MARENGO Matteo
matteo.marengo@ens-paris-saclay.fr

Part I
Analyzing a Real-World Graph

1 Question 1

1.1 Maximum number of edges

An undirected graph without self-loops means that no node is connected to itself. Therefore, one node can be connected to n-1 nodes. If you then take an other one it can be connected to n-2 nodes (not n-1 because it is an undirected graph). Therefore it means that:

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

1.2 Maximum number of triangles

Basically, a triangle is made of 3 nodes. So with the same idea than with edges if you take a determinate node, you have to choose 2 nodes between n-1 other nodes. In addition, it should be divided by three as it is the same triangle for the three distinct nodes.

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

We hence have to multiply that result for one node for every n nodes of the undirected graph.

$$Ntrianglesmax = n * Ntripernode = \frac{n(n-1)(n-2)}{6} \tag{3}$$

2 Question 2

For two graphs to be isomorphic, there should exist a bijective mapping $f : V_1 \rightarrow V_2$ such that an edge (v_i, v_j) exists in E_1 if and only if an edge $f(v_i), f(v_j)$ exists in E_2. It is shown in Fig 1.

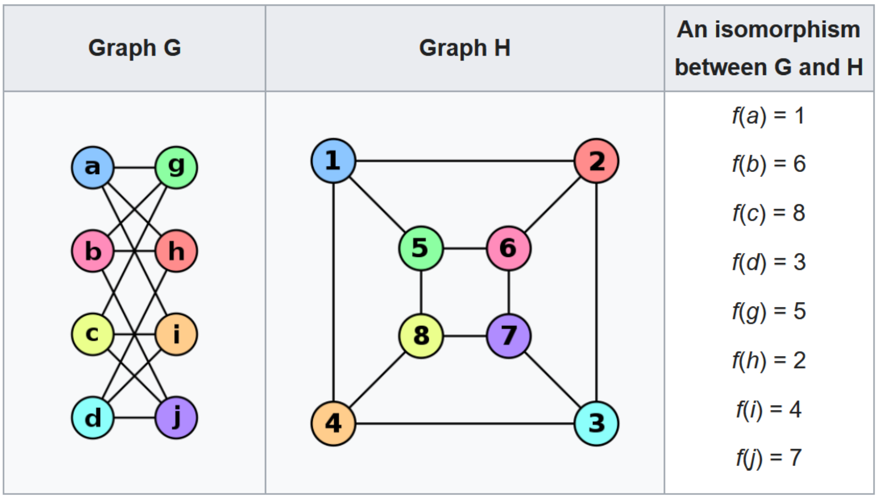


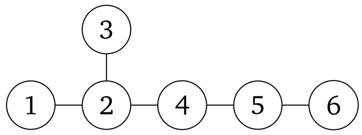
Figure 1: Simple example of two isomorphic graphs. Adapted from Wikipedia.

We can easily proove with a counterexample that even if two graphs have identical degree distributions,it does not imply that the two graphs are isomorphic to each other.

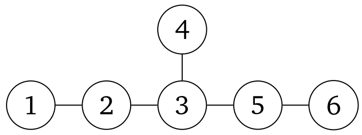
Given Graphs:

- **Graph** G_1: A path graph with 5 vertices with an extra node on node 2.
- **Graph** G_2: A path graph with 5 vertices with an extra node on node 3.

Graph G_1**:**



Graph G_2**:**



These graphs have the same degree distribution (3 nodes of degree 1, 2 nodes of degree 2 and 1 node of degree 3) but there are not isomorphics.

Attempt : Let's map the vertices with the same degrees:

- $f(1) = 1$
- $f(2) = 3$
- $f(3) = 4$
- $f(4) = 2$
- $f(5) = 5$
- $f(6) = 6$

Using this mapping, $1, 2)$ in G_1 maps to $(1, 3)$ in G_2 which is not correct. So, this mapping fails.

Since G_1 has an edge $(1, 2)$ and no possible bijective mapping can produce a corresponding edge in G_2, it's clear that no such bijective function f can exist that satisfies the given definition of graph isomorphism.

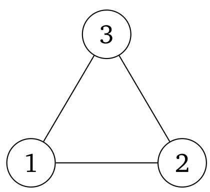
Therefore, G_1 and G_2 are not isomorphic even if they have the same degree distribution.

3 Question 3

The global clustering coefficient of a graph is given by:

$$C(G) = \frac{3 \times \text{number of triangles in } G}{\text{number of connected triples of vertices in } G}$$

In a cycle graph C_n, there are no triangles for $n > 3$. This is because a triangle requires three nodes to be connected, while in a cycle graph each node is connected to just its two neighbors. So, the numerator of our formula is always 0 for C_n where $n > 3$.



For the denominator, a connected triple is a set of three nodes where at least two of them are connected by an edge. For each vertex in C_n, it is part of two connected triples: one with its two neighbors. So, for n nodes, there are $2n$ connected triples in total.
Substituting in our values:

$$C(C_n) = \frac{3 \times 0}{2n} = 0$$

So, the global clustering coefficient of C_n for $n > 3$ is always 0. The only situation that differs is when n = 3, when we have a triangle therefore **C(C_3) = 1**.

