find clusters C_1, C_2, \dots, C_k
-

We will now implement the Spectral Clustering algorithm.

Figure 3: Pseudo code for the spectral clustering algorithm

We first note that the first three steps are mathematical computations that do not include any stochasticity (no approximation or randomness). In the step 4, the k -means algorithm is implemented. The implementation of the k -mean algorithm do not allow to find a global optimum, and therefore carry some stochasticity. This means that at each run, variability in the output is probable. With this respect, we can consider the output of the spectral clustering algorithm as being stochastic.