Part II
Community Detection

4 Question 4

Given:

- $u_1 \in \mathbb{R}^n$ is the eigenvector associated with the smallest eigenvalue of L_{rw}.
- $[u_1]_i$ is the i-th element of u_1.
- A is the adjacency matrix of the graph.

We need to compute:

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

As presented in the lectures and in [2], L is symmetric, positive and semi-definite and the smallest eigenvalue of L is 0 with a corresponding eigenvector that is 1. Therefore, it means we have that

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

We can prove it more formally thanks to [2]. We have that :

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

The proof is presented in Fig 2.

$$f'Lf = f'Df - f'Wf = \sum_{i=1}^{n} d_i f_i^2 - \sum_{i,j=1}^{n} f_i f_j w_{ij}$$
$$= \frac{1}{2}\left(\sum_{i=1}^{n} d_i f_i^2 - 2\sum_{i,j=1}^{n} f_i f_j w_{ij} + \sum_{j=1}^{n} d_j f_j^2 \right) = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij}(f_i - f_j)^2.$$

Figure 2: Demonstration of the previous sum. Adapted from [2].

Furthermore u_1 is the eigenvector related to the lowest eigenvalue, so

$$u_1^T L_{rw} u_1 = u_1^T \lambda_1 D u_1$$

As $\lambda_1 = 0$ we return to the sum is equal to 0.

5 Question 5

5.1 Computation for graph 1

$m = 14$, $n_c = 2$, $l_1 = 6$, $l_2 = 6$, $d_1 = 14$, $d_2 = 14$

$$Q = 2 * (\frac{6}{14} - (\frac{14}{28})^2) = 0.357 \tag{4}$$

5.2 Computation for graph 2

$m = 14$, $n_c = 2$, $l_1 = 5$, $l_2 = 2$, $d_1 = 11$, $d_2 = 17$

$$Q = (\frac{5}{14} - (\frac{17}{28})^2) + (\frac{2}{14} - (\frac{11}{28})^2) = -0.0229 \tag{5}$$

I obtain that modularity of graph 1 is higher than modularity of graph 2. It seems logical as you would intuitively classify the graphs such as the one in graph 1.

Part III
Graph Classification using Graph Kernels

6 Question 6

6.1 shortest path kernel (P_4, P_4)

After Floyd transformations [1] for P_4:

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

$$k(P_4, P_4) = 3 \cdot 3 + 2 \cdot 2 + 1 \cdot 1 = 14 \tag{6}$$

6.2 shortest path kernel (P_4, S_4)

After Floyd transformations for S_4:

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

$$k(P_4, S_4) = 3 \cdot 3 + 2 \cdot 3 + 1 \cdot 0 = 15 \tag{7}$$

6.3 shortest path kernel (S_4, S_4)

$$k(S_4, S_4) = 3 \cdot 3 + 3 \cdot 3 = 18 \tag{8}$$

7 Question 7

If the graphlet kernel $k(G, G')$ is defined as

$$k(G, G') = f_G^\top f_{G'} = 0$$

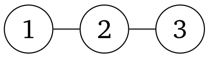
then a kernel value of 0 indicates that the two graphs G and G' have no graphlets of size 3 in common if we look at the sampled graphlets.

In other words, in the context of the graphlets of size 3 that were sampled, the two graphs do not share common substructures.

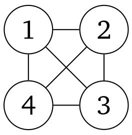
Example:

Given the two graphs G and G':

Graph G:



Graph G':



We want to compute the graphlet kernel $k(G, G')$, which is the dot product of the feature vectors f_G and $f_{G'}$.

1. **Identify Graphlets of Size 3:** The graphlets of size 3 are all the possible connected subgraphs that have 3 nodes. For example:
   - A path or a cycle with 3 nodes
2. **Compute Feature Vectors:** For each graph, we count the number of subgraphs that are isomorphic to each graphlet. In our example, we have two graphlets of size 3 (path and cycle). So, each feature vector will have 2 entries.

   For graph G:

   - There is 1 path with 3 nodes ($1 - -2 - -3$), so the first entry is 1.
   - There are no cycles with 3 nodes, so the second entry is 0.

   Hence, the feature vector for G is $f_G = [1, 0]$.

   For graph G':

   - There are no paths with 3 nodes, so the first entry is 0.
   - There is 4 cycle with 3 nodes, so the second entry is 4.

   Hence, the feature vector for G' is $f_{G'} = [0, 4]$.
3. **Compute Dot Product:** The dot product of two vectors $a = [a_1, a_2, \ldots, a_n]$ and $b = [b_1, b_2, \ldots, b_n]$ is calculated as:
$$a \cdot b = a_1 \times b_1 + a_2 \times b_2 + \ldots + a_n \times b_n$$
   For our example, the dot product of f_G and $f_{G'}$ is:
$$f_G \cdot f_{G'} = 1 \times 0 + 0 \times 4 = 0$$

Hence, the graphlet kernel $k(G, G')$ is 0; there are no common graphlets of size 3 between the two graphs G and G'.

References

[1] Karsten M Borgwardt and Hans-Peter Kriegel. Shortest-path kernels on graphs. In *In Proceedings of the 5th IEEE International Conference on Data Mining*, pages 74–81, 2005.

[2] Ulrike Von Luxburg. A tutorial on spectral clustering. In *Statistics and computing*, pages 17(4):395–416, 2007